Predicting Early Indicators of Dropout in Online Therapy for Problem Drinkers:

Using LIWC to analyse email contact between client and counsellor

Tinka S. van den Hazel
s1864106

Date: 19-02-2020

University of Twente
Supervisor: Wouter Smink
Second supervisor: Anneke Sools
Third supervisor: Erik Tjong Kim Sang
Faculty: BMS
Positive Psychology and Technology
Course code: 201300125

ABSTRACT

Mental disorders such as Alcohol Use Disorders (AUD), including alcohol abuse and dependence, are highly prevalent in the Netherlands. Problem drinking affects the mental well-being of those suffering from it and often cause other psychological issues, mainly anxiety and depression disorders. Online therapy treatments designed for problem drinkers are effective on average, though do not consider individual differences, and are characterised by high dropout rates. Moreover, premature quitting the intervention often results in worsening or chronic symptoms. It is important to note that there exists a literature gap in predicting dropout in online interventions using linguistic predictors, as non-linguistic predictors for dropout such as gender, age and education have been researched before. Therefore, the purpose of this research was to identify what factors can predict dropout in online therapy, by means of using non-linguistic and linguistic factors. It is important to identify those factors; therapists are able to tailor their treatment in order to facilitate completion for their clients, which is essential for better therapy and long-term outcome.

Method: A sample (n = 990) of clients was taken from the existing intervention found on alcoholdebaas.nl developed by Tactus. Only clients who followed the intensive program were included in the sample.. The client data was first anonymised, then used to read conversations between the client and counsellor to label clients as *dropout* or *completer*. Clients were labelled as dropout when they did not complete all assignments. An explorative approach was used to investigate whether linguistic and non-linguistic factors could predict for dropout. Two logistic regression tests were performed to predict for dropout with demographic characteristics and LIWC dictionaries.

Results: Predictive factors for dropout were being male, younger of age, lower education, smoking, and a higher baseline of alcohol intake per week. A number of dictionaries of LIWC were found to predict for completion: *impersonal pronouns*, *common adverbs, male references*, *tentative*, *biological processes*, *affiliation*, *focus past*, *informal*, *1st person plural*, *3rd person singular*, *positive emotion words, negative emotion words, perceptual processes*, *leisure,* and *death* words. The ability to reflect, having a social network, and sharing personal details with the counsellor could be influencing the word use of the client, and predict for completion.

Conclusion: It is possible to predict for completion and dropout using linguistic factors and might give more insight on dropout in online therapy in combination with non-linguistic predictors which have been researched before. A main limitation of this study was that only the average of the first four mails were used. Notably, linguistic and the non-linguistic predictors alike had small differences between the groups of dropout and completers and small predictive values. However, with no literature existing on using linguistic factors to predict for dropout in web-based treatments, more research is recommended. It could be useful to look at the first few mails separately, and include more non-linguistic factors regarding other psychological problems to broaden the scope of the current study.

Key words: Dropout, LIWC, Web-based treatment, Therapeutic Change Process Research (TCPR), Natural Language Processing (NLP), What Works When for Whom.

Predicting Early Indicators of Dropout in Online Therapy for Problem Drinkers

Mental disorders are a worldwide problem as it affects all facets of one's life, such as career and social relationships (Wykes et al., 2015). Around 25% of the global population (Formánek et al., 2019) and 33.3% of Europe's population (Holm-Hadulla & Koutsoukou-Argyraki, 2015) are estimated to develop a mental disorder every year, and numbers are predicted to grow further coming years (Baingana et al., 2019). Specifically, alcohol use disorders (AUD), including alcohol abuse and alcohol dependence, are reported to be highly prevalent (Jeanblanc, 2015). In 2018, around 8 per cent of the population in the Netherlands has stated to drink excessively; for men, this means that they consume 21, and for women 14 standard alcohol units or more each week (Trimbos.nl, n.d.). Alcohol is measured in units because not every drink has the same amount of alcohol (Day, Copello, & Hull, 2015). Drinking too much alcohol has an impact on one's health, mental well-being, and others affected by the actions of people with AUD (Grant et al., 2015). Health risks often apparent with excessive drinking are for example cancer, diabetes, and cardiovascular diseases (Rehm, 2011).

Furthermore, people with alcohol related issues often endure other mental problems as well. Approximately 50 – 70% of clients with AUD suffer from other psychological disorders as a result of poorer self-control found in clients with AUD; mainly depressive and anxiety disorders are co-occurring with AUD, increasing the risk of suicide attempts (Jeanblanc, 2015; Yang et al., 2018). Alternatively, psychological disorders such as schizophrenia, mood and personality disorders may increase the risk of alcohol dependence (Yang et al., 2018). Good mental well-being is needed to function well in all aspects of life, and flourishing mental health prevents mental disorders and extends a person's lifespan (Schotanus-Dijkstra et al., 2017). People are encouraged to seek professional help when they experience symptoms of mental health disorders as it can get worse if they are left untreated; There are various forms of treatment for mental health problems, with psychotherapy and medication being only just a few examples of the many options available (Holland, 2018).

Online counselling/therapy

With the rise of the internet 20 years ago, therapy was also available in forms of communicating via the phone, webcam or emailing (Novotney, 2017). The number of internet users have been increasing globally since then, and many services, including online counselling and therapy, are nowadays more available and accessible to a wider audience (Hoogendoorn, Berger, Schulz, Stolz, & Szolovits, 2017; Lau, Jaladin, & Abdullah, 2013). However, it is unclear and debated over how *online counselling/therapy* should be defined (Lau et al., 2013); as a result, many researchers have their own definition of what online-counselling/therapy entails (see for example Barak 2009; Mallen & Vogel 2005; Bloom 1998).

Several attempts have been made to use specific terms for online therapy related to their specific activities, such as *online therapy*, *cyber therapy*, or *web counselling* (Lau et al., 2013), however, many definitions are used by researchers and laypeople interchangeably (Barak, Hen, Boniel-Nissim, & Shapira, 2008). With unifying definitions, it can be ensured that reviews and meta studies include the appropriate studies (Ritterband & Tate, 2009), which creates clarity what terms actually include and exclude in their definitions. Unfortunately, there is no clear-cut definition (yet) on online therapy and counselling. The definition that I adopt in the current work of online counselling and therapy includes that a) the client and therapist are in different locations, b) the client follows a (semi-)structured program in an online environment for x given weeks, and c) communication between client and therapist is asynchronous or enhanced with synchronous contact.

E-mail therapy. Email therapy is a form of asynchronous contact[1], and can be enhanced with synchronous contact, or entirely used on its own though this is less common. Synchronous contact includes private chatting, video conferencing, and phone calling (Barak et al., 2008), though email-therapy is the most popular used method for online counselling (Chester & Glass, 2006). There is still limited literature regarding email therapy and a clear definition of *email therapy* is not defined yet (Francis-Smith, 2014). Based on the work of Hoogendoorn et al. (2017) email therapy can be defined as an asynchronous form of online counselling where the client and therapist communicate through emails in a secured environment.

Advantages of online therapy

Studies show that online therapy is effective and able to improve behavioural change outcomes (Gainsbury & Blaszczynski, 2011; Kumar, Sattar, Bseiso, Khan, & Rutkofsky, 2017; Wantland, Portillo, Holzemer, Slaughter, & McGhee, 2004), and there are various advantages of online therapy, the most prominent ones are discussed. First, studies on the effectiveness of online therapy found that online therapy is as effective as “traditional” face-to-face therapy (Andersson et al., 2012; Barak et al., 2008; Hedman et al., 2014; Howes, Purver, & McCabe, 2015). Second, e-therapy is more accessible than face-to-face therapy (Barak et al., 2008; Howes et al., 2015). People with a (physical) illness or those who are living in remote areas can stay at home and do not require to travel to the therapist. Third, the costs of e-therapy are lower which makes it possible for people with lower incomes to make use of online treatment (Fernández-Álvarez et al., 2017; Gainsbury & Blaszczynski, 2011). Fourth, there are usually no – to short waiting lists to start e-therapy (ter Huurne, Postel, de Haan, van der Palen, & DeJong, 2017). This can lower the initial hesitation of a person with feelings of shame or anxiety to seek professional help. Furthermore, e-therapy can be a convenient alternative to face-to-face treatment if a person is struggling with social anxiety, because all interaction is conducted via the internet (Bennett, Bennett, & Griffiths, 2010; Kumar et al., 2017; Lau et al., 2013). Lastly, clients can feel more safe and secure online when they have an anonymous identity and may be willing to disclose more about themselves (Fernández-Álvarez et al., 2017).

[1] Asynchronous contact is the most common form of communication in web-based-treatment; the client and therapist talk in turns and their communication is time-delayed (Gainsbury & Blaszczynski, 2011).

Disadvantages of online therapy

In contrast, online therapy also has several disadvantages, the most relevant ones are discussed. First, insufficient protection and inappropriate or inadequate handling of patient data (Amichai-Hamburger, Klomek, Friedman, Zuckerman, & Shani-Sherman, 2014). Protection of hardware breaches is often not prioritized and prone to unauthorised physical access, with client data becoming even more vulnerable when staff members act inappropriately in terms of handling the data, losing external hard drive, or uploading data to a shared drive (Lau et al., 2013). Furthermore, the absence of physical presence limits the therapist to see and react on non-verbal and voice cues of the client (Ball, Carroll, Canning-Ball, & Rounsaville, 2006; Postel, de Haan, ter Huurne, Becker, & de Jong, 2010). This also has an impact on the working relationship between client and therapist. For instance, with written text it is harder to create an emphatic and trusting atmosphere as written text by email can be perceived as cold and distant as a result of formal writing, and some words can be perceived harsher than intended.

Dropout in online therapy. However, even though online counselling is proven to be effective, the biggest drawback of online therapy seems to be the high dropout rates ranging from 8% to 99% (Andrade et al., 2016; Christensen, Griffiths, & Farrer, 2009; Karyotaki et al., 2015; Postel et al., 2011). Dropout rates in online counselling seem to be higher on average compared to face-to-face treatment (ter Huurne et al., 2017), as dropout rates are approximately around 50% (Palmer, Murphy, Piselli, & Ball, 2009; Postel et al., 2011). For web-based treatment, the dropout rates in the research of (Postel et al., 2011) 54% and in the research of (Linke, Murray, Butler, & Wallace, 2007) 84.5% of respondents dropped out of the interventions designed for problem drinkers. For face-to-face treatment, the dropout rates in the research of de Weert-van Oene (2007) were 50% and in Palmer et al. (2009) around 30-50% of people stopped prematurely in the intervention targeted for substance abusers.

Eysenbach’s Law of Attrition distinguishes two types of dropout: dropout attrition and nonusage attrition. Dropout attrition refers to participants lost to follow up; they do no longer have contact with the therapist and do not fill in follow-up questionnaires. Nonusage attrition refers to participants who stopped with the intervention, but still fill in follow-up questionnaires (Eysenbach, 2005). However, not every author uses these terms, or describe them separately (Postel et al., 2011). In my view, dropout can be defined as a participant who does not complete all sessions within a treatment. Furthermore, because of scarce empirical research, most studies regarding dropout in online therapy target different sample groups, with different treatments, different subtypes of disorders, and have different definitions of dropout (Khazaie, Rezaie, Shahdipour, & Weaver, 2016); this makes it even harder to compare dropout studies.

Consequences of dropout. As mentioned before, online therapy is as effective as traditional face-to-face therapy (Andersson et al., 2012; Barak et al., 2008; Hedman et al., 2014; Howes et al.,

2015) and has shown to be effective for substance abuse disorders, however, dropout rates are very high (Ball et al., 2006; Khazaie et al., 2016; Palmer et al., 2009; Postel et al., 2010). Watson et al. (2017) add that “even though the past 25 years have brought innovation in CBT delivery and accessibility, failure to engage in treatment remains problematic and dropout is unacceptably high” (p. 2). Many studies that conducted research on dropout in (online) counselling have emphasised the major negative results on the client’s long-term outcomes when quitting the treatment prematurely, and symptoms of drop-outs are likely to become worse and could become chronic (Ball et al., 2006; Fernández-Álvarez et al., 2017; Postel et al., 2010; ter Huurne et al., 2017). Furthermore, some authors have seemingly difficulties publishing their work when they experienced high dropout rates in their studies (Eysenbach, 2005). Therefore, it is easily assumed that low dropout rates equal therapy success and credibility, and high dropout rates equal failure and implausibility, even though this does not have to be the case.

Text Mining and Natural Language Processing

Email therapy generates a lot of textual data (Basit, 2003), which in turn, gives room to explore online language and interaction in a therapeutic setting. With Text Mining (TM) and Natural Language Processing (NLP) it is possible to extract meaningful information from large amounts of unstructured text (Cohen, 2013; Kotu & Deshpande, 2015). Dreisbach, Koleck, Bourne, and Bakken, (2019) explain that text mining has techniques used for “characterizing and transforming text” (p. 1); within text mining, NLP is a “collection of syntactic and/or semantic rule- or statistical-based processing algorithms” (p. 1) which also can be used to analyse data. With both approaches, one can analyse textual data, but with different intentions. Thus, text mining structures and analyses text, and turns text into numbers in order to apply algorithms to the large document databases (Miner et al., 2012). With NLP, it is possible to analyse content and phrase patterns (Dreisbach et al., 2019). The use of TM and NLP have been successfully used in previous research literature to predict therapeutic process outcomes, symptom severity, and warning signs of depression in textual data (Chung & Pennebaker, 2013; Tausczik & Pennebaker, 2010). One possible NLP approach to analyse textual data is the Linguistic Inquiry Word Count (LIWC) program developed by Pennebaker, Francis, and Booth in 2001. The program counts words in categories and allows for predictions of behavioural outcomes, as well as identifying words that reflect the underlying psychological state of a person (Chung & Pennebaker, 2013; Tausczik & Pennebaker, 2010).

What works when for whom?

Literature has covered therapy effectiveness for online therapy overall, however, it only considers average scores and individual differences among patients are often overlooked (Ball et al., 2006; Fernández-Álvarez et al., 2017; Postel et al., 2010). Moreover, cognitive behavioural therapy works for many, however it does not work equally for everyone (Newman, Jacobson, Erickson, &

Fisher, 2017). Differences in effectiveness could possibly explain the high percentages of dropout because individuals respond differently to treatment (Newman et al., 2017; Postel et al., 2011). The field of Therapeutic Change Process Research (TCPR) aims to identify what mechanisms bring positive and relevant therapeutic change (Greenberg, 1986; Smink et al., 2019). There is a need of researchers and practitioners to look at individual differences of clients in online therapy to see what treatment is most effective for one individual, with a specific problem, under which circumstances; "*what works when for whom?"* (Smink et al., 2019). Thus, if it is possible to personalise treatment per individual, it could lead to lower dropout rates and better therapy outcome for the client.

Research aim

Research has already been done on the reasons to dropout of (online based) interventions (Postel et al., 2010; ter Huurne et al., 2017) as well as predictors for dropout related to demographic and personal factors (Karyotaki et al., 2015; Watson et al., 2017). To the best of my knowledge, no literature exists yet on predicting dropout in email therapy with use of NLP. This research aims to investigate what factors can predict dropout, trying to answer the underlying research question "What are linguistic and non-linguistic factors in predicting dropout in a web-based intervention for problem drinkers?"

This study took an exploratory approach to search for early predictors of dropout in online therapy mails in the first phase of the intervention Alcohol de Baas of Tactus (i.e., in the first 4 emails sent by the client) to compare clients who dropped out with the clients who finished the intervention, based on demographic characteristics and LIWC dictionaries. If predictors for dropout can be discovered at a the early stage of the intervention, it could be helpful for the therapist and client alike; the therapist could adapt to their client's needs to prevent a potential dropout, because early dropout will most likely result in a negative long-term outcome and worsen the client's symptoms (Postel, 2011).

METHOD

Participants

A sample of 1997 clients was used in this study. After excluding respondents who did not complete the intake questionnaire, did not send at least one email, and who followed another treatment variation (short treatment, self-help module, or aftercare contact) 990 clients remained (median age = 46, age range: 17-83 years). The flow of excluded respondents is illustrated in Figure 1. From the 990 clients, only 353 clients completed the treatment, providing a dropout rate of 64.3%.

The majority of the sample reported to be women, of Dutch nationality, married, and to have finished a higher vocational degree. Furthermore, the majority indicated to smoke (occasionally to daily), but not use drugs or gamble. In addition, the majority started with the treatment because they thought they were drinking too much. The median of standard units of alcohol the clients reported to have consumed in the week of filling in the questionnaire was 36. Lastly, almost half of the sample stated that they experienced feelings of depression occasionally to frequently and did get treatment before for psychological or emotional problems. The median of mails sent by the clients was 9.5. There were 45 counsellors who treated the client sample. The distribution of clients per counsellor was very uneven; for example, counsellor 1 treated 70 clients, whereas counsellor 7 treated 6 clients. Demographic characteristics are summarised in Table 1. In Table 2, demographic characteristics are summarised and split to dropout and completer. It is important to note that there were (at least three) different intake questionnaires, and not every question was included in all questionnaires. Moreover, it was optional to answer a question; this means that the meta data had a lot of missing values, hence some questions had low answer rates, some as low as approximately 25% (the questions regarding nationality and marital status).

Design

The data used for this research was provided by Tactus, and used the same design and data as used in (Postel et al., 2011), therefore no ethical approval was needed to conduct this research**.** In addition, when clients started with the treatment, they gave automatically their consent that the data would be used in research by Tactus to evaluate and improve the intervention. The sample of clients took part in the web-based intervention alcoholdebaas.nl[2] around 2010. The web-based treatment is for everyone over 16 years-old who is concerned about or wants to change their drinking behaviour. It is also suitable for people who relapsed or who want to prevent excessive drinking habits (Postel et al., 2011).

Based on the outcomes of the intake questionnaire, the client is advised to take the intensive or short program depending on the client's motivation and degree of problems related to his/her drinking habits. The intensive variation of the program takes around 16 weeks to complete, depending on how fast the client finishes the tasks given by the therapist. It consists of two parts. In the first part of the intervention, the client explores his/her drinking habits in the first five assignments. The end of the first part consists of a summary of the obtained information and a personalised advice for the second part of the intervention. The second part focuses on how to change the client's drinking behaviour, consisting of six assignments. The intervention offers an aftercare service for six weeks where the client has the possibility to have contact with the therapist once a week. Including the extension, the whole intervention takes around 22 weeks (Postel et al., 2011). An example of mail contact with typical quotes from the client and counsellor can be seen in Table 3. During the program, the client has (usually) the same therapist with whom they communicate via mail contact, and in some cases also have contact by phone or face-to-face. All email contact is established in a secured web application,

[2] Lookatyourdrinking.com is the English equivalent (Postel, 2011).

clients have to log in with a username and password on the website www.alcoholdebaas.nl, thus the client does not use his/her own email address to contact the therapist.

Procedure

Anonymising the client data. To ensure anonymity and security of the data Tactus provided for this study, all words containing personal information in the emails were anonymised with the program Frog and manually checked (Van den Bosch, Busser, Daelemans, Canisius, 2007; Tjong Kim Sang et al., 2019). Frog uses machine learning to identify words of several categories, another program replaces them with the name of the type of the category. This procedure is called *named entity recognition* (Rømcke, 2008; Tjong Kim Sang et al., 2019). In the mails, a name, date, number, location, medical problems, and miscellaneous entities were changed to 'PER', 'DATE', 'NUM', 'LOC', 'PRO', and 'MISC' respectively. An example sentence from a client mail: "PER went home earlier for PER and for PER, but also for PER, because she had her NUM day of exams". When reading the email conversation, it is possible to get the gist of the text, though without reading personal information to make it impossible to track it back to the person. On the places of PER could have been family member or friend names, on the place of NUM it could refer to the 'first day of exams'. After anonymising the data, every client was assigned to a random number to keep the data manageable; answers to questionnaires and email data corresponded to their assigned number. Furthermore, to safeguard the data, it was transferred to an USB-stick and only accessible with a password.

Labelling client data. Mail contact from every client was read in Orange. Orange is a software package that can be used as a tool for data mining. The program works with *widgets* and *pipelines*; an example of a pipeline can be seen in Figure 2. A pipeline exists of *widgets*, which are programs in Orange that have a certain task or function. Every widget can be connected to another widget and form a working line, a pipeline. For this research, the mail contact of one client was loaded in the widget *Mail loader*. The content of the mails could be read in the widget *Corpus viewer*. However, not all mails were sorted in chronological order, thus the widget *Sort emails* was used to sort the mails in chronological order. Many mails contained copied text as a result of replying on previous emails. With the widgets *Mark duplicates* and *Remove marked text*, the duplicate text was first marked and then removed. The widget *Corpus viewer* is used again to read the mail contact without duplicate text in chronological order. The mail contact for every client was read to label the client as 'dropout' or 'completer'. The client was labelled as 'dropout' when he/she had not received a mail from the counsellor with the mail subject *Afsluiting* [wrapping up] and had received an email with the notification that their status had been changed to *niet-actief* [non-active]. The client was labelled as 'completer' when he/she received the email regarding *Afsluiting* and thus finished the whole treatment. Not responding to the aftercare mail or not filling in the last questionnaire was not considered being a dropout because the client had finished the initial treatment.

Inclusion criteria. The inclusion criteria for the sample was that the client needed to have finished the intake questionnaire, followed the intensive treatment of alcoholdebaas.nl, and sent at least one email. Clients were excluded when they followed another variant of the program; the short program which takes less weeks to complete, a self-help module where the client only fills in the alcohol diary and the counsellor gives feedback once after a few weeks, and aftercare contact for people who already completed the program before and do not want to follow the extensive program again. The sample of (Postel et al., 2011) also only included clients who took part in the extensive program.

It was decided to calculate the mean of LIWC dictionaries on the first four emails to be able to predict for dropout in an early stage of the intervention. Less than four mails could result in less predictive value of the text in mails, though more than four mails would decrease the value of *early* indicators. Clients who sent at least one mail were included in the analysis.

Language Inquiry Word Count 2015 (LIWC). The dictionaries are the most important component of LIWC (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007). A dictionary is the collection of words within one category (Chung & Pennebaker, 2011; Tausczik & Pennebaker, 2010), for example: the dictionary *positive* and *negative emotion words* fall under the dictionary of *affective processes*. Example words of *positive emotion* words are *love*, *nice*, and *sweet*. Example words for *negative emotion* words are *hurt*, *ugly*, and *nasty* (Pennebaker, Boyd, Jordan, & Blackburn, 2015). The program analyses each text file, then each target word (the words in the text file) is processed; it proceeds to look for a match within every dictionary for the current word. If a match is present, the dictionary scale(s) for that word is incremented (Pennebaker et al., 2015). The following text is a translated segment of the first mail a client sent:

> "I do have a lot of **hobbies**. I am an emotional person; I **paint**, I **craft** with silver, **felting**, making jewellery from rubber. I also give **workshops** about these crafts and I take part in **art** fairs. I also **exhibit** in galleries. I have many contacts, and I also have good relationships with my family. I have to say that I rescheduled many appointments, because of headaches; meaning I drank too much. My **husband** knows that I drink too much, we have fights occasionally. **He** always says: "that is your problem, I don't want to be involved in this". Now **he** notices that I want to tackle my problem, and **he** has become more benevolent, but **he** will end his support if I drink one glass!"

The green bolded words are words within the dictionary *leisure*. The red bolded words are words within the dictionary *male references*. This segment has more matches with dictionaries, for example *social processes, family,* and *1st person singular*.

Furthermore, a word can belong to more (sub)dictionaries. For example, the word *cried* falls into the dictionary of *sadness*, *negative emotion*, and *verbs* amongst other things. For each text file that is analysed, the output contains approximately 90 variables which matched with all the LIWC dictionaries and contain values for each dictionary (Pennebaker et al., 2015). The fourth and current version, LIWC2015, has an accordingly updated dictionary and software. The Dutch translation of LIWC has an average correlation coefficient of 0.69 between variables when translated automatically. When manually translated, the correlation appeared to be 0.73 (Wissen & Boot, 2017).The LIWC values were created for the first four mails, then one average score on all LIWC dictionaries was calculated for every client. The values were divided with the number of words in a LIWC dictionary. This means that if the value for a dictionary is 0.25, a quarter of the words in the text is in the LIWC dictionary. The LIWC output was combined with the intake questionnaire in Excel. A column was added to the data, indicating how many emails the client had sent. Every client now contained variables with averages of LIWC dictionaries and demographic information. The data was further analysed in SPSS.

Data analysis

Looking into current literature, no other study seemingly focused on using LIWC dictionaries for predicting dropout in email therapy, therefore all LIWC dictionaries were included. Furthermore, only the LIWC averages of the first four mails were used in the analyses. To compare means between the two groups 'dropout' and 'completers' on one variable, an ANOVA was conducted on demographic characteristics and LIWC variables. A logistic regression was performed to predict dropout on demographic characteristics and LIWC dictionaries.

RESULTS

An ANOVA was conducted to investigate possible differences between dropouts and completers on one dependent variable. Significance levels were reported as p <0.05. There were many differences found between the groups, only the variables *gambling*, *depressive thoughts*, and *reason for application* were p >0.05. The results of the variables *gender*, *education*, *smoking*, *drugs*, and *treatment for psychological or emotional problems* indicated significant differences between the groups dropout and completer. However, the averages did not indicate a difference in some cases as a result of nominal measurement levels; therefore, these findings of statistical significance do not appear to have much relevance.

Number of mails. As expected, clients who dropped out of the treatment wrote on average less mails (M = 6.9, SD = 8.9), than clients who completed the treatment (M = 28.4, SD = 13.9). This difference was significant F (2, 988) = 871.4, p <.001. The range of written mails by clients who dropped out ranged between one and 87, for completers this was between three and 116.

Age. Clients who completed the treatment were on average older (M = 47.7, SD = 10.2) than clients who dropped out of the treatment (M = 45.1, SD = 11.4). This difference was significant F (1, 988) = 12.8, p <.001.

Baseline alcohol intake. Clients who dropped out of the treatment had a higher baseline alcohol intake (*M* = 39.0, *SD* = 21.7) than clients who completed the treatment (*M* = 33.1, *SD* = 21.2). This difference was significant F (1, 700) = 13.2, p <.001.

Marital status. There was a significant difference found in marital status between dropouts (*M* = 2.5, *SD* = 1.2) and completers (*M* = 3.0, *SD* = 1.2), F (1, 236) = 4.1, p = .044. Completers were married more often, while dropouts indicated more often to live together with someone.

Summarised, dropouts were on average younger, sent less mails, had a higher baseline of alcohol intake, and were living together with someone but not married; likewise, completers were relatively older, sent more mails, had a lower baseline of alcohol intake, and were more often married. The two groups differed on other demographic characteristics as well, though it was not possible to indicate for those factors as a result of nominal measurement levels.

LIWC dictionaries

Another ANOVA was conducted to investigate whether dropouts and completers differed in use of LIWC dictionaries. There were approximately 70 dictionary matches, 19 of them differed significantly between dropouts and completers. Because the numbers are very small, they are rounded to four decimal places. Two pieces of a client mail are used to illustrate the following results: words are marked in colours with their respective LIWC dictionary. It is important to note that some words can belong to multiple dictionaries.

“Dear [PER], my name is [PER]. I am going to try my best to answer your questions, **though** there are **quite** many! I think the quantity has **stayed** the same last years. **Of course**, I have looked on this website before, **and** I know it is **too** much. I think that it has fluctuated in the **past** from [NUM] to [NUM] glasses per week. **So** I actually **tried** to monitor it for a while **and** that is why the quantity is stable **though** still **too** much. For the onset of my **drinking** habits I need to go back to my **past**. I am from a traditional **family** where smoking and **drinking** is **really** ‘not done’ for a girl. **So** I **did** not **drink** much in my childhood. From my [NUM] year I **drank** occasionally a beer or wine, not **too** much. From [NUM] I **socialised** with ‘other groups’ where drinking (much) **was** the norm.”

“I am already **married** for [NUM] years with the same **man** **and** we have [NUM] **daughters** together, from [NUM] years old **and** almost [NUM] years old. My relationship with the **children** has not **suffered** because of my **drinking** behaviour, I cannot imagine. My eldest **daughter** **and** I sometimes joke about it, that is all. My alcohol behaviour is a topic for me **and** my **husband**. I discuss with him that we **drink** **too** much, both of us (he drinks even more than I do) **and** that I **really** **would** like to quit. Sometimes I ask him to **drink** nothing at all together with me, just to prove that we can do it. He thinks that it does not make sense. He can do it he thinks but I doubt it. Other proposals, as not

drinking on working days, is also not taken into consideration. I think I am quite weak when he asks me in the evening if I want a glass of wine. Then I still say 'yes' **and** the **whole** bottle is empty again."

1st person plural. Clients who completed the treatment wrote more with *we* words (e.g. we, us; *M* = .0028 , *SD* = .0046) than clients who dropped out of the treatment (*M* = .0019, *SD* = .0061), this difference was significant *F* (1, 989) = 5.3, *p* = .021.

3rd person singular. Clients who completed the treatment wrote more with *she* and *he* words (*M* = .0028, *SD* = .0058) than clients who dropped out of the treatment (*M* = .0042, *SD* = .0074), this difference was significant *F* (1, 989) = 11.2, *p* = .001.

Adverb. Clients who completed the treatment wrote more with *adverb* words (e.g. very, really; *M* = .1262, *SD* = .0354) than clients who dropped out of the treatment (*M* = .1190, *SD* = .0475), this difference was significant *F* (1, 989) = 6.2, *p* = .013. *Adverb* words are marked with purple in the texts.

Conjunctions. Clients who completed the treatment wrote more with *conjunction* words (e.g. but, and; *M* = .0772, *SD* = .0235) than clients who dropped out of the treatment (*M* = .0730, *SD* = .0327), this difference was significant *F* (1, 989) = 4.7, *p* = .030. *Conjunction* words are marked with indigo in the texts.

Negative emotion. Clients who completed the treatment wrote more with *negative emotion* words (e.g. hurt, nasty; *M* = .0195, *SD* = .0124) than clients who dropped out of the treatment (*M* = .0167, *SD* = .0173), this difference was significant *F* (1, 989) = 7.5 , *p* = .006.

Family. Clients who completed the treatment wrote more with *family* words (e.g. mother, father; *M* = .0055, *SD* = .0061) than clients who dropped out of the treatment (*M* = .0038, *SD* = .0058), this difference was significant *F* (1, 989) = 17.6, *p* < .001. *Family* words are marked with green in the texts.

Male references. Clients who completed the treatment wrote more with *male reference* words (e.g. he, husband; *M* = .0106, *SD* = .0085) than clients who dropped out of the treatment (*M* = .0087, *SD* = .0122), this difference was significant *F* (1, 989) = 7.1, *p* = .008.

Discrepancies. Clients who dropped out of the treatment wrote more with *discrepancy* words (e.g. should, would; *M* = .0294, *SD* = .0250) than clients who completed the treatment (*M* = .0244, *SD* = .0127), this difference was significant *F* (1, 989) = 12.6, *p* < .001. *Discrepancy* words are marked with orange in the texts.

Perceptual processes. Clients who completed the treatment wrote more with *perceptual processes* words (e.g. look, see, feel; *M* = .0205, *SD* = .0132) than clients who dropped out of the treatment (*M* = .0158, *SD* = .0156), this difference was significant *F* (1, 989) = 23.1, *p* < .001.

Hear. Clients who completed the treatment wrote more with *hear* words (e.g. listen, hear; *M* = .0039, *SD* = .0043) than clients who dropped out of the treatment (*M* = .0031, *SD* = .0057), this difference was significant *F* (1, 989) = 4.8, *p* = .029.

Feel. Clients who completed the treatment wrote more with *feel* words (e.g. feel, touch; *M* = .0079, *SD* = .0076) than clients who dropped out of the treatment (*M* = .0051, *SD* = .0078), this difference was significant *F* (1, 989) = 29.8, *p* < .001.

Biological processes. Clients who completed the treatment wrote more with *biological processes* words (e.g. blood, pain; *M* = .0314, *SD* = .0154) than clients who dropped out of the treatment (*M* = .0278, *SD* = .0234), this difference was significant *F* (1, 989) = 6.7, *p* = .010.

Ingestion. Clients who completed the treatment wrote more with *ingestion* words (e.g. food, drink; *M* = .0195, *SD* = .0123) than clients who dropped out of the treatment (*M* = .0158, *SD* = .0170), this difference was significant *F* (1, 989) = 12.8, *p* < .001. *Ingestion* words are marked with blue in the texts.

Achieve. Clients who completed the treatment wrote more with *achieve* words (e.g. win, success; *M* = .0210, *SD* = .0149) than clients who dropped out of the treatment (*M* = .0187, *SD* = .0178), this difference was significant *F* (1, 989) = 4.4, *p* = .036.

Focus past. Clients who completed the treatment wrote more with *focus past* words (e.g. slept, did; *M* = .0643, *SD* = .0275) than clients who dropped out of the treatment (*M* = .0595, *SD* = .0312), this difference was significant *F* (1, 989) = 5.8, *p* = .016. *Focus past* words are marked with red in the texts.

Focus future. Clients who dropped out of the treatment wrote more with *focus future* words (e.g. will, soon; *M* = .0428, *SD* = .0308) than clients who completed the treatment (*M* = .0361, *SD* = .0159), this difference was significant *F* (1, 989) = 14.4, *p* < .001.

Leisure. Clients who completed the treatment wrote more with *leisure* words (e.g. read, chat; *M* = .0243, *SD* = .0137) than clients who dropped out of the treatment (*M* = .0194, *SD* = .0182), this difference was significant *F* (1, 989) = 19.3, *p* < .001.

Money. Clients who dropped out of the treatment wrote more with *money* words (e.g. pay, cash; *M* = .0049, *SD* = .0116) than clients who completed the treatment (*M* = .0032, *SD* = .0039), this difference was significant *F* (1, 989) = 6.8, *p* = .009.

Death. Clients who completed the treatment wrote more with *death* words (e.g. die, kill; *M* = .0007, *SD* = .0017) than clients who dropped out of the treatment (*M* = .0004, *SD* = .0016), this difference was significant *F* (1, 989) = 6.7, *p* = .010.

Thus, dropouts and completers differed significantly on many LIWC dictionaries, however differences were quite small. See Table 4 for an overview of all ANOVA results with a significant difference between the groups.

Predicting dropout

After analysing differences between clients who dropped out and completed the treatment, it was investigated which variables could be possible predictors for dropout or completion[3] by conducting several logistic regression tests. The baseline for predicting dropout without using any predictors was 50% as a result of only two possible outcomes: dropout or completion. Because an explorative approach was used, two models were built including all demographic variables and LIWC dictionaries. Furthermore, more models were built to predict dropout according to dictionaries, for example the dictionary *Perceptual processes* includes the subdictionaries *see*, *hear,* and *feel*. Only the significant model fits are described. Classification plots for the models are described in Table 5. An overview of Model 1 and 2 are presented in Table 6 and 7.

Model 1, demographic characteristics. The variables *gender, age, education, smoking behaviour, gambling behaviour, depressive symptoms, reason for starting treatment, treated before for psychological and emotional problems,* and *baseline alcohol intake* were used in a logistic regression to predict for dropout. The model explained 12.5% (Nagelkerke R^2) of the variance of the variables and classified 64.6% of cases correctly. The model had a $\chi2$ of 58.0 (df = 21, p < .001).

Being male, having a lower education and a higher baseline of alcohol intake were significant predictors for dropout. On the other hand, having an older age, and not smoking were significant predictors for completion. Notably, people who indicated to have some to moderate depressive symptoms were more likely to drop out of the treatment, however, when they indicated to have no depressive symptoms at all, they were also likely to drop out of the treatment. The variable *depressive symptoms* is therefore not of much relevance for predicting dropout.

Model 2, LIWC dictionaries. All LIWC variables were used in a logistic regression to predict for dropout. The model explained 22.9% (Nagelkerke R^2) of the variances of almost al LIWC dictionaries (not every dictionary had a match within the texts) and classified 68.8% of cases correctly. The model had a $\chi^2$of 181.0 (df = 80, p = .001). There were 17 LIWC variables found to be a significant predictor. The variables which are likely to predict completion of the treatment were *3rd person singular, impersonal pronouns, common adverbs, positive emotion, negative emotion, male references, tentative, biological processes, affiliation, focus past, leisure,* and *informal words*. The more a client wrote words within these dictionaries, the more likely that the client completed the treatment. The variables which are likely to predict dropout were *affective processes, social processes, health, ingestion, focus present, net speak,* and *filler words*. The more the words in the dictionaries were written, the more the client is at risk for not completing the treatment.

Odds Ratio (OR) are used in logistic regressions. As can be seen in Table 6, the OR for some LIWC dictionaries are .000. (Szumilas, 2010) explains that "*the regression coefficient (b1) is the estimated increase in the log odds of the outcome per unit increase in the value of the exposure*". This means that the odds ratio is associated with every unit increase in a variable of a model. When OR = 0,

[3] Predicting for completion is the same as for predicting dropout, as it works the other way around; when a variable predicts for completion when it is more written, it also predicts for dropout if a dictionary is less written.

the variable does not affect the odds of outcome. When OR > 1, the variable is associated with higher odds of outcome; likewise, when OR < 1, the variable is associated with lower odds of outcome. For example, if a LIWC dictionary has a positive regression coefficient with a > 1 OR, it means that for every unit the variable increases, the client is more likely to complete the treatment. Thus, if the OR = 0, the predictive value whether the client is likely to dropout or complete is zero, the variable (which could be a significant fit for the model) is non-predictive, hence the LIWC predictors with .000 values are not of much relevance.

The dictionaries *affective processes*, *social processes*, *health*, *ingestion*, *focus present*, *net speak*, and *filler words* had an OR of .000, which does not indicate whether a client is likely to drop out more or less of the intervention, therefore these significant predictors were not relevant. This model contained all the LIWC dictionaries, now the significant models with individual predictors will be discussed. Predictors with an OR of .000 are not reported.

Model 3, Linguistic dimensions. The variables *total function words, total pronouns, personal pronouns, 1st person singular, 1st person plural, 2nd person, 3rd person singular, 3rd person plural, impersonal pronouns, articles, prepositions, auxiliary verbs, common verbs, conjunctions,* and *negations* were used in a logistic regression to predict for dropout. The model explained 6.0% (Nagelkerke R^2) of the variance of the variables and classified 63.1% of cases correctly. The model had a $\chi 2$ of 44.4 (df = 15, p < .001). The variables which were a significant predictor for completion are *1st person plural*, *3rd person singular, impersonal pronouns,* and *adverbs*.

Model 4, Affective processes. The variables *affective processes, positive emotion words, negative emotion words, anxiety, anger,* and *sad* were used in a logistic regression to predict for dropout. The model explained 2.8% (Nagelkerke R^2) of the variance of the variables and classified 64.2% of cases correctly. The model had a $\chi 2$ of 20.1 (df = 6, p = .003). The variables which were a significant predictor for completion are *positive* and *negative emotion words*.

Model 5, Social processes. The variables *social processes, family, friend, female* and *male references* were used in a logistic regression to predict for dropout. The model explained 3.4% (Nagelkerke R^2) of the variance of the variables and classified 63.4% of cases correctly. The model had a $\chi 2$ of 24.7 (df = 5, p < .001). However, only the variable *male references* was a significant predictor for completion of the treatment.

Model 6, Perceptual processes. The variables *perceptual processes, see, hear,* and *feel* were used in a logistic regression to predict for dropout. The model explained 6.2% (Nagelkerke R^2) of the variance of the variables and classified 63.0% of cases correctly. The model had a $\chi 2$ of 45.8 (df = 4, p < .001). however, only the variable *perceptual processes* was a significant predictor for completion of the treatment.

Model 7, Biological processes. The variables *biological processes, body, health, sexual,* and *ingestion* were used in a logistic regression to predict for dropout. The model explained 3.2% (Nagelkerke R^2) of the variance of the variables and classified 63.2% of cases correctly. The model had a $\chi 2$ of 26.8 (*df* = 5, *p* < .001). However, only the variable *biological processes* was a significant predictor for completion of the treatment.

Model 8, Time orientations. The variables *focus past, focus present*, and *focus future* were used in a logistic regression to predict for dropout. The model explained 3.6% (Nagelkerke R^2) of the variance of the variables and classified 63.6% of cases correctly. The model had a $\chi 2$ of 26.2 (*df* = 3, *p* < .001). However, only the variable *focus past* was a significant predictor for completion of the treatment.

Model 9, Personal concerns. The variables *work, leisure, home, money, religion,* and *death* were used in a logistic regression to predict for dropout. The model explained 4.9% (Nagelkerke R^2) of the variance of the variables and classified 63.0% of cases correctly. The model had a $\chi 2$ of 35.7 (*df* = 6, *p* < .001). The variables *leisure* and *death* were significant predictors for completion of the treatment.

Summarised, predictors for dropout are being male and a younger age, having a lower education level, smoking behaviour, and a higher baseline of alcohol intake prior to the treatment. Writing with less words in the dictionaries of *impersonal pronouns, common adverbs, male references, tentative, biological processes, affiliation, focus past, informal, 1st person plural, 3rd person singular, positive emotion words, negative emotion words, perceptual processes, leisure, and death* increased the chance of dropout.

DISCUSSION

The sample consisted of 990 clients, from which 351 clients completed the treatment. The dropout rate was 64.3%, which illustrates the high dropout rates among web-based treatments and is comparable with the works of Postel (2011) and Postel et al. (2010) who also conducted studies targeting problem drinkers with the intervention from Tactus; dropout rates reported to be 35%, 54%, and 55% respectively. However, this dropout rate is still lower compared to the work of Linke et al. (2007) who experienced 84.5% dropout rate. Still, compared with other comparable studies, the dropout rate of 64.3% remains high, and further illustrates the fact that online therapy may work for many, but not for everyone (Newman et al., 2017). For the studies of Postel et al. (2010, 2011) and this study, the same criteria was used to determine when a client was dropout or not, when the client had not completed all the exercises of the structured treatment. However, the intervention of Postel et al. (2010, 2011) was offered for free, and thus those clients did not have to worry whether they could pay for the treatment. It was observed several times in the mail contact of clients that the first few mails contained some questions of the client, asking if the insurance did cover the intervention. When the counsellor replied that this was not the case, some clients then thanked the counsellor for his/her time and stated that they could not afford the intervention. Some clients did not reply at all. A significant difference between the groups dropout and completer was observed using an ANOVA in the LIWC dictionary *money* which was also a significant predictor for dropout*;* people who dropped out of the intervention used more financial related words than completers. It is likely that money played a major role in the reasons to dropout for this study which was not present in the studies of Postel (2011) and Postel et al. (2010, 2011) and may have caused a higher dropout rate.

Predictors for dropout

Demographic characteristics which were significant predictors for dropout were being male, having a younger age, lower education levels, smoking behaviour, and a higher baseline of alcohol intake. On the other hand, older women with a higher education who do not smoke and have a lower alcohol intake prior to the treatment were most likely to complete the intervention successfully. These findings are in line with the works of (Karyotaki et al., 2015; Postel, 2011; Postel et al., 2011; Riper et al., 2008). Smoking behaviour was not found in these works to be a predictor for dropout, though smoking is highly correlated with alcohol use, and treatment outcomes for addiction interventions are worse when a drinker also smokes compared to drinkers who do not smoke (Drobes, 2002). Correlation between alcohol use and smoking can also be observed in the sample as more than half of the respondents indicated to smoke occasionally or daily. Even though no predictive value was found for the variable *treatment before for psychological or emotional problems,* comorbidity of psychological and emotional problems were present with clients.

Linguistic predictors. Researchers have conducted research to predict on dropout with demographic factors, though it has not yet been considered to look at linguistic predictors for dropout in online treatment, targeted to find predictors in an early stage. A logistic regression test in the current study revealed that writing with less words in the dictionaries of *1st person plural, 3rd person singular, impersonal pronouns*, *common adverbs, positive emotion*, *negative emotion, male references*, *tentative*, *perceptual processes*, *biological processes*, *affiliation*, *focus past*, *leisure*, *death* and *informal words* increased the chance of dropout; likewise, when clients wrote more with words in these dictionaries, the chance of completion of the treatment increased.

When the dictionaries *impersonal pronouns, common adverbs*, *tentative, perceptual processes, focus past*, *positive emotion words* and *negative emotion words* are written more often, it might be an indication that some clients (possibly completers) are able to reflect and describe their thoughts more often than other clients (possibly dropouts). The ability and willingness to reflect on past and current situations and behaviour plays a major part in learning and changing behaviour (Higgins, 2011). If a client is able to reflect and learn new behaviour (i.e. getting control over drinking behaviour), it might motivate the client even more to complete the program. In addition, the results that emotion words can predict for completion is partly in line with Smink et al. (2019) who state that a high rate of positive emotion words is associated with positive therapy outcome, however including a low rate of negative emotion words and an increasing number of cognitive words throughout the treatment. Perhaps negative emotion words are also needed in order to reflect on the current or a past situation, in combination with the aforementioned LIWC dictionaries it might work in favour of the clients, instead of only looking at positive and negative emotion words alone. Furthermore, *cognitive processes* words did not predict for dropout in the current study, and due to the analysis design of this study, it is impossible to see an increase of *cognitive processes* words in the intervention because only the first four mails were used.

Furthermore, when the client wrote often with the dictionaries *1st person plural*, *3rd person singular*, *male references,* and *affiliation*, it may indicate a strong social network with social support. Social support is proven to be beneficial for mental health (Cobb, 1976); there is a positive correlation with social interaction with friends and family and treatment adherence (Spence, March, & Donovan, 2019). Thus, if a client is supported by people around him/her, he/she might write more about that during the treatment, while at the same time it works as a facilitator for completion. This could explain the dictionaries aforementioned being a predictor for completion.

Lastly, the dictionaries of *leisure, death* and *informal* predicting for completion might mean that if the client shares more personal details about his/her life, the client feels safe to disclose which may benefit the working relationship between counsellor and client. If a client does not perceive an enjoyable working relationship, he/she might disclose less or not at all and quit the treatment as a result of poor connection between the counsellor and client. Summarised, several linguistic and non-linguistic predictors were found to predict for dropout. It is important to note that the significant predictive values were small, and for this reason may not have too much statistical relevance. The next paragraph will briefly discuss the differences between dropouts and completers.

Differences between dropouts and completers. Dropouts and completers differed in the number of mails they sent and the baseline of alcohol intake. Differences were also found for gender, age, education level, marital status, drugs, and treatment for psychological and emotional problems. Regarding linguistic differences, clients who dropped out of the intervention wrote more with words in the LIWC dictionaries *discrepancies, focus future,* and *money*. Thinking about the future can be beneficial in terms of setting goals and planning how to reach them (Draghici, 2015). Moreover, thinking about and describing positive possible selves is proven to work as a protective against risk behaviours (Lee et al., 2016), such as excessive alcohol use. However, *discrepancy* words, such as ‘should’ and ‘would’ combined with *focus future* words could indicate that clients who dropped out of the intervention often described their ‘possible selves’ in a discrepant, unsure way, which could be a facilitating risk for dropout. Clients who completed the treatment wrote more often with words in the dictionaries of *1st person plural*, *3rd person singular, adverb, conjunctions, negative emotion*, *family, male references, perceptual processes*, *hear, feel, biological processes*, *ingestion*, *achieve*, *focus past*, and *leisure*. This might be a result of reflecting more often and having a social support system. Differences for the variables gender, age, education, marital status, smoking, drugs use, and whether the clients had received treatment before for psychological and emotional problems, and several LIWC dictionaries were statistically significant, however, differences were quite small which may indicate that the differences might not be of too much relevance.

Strengths and limitations

This research took an explorative approach to search for predictors in mail contact of online counselling using NLP. It is important to highlight that, to the best of my knowledge, there is no existing literature yet that covers prediction of dropout in an early stage of a web-based treatment for problem drinkers especially focused on using NLP in email therapy to predict for dropout (or completion). This can be seen as a strength, as this research can be used to build upon further, as well as a limitation. It is recommended to carry out more research on predictive factors of web-based treatment for problem drinkers, as much of it is still unknown, and thus incomparable with other studies. It is of importance to see whether certain factors are linked to dropout, and which LIWC dictionaries (or any other NLP program) can predict and give more insight of dropout in email therapy and will create the opportunity for counsellors to adapt to the client.

Another main limitation was the choice to only analyse the mean of the first four mails. Choosing the first four mails was in retrospect perhaps not the optimal choice, as it provides less information if the first four mails were analysed individually. The reason to analyse only the first four mails was a result of outsourcing the tasks of anonymising and processing the mails in LIWC; obtaining the data took longer than was expected, and it was therefore considered more feasible to take the mean of LIWC dictionaries of the first four mails. Therefore, it was not possible to look at the progress of the intervention and predict for dropout each mail per client.

Moreover, the data Tactus provided did not come with labels whether the client completed the treatment. With the given options of software, it was not possible to process multiple clients at the same time, thus the client's email correspondence with the counsellor were manually read. Besides that it took much time to label the clients as a result of reading all client conversations, manual coding is prone to human error which can have caused mistakes in labelling the clients. It could have been useful and perhaps more precise if the counsellors had indicated whether the client stopped the treatment and whether he/she had completed all assignments, linked to their ID number.

Future recommendations. A recommendation for future research could be to look at the first four mails individually instead of calculating a mean for the LIWC dictionaries. This way it would be possible to see if there is any in/decrease in LIWC variable values per mail. Moreover, it is possible that an average of four mails was not the optimal choice to predict for dropout. Riper et al. (2008) found different predictive power after 6- and 12-months follow-up, suggesting that "different factors operate at different points during the post-intervention period". This can be the case as well during the treatment that there are certain 'peak moments' when clients drop out more than other points of time during the intervention.

Another recommendation is to include more intake questions from the intake questionnaire to broaden the scope and make it possible to compare with other literature. The variables comorbid disorders/issues, and the degree social support are considered to be insightful to research whether they can predict for dropout. For example, as was mentioned by Jeanblanc (2015) other psychological problems are often related to excessive alcohol use. It might give more insight if the existence of psychological issues creates more risk of dropout and investigate if different disorders have differences in the risk of predicting dropout. Furthermore, support in the environment could be associated with the LIWC dictionaries regarding personal pronouns, and *female/male references*, as some of those LIWC dictionaries were a predictor for completion. It should be assessed if perceived social support does indeed increase the likelihood of completion.

Dropout types. When the client perceives improvement, he/she may decide to quit treatment, but in reality is not ready yet to leave. As a result, the client is more likely to experience poorer therapy outcome (de Weert-Van Oene, Schippers, de Jong, & Schrijvers, 2001). However, dropping out of the treatment does not always mean poorer outcome, as Stark (1992) suggests that there are different groups of dropouts. A client could quit the treatment because he/she only need a little nudge into the right direction, or a client could quit the intervention though still continues to use the alcohol diary to keep track of alcohol intake. Furthermore, this study distinguished only one group of dropout, according to Eysenbach's Law of Attrition called *dropout attrition* (Eysenbach, 2005); the clients did not finish the treatment and did not fill in the follow-up questionnaire. However, there were various reasons for quitting the treatment, for example "I changed my mind", "I don't feel motivated to continue", or "I don't have the money and my insurance does not cover for it". These reasons for discontinuing treatment have been researched before (Ball et al., 2006), though it could indicate different dropout groups, which were not considered in this research. It might give more insight if those factors are included as well when predicting for dropout.

Conclusion

Existing online treatment for problem drinkers are effective, though characterised with high dropout rates. It is important to note that premature quitting of the treatment is likely to worsen symptoms and may cause symptoms to become chronic. Therefore, it is needed to look into individual client differences, as online therapy works for many but not for everyone, which is most probably the main contributor to high dropout rates. The option to tailor the treatment to an individual client could reduce dropout rates of online interventions and prevent worsening of the client's symptoms. This research took an exploratory approach, aimed to investigate what linguistic and non-linguistic factors could predict for dropout in a web-based intervention for problem drinkers. Non-linguistic predictors for dropout were being a male, younger age, smoking behaviour and a lower education level. Linguistic predictors for dropout were the LIWC dictionaries *3rd person singular, impersonal pronouns, common adverbs, positive emotion, negative emotion, male references, tentative, biological processes,* *affiliation, focus past, leisure, and informal words.* However, it is important to note that even though statistically significant predictors and differences between groups were found, the differences and the predictive values were small. This indicates that the statistically significant results might not be too much of practical relevance.

The main limitation of the study was only including the average scores of LIWC dictionaries of the first four mails. Recommendations are to look at the first few mails individually sent by the client to analyse the progress in LIWC dictionaries, and to analyse whether LIWC dictionaries differ significantly per mail; it might give more insight in knowing when clients are likely to dropout. Moreover, it could widen the scope of the research focus if more non-linguistic factors were included predict for dropout, which would make it easier to compare with more studies. In addition, research for different dropout types could give more insight in the high dropout rates of online therapy. All in all, the current research has laid a groundwork for future studies to build upon by taking an explorative approach to predict for dropout with linguistic features.

TABLES AND FIGURES

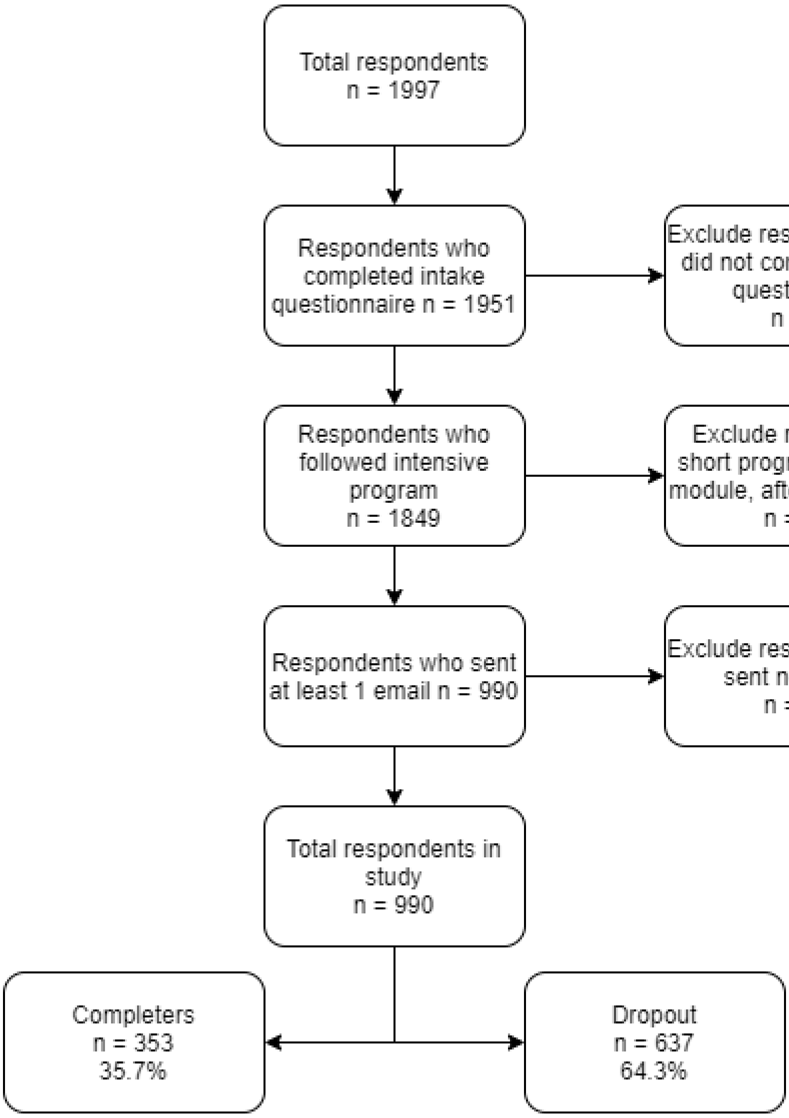

Figure 1. Flowchart of excluded respondents. *This figure shows a flowchart of excluded respondents who did not complete the intake questionnaire, wrote no mails and did not follow the extensive version of the intervention*.

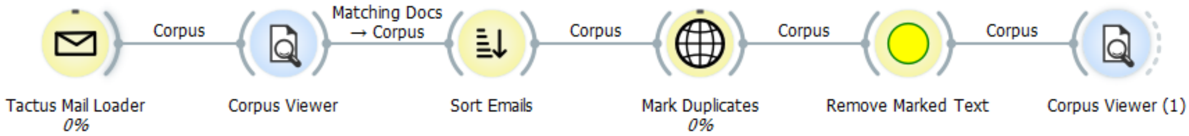

Figure 2. Example Orange pipeline. *This figure shows a pipeline created in Orange to read the email conversation between the client and counsellor.*

Table 1

Overview demographic characteristics intake questionnaire.

		n = 990	%
Gender	Man	439	44.3
	Woman	551	55.7
Nationality	Dutch	239	24.1
	Belgian	3	0.3
	Other	3	0.3
	No answer given	745	75.3
Education[4]	Primary school	11	1.1
	Lower Vocational Education	143	14.4
	School of higher general secondary education/ Pre-university education	117	11.8
	Intermediate Vocational Education	227	22.9
	Higher Vocational Education	322	32.5
	University	135	13.6
	No answer given	35	3.5
Marital status	Unmarried	59	6.0
	Living together	38	3.8
	Married	93	9.4
	Divorced	39	3.9
	Widow/widower	4	0.4
	Living apart together	5	05
	No answer given	752	76.0
Label	Dropout	637	64.3
	Completer	353	35.7
Reason for starting treatment	I think I am drinking too much	762	77.0
	I want advice about alcohol usage	38	3.8
	Something negative happened, I want to change something about my drinking behaviour	91	9.2
	Others think I am drinking too much	28	2.8
	No answer given	71	7.2

[4] Education levels are according to the Dutch schooling system.

Smoking	No	337	34.0
	Yes, occasionally	58	5.9
	Yes, every day	350	35.4
	No answer given	245	24.7
Drugs	No	667	67.4
	Yes	78	7.9
	No answer given	245	24.7
Gambling	No	717	72.4
	Yes, less than once per week	21	2.1
	Yes, once to three times per week	7	0.7
	No answer given	245	24.7
Depressive feelings	Never	85	8.6
	Almost never	110	11.1
	Sometimes	267	27.0
	Often	203	20.5
	All the time	42	4.2
	No answer given	283	28.6
Ever treated before for psycho-logical/emotional problems	No	371	37.5
	Yes	466	47.1
	Yes, in the last month	19	1.9
	Yes, in the last year	39	3.9
	Yes, (no timeframe indicated)	95	9.6
	No answer given	371	37.5

Table 2

Overview demographic characteristics intake questionnaire split to drop-out and completer.

	Drop-out		Completer	
Gender	*N*	%	*N*	%
Men	303	69.0	136	31.0
Woman	334	60.6	217	39.4
Nationality				
Dutch	207	86.6	32	13.3
Belgian	3	100.0	0	0.0
Other	3	100.0	0	0.0
Education				
Primary	7	63.6	4	36.4
Lower vocational education	102	71.3	41	28.7
School of higher general secondary education/ Pre-university education	82	70.1	35	29.9
Intermediate vocational education	157	69.2	70	30.8
Higher vocational education	197	61.2	125	38.8
University	75	55.6	60	44.4
Marital status				
Unmarried	54	91.5	5	8.5
Living together	35	92.1	3	7.9
Married	79	84.9	14	15.1
Divorced	31	79.5	8	20.5
Widow/widower	3	75.0	1	25.0
Living apart together	4	80.0	1	20.0
Reason for treatment				
I think I am drinking too much	493	64.7	269	35.3
I want advice about alcohol usage	24	63.2	14	36.8
Something negative happened, I want to change something about my drinking behaviour	60	65.9	31	34.1
Others think I am drinking too much	21	75.0	7	25.0
Smoking				
No	166	49.3	171	50.7
Yes, occasionally	36	62.1	22	37.9
Yes, every day	222	63.4	128	36.6
Drugs				
No	370	55.5	297	44.5
Yes	54	69.2	24	30.8

Gambling					
	No	404	56.3	313	43.7
	Yes, less than once per week	14	66.7	7	33.3
	Yes, once to three times per week	6	85.7	1	14.3
Depressive symptoms					
	Never	51	60.0	34	40.0
	Almost never	62	56.4	48	43.6
	Sometimes	155	58.1	112	41.9
	Often	125	61.6	78	38.4
	All the time	17	40.5	25	59.5
Ever treated before for psychological/emotional problems					
	No	246	66.3	125	33.7
	Yes	258	55.4	208	44.6
	Yes, in the last month	19	100.0	0	0.0
	Yes, in the last year	34	87.2	5	12.8
	Yes, (no timeframe indicated)	80	84.2	15	15.8

Table 3

Overview of the intervention, with some quotes from each of the weeks.

Week	*Content / description of the week*	*Counsellor*	*Client*
1	Starting	Dear participant, Thank you for your application. I read your intake form; you are a suitable candidate to start the web-based treatment. Within [number] working days you will get a counsellor assigned to you. He/she will go through your intake form extensively. With kind regards,	Thank you for your reply. I did not have time yet to reply, but I read your questions and I will try to answer them as soon as possible. I already filled in the alcohol diary, and I will to that tonight as well.
2	Pros and Cons of alcohol consumption	To gain a better understanding of your situation, I would like to ask you a few more questions.	
3	Keeping up with the alcohol diary	During this treatment it is important to keep up using the diary every day. That requires some discipline. The easiest way to keep up is to write on the same time every day, for example after dinner.	I started drinking secretly because I was sleeping poorly. I lived together with my ex-boyfriend who snored a lot. Furthermore, he put high pressure on our relationship, I was extremely miserable.
4	Pros and cons of alcohol consumption		Here the top pros. It tastes good (I am a wine lover, there is no other drink who can replace this). I stop overthinking, being consciously aware the whole day is very hard for me. I become tranquil, it relaxes me, certainly after a hard day of work, other things like sports don’t have that effect on me. The top cons. I get forgetful, it can also relate to my age, and my

			achievements at work decrease. I become dependent, I think it's annoying and weak to be dependent on alcohol. My stamina, I get physical complaints, high blood pressure, sometimes stomach/liver complaints and an increase in weight.
5	Analysing situations	In this part of the treatment you will explore your drinking habits. For the next time, use your alcohol diary to analyse situation. Write down [number] situations in more detail when you were craving or had drunk alcohol. Try to give answers to the following questions.	I am not too well at the moment, I drink daily and more than is good for me. The reward is high; peace, not thinking, sleeping. If I think back to my alcohol-free period I only remember being depressed, which really demotivates me to stop drinking again. I sometimes succeed in drinking less when I really have to be clear headed the next day, for example when I have to attend a meeting. But then the pressure is gone, and I start again (thus after the meeting). My forgetfulness and information processing time has increased. I forgot my password for this website for example, fortunately I saved it somewhere.
6	The numbers tell the tale	I want to ask you to compare your alcohol intake with the low alcohol intake guidelines. The rule of thumb is a maximum of [number] glasses per week. Preferably not more than [number] glasses a day.	It was worse in the past, but over time I became mild towards myself. I am allowed to fail and not everything has to be perfect. […] You ask me if I know what responsible drinking is. Dear PER, I know that already, I think I have read almost everything related to alcohol intake and its consequences.

7	Setting goals	You are going to formulate a concrete and feasible goal with help of the exercise attached to this mail. As in part [number] you will receive multiple exercises. I will also send you informative texts regarding your drinking habits. The texts are called ‘something to think about’.	I want to quit. Controlled drinking is still taking too much energy. With the psychologist I agreed to stop when autumn is over and see what happens. Will I get depressed again and sleep poorly, or not? And what can we then do about it?
8	Breaking habits	It seems that you sometimes drink to experience fewer negative feelings, for example when you are angry, agitated or feeling stressed. Those are risk feelings you described.	I still have a few days to think about it. I will decide DATE what I am going to do. I think decrease my alcohol intake in a week, otherwise quitting. Again? Weak, weak, weak.
9	Thinking differently	Now follows step [number]. You are going to find and explore your risk/non-helping thoughts and change them in more positive, helping thoughts. Thoughts are differentiated in helping and non-helping thoughts	I admire the positive way you summarise, reply, paraphrase, without being dismissive or judgemental. Negative things are still turned into something positive. I enjoy reading your mails and it helps me to keep going. It keeps me alert to analyse my situations. […] That is also a frightening though to me, giving up my structured life of working and drinking. What will replace it when I give up something so secure and structured? The only thing I can think of is getting more feelings of self-worth, something I actually achieve and maintain.

10	Doing differently	Describe the moments when you dealt with moments you were craving alcohol, without drinking alcohol. Remember to describe the moment and the helping thoughts and behaviour.	I already looked at the assignment. It is about coming up with excuses to drink, but I will not drink, so I thought back to the “past”.
11	Something to think about, low intake	You are doing so well, thumbs up! Attached you find a text with the subject ‘tips for drinking moderately’, perhaps you will find some helpful tips!	Thank you for your reply and the thumbs up! When I am reading these tips, I realise that I am already doing many of them.
12	Decisions	With every decision there is a possibility to step out of the situation. Read this carefully and try to imagine what you would do in different decision situations. Write down and describe the moments of success, moments when you were craving alcohol but did not drink.	Also, the weekend went by quite well. It was a nice weekend away, unfortunately not that nice weather in the morning, though a walk on the beach with an umbrella is doable. Met NUM people in the hotel and spoken a lot with them while drinking alcohol and eating snacks. Of course, they drank, I drank glasses of coca cola, shame yes, because it did not taste that great. Later that day I even drank tea.
13	Action plan	In this last exercise you make a short overview of your personal reason to start treatment, your helping thoughts and behaviour. This is a summary of what you have learned.	The party of my mother was a success. I laughed often and I was funny if I may say so myself. Alcohol was not a problem at all, I drank water and soft drinks, the others drank wine. I did sniff my sister’s liquor, but I did not even think it smelled good! […] It’s a foreign idea to me to pass the Holidays without drinking. I do not look forward to it.

15	Wrapping up	See this treatment as a starting point for behaviour change. You made a great start and you can have faith in that you are capable to reach your goal.	Hello PER, I am reading this last message with tears in my eyes. This is because I am finishing something, and I have to do things on my own from now on. But also, because I achieved something, which is something I thought to be impossible. Do not forget your own contribution as well! You were patient, understanding, empathetic, just, and almost loving approach and guidance. I did not expect that at all in an online treatment. It touched me. Maybe this is also something I have been missing in my life, love and understanding. […] I want to follow the aftercare trajectory, certainly with the Holidays around the corner.

Table 4

Results ANOVA demographic characteristics and LIWC variables split on dropout label.

	Dropout		Completer			
Demographic characteristic	M	SD	M	SD	*F*	*p*
Number of mails	6.9	8.9	28.4	13.9	871.4	<.001
Gender[5]	1.5	0.5	1.61	0.5	7.6	.006
Age	45.5	11.4	47.7	10.2	12.8	<.001
Education[6]	4.1	1.3	4.3	1.3	-2.6	.009
Marital status[7]	2.5	1.2	3.0	1.2	4.1	.044
Smoking[8]	2.1	0.9	1.9	1.0	14.3	<.001
Drugs[9]	1.1	0.3	1.1	0.3	5.4	.020
Treatment[10] psychological/emotion-al problems	2.1	1.3	1.8	0.9	17.3	<.001
Baseline alcohol	39.0	21.7	33.1	21.2	13.2	<.001
LIWC variable						
1st person plural	.0019	.0061	.0028	.0046	5.3	.021
3rd person singular	.0042	.0074	.0057	.0058	11.2	.001
Adverb	.1190	.0475	.1262	.0354	6.2	.013
Conjunctions	.0730	.0326	.0772	.0235	4.7	.030
Negative emotion	.0167	.0173	.0195	.0124	7.5	.006
Family	.0038	.0058	.0055	.0061	17.6	<.001
Male references	.0087	.0122	.0106	.0085	7.1	.008
Discrepancies	.0294	.0250	.0244	.0127	12.6	<.001
Perceptual processes	.0158	.0156	.0205	.0132	23.1	<.001
Hear	.0031	.0057	.0039	.0043	4.8	.029
Feel	.0051	.0078	.0079	.0076	29.8	<.001
Biological processes	.0278	.0234	.0314	.0154	6.7	.010
Ingestion	.0158	.0170	.0195	.0123	12.8	<.001

[5] Labels: 1 Man, 2 woman
[6] 1 Primary school labelled, 2 Lower vocational education labelled, 3 School of higher general secondary education/Pre-university education labelled, 4 Intermediate vocational education labelled, 5 Higher vocational education labelled, 6 University
[7] 1 unmarried, 2 living together, 3 married, 4 divorced, 5 widow/widower, 6 living apart together
[8] 1 no; 2 yes, occasionally; 3 yes, every day
[9] 1 no, 2 yes
[10] 1 no; 2 yes; 3 yes, in the last month; 4 yes, in the last year; 5 yes (no timeframe indicated)

Achieve	.0187	.0178	.0210	.0149	4.4	.036
Focus past	.0595	.0312	.0643	.0275	5.8	.016
Focus future	.0428	.0308	.0361	.0159	14.4	<.001
Leisure	.0194	.0182	.0243	.0137	19.3	<.001
Money	.0032	.0039	.0049	.0116	6.8	.009
Death	.0004	.0016	.0007	.0017	6.8	.010

Note. Only significant findings are presented, other LIWC variables had a significance level of p >.05 or there were no matches found in the text.

Table 5
Classification plots.

Model 0[11]		Observed		
		Dropout	Completer	Percentage correct
Predicted	Dropout	357	0	100.0
	Completer	242	0	0
	Overall percentage			59.6
Model 1, demographic characteristics[12]		Observed		
		Dropout	Completer	Percentage correct
Predicted	Dropout	294	63	82.4
	Completer	149	93	38.4
	Overall percentage			64.6
Model 2, LIWC variables[13]		Observed		
		Dropout	Completer	Percentage correct
Predicted	Dropout	542	95	85.1
	Completer	214	139	39.4
	Overall percentage			68.8

[11] No variables were used to predict for dropout.
[12] Variables gender, age, education, marital status, depressive symptoms, reason for starting treatment, smoking, gambling, drugs, treated before for psychological or emotional problems, and baseline alcohol intake were used to predict for dropout.
[13] All LIWC variables were used to predict for dropout.

Table 6

Logistic regression, Model 1 demographic characteristics.

Variable	Model 1	
	B	OR
Constant	-.777	.460
Gender	* -.467	1.022
Age	* .022	1.022
Education		
Primary school	.332	1.393
Lower Vocational Education	-.652	.521
School of higher general secondary education/ Pre-university education	-605	.546
Intermediate Vocational Education	*-.678	.507
Higher Vocational Education	-.395	.716
Smoking		
Never	*.460	1.584
Occasionally	-.031	.969
Gambling		
Never	.567	1.762
Yes, less than once per week	.080	1.084
Drugs		
No	.054	1.056
Depressive symptoms		
Never	*-1.097	.334
Almost never	-.883	.414
Sometimes	*-1.142	.319
Often	-1.130	.323
Reason starting treatment		
I think I am drinking too much	.817	2.264
I want advice about alcohol usage	.678	1.970
Something negative happened, I want to change something about my drinking behaviour	1.052	2.864
Nagelkerke R^2	12.5%	
χ^2	58.0, df = 21, p < .001	

Note. *p < .05. Gender compared to men.

Table 7
Logistic regression, model 2 with all LIWC dictionaries.

Variable	Model 2	
	B	OR
Constant	*-.812	.444
Linguistic dimensions		
Total function words	-.833	.435
Total pronouns	-11.750	.000
Personal pronouns	-18.671	.000
1st person singular	32.493	129215539317867.860
1st person plural	58.036	160192996454489280000000 00.000
2nd person	46.743	199564034107637960000.000
3rd person singular	*66.234	582089544950451950000000 00000.000
3rd personal plural	-5.238	.005
Impersonal pronouns	*16.658	17152978.628
Articles	8.978	7926.919
Prepositions	-1.203	.300
Auxiliary verbs	-.421	.656
Common adverbs	*12.091	178309.445
Conjunctions	3.768	43.287
Negations	-2.600	.074
Other grammar		
Common verbs	3.954	52.169
Common adjectives	-2.713	.066
Comparisons	-2.036	.131
Interrogatives	-13.231	.000
Numbers	-2.231	.107
Quantifiers	-2.005	.135
Affective processes	*-194.879	.000
Positive emotion	*198.294	1.312E+86
Negative emotion	*207.662	1.537E+90
Anxiety	7.324	1516.987
Anger	-4.861	.008
Sad	-33.689	.000
Social processes	*-12.697	.000
Family	21.774	2861150292.023
Friend	-17.132	.000
Female references	-8.157	.000
Male references	*24.979	70530702836.040
Cognitive processes	-14.956	.000
Insight	16.367	12826898.443

Causation	8.684	5907.559
Discrepancy	4.533	93.056
Tentative	*16.976	23583616.920
Certainty	9.413	12249.127
Differ	9.784	17740.822
Perceptual processes	28.689	2881554814313.442
See	-25.548	.000
Hear	-10.060	.000
Feel	16.906	21998693.333
Biological processes	*45.578	62247627084100180000.000
Body	-42.834	.000
Health	*-57.898	.000
Sexual	33.618	398193708307667.560
Ingest	*-57.369	.000
Drives	-7.828	.000
Affiliation	*25.946	185360442803.104
Achieve	12.267	212533.606
Power	6.166	476.485
Reward	-6.775	.001
Risk	5.051	156.162
Time orientation		
Focus past	*9.596	14706.116
Focus present	*-12.164	.000
Focus future	-9.963	.000
Relativity	3.183	24.128
Motion	-3.828	.022
Space	-1.035	.355
Time	-3.830	.022
Personal concerns		
Work	-10.217	.000
Leisure	*18.606	120375851.267
Home	-8.410	.000
Money	-19.451	.000
Religion	30.211	13192147710490.180
Death	60.966	300108907619342060000000 000.000
Informal language	*21.888	3204777349.998
Swear	-97.778	.000
Netspeak	*-26.705	.000
Assent	5.111	165.912
Nonfluencies	-253.158	.000

Fillers	*-24.441	.000
Nagelkerke R^2	22.9%	
χ^2	181.03, $df = 73$, $p < .001$	

REFERENCES

Amichai-Hamburger, Y., Klomek, A. B., Friedman, D., Zuckerman, O., & Shani-Sherman, T. (2014). The future of online therapy. Computers in Human Behavior, 41, 288–294. https://doi.org/10.1016/j.chb.2014.09.016

Andersson, E., Enander, J., Andrén, P., Hedman, E., Ljótsson, B., Hursti, T., … Rück, C. (2012). Internet-based cognitive behaviour therapy for obsessive-compulsive disorder: A randomized controlled trial. Psychological Medicine, 42(10), 2193–2203. https://doi.org/10.1017/S0033291712000244

Andrade, A. L. M., de Lacerda, R. B., Gomide, H. P., Ronzani, T. M., Sartes, L. M. A., Martins, L. F., … Schaub, M. P. (2016). Web-based self-help intervention reduces alcohol consumption in both heavy-drinking and dependent alcohol users: A pilot study. Addictive Behaviors, 63, 63–71. https://doi.org/10.1016/j.addbeh.2016.06.027

Baingana, F., Becker, A. E., Pringle, B., Medical, D., Medicine, S., Health, G. M., & Boulevard, E. (2019). HHS Public Access. 527(7578). https://doi.org/10.1038/nature16032

Ball, S. A., Carroll, K. M., Canning-Ball, M., & Rounsaville, B. J. (2006). Reasons for dropout from drug abuse treatment: Symptoms, personality, and motivation. Addictive Behaviors, 31(2), 320–330. https://doi.org/10.1016/j.addbeh.2005.05.013

Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. (2008). A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. Journal of Technology in Human Services, 26(2–4), 109–160. https://doi.org/10.1080/15228830802094429

Barak, A., Klein, B., & Proudfoot, J. G. (2009). Defining internet-supported therapeutic interventions. *Annals of Behavioral Medicine*, *38*(1), 4–17. https://doi.org/10.1007/s12160-009-9130-7

Basit, T. N. (2003). Manual or electronic? The role of coding in qualitative data analysis. Educational Research, 45(2), 143–154. https://doi.org/10.1080/0013188032000133548

Bennett, K., Bennett, A. J., & Griffiths, K. M. (2010). Security considerations for e-mental health interventions. Journal of Medical Internet Research, 12(5). https://doi.org/10.2196/jmir.1468

Bloom, J. W. (1998). The ethical practice of web counseling. British Journal of Guidance and Counseling, 26(1), 53-59.

Chester, A., & Glass, C. A. (2006). Online counselling: A descriptive analysis of therapy services on the Internet. British Journal of Guidance and Counselling, 34(2), 145–160. https://doi.org/10.1080/03069880600583170

Christensen, H., Griffiths, K. M., & Farrer, L. (2009). Adherence in internet interventions for anxiety and depression. Journal of Medical Internet Research, 11(2), 1–17. https://doi.org/10.2196/jmir.1194

Chung, C. K., & Pennebaker, J. W. (2011). The psychological functions of function words. Social Communication, (February), 343–359. https://doi.org/10.4324/9780203837702

Chung, C. K., & Pennebaker, J. W. (2013). Linguistic Inquiry and Word Count (LIWC). Applied Natural Language Processing, (January 1999), 206–229. https://doi.org/10.4018/978-1-60960-741-8.ch012

Cohen, K. B. (2013). Biomedical Natural Language Processing and Text Mining. In Methods in Biomedical Informatics: A Pragmatic Approach. https://doi.org/10.1016/B978-0-12-401678-1.00006-3

Day, E., Copello, A., & Hull, M. (2015). Assessment and management of alcohol use disorders. *BMJ (Online)*, *350*(March). https://doi.org/10.1136/bmj.h715

Draghici, A. (2020). Future or Present? - Is It Better to Be Future-Oriented or Present-Oriented?. Retrieved from https://explorable.com/e/future-or-present

Dreisbach, C., Koleck, T. A., Bourne, P. E., & Bakken, S. (2019). A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. International Journal of Medical Informatics, 125(December 2018), 37–46. https://doi.org/10.1016/j.ijmedinf.2019.02.008

Drobes, J. (2002). Concurrent Alcohol and Tobacco Dependence Mechanisms and Treatment. National Institute of Alcohol Abuse and Alcoholism. Retrieved from: https: //pubs.niaaa.nih.gov/publications/arh26-2/136-142.htm

Eysenbach, G. (2005). The law of attrition. Journal of Medical Internet Research, 7(1), 1–11. https://doi.org/10.2196/jmir.7.1.e11

Fernández-Álvarez, J., Díaz-García, A., González-Robles, A., Baños, R., García-Palacios, A., & Botella, C. (2017). Dropping out of a transdiagnostic online intervention: A qualitative analysis of client's experiences. Internet Interventions, 10(August), 29–38. https://doi.org/10.1016/j.invent.2017.09.001

Formánek, T., Kagström, A., Cermakova, P., Csémy, L., Mladá, K., & Winkler, P. (2019). Prevalence of mental disorders and associated disability: Results from the cross-sectional CZEch mental health Study (CZEMS). European Psychiatry, 60, 1–6. https://doi.org/10.1016/j.eurpsy.2019.05.001

Francis-Smith, C. (2014). Email Counselling and the Therapeutic Relationship: a Grounded Theory Analysis of Therapists' Experiences. (December). Retrieved from http://eprints.uwe.ac.uk/24554/7/Thesis amended for repository.pdf

Gainsbury, S., & Blaszczynski, A. (2011). A systematic review of Internet-based therapy for the treatment of addictions. Clinical Psychology Review, 31(3), 490–498. https://doi.org/10.1016/j.cpr.2010.11.007

Grant, B. F., Goldstein, R. B., Saha, T. D., Patricia Chou, S., Jung, J., Zhang, H., … Hasin, D. S. (2015). Epidemiology of DSM-5 alcohol use disorder results from the national epidemiologic survey on alcohol and related conditions III. JAMA Psychiatry, 72(8), 757–766. https://doi.org/10.1001/jamapsychiatry.2015.0584

Greenberg, L. S. (1986). Change Process Research. Journal of Consulting and Clinical Psychology, 54(1), 4–9. https://doi.org/10.1037/0022-006X.54.1.4

Hedman, E., Ljótsson, B., Kaldo, V., Hesser, H., el Alaoui, S., Kraepelien, M., … Lindefors, N. (2014). Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. Journal of Affective Disorders, 155(1), 49–58. https://doi.org/10.1016/j.jad.2013.10.023

Higgins, D. (2011). Why reflect? Recognising the link between learning and reflection. Reflective Practice, 12(5), 583–584. https://doi.org/10.1080/14623943.2011.606693

Holland, K. (2018). Mental Health Basics: Types of Mental Illness, Diagnosis, Treatment, and More. Retrieved from https://www.healthline.com/health/mental-health

Holm-Hadulla, R. M., & Koutsoukou-Argyraki, A. (2015). Mental health of students in a globalized world: Prevalence of complaints and disorders, methods and effectivity of counseling, structure of mental health services for students. Mental Health and Prevention, 3(1–2), 1–4. https://doi.org/10.1016/j.mhp.2015.04.003

Hoogendoorn, M., Berger, T., Schulz, A., Stolz, T., & Szolovits, P. (2017). Predicting Social Anxiety Treatment Outcome based on Therapeutic Email Conversations. IEEE J Biomed Health Inform, 21(5), 1449–1459. https://doi.org/10.1109/JBHI.2016.2601123.

Howes, C., Purver, M., & McCabe, R. (2015). *Linguistic Indicators of Severity and Progress in Online Text-based Therapy for Depression*. (611733), 7–16. https://doi.org/10.3115/v1/w14-3202

Howes, C., Purver, M., & McCabe, R. (2015). Linguistic Indicators of Severity and Progress in Online Text-based Therapy for Depression. (611733), 7–16. https://doi.org/10.3115/v1/w14-3202

Jeanblanc, J. (2015). Comorbidity Between Psychiatric Diseases and Alcohol Use Disorders: Impact of Adolescent Alcohol Consumption. Current Addiction Reports, 2(4), 293–301. https://doi.org/10.1007/s40429-015-0076-5

Karyotaki, E., Kleiboer, A., Smit, F., Turner, D. T., Pastor, A. M., Andersson, G., … Cuijpers, P. (2015). Predictors of treatment dropout in self-guided web-based interventions for depression: An “individual patient data” meta-analysis. Psychological Medicine, 45(13), 2717–2726. https://doi.org/10.1017/S0033291715000665

Khazaie, H., Rezaie, L., Shahdipour, N., & Weaver, P. (2016). Exploration of the reasons for dropping out of psychotherapy: A qualitative study. Evaluation and Program Planning, 56, 23–30. https://doi.org/10.1016/j.evalprogplan.2016.03.002

Kotu, V., & Deshpande, B. (2015). Predictive Analytics and Data Mining. 275–303. https://doi.org/10.1016/B978-0-12-801460-8.00009-4

Kumar, V., Sattar, Y., Bseiso, A., Khan, S., & Rutkofsky, I. H. (2017). The Effectiveness of Internet-Based Cognitive Behavioral Therapy in Treatment of Psychiatric Disorders. Cureus, 9(8). https://doi.org/10.7759/cureus.1626

Lau, P. L., Jaladin, R. A. M., & Abdullah, H. S. (2013). Understanding the Two Sides of Online Counseling and their Ethical and Legal Ramifications. Procedia - Social and Behavioral Sciences, 103, 1243–1251. https://doi.org/10.1016/j.sbspro.2013.10.453

Lee, C., Corte, C., Stein, K. F., Park, C. G., Finnegan, L., & Mccreary, L. L. (2016). Prospective Effects of Possible Selves on Alcohol Consumption in Adolescents. 38(1), 71–81. https://doi.org/10.1002/nur.21641.Prospective

Linke, S., Murray, E., Butler, C., & Wallace, P. (2007). Internet-based interactive health intervention for the promotion of sensible drinking: Patterns of use and potential impact on members of the general public. Journal of Medical Internet Research, 9(2), 1–13. https://doi.org/10.2196/jmir.9.2.e10

Mallen, M.J., & Vogel, D.L. (2005). Introduction to the major contribution: Counseling psychology and online counseling. The Counseling Psychologist, 33(6), 761-775.

Miner, G., J. Elder, A. Fast, T. Hill, R. Nisbet and D. Delen (2012). Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications, Academic Press.

Newman, M. G., Jacobson, N. C., Erickson, T. M., & Fisher, A. J. (2017). Interpersonal Problems Predict Differential Response to Cognitive Versus Behavioral Treatment in a Randomized Controlled Trial. Behavior Therapy, 48(1), 56–68. https://doi.org/10.1016/j.beth.2016.05.005

Novotney, A. (2017). A growing wave of online therapy. Retrieved from https://www.apa.org/monitor/2017/02/online-therapy

Palmer, R. S., Murphy, M. K., Piselli, A., & Ball, S. A. (2009). Substance user treatment dropout from client and clinician perspectives: A pilot study. Substance Use and Misuse, 44(7), 1021–1038. https://doi.org/10.1080/10826080802495237

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The Development and Psychometric Properties of LIWC2015. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The Development and Psychometric Properties of LIWC2007. (January).

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates, *71*(1)

Postel, M. G. (2011). Well connected. Web-based treatment for problem drinkers.

Postel, M. G., de Haan, H. A., ter Huurne, E. D., Becker, E. S., & de Jong, C. A. J. (2010). Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: Randomized controlled trial. Journal of Medical Internet Research, 12(4). https://doi.org/10.2196/jmir.1642

Postel, M. G., de Haan, H. A., ter Huurne, E. D., van der Palen, J., Becker, E. S., & de Jong, C. A. J. (2011). Attrition in web-based treatment for problem drinkers. Journal of Medical Internet Research, 13(4). https://doi.org/10.2196/jmir.1811

Rehm, J. (2011). The risks associated with alcohol use and alcoholism. Alcohol Research and Health, 34(2), 135–143.

Riper, H., Kramer, J., Keuken, M., Smit, F., Schippers, G., & Cuijpers, P. (2008). Predicting successful treatment outcome of web-based self-help for problem drinkers: Secondary analysis from a randomized controlled trial. Journal of Medical Internet Research, 10(4). https://doi.org/10.2196/jmir.1102

Ritterband, L. M., & Tate, D. F. (2009). Introduction: The Science of internet interventions. Annals of Behavioral Medicine, 38(1), 1–3. https://doi.org/10.1007/s12160-009-9132-5

Rømcke, A. (2008). Named Entity Recognition Using the Web. 1–4.

Schotanus-Dijkstra, M., Drossaert, C. H. C., Pieterse, M. E., Boon, B., Walburg, J. A., & Bohlmeijer, E. T. (2017). An early intervention to promote well-being and flourishing and reduce anxiety and depression: A randomized controlled trial. Internet Interventions, 9, 15–24. https://doi.org/10.1016/j.invent.2017.04.002

Smink, W. A. C., Fox, J. P., Sang, E. T. K., Sools, A. M., Westerhof, G. J., & Veldkamp, B. P. (2019). Understanding therapeutic change process research through multilevel modeling and text mining. Frontiers in Psychology, 10(MAY), 1–14. https://doi.org/10.3389/fpsyg.2019.01186

Spence, S. H., March, S., & Donovan, C. L. (2019). Social support as a predictor of treatment adherence and response in an open-access, self-help, internet-delivered cognitive behavior therapy program for child and adolescent anxiety. Internet Interventions, 18(April), 100268. https://doi.org/10.1016/j.invent.2019.100268

Stark, M. J. (1992). Dropping out of substance abuse treatment: A clinically oriented review. Clinical Psychology Review, 12(1), 93–116. https://doi.org/10.1016/0272-7358(92)90092-M

Szumilas, M. (2010). Explaining odds ratios. Journal of the Canadian Academy of Child and Adolescent Psychiatry, 19(3), 227–229.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29(1), 24–54. https://doi.org/10.1177/0261927X09351676

ter Huurne, E. D., Postel, M. G., de Haan, H. A., van der Palen, J., & DeJong, C. A. J. (2017). Treatment dropout in web-based cognitive behavioral therapy for patients with eating disorders. Psychiatry Research, 247(October 2016), 182–193. https://doi.org/10.1016/j.psychres.2016.11.034

Tjong Kim Sang, E., Vries, B. de, Smink, W., Veldkamp, B., Westerhof, G., & Sools, A. (2019). De-identification of Dutch Medical Text. 1–4.

Trimbos.nl | Cijfers alcohol. (n.d.). Retrieved from https://www.trimbos.nl/kennis/cijfers/cijfers-alcohol

Wantland, D. J., Portillo, C. J., Holzemer, W. L., Slaughter, R., & McGhee, E. M. (2004). The effectiveness of web-based vs. non-web-based interventions: A meta-analysis of behavioral change outcomes. Journal of Medical Internet Research, 6(4). https://doi.org/10.2196/jmir.6.4.e40

Watson, H. J., Levine, M. D., Zerwas, S. C., Hamer, R. M., Crosby, R. D., Sprecher, C. S., … Bulik, C. M. (2017). Predictors of dropout in face-to-face and internet-based cognitive-behavioral therapy for bulimia nervosa in a randomized controlled trial. International Journal of Eating Disorders, 50(5), 569–577. https://doi.org/10.1002/eat.22644

de Weert-Van Oene, G. H., Schippers, G. M., de Jong, C. A. J., & Schrijvers, G. J. P. (2001). Retention in substance dependence treatment: The relevance of in-treatment factors. Journal of Substance Abuse Treatment, 20(4), 253–261. https://doi.org/10.1016/S0740-5472(01)00160-X

de Weert-van Oene, G., Burger, H., Grobbee, D., & Schrijvers, A. (2007). Identification of Patients at High Risk of Noncompliance in Substance Dependence Treatment. *European Addiction Research*, *13*(2), 74-80. doi: 10.1159/000097936

Wissen, L. van, & Boot, P. (2017). An Electronic Translation of the LIWC Dictionary into Dutch. ELex 2017: LEXICOGRAPHY FROM SCRATCH, 703–715. Retrieved from https://pure.knaw.nl/portal/en/publications/an-electronic-translation-of-the-liwc-dictionary-into-dutch(de9c8272-0df1-4c92-bcb3-d789ad793603)/export.html

Wykes, T., Haro, J. M., Belli, S. R., Obradors-Tarragó, C., Arango, C., Ayuso-Mateos, J. L., … Wittchen, H. U. (2015). Mental health research priorities for Europe. The Lancet Psychiatry, 2(11), 1036–1042. https://doi.org/10.1016/S2215-0366(15)00332-6

Yang, P., Tao, R., He, C., Liu, S., Wang, Y., & Zhang, X. (2018). The risk factors of the alcohol use disorders-through review of its comorbidities. Frontiers in Neuroscience, 12(MAY), 1–7. https://doi.org/10.3389/fnins.2018.00303