M.Sc. Thesis

Can Text Mining Enrich Data on Patient Characteristics on Dropout Cases in an Email-Based Alcohol Intervention?

Lukas Bern Libbertz-Mohr
s1692925

24.7.2020

Supervisors:
Dr. A.M. Sools
Dr. ing. G.-J. Prosman

Faculty of Behavioral, Management and Social Sciences

Course Code: 201300102

Positive Psychology & Technology

UNIVERSITY OF TWENTE.

Abstract

Previous studies on patient characteristics associated with dropout have largely produced ambiguous findings when relying on patient information that existed pre-treatment. To clarify this ambiguity, it is proposed to supplement pre-treatment information with linguistic data from within therapy sessions. Here, the rise in online therapy provides an ideal surrounding since the client texts written during online therapy form a transcript of the therapy process that is readily available for analysis. With *Text Mining* (TM) software such as *Linguistic Inquiry and Word Count* (LIWC), these client texts can be analyzed for differences in word choice between dropouts and completers. If clear dropout characteristics indeed exist, they might be present in the clients written texts. To investigate this, data from the Dutch email-based treatment program *Alcoholdebaas.nl* targeted at alcohol abuse disorder (AUD), was used. The sample consisted of 1987 Dutch-speaking participants of which 770 met the inclusion criteria and total of 2,793 patient emails were analyzed. Of these participants, the intake-questionnaire data was analyzed for differences in patient characteristics between completers and dropouts by using a series of *t*-test and a Chi-square test. For differences in linguistics, measured by word choice, the text mining software LIWC was used. Significant differences between the groups were found for *age*, with dropouts being younger on average, for *gender* with more females completing therapy, *educational level*, with completers being more concentrated in higher forms of education as well as for *smoking* and taking *drugs* with dropouts being more represented in both. For differences in choice of words, the LIWC dictionary *work* reached significance with dropouts using more words related to work, being busy and jobs. The findings on the patient characteristics partly confirmed and substantiated prior research on patient characteristics, thus helping in clearing the ambiguity around them. Noteworthy, word use related to work among dropouts is a new discovery, highlighting the potential of LIWC in dropout research. Thus, the application of LIWC proved to be an effective and efficient methods to handle the large amount of data generated in online therapy. Future refinements in dictionary-based TM methods, tailored for dropout research, will benefit from the wealth of textual data that online therapy provides and ultimately help in clarifying the phenomenon of client dropout.

Keywords: dropout, premature termination, alcohol abuse disorder, AUD, text mining, LIWC, linguistic analysis, online therapy, online counselling, email therapy

Table of Contents

Introduction

Internet-based psychotherapy interventions have become a valid alternative to classic face-to-face therapy, realizing their full potential is hampered by the high numbers of treatment dropouts. With more than one in three Europeans experiencing mental health problems each year (György Purebl el al., 2017; Wykes et al., 2015), the need for accessible therapy becomes apparent. Especially diseases that pose major health related, societal and economic challenges such as alcohol use disorder, abbreviated *AUD,* require extensive treatment coverage. Moving therapy online can provide easy to access treatment possibilities that scale up to this demand. Concerning the effectiveness of online treatments, online versions of classic therapy methods like cognitive behavioral therapy, were found to be on par with traditional methods (Carlbring, Andersson, Cuijpers, Riper, & Hedman-Lagerlöf, 2018; Richards & Richardson, 2012).

While effective, many patients decide to forego therapy and do not terminate their online treatment (Alfonsson, Olsson, & Hursti, 2016). Researchers investigate this phenomenon, called *dropout*, present in both online and traditional therapy by focusing on patient characteristics to understand which individuals drop out of treatment prematurely (Belleau et al., 2017; Yeung et al., 2015). Yet studies on patient characteristics that can predict dropout, either for AUD therapy specifically or other forms of therapy, are ambivalent in their findings (Swift & Greenberg, 2012; Zandberg et al., 2016). Additionally, it is criticized that the focus on personal characteristics misses out on within therapy information that might hold additional explanatory value on dropout (Fernández-Álvarez et al., 2017). Understanding the individuals that exit therapy prematurely is important so that interventions can be tailored to their needs and to support them in achieving symptom relief.

Here, online therapy interventions provide a basis to further the dropout research topic. The large patient samples in online therapies allow for thorough and robust analyses of patient characteristics that might be related to dropout. Next, the written exchanges generated by patients in online therapy provide data from within treatment in linguistic form (Smink et al., 2019). With text mining software (TM), their texts can be analyzed for linguistic properties that hold the potential to enrich the information on dropout patients in a way that the sole analysis of characteristics cannot. To that end, this study applied TM on patient emails of an online alcohol intervention to supplement possible differences in patient characteristics between dropouts and completers with possible distinctions in linguistic features.

The Online Shift of Therapy and Online Alcohol Therapy

Today, one can find a large variety of treatment possibilities for numerous psychopathologies online. Especially the more common psychopathologies depression, PTSD, eating disorders and AUD have been covered by online interventions (Rogers et al., 2017). Of these online interventions, the ones treating AUD and other forms of substance abuse form the majority of evidence-based online programs (Rogers et al., 2017). These interventions range from purely informational websites and self-guided interventions to fully counsellor guided therapy. The latter is of special interest, as it often emulates traditional face-to-face treatment programs (Carlbring et al., 2018; Crombie et al., 2018) and the online support by trained professionals makes this form of online therapy more effective than self-guided online interventions (Karyotaki et al., 2015; Mohr, Cuijpers, & Lehman, 2011). Online therapy is a broad field that is not yet clearly defined since multiple definitions for the topic are used interchangeably such as *online counselling*, *web counselling* or *cyber therapy* (Hanley, Ersahin, Sefi, & Hebron, 2017; Li,Lau, Jaladin, & Abdullah, 2013). In this study, it was defined as therapy provided by trained professionals that establish client-therapist communication primary via digital online means (e.g. using email or voice over IP services like Skype) and follow a structured approach to pathology treatment. The majority of counsellor guided online therapies rely on written client-counsellor exchanges as in emails or chat services. (Chester & Glass, 2006). This is called an asynchronous way of communication since the parties can postpone answering to their convenience as opposed to synchronous communication where answers are given immediately (Reynolds, Stiles, Bailer, & Hughes, 2013).

When it comes to text based treatment methods, a commonly argued disadvantage is that the client-therapist working alliance would be impaired as therapists cannot react to non-verbal cues of their clients (Rochlen, Zack, & Speyer, 2004). For AUD therapy, this would pose a problem, since a functional working-alliance is related to positive treatment outcomes in the sense of increasing the motivation to change problematic drinking behavior (Cook et al., 2015). However, other research in online therapy found no impairment of therapist-client working alliance (Cook & Doyle, 2002) and others found it to be possible to create a functional working alliance via email exchange (Anderson et al., 2012; Reynolds et al., 2013). Yet another point of criticism is that online therapy might be less appealing to groups that are not eager to adopt to new technologies. For instance the elderly are said to be less willing to use internet services due to their lack of knowledge on how to navigate it or out of fear of being exploited online (Hussain, Ross, & Bednar, 2018). Yet, internet usage among US elder citizens went up from 22% of the elderly in 2004 (Fox, 2004) to 67% in 2016 (Anderson & Perrin, 2017). Additionally, around 30% of the elderly regularly access the internet for health and medical issues (Levy, Janke, & Langa, 2015), meaning that this group is picking up on the trend. With 90% of the general European population having internet access (Johnson, 2019) and many using the internet for health-related information (Fox & Duggan, 2013; Tan & Goonawardene, 2017) a major advantage of online therapy is its reach. Especially pervasive psychopathologies such as AUD (Table 1), can profit from such far reaching and easy to access treatment possibilities. This is crucial as AUD therapy is hindered by the general low tendency of help seeking behavior by the affected, ranging from 13 to 24 years until making first contact with a professional (Chapman, Slade, Hunt, & Teesson, 2015; Kay Lambkin, 2014). Regarding that, online therapy can lower the barriers for individuals to gain treatment access.

Online therapy is accessible at any time, from any place with internet access, making it appealing for patients that cannot attend therapists at business hours or that live in remote areas (Moritz, Schröder, Meyer, & Hauschildt, 2013; Pedersen, Marshall, & Schell, 2016). Adding to that, there are none to only short wait lists for online treatments (ter Huurne, Postel, de Haan, van der Palen, & DeJong, 2017), making it a valuable on demand tool for individuals with AUD (Cloud & Peacock, 2001). This is advantageous for AUD patients, as their willingness to change is volatile, caused by a recent negative event due to their drinking and tends to diminish over time (Cunningham, Sobell, Sobell, & Gaskin, 1994). Yet another barrier to seeking AUD treatment is the stigmatization of being labelled as an alcoholic, often associated with a loss in status and discrimination (Schomerus et al., 2011). In anonymous online treatments however, individuals feel less stigmatized (Berger, Wagner, & Baker, 2005; Marloes, et al., 2010; Rooke et al., 2010) and perceive it as easier to disclose their problems and feelings (Fotheringham et al., 2000), making online therapies an appealing alternative to traditional face-to-face treatments. Attesting to the general usage of online AUD websites, a meta-analysis of 7 AUD websites by Michael L. Vernon (2010) showed that 60.000 people visited such websites in a period of 6 months and on average 56.4% took assessments of drinking behavior that were offered. However, there are hindrances to the effectiveness of online therapy as well since reaching high numbers of clients does not translate well into successful treatment completions.

Table 1

Burdens and Risk Factors of AUD

Societal*	Physiological**	Psychological***	Risk Factors****
100.4-283 Million Cases Globally	Cancer	Increased Suicidality	Poverty
> 11 Million Alcohol Dependents in the EU	Increased Mortality	Depression	Lower Education
Associated with 5.3% of all Deaths Globally	Cardiovascular Disease	Anxiety Disorders	Unemployment
Associated with 7.3% of all Global Premature Deaths	Early Onset of Dementia	PTSD	All Age Groups
99.2 Million DALYs in 2016	Decrease in Male Fertility	Schizophrenia	
24.5 Billion € in Annual Costs of Alcohol Related Hospitalizations			

Note. DALYs = Disability-Adjusted Life Years; *Premature Deaths* were considered deaths below 69 years of age.
*(Degenhardt et al., 2018; Hammer et al. 2018; Olesen et al., 2012; Rehm et al., 2015)
**(Bagnardi et al., 2013; Graff-Iversen et al., 2013; Sansone et al., 2018; Schwarzinger et al., 2018)
***(Goldstein, Bradley, Ressler, & Powers, 2017; Subramaniam et al., 2017; Wiener et al., 2018)
****(Baumann et al., 2007; Dauber, Pogarell, Kraus, & Braun, 2018; Degni, Vaherkylä, & Hurme, 2017; Ryan & Kokotailo, 2019; Teixidó-Compañó et al., 2018)

Psychotherapy and Dropouts

The problem that is curbing the potential of online therapy is the high number of patients who do not finish their treatment. It is referred to by *attrition*, *dropout*, *non-usage* or *premature termination* interchangeably (Elisabeth, Stacey, & Maintaining, 2009). In this study, the term *dropout* is used. Labelling a client as such is not universally agreed upon. It is argued that a dropout can be defined as someone who either did not finish the complete intervention, did not reach a certain cap of required attended sessions or based on therapist judgement (Swift &

Greenberg, 2012). Yet, such distinctions matter, since depending on how extensive the requirements for treatment completion are, the number of cases labelled as dropout changes (Yeung et al., 2015). For the purpose of this study, dropouts are considered to be therapy participants who did not finish all treatment sessions required to complete the treatment protocol. It has to be pointed out that dropouts do not necessarily have to be understood as treatment failures. Some drop out prematurely because they feel that their problem has improved sufficiently (Krishnamurthy et al., 2015). Nonetheless, dropping out of treatment can be problematic for clients. It prevents clients from achieving symptom reduction or return to normal functioning (Luedke et al., 2017) since symptom improvement is related to frequent attendance in therapy (Lambert, 2013). Additionally, dropouts strain health services as they require administrative resources and costs accrue, while at the same time access to treatment for waiting patients is delayed (Watson, Fursland, & Byrne, 2013). Concerning online therapy istelf, there are factors that increase dropout rates. For instance, visual appeal of the online contents influences clients in re-visiting the online interventions (Brouwer et al., 2009). Asking for extensive personal information (O'Neil, Penrod, & Bornstein, 2003) and general motivation (Coa & Patrick, 2016) were also found to be associated with dropout from online therapy.

Generally, dropouts are a common occurrence in both classic therapy as well as online therapy with higher dropout averages for the latter. A general comparison of dropout rates can be found in Table 2. To understand the dropout phenomenon many studies were dedicated to identify patient characteristics that existed pre-treatment which can reliably predict patient dropout (Belleau et al., 2017). These predictor studies were, among others, focused on patient characteristics like *age*, *gender*, *educational level*, *ethnicity*, *relationship status* and *employment status* (Karyotaki et al., 2015; Stelzhammer et al., 2015; Watson et al., 2017). Concerning online AUD and substance abuse treatments specifically, younger age was found to be related to dropout (Elbreder et al., 2011; Vuoristo-Myllys et al., 2013). For gender, some found males to dropout less (Darke, Campbell, & Popple, 2012), while others found females tending to dropout less (Radtke et al., 2017). Again others found no association between gender and dropout (Elbreder et al., 2011). For relationship status, some theorized an association to dropout but did not find a significant relationship (Corrêa Filho & Baltieri, 2012). Comorbid depression as predictor was identified by some (Corrêa Filho & Baltieri, 2012), but again, others did not find this connection (Kavanagh et al., 2006). Even more relevant to AUD, baseline amount of alcohol consumption as measurement of AUD severity has not been conclusively shown to predict dropout. Some found it does (Radtke et al., 2017) and others found it to predict lower dropout instead (Ray, Hutchison, & Bryan, 2006). Generally, it is reported that consistent findings on predictors of AUD treatment are scarce (Zandberg et al., 2016) and literature investigating AUD treatment dropout is limited (Vuoristo-myllys et al., 2013). This indicates that there is still a need to further explore dropout from AUD treatment to obtain reliable predictors. Given the ambiguity among identified predictors, extending the view beyond AUD therapy may yield new insights on alternative predictors, priorly unconsidered in the AUD dropout context.

When looking at other forms of treatment, finding reliable predictors of dropout has not yielded conclusive results either, especially for demographic characteristics (Barrett et al., 2008; Swift & Greenberg, 2012). For anxiety, Krishnamurthy and colleagues (2015) demonstrated that high anxiety symptom severity was predicting dropout from cognitive behavioral treatment. However, others did not find anxiety to be associated to dropout (Back et al., 2001). Others argue that looking at negative emotional state (feeling any form of negative feeling) as sign of general psychological distress can be a better predictor than specific negative emotions (Deane et al., 2012). It is also theorized that increasing positive affect in participants might increase client retention (Geraghty, Wood, & Hyland, 2010). Another interesting predictor of dropout is the lack of insight, found to be associated with dropout during treatment assessment (Lincoln et al., 2014). In another study, dropouts were found to be less focused on the future and their goals (Alfonsson et al., 2016). Additionally, dropout was found to be related with lower socio-economic status (Barrett et al., 2008). When it comes to education both medium as well as lower education were found related to dropout (Reinwand et al., 2015; Rizvi, Vogt, & Resick, 2009; Watson et al., 2017). While these results are promising, many researchers state that findings on generalizable and consistent predictors are scarce (Belleau et al., 2017; Oldham et al., 2012). Nevertheless, identifying characteristics of dropout in patients is crucial, so that individuals at risk can be focused on by therapists and interventions can be tailored to their needs which reduces dropout rates (Geraghty et al., 2010). With the general lack of reliable predictors for dropout, identifying them in online AUD therapy would not only help AUD therapy alone, but could proof to be a valuable orientation for other forms and styles of therapy as well.

Table 2

Reports of Dropout Rates in Meta Analyses from Traditional and Online Treatments

Traditional Therapy			**Online Therapy**		
Study	*N*	*M*	Study	*N*	*M*
Hans & Hiller, 2013	1,880	24.63	Karyotaki et al., 2015	2,705	70
Swift & Greenberg, 2012	83,834	19.7	Richards & Richardson, 2012	10,395	57
Swift & Greenberg, 2014	--*	18.84**	Van Ballegooijen et al., 2014	477	24.9
Gersh et al., 2017	2224	16.99	Zachariae, 2016	1,460	24.7
Van Ballegooijen et al., 2014	504	15.3	Van Beugen et al., 2014	4,340	18

Note. N = number of participants in the meta analyses; *M* = mean dropout expressed in percentages.
*Swift & Greenberg (2014) did not report a participant number. Their dropout rate calculation was based on 587 studies.
** The Mean Dropout rate was computed using the weighted average dropout rate per disorder category in Swift & Greenberg (2014), Table 1.

Linguistic Analysis in Psychology & LIWC

Because psychotherapy is mainly conducted as a conversation between client and therapist, it has been argued that this linguistic interaction contains the factors that elicit change in clients (Imel, Steyvers, & Atkins, 2015; W. Smink et al., 2019). This interaction produces a continuous stream of linguistic data from within therapy that can be textualized to tie linguistics to a multitude of themes relevant to psychology research like self-report data, behavioral data, personality, social behavior and cognitions (Tausczik & Pennebaker, 2010). Especially today, when information is communicated and stored digitally at unprecedented scale (Hilbert & López, 2011), the analysis of language can be of special value to the field of psychology. However, traditional textual analysis relies on human coding and analysis of therapy transcripts which is a labor intensive and time-consuming task. Coding a standard 50 minutes therapy session can take up to several hours depending on the complexity of coding (Tanana et al., 2016). Therefore, manual coding is not well suited as a method to analyze the vast amounts of digital texts (Can et al., 2016), especially since this task can be performed by computers more efficiently.

A main methodology of computerized textual analysis is TM. For TM, software algorithms from the field of statistics and machine learning are employed to turn machine readable text into data that is suitable for statistical calculations (Hotho, Nürnberger, & Paaß, 2005). It is mainly used to count the number of words for e.g. frequency calculations (Dreisbach et al., 2019). TM can analyze pages, books, forum posts or emails and doing so at a speed that no human coder could achieve by hand. In fact, the efficiency of TM algorithms has made some researchers argue that, considering processing time and costs, human coders would no longer be practicable (Imel et al., 2015; Snow et al., 2008). In the field of psychology, TM approaches are widely used to identify key words and themes as well as sorting them according to different categories and concepts (Hoogendoorn et al., 2017). Generally, the application of TM has become a staple method in the scientific community and due to constant methodological refinements TM as a methodology is ready for widespread utilization (Abbe et al., 2015).

One of the most renown TM software is the Linguistic Inquiry and Word Count, abbreviated LIWC (Tausczik & Pennebaker, 2010; Pennebaker et al., 2015). LIWC features a multitude of hierarchically ordered dictionaries comprising of 6400 words and signs that identify expressions related to various categories such as emotional affect, work, family, grammatical structures and others. LIWC has been shown to be a valuable tool for analysis of substance abuse, alcohol abuse and therapies aimed at their treatment. Using LIWC, Liehr et al. (2010) found that residents in a mindfulness based therapeutic community aimed to treat substance abuse used fewer negative emotion indicating words than a control group. Jensen & Hussong (2019) used LIWC to investigate alcohol related talk in student text-messages and found alcohol related word usage to predict the risk of engaging in alcohol drinking. In another study investigating texts of an open ended writing task, participant texts were analyzed by LIWC and other methods to understand when and why people drink alcohol (Lowe et al., 2013). Among the 7 themes related to drinking were children and family, consequences of consumption as well as special occasions. Patient stories were also analyzed by LIWC in the study of Dunlop and Tracy (2013) investigating writings of recovering. alcoholics and it was found that self-redemptive narratives stimulate long term behavioral change regarding consumption. Self-stigmatization and negative emotions have also been analyzed by LIWC in the context of AUD in the study of Bliuc, Doan & Best (2019). Both were negatively associated with identifying oneself as being part of a recovery network, though increased social identification with the group was contributing to sobriety. Despite the capability of LIWC to identify discussion topics and risk of alcohol drinking, no studies were found that used LIWC in identifying linguistic markers related to dropout in AUD therapies. In the past, LIWC has been successfully used to meaningfully differentiate between groups (Lyons, Mehl, & Pennebaker, 2006), which is why this is surprising. In this study, it is argued that LIWC can be a valuable addition to investigate possible linguistic differences in texts of treatment dropouts and completers.

Aim of the Research

So far, the findings on individuals completing or dropping out of online interventions have been ambivalent and overly focused on patient characteristics existing prior to psychological interventions (Fernández-Álvarez et al., 2017; Swift & Greenberg, 2012). Specifically, for AUD treatment, information on patient characteristics are ambiguous and though TM was used for AUD research, it was not used to investigate dropout. In this study it is propose that textual data analysis of patient emails from an online AUD treatment can create new insights on within therapy information to discern dropouts from completers. Additionally, it is argued that differentiating between the groups based on pre-existing characteristics can be effectively supplemented by identifying differences in choice of words. For patient characteristics *age*, *gender*, *education*, *nationality*, *relationship status*, *employment*, *smoking*, *drugs*, *gambling*, *depression*, *prior psychological treatment, reason for participation, goal of the therapy, years lived with problematic consumption* and *daily average alcohol intake* were compared. Additionally, to the patient characteristics, 9 LIWC dictionaries were compared in linguistics between dropout and completers. Even if completers and dropouts do not differ in terms of characteristics, they might differ in the words they employ to describe their respective situations which could yield new insights into dropout. Subsequently, the aim of this study was to investigate whether the patient characteristics theorized to be associated with dropout can be confirmed by this study and whether differences in LIWC dictionaries can be identified that reflect differences between dropouts and completers as well.

Methodology

Data Sample

The data sample used in this study was provided by Tactus, a mental health institution in the Netherlands with a focus on addiction treatment. The data consisted of individuals who had participated in the online program Alcoholdebaas.nl. The sample was obtained by convenience sampling methods since participants reached the website via personal online research, TV or radio advertisement, referral by professionals or recommendations by friends and family. For research and quality management purposes Tactus had asked all participants of this program for their consent to collect their data. No data was included of participants that had not given their consent or have revoked their consent. The website features two different treatment interventions, the '*intensive*' and the '*short*' one. This study only used data from the intensive treatment program.

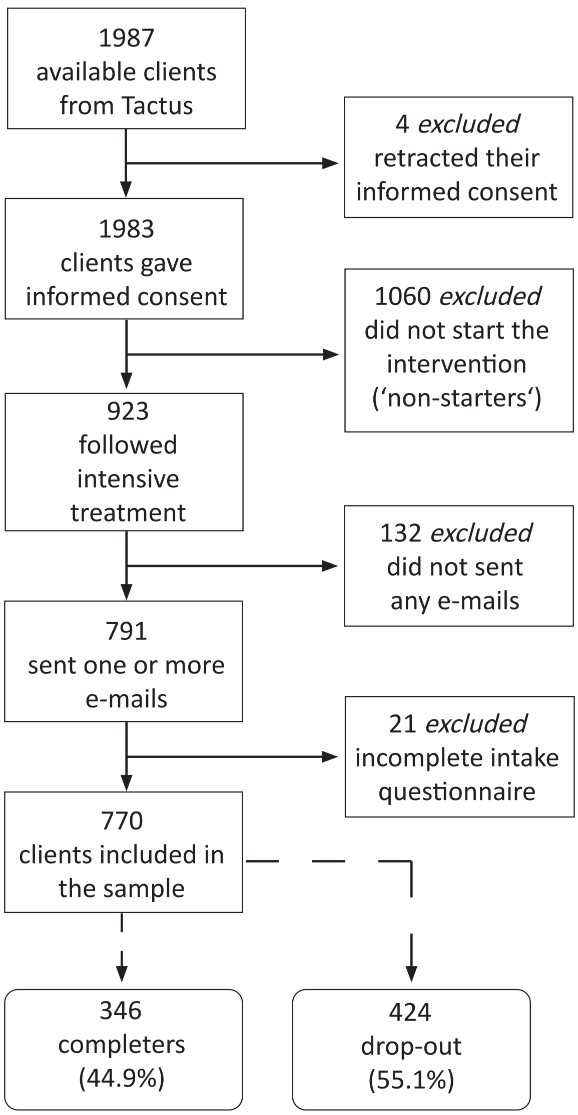
In total, the sample consisted of 1987 participants. All participants were Dutch speaking. From the original 1987 participant cases, some had revoked their consent, did not start the intervention or started the intervention but sent no email to their therapist and for others, data on the intake questionnaire was incomplete. These cases have been excluded as can be seen in the patient flow (Figure 1). The 770 remaining client cases included 428 females (55.58%) and 342 males (44.42%) with a mean age of 46 (*SD*=10.76), a minimum age of 17 years and a maximum age of 78 years. These cases were split into two groups, the ones who had completed the intervention and the ones who dropped out prematurely. For both groups, the average daily alcohol consumption met the criteria for hazardous drinking (Stolle, Sack, & Thomasius, 2009), set at drinking 4 standard glasses of alcohol for females and 5 for males (e.g. 0.3 liter of beer or 0.2 liter of wine) with dropouts drinking 8.14 (*SD*=7.02) glasses daily and completers drinking 6.96 (*SD*=4.96). A client was labelled as completer if their therapist had sent them the e-mail informing them about the conclusion of the therapy when all assignments were passed. Clients were labelled as dropouts when they had written at least one email to their therapist but stopped writing to their therapist before the end of therapy. When they did not respond to two reminders by their therapists, their status was switched to '*inactive*'. Overall, 346 clients were identified as completers and 424 as dropouts.

Concerning the emails, the 424 dropouts sent on average 9.4 emails (*SD*=9.4) and the 346 completers sent 28 (*SD*=14). This study included emails only up to the fourth per patient of both groups. This was done because among the first four emails, both groups were represented more balanced as in later stages of the treatment when the majority of dropouts had stopped their participation. Not posing a limitation might have skewered the data towards completers. Among these first four mails, the dropouts wrote 1410 emails comprising of a total of 323,057 words and the completers wrote 1383 emails comprising of 415,545 words.

Alcoholdebaas.nl has been online since 2005 and that slight alterations to the program have been made. The intake questionnaire has been revised several times and not all questions were included in every iteration. Next to that, answering the questions was optional, which resulted in low answer rates for some questions like a 4% answer rate on *prior psychological treatments*. Additionally, clients were unevenly distributed between the 45 therapists with some treating substantially more patients than others.

Figure 1

Patient Flow



Note. This figure shows the flow of participants that were included in the study.

The Alcoholdebaas.nl Online Intervention

The intervention Alcoholdebaas.nl (Dutch for 'Alcohol the Boss') is an online alcohol intervention (Alcoholdebaas.nl, n.d.) that was developed in 2005 to provide an easy-access treatment possibility for problem drinkers over 16 years of age (Postel, 2011). It is noteworthy that the program not necessarily puts the sobriety of the participants as the goal of the intervention, but lets the client decide whether he or she wants to stop, change or reduce the personal drinking behavior. The program was developed by *tactive*, a subsidiary institution of Tactus, and is a Dutch mental health institution specialized in addiction treatment that is ISO as well as HKZ certified. The intervention can be accessed via the internet and allows visitors to sign up for two treatment interventions, the 'intensive' treatment taking approximately 16 to 22 weeks and the 'short' treatment program, taking 6 weeks. Participants signing up for either intervention do so by creating a personal account on the website and start out with an intake questionnaire. During these interventions, participants are assigned to a therapist and they keep the assigned therapist until the conclusion of the intervention. According to the website, the therapists are social workers with experience in face-to-face therapy and are supported by a psychologist, a dietician, as well as a physician. In the course of the intervention, the therapist-participant pair communicates asynchronously via a protected web-based email application embedded in the website. Generally, one contact with the therapist is planned per week. The costs of both interventions are payed for by the participants personal health insurance, though private payment is possible, too.

The treatment programs start out with the participant filling in the intake questionnaire and a professional giving a recommendation for treatment based on the intake. Then, participants continue with the treatment program. It consists of two parts. The first part includes four assignments and two assessments and is generally aimed at analyzing the participants' drinking habits. The assignments and assessments follow a distinct order and progression is only possible by completing the current task. The second part is focused on changing the participants' dysfunctional drinking behavior and thoughts as well as to replace them with helpful ones. After 16 weeks, the second part concludes with the formulation of an action plan to maintain the new drinking behavior and for relapse prevention. The intervention ends with a '*wrapping up*' email by the therapist, summing up the therapy process and explaining further organizational details. If the participant wants it, he or she can participate in an aftercare program for another 6 weeks, during which the participant can have one contact with their therapist each week. When the aftercare concludes, the participant retains access to the personal account and files for another 6 months before the account is closed. The interventions are scientifically rooted in cognitive behavioral therapy and motivational interviewing (Postel, 2011, p.14), both empirically substantiated approaches for substance disorder treatment (Magill et al., 2019; Miller & Rollnick, 2012). An outline of the intensive treatment protocol with the activities for each week as well as exemplary client-therapist quotes can be seen in Appendix A.

Software

Frog

Frog is an assortment of natural language processing modules developed for the pre-processing of Dutch texts to prepare them for text mining analysis (van der Sloot et al., 2018). Frog can be used for, among others, pre-processing methods and tokenization. Tokenization refers to the process by which the individual words a text comprises of are recognized as own entities instead of a continuous string of text (Hotho, Nürnberger, & Paaß, 2005). This step is required so that every word included in a text can be counted by text-mining algorithms. This also involves separating words from punctuation (van der Sloot et al., 2018). Another property of Frog is, that it can be used to anonymize sensitive data included in the texts while still preserving the general information (Tjong Kim Sang et al., 2019). In a process called *named entity recognition*, Frog uses machine learning techniques to identify words that are required to be anonymized to safeguard the authors privacy. These are dates, numbers, locations, names, product names and miscellaneous entities which are replaced by the representatives 'DATE', 'NUM', 'LOC', 'PER', 'PRO' and 'MISC'. For this study, the initial anonymization by FROG was manually checked by Dr. Tjong Kim Sang to correct wrongfully anonymized words as well as to manually anonymize words that were not processed by the software (Tjong Kim Sang et al., 2019). The following is an example excerpt from an anonymized client email translated from Dutch:

'*My daughter will return on DATE from LOC, then she will fly to LOC and on DATE she will fly back to LOC.*'

Linguistic Inquiry and Word Count

The program LIWC is a text analysis tool for written and digitalized text, which was developed by James W. Pennebaker and his colleagues (Tausczik & Pennebaker, 2010; Pennebaker et al., 2015). LIWC comprises of a variety of hierarchically structured dictionaries, 73 in total, that classify certain words and categorize them (Pennebaker et al., 2015; Wissen & Boot, 2017). For instance, LIWC2015 has the dictionary '*Cognitive Processes*' and as a sub-dictionary to it a dictionary called '*Insight*' to both of which words like 'understanding' and 'realizing' are counted. LIWC processes each single word in a given text and checks if they match entries in its dictionaries. For each match, the count of the dictionary is raised by one. Then, LIWC calculates the fraction of the words of a specific dictionary to all the words in the text. The following sentence is an example excerpt from a client email:

'Every day I feel remorse and am disgusted of myself that it did not work out again. I think it comes from forgetfulness and searching for relaxation. Obviously in vain. I know very well that my health is at stake.'

The words *remorse* and *disgust* are counted towards the *negative emotions* dictionary and since the entire segment of text comprises of 40 words, the total LIWC score for *negative emotions* in this segment is the number of words counted towards the dictionary divided by 40 (0.05). LIWC calculates these values for each of its dictionaries. Additionally, LIWC can give out the usage frequency of every word in a text and rank them accordingly. Depending on the words one is interested in limitations can established to e.g. only analyze nouns. If no limitations are established, words like '*th*e' or '*and*' or punctuation achieve high frequencies yet, depending on the research question, do not hold much explanatory value. The keyword lists for the LIWC dictionaries used in this study can be found in Appendices B to J.

For this study, the current fourth version, LIWC2015, was used with updated dictionaries and software. Originally, the LIWC dictionaries are in English, yet users can add their own dictionaries to accommodate other languages as well. This study used the Dutch translation by Wissen & Boot (2017). The correlations between the English and Dutch dictionaries was found to be acceptable, with an average correlation of 0.69.

Choice of LIWC Categories. When it comes to the LIWC categories, this study did not investigate possible differences between dropouts and completers in all 73 dictionaries. Based on the findings of other researchers discussed earlier, 9 LIWC dictionaries were chosen that can be considered related to the information presented on dropouts in the introduction.

The LIWC dictionary *positive emotions* was used based on the findings of Geraghty and colleagues (2010) that positive emotions are related to improved client retention. The dictionary *negative emotions* was included based on the theory by Deane and colleagues (2012) that negative emotional state might better predict dropout than specific negative emotions. The dictionary *anxiety* was included because of the findings of Krishnamurty and colleagues (2015) that high anxiety severity was related to dropout. *Sadness* was included as an approximation of depressive symptoms that were theorized to predict dropout (Corrêa Filho & Baltieri, 2012). The dictionary *family* was included based on the findings of Lowe and colleagues (2013) that related alcohol drinking to words associated with family in their analysis of patient writings. *Insight* was included based on the role of insight into personal problems and dropout (Lincoln et al., 2014). *Focus past* and *focus future* were included because of the finding by Alfonsson and colleagues (2016), that dropouts were less focused on the future than completers. With *focus past*, the reverse relationship was tested. The dictionary *work* was included because of the found association of dropout and employment status (ter Huurne et al., 2017; Watson et al., 2017).

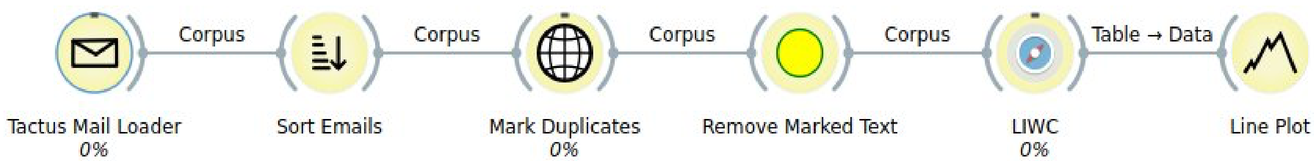
It has to be pointed out that the relationships between the LIWC dictionaries and their possible association to factors contributing to dropout or completion are difficult to establish due to the ambivalence and scarcity in reliable findings. Therefore, only these 9 of the 73 were included since based on the presented studies, they have a theoretic background to warrant their inclusion. Another reason for a limitation of dictionaries was, that when testing all 73 dictionaries using a standard level of significance (p=.05), for at least 3 to for 4 results, chance cannot reliably be ruled out despite a possible significant relationship. While for TM, it would not be a problem to handle all 73 dictionaries, this study was aimed at clarifying ambiguities around predictors theorized to relate to dropout by using linguistic measurements. Still, the remaining LIWC dictionaries hold potential for interesting findings, but including them all and discuss their potential background as well as providing an explanation for their relationship with dropout, would surpass the scope of this study.

Orange

The program Orange is developed by the Bioinformatics Lab of the University of Ljubljana in Slovenia in collaboration with the GitHub open source community. It is an open source program that can be downloaded at the Orange homepage (Orange, n.d.). It can be described as a toolbox for machine learning and data mining components (Demsar et al., 2013). The program allows for building analysis or processing routines for various data mining and machine learning approaches by interconnecting the different analysis modules into a workflow, called a *pipeline*. The pipeline used for this study can be seen in Figure 2. For this study and the *What Works When for Whom* project, Orange was extended with 6 additional processing modules (Tjong Kim Sang et al., 2020). The core software of Orange was built with C++ and it utilizes Python scripts. For this study, the Orange version 3.23.1. was used.

Figure 2

Orange Pipeline



Note. The figure shows the used Orange pipeline. From left to right, the widgets are shown that loaded, pre-processed, analyzed and visualized the data derived from client emails.

IBM-SPSS

SPSS is a program developed and distributed by the International Business Machines Corporation (IBM) and is designed to handle statistical calculations. The SPSS license that was used was provided by the University of Twente. For this study, the SPSS version 26 was used.

Design and Procedure

The study featured a descriptive observational cohort design by investigating a group of participants of an online alcohol intervention to analyze demographic as well as linguistic differences between the ones completing the intervention and the ones dropping out prematurely. The demographic data was sampled by an intake questionnaire and the linguistic data was sampled by text-mining the e-mails that participants wrote to their therapists.

The results of the intake questionnaire as well as the email conversations were provided by Tactus. For reasons of security they were kept on a password protected, encrypted USB and only handed to the researchers. Each client was given an identification code to replace their names (AdB0001 to AdB1987). The emails were anonymized and tokenized by the program FROG as described. After the emails had been pre-processed, they were loaded into Orange. The first widget of the Orange pipeline loaded all e-mails into Orange. The second widget sorted the emails in chronological order. The third widget highlight any copied text that the client used or cited from previous emails, so it would not be analyzed again. Duplicates were considered to be all strings of words equal to or greater than 20 words that had been included in any prior email. Then, the fourth widget was used to exclude the highlighted text duplicates. The fifth widget, *LIWC,* analyzed the emails in the fashion described earlier. Lastly, a data visualization widget was used to visualize the LIWC data as needed. Though the pipeline features a line plot widget, the data could be filled in tables or other form of visualization using the corresponding widgets. Once the data was analyzed by LIWC, the values per LIWC dictionary were saved. LIWC created a value for each category and for each single email written. Therefore, for each client the mean LIWC score per category was computed by summing up all LIWC scores per category throughout the first four emails and then divided by the number of emails. Additionally, for each of the selected LIWC categories, the 15 most frequently used words were compiled into frequency tables for comparison. Then, patient data from the intake questionnaire, the LIWC data per client and per category as well as a list of the number of emails sent by each client were put into an Excel spreadsheet for further analysis in SPSS. While the number of emails and LIWC scores were ready to be used in SPSS, the intake questionnaire data had to be transformed first into continuous numbers.

Data Analysis

For the variables *number of emails*, *age*, *alcohol consumption* and *years lived with problematic alcohol consumption*, the mean, standard deviation, minimum and maximum values were calculated. Then, except for *number of mails*, these characteristics were compared between the groups using independent sample *t*-tests (Table 4). Since number of mails was adjusted manually, they could no longer be compared meaningfully as all completers had a score of 4 by default. For the nominal patient characteristics *gender*, *education*, *nationality*, *relationship status*, *employment*, *smoking*, *drugs*, *gambling*, *depression*, *prior psychological treatment, reason for participation* and *goal of the therapy*, the total values and percentages were calculated and split to dropout and completers (Table 3). Then a Chi-Square test was done to test for differences in patient characteristics between the groups.

For the linguistic data the mean LIWC scores per LIWC category were computed for each participant based on the first four emails. Then an independent sample *t*-test was conducted to investigate the difference in LIWC category means between dropout and completer patients (Table 4). Since this comparison does not cover the context in which the words counted towards the LIWC category were used, a qualitative assessment of client mails was done for each significant difference in LIWC category means that was found. For this qualitative assessment, the emails of one patient per group that ranked high in the LIWC category mean were compared based on the context the words were used. Then, representative email excerpts were presented. Additionally, for the analysis of the LIWC categories, the keywords tables were compared between dropouts and completers (Appendices B to J).

Results

The intention in this section was to analyze patient characteristics, as well as patient linguistics to find possible differences between the completer and dropout group. First, a typical dropout and completer is described based on the frequency and percentages derived from the intake. Secondly the differences in patient characteristics are assessed for significance. Thirdly, it is presented how the groups differ in linguistics based on the LIWC dictionary scores. Additionally, email excerpts from dropouts and completers are presented and qualitatively assessed for each significant difference in LIWC dictionaries.

Description of Typical Dropout and Completer Patients

In terms of likeliness (Table 3), a completer patient is a female (61.6%) around the age of 48 (*M*=47.74, *SD*=10.23) and of higher education (52.9%, WO & HBO) who is employed full-time (29.8%) and lives with family (33.3%). She would have lived 19 years (*M*=19.26, *SD*=11.5) with a problematic level of alcohol consumption and has a daily alcohol intake of 7 units (*M*=6.96, *SD*=4.96). Neither would she smoke, gamble nor take drugs and she would experience depressive symptoms at least sometimes (62.1%). Her reasons for participation would be the insight of her own alcohol overconsumption (76.3%) and she would have found the program by personal web research (12.7%). Her goal would be any combination of stopping and reducing consumption, relapse prevention or seeking advice and information (40.5%).

A typical dropout patient is almost equally likely male (49.3%) or female (50.7%), 44 years of age (*M*=44.43, *SD*=10.96), of higher education (42.2%, WO & HBO), is employed full-time and lives with family (29.8%). He or she would likely smoke (57.3%) and take drugs but neither take drugs nor gamble. He or she would experience depressive symptoms at least sometimes (64.9%). The reasons for participation would be the personal insight of alcohol overconsumption (79%), and the program would have been found by personal web research (21.9%). The goal would again be a combined one (44.8%). The dropout patient would have lived 18 years (*M*=17.91, *SD*=10.47) with problematic consumption and has a daily consumption of 8.14 units of alcohol (*M*=8.14, *SD*=7.03).

Table 3

Overview on the Self-Report Intake Questionnaire Split to Dropouts and Completers with the Chi-Square comparison

Variables		Dropout		Completer		Chi-Square	
		N	%	*N*	%	χ^2 *(df, N)*	*p*
Gender							
	Male	209	49.3	133	38.4	9.09 (1, 770)	.003*
	Female	215	50.7	213	61.6		
Nationality							
	Dutch	22	5.2	27	7.8	--	--
	No answer	402	94.8	319	92.2		
Housing							
	Alone	110	25.9	78	22.5	3.4 (3, 693)	.334
	With partner	102	24.1	99	28.6		
	With family	141	33.3	103	29.8		
	With children	31	7.3	29	8.4		
	No answer	40	9.4	37	10.7		
Education**							
	WO	42	9.9	59	17.1	15.1 (5, 740)	.01*
	HBO	137	32.3	124	35.8		
	MBO	103	24.3	68	19.7		
	IBO/MAVO/VMBO	69	16.3	40	11.6		
	HAVO/VWO	56	13.2	33	9.5		
	Grammar school	5	1.2	4	1.2		
	No answer	12	2.8	18	5.2		
Relationship status							
	Unmarried	6	1.4	4	1.2	--	--
	Divorced	3	.7	7	2		
	Widowed	0	0	1	.3		
	Married	4	.9	11	3.2		
	Living together	6	1.4	3	.9		
	LAT	1	.2	1	.3		
	No answer	404	95.3	319	92.2		
Employment							
	Self-employed	47	11.1	36	10.4	5.94 (7, 689)	.547
	Full-time	129	30.4	103	29.8		
	Part-time	81	19.1	78	22.5		

Variables		Dropout		Completer		Chi-Square	
	Unemployed	31	7.3	15	4.3		
		N	*%*	*N*	*%*	χ^2 *(df, N)*	*p*
	Unfit for work	28	6.6	26	7.5		
	Stay-at-home	35	8.3	21	6.1		
	Pensioner	17	4	17	4.9		
	Other Forms****	13	3.1	12	3.5		
	No answer	43	10.1	38	11		
Reason for participation							
	I think I drink too much	335	79	264	76.3	1.79 (3, 718)	.617
	Bad event	38	9	34	9.8		
	Seeking advice	15	3.5	10	2.9		
	Others think I drink too much	15	3.5	7	2		
	No answer	21	5	31	9		
Finding Alcohold-ebaas.nl							
	Internet search	93	21.9	44	12.7	9.46 (6, 368)	.149
	TV advertisement	16	3.8	13	3.8		
	Family or knowledge	19	4.5	10	2.9		
	Link on another website	13	3.1	14	4		
	Tactus	28	6.6	24	6.9		
	Recommended by professional	32	7.5	26	7.5		
	Newspaper or journal	17	4	19	5.5		
	No answer	206	48.6	196	56.6		
Smoking							
	Yes	243	57.3	149	43.1	13.53 (1, 721)	<.001*
	No	159	37.5	170	49.1		
	No answer	22	5.2	27	7.8		
Drugs							
	Yes	49	11.6	23	6.6	4.91 (1, 721)	.027*
	No	353	83.3	296	85.5		
	No answer	22	5.2	27	7.8		

Variables		Dropout		Completer		Chi-Square	
		N	%	*N*	%	χ^2 (*df, N*)	*p*
Gambling	Yes	19	4.5	8	2.3	2.429 (1, 721)	.119
	No	383	90.3	311	89.9		
	No answer	22	5.2	27	7.8		
Depression							
	Never to rarely	113	26.7	80	23.1	.33 (1, 683)	.564
	At least sometimes	275	64.9	215	62.1		
	No answer	36	8.5	51	14.7		
Goal							
	Stopping	50	11.8	28	8.1	5.4 (4, 587)	.249
	Reducing	60	14.2	56	16.2		
	Relapse prevention	31	7.3	30	8.7		
	Advice & Information	2	.5		0		
	Multiple reasons	190	44.8	140	40.5		
	No answer	91	21.5	92	26.6		
Psy. treatment							
	Depression related	3	.7	6	1.7	--	--
	Anxiety related	0	0	0	0		
	Other reasons****	14	3.3	14	4		
	No answer	407	96	326	94.2		

Note. For the Chi-Square test, the *No answer* category was excluded, hence the changes in *df*.
*Significant results for a standard level of significance of *p* =.05.
**The education variable is based on the Dutch school system. Their order represents the educational level.
****Other forms* of employment were students (*N*=10), voluntary workers (*N*=12) and '*mantelzorgers'* (*N*=3), which is the Dutch definition for people who stay at home to care for e.g. relatives
*****Other reasons* were considered to be any prior psychological treatment other than depression-, anxiety- or alcohol-related e.g. family therapy.

Differences in Patient Characteristics

The independent sample *t*-test for differences between the groups for the characteristics *age*, *years of problematic consumption* and *average daily alcohol intake,* showed a significant difference only for *age* (Table 4). It was found that dropouts (M=44.43, SD=10.96) compared to completers (M=47.74, SD=10.23), were significantly younger, $t(768) = -4.294$, p = <.001. The findings do not support the conclusion that there is a difference in *years of problematic alcohol consumption* or *average daily alcohol intake* between the groups.

Table 4

T-test for differences in Age, Years Lived with Problematic Consumption and Average Daily Consumption

	Dropout (N = 424)		Completer (N = 346)				95% CI	
Characteristics	*M*	*SD*	*M*	*SD*	*t* (768)	*p*	*LL*	*UL*
Age	44.43	10.96	47.74	10.23	-4.294	<.001*	-.4.82	-1.8
Years of Problematic Consumption	17.91	10.47	19.26	11.5	-.425	.673	-7.74	5.04
Average Daily Alcohol Intake	8.14	7.02	6.96	4.96	-.684	.497	-2.28	4.77

*Significant result for a standard level of significance of p =.05.

For the patient characteristics *gender*, *education*, *nationality*, *relationship status*, *employment*, *smoking*, *drugs*, *gambling*, *depression*, *prior psychological treatment, reason for participation* and *goal of the therapy* a Chi-square test was conducted (Table 3). For *gender,* significant differences between dropouts and completers were obtained, χ^2 (1, 770) = 9.09, p=.003. As shown in table 3, more females were in the completion group and more males among dropouts. *Education* showed also significant differences between the groups, χ^2 (5, 740) = 15.01, p=.01. Completers were found to be more represented in higher forms of education (HBO & WO) than dropouts. Additionally, *smoking* showed significant differences between the groups, χ^2 (1, 721) = 15.01 p<.001. As well as taking *drugs,* χ^2 (1, 721) = 4.91, p=.027. In both characteristics, dropouts were more represented than completers. None of the remaining patient characteristics showed significant results concerning differences between the dropout and completer group. The characteristic nationality had to be excluded from analysis. Only 22 participants replied being of Dutch nationality and because no other nationality appeared while the answer rate was low, a reliable Chi-square calculation could not be computed. The characteristics *relationship status* and *psychological treatment* had too much missing values as well to run a reliable Chi-square analysis and were therefore excluded as well.

Comparison of Linguistics

Comparison of LIWC Dictionaries

To analyze whether dropouts and completers differ in their choice of words, a series of independent samples *t*-tests was conducted to compare whether there is a relevant difference in means in the usage of words corresponding to the LIWC categories in question (Table 5). A significant result was obtained for the LIWC category *work* with the dropout group (M=0.019, SD=0.016) compared to the completer group (M=0.016, SD=0.011) having used significantly more words of said category in their emails, $t(749.7) = 2.534$, $p = .011$. Thus, dropouts tended to write more about work, being busy and jobs in their emails than completers did. The words with the highest number of matches was for the dictionary dictionary *focus past* and the dictionary with the least matches was *negative emotions* as can be seen in table 5.

The results for all other LIWC categories were not significant. A comparison between specific key words per LIWC category and their frequency, split to dropouts and completers, can be seen in Appendices B to J. The comparison of keywords showed a high overlap in words represented in the top 15 between the two groups and no substantial differences in ranking. Even for the LIWC category *work* that showed a significant difference, the top 15 words are identical between dropouts and completers, except for slight shifts in rank. While the frequency counts of words were usually higher for completers (Appendices B to J), this finding is not considered to be meaningful, as completers used more words in their emails. Therefore, higher word frequencies are attributable to the higher word use in general, yet not to differences in usage numbers compared between dropouts and completers.

Table 5

Comparison of Means for LIWC Categories Between Dropouts and Completers

	Dropout (N = 424)		Completer (N = 346)				95% CI	
LIWC	*M*	*SD*	*M*	*SD*	*t* (*df*)*	*p*	*LL*	*UL*
Positive Emotions	.034	.027	.036	.02	-.865 (768)	.387	-.005	-.002
Negative Emotions	.02	.014	.02	.012	-.11 (767.92)	.913	-.002	.002
Anxiety	.004	.005	.005	.002	-.405 (768)	.685	-.001	.001
Sadness	.005	.008	.005	.004	1.378 (660.12)	.169	.00	.001
Family	.005	.006	.005	.006	-.892 (768)	.373	-.001	.00
Insight	.036	.02	.037	.02	-.864 (768)	.388	-.004	.002
Focus Past	.06	.027	.064	.027	-1.82 (768)	.069	-.007	.00
Focus Future	.04	.022	.038	.017	1.781 (763.48)	.075	.00	.005
Work	.019	.016	.016	.01	2.534 (749.73)	.011**	.001	.004

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit. All values rounded to the third decimal.
*Non-integer *df* were reported when equal variance could not be assumed in the respective categories. SPSS uses the Satterthwaite approximation to compute *df* for unequal variances.
**Significant results for a standard level of significance of p =.05.

Qualitative Assessment of Email Excerpts

Concerning the textual representation of the significant finding for the LIWC dictionary work, the following excerpts from exemplary client emails will be presented. The first one is from a dropout participant, the second from a completer participant. Both were translated from Dutch by the researcher. Based on the intake questionnaire, the dropout was a 42 years old male, who lived with a partner, had an MBO education and was full-time employed. He participated because he felt that he was drinking too much. He was a smoker but did neither take other drugs nor did he gamble. He had indicated to feel depressed at least sometimes. He gave multiple reasons for his participation. For the other characteristics he did not provide the answers. His excerpt reads:

‘*I started drinking around one year after I had a hernia surgery and while I spent time at home for recovery. I had an understanding manager who proposed to change my position from service technician to planning technician which meant that I went from outside to inside. The reason for this was that I have a herniated disc and no more feeling in my right foot, which means that I cannot do hard labor. After a few months, my manager has resigned from his function and I got a new one, who could not empathize with my problems. The working atmosphere became unacceptable after some time and I started to go to work with lead in my shoes, every day. After a while they have given me a cancellation agreement so from DATE, I had no more work. PER have started since DATE to work for my new boss and up until now, it seems to be a very fun job. I do not have many hobbies, but I like to tinker, ride a bike and going for strolls* […]’

According to the intake questionnaire, the completer was a 37 years old female who lived with her family, had a HAVO or VWO education and was employed in part-time. She participated because she had experienced a bad situation in combination with her alcoholism and had the program recommended by Tactus. She said that she would neither smoke, take drugs or gamble. She said that she was experiencing depression at least sometimes and her goal for the treatment was relapse prevention. No answers were provided for the other characteristics. Her excerpt reads:

‘*There are periods where I just drink a few glasses of wine and there are periods where I cannot stop and then I continue until the next day, even in the morning. In the past days this happened often, for around NUM to NUM weeks. If I have such a period, I drink NUM bottles of wine or just whatever is available. I find it difficult to think about a real reason why, often it is after a bad event in the family, but then sometimes there is no real reason* […]. *The bad event was that my partner and child were very sad after I had again drunken alcohol for NUM days and I do not want to have this on my conscience. We try to split* […] *as good as possible, but I am doing the most of it, also because I am the one who works part-time. My job is secretary. Up until now, it did not have much influence on my job, but I have taken sick-leave NUM times and that is not right*. […]’

Given the two excerpts, it becomes apparent that the dropout patient extensively described his medical problems and how they affected his job. He extensively used work related words in the course of his descriptions of the situation deteriorating at his old job by describing the job, job titles, the kind of work he does as well as the relationship with his changing superiors. Adding to that, working in the sense of tinkering was a hobby of his, so work does not only play a role in his professional life, but also in his private and recreational time. Additionally, he began drinking when he could not work due to his surgery, and his medical implications contributed to the loss of his first job, presumably putting health, work and AUD in a triangular relationship. All these aspects contributed much to entries in the LIWC dictionary work.

For the completer patient, descriptions of her work and activities were less rich. She glanced over her job as a secretary and stated that her addiction did not yet affect the job. However, she presumably referred to work related to her role in house chores that she does the majority off since explains this with her a part-time job probably referring to her husband who might work full-time. Unfortunately, this section was not fully included in the data. Nevertheless, she indicated that her addiction might stem from problematic family affairs, though she also states that she does not really know why she continues drinking excessively. There are parts were work related words are used, yet they are fewer than with the dropout participant emphasizing the different foci these participants placed on describing their problem and situation.

Discussion

For this study, it was argued that LIWC can supplement dropout research by identifying possible linguistic differences between completers and dropouts of an online based alcohol therapy program. Significant linguistic differences between the groups were found for one LIWC dictionary, *work*, out of 9 and the comparison of keyword tables showed little divergence in choice of words in all dictionaries. The comparison of client characteristics showed a significant difference for the client characteristics *age*, *education*, *smoking* and taking *drugs* between the groups. Regarding these findings, the aim of the research is considered achieved. The significant differences found in characteristics further substantiate their role in dropout from therapy, while with work related writing a new linguistic differentiation between dropout and completers was discovered, successfully adding linguistic forms of differentiation in dropout research to non-linguistic patient characteristics.

Findings on Patient Characteristics

Drawing from the intake questionnaire from Alcoholdebaas.nl, this study aimed to generate findings that characterize and differentiate dropouts from completers based on patient characteristics theorized to predict dropouts by other studies. For the characteristic *age* it was found that dropouts were meaningfully younger than completers. This does substantiate the findings of Elbreder and colleagues (2011) and Vuoristo-Myllys and colleagues (2013) who found younger age to predict dropout. A potential explanation to this can be the suggestions of McKellar and colleagues (2006) that younger participants may have experienced less adverse consequences due to alcohol consumption and therefore see no necessity for change.

Concerning *education* this study found a meaningful difference between the groups, with completers being concentrated more in the higher levels of education than dropouts. Yet placing this finding in the context of prior studies is intricate. While Reinwand and colleagues (2015) found dropout to be more present in middle levels of education, Rizvi and colleagues (2009) as well as Watson and colleagues (2017) found lower education to be associated with dropout. Yet, with 42.2% of dropouts and 52.9% of completers in this study who have completed an applied sciences program, engineering or university program (HBO or WO), the educational levels in both groups were substantially higher than the Dutch population average of 32.5% in 2019 as shown by a publication by the Dutch Ministry of Education, Culture and Science on the Dutch website ‘Education in Numbers’ (Original Dutch title: *Onerdwijs in*

Cijfers) titled ‘Highest achieved level of education’ (Ministry of Education, Culture and Science, n.d.). While technically the finding of this study is in line with the ones of Rizvi and colleagues (2009) and Watson and colleagues (2017) since dropouts were found with lesser educational level, dropouts cannot be understood as less educated, since a major share of them was represented in higher forms of education. Since Dutch HBO degrees are called bachelor and engineering degrees (but not M.Sc.), this study does not necessarily contradict the findings of Reinwand and colleagues (2015) since a large share of dropouts was indeed concentrated in medium to high (MBO & HBO) forms of education. This in turn indicates, that the online therapy of Alcoholdebaas.nl catered more towards individuals of higher education than lower education, substantiating the findings of Hall and colleagues (2015) that the higher educated engage in more health related internet usage. Yet since the lower educated were presented to be a high-risk group for AUD, it would have been better if this target group would have made more use of the Alcoholdebaas.nl intervention. The top-heavy distribution in educational level among both groups points out that the lower educated groups were underrepresented.

The findings that there were substantially more smokers among dropouts than completers is in line with the findings of others (Roberts, Murphy, Turner, & Sharman, 2019; ter Huurne et al., 2017) that found lifetime smoking and baseline smoking to be a predictor of dropout. However, none provided an explanation for this finding. It might be explained by looking at promising findings in gene research that try to identify genetic characteristics that might explain proneness to substance addictions (Brewer & Potenza, 2008; Yang & Li, 2016). If there indeed were genetic factors making individuals more susceptible to substance addictions, dropouts from AUD and other substance abuse treatments might potentially show these genetic characteristics. In turn, this would hold explanatory value for the significant finding that dropouts have more often taken drugs than completers as found in this study. Adding to that, others have theorized drug-use to be related to dropout as withdrawal symptoms and not being able to drown arising emotions in substances may cause relapse and discontinuation of treatment (Lejuez et al., 2008).

For *gender*, a meaningful difference between the groups with more females completing the intervention was found. This finding is in line with the studies of Radtke and colleagues (2017) and Postel and colleagues (2011) who found females to be more represented in the completer group in online AUD therapy. This result is opposing the findings of Darke and colleagues (2012) who found males to complete more often and other studies who found no relationship between dropout and gender (Elbreder et al., 2011;

Greenberg & Swift, 2012).

For the variable *relationship status* answer rates of dropouts and were not sufficient for any conclusions to be drawn from this study. Therefore, findings of other studies can neither be considered substantiated nor opposed.

On *employment status* completers and dropouts showed no meaningful differences. While the overall fraction of unemployed participants was low, the difference between dropouts and completers was the largest in this category and more dropouts were unemployed than completers. This would have been in line with the findings of ter Huurne and colleagues (2017), who found unemployment to be associated with dropout, if the difference would have been significant.

No meaningful difference in *average daily alcohol intake* was found between the groups. This is opposing the findings of Radtke and colleagues (2017) and Ray and colleagues (2007) that either dropout or completion would be related to general alcohol consumption. Considering daily consumption a benchmark of AUD severity, the insignificant difference opposes the findings of McMurran, Huband and Overton (2010) who found symptom severity to be related to dropout in various other psychopathologies.

It is worth pointing out, that substantially more dropouts found the program via individual online search. One could interpret that in the sense that dropouts sought for help more spontaneously once they became aware of their problematic consumption. Since the willingness to seek help for AUD often arises only momentarily (Cunningham et al., 1994), they might have signed up to the intervention by individual internet research out of momentary pressure, yet when the intervention began, their willingness to receive help had already diminished again. Further substantiation to this idea can be seen in the high number of individuals who signed up for the treatment, though never sent an email to their therapist. However, this is just theorized and would require additional testing to confirm.

Linguistic Comparison Between Dropouts and Completers

This study aimed to analyze patient emails to identify linguistic differences between dropouts and completers for several LIWC categories that were hypothesized to relate to dropout. Yet meaningful differences in linguistics were only found for the LIWC category *work*, with dropouts using more words related to work, job or being busy than completers. By further investigating the LIWC categories it was found that not only were the majority of differences in LIWC category averages negligible, but the exact choice of words was remarkably similar.

Except for *work* related linguistics, the general findings indicate that there are no substantial linguistic differences between the groups in the selected LIWC dictionaries.

The meaningful difference in usage of words related to *work*, with dropouts using more work-related words than completers, is a novelty to the knowledge of the author. Though other researchers had found differences between dropouts and completers in employment status (ter Huurne et al., 2017; Watson et al., 2017), this study was the first to find a linguistic representation of a difference related to work between the groups. Unfortunately, the mentioned studies did not include a theoretic reflection on why employment would be associated with dropout. It can be argued that the significant difference for the LIWC category *work* could be related to the educational levels of dropouts and completers. With completers being represented in higher, academic education and dropouts being represented in higher to medium education, dropouts might be more likely to work in practical jobs. This practical work might be more reflected in the LIWC category *work*, than academic work. Looking at the qualitative assessment of the dropout patient mail, this tendency is worth considering for further investigation. Yet the nature of the work the two groups did, was not analyzed and the keywords do not suggest a difference or that dropouts did more practical work. Therefore, this assumption cannot be considered substantiated and is only hypothetical. A more straightforward explanation of the finding can be that the dictionary *work* picks up words related to being busy and leading a laborious life. Having no time, is a factor found in preventing individuals from seeking treatment (Clough et al., 2019). Presumably, dropouts may drop-out because they perceive their life as too stressful and busy to continue therapy. Definitely, this new finding needs further exploration in other studies, as well as in the context of why these groups wrote about *work* differently.

Concerning the other LIWC dictionaries, no further meaningful differences between the groups were found, despite the theoretic background of their inclusion. While Geraghty and colleagues (2010) had hypothesized that increasing positive affect in participants increases client retention, this study did not find that speech related to *positive emotions* was meaningfully different between dropouts and completers. Reversely, while Deane and colleagues (2012) argued for negative affectional states being potentially able to predict dropout, words indicating *negative emotions* were not found to differ between the groups in this study. Related to this, comorbid depression was found to predict dropout (Corrêa Filho & Baltieri, 2012; Markowitz et al., 2015), yet this study did not find a meaningful difference in *sadness* related word use. In the light of this finding, substantiation is given to the findings of Kavanagh and colleagues (2006), who did not find depression being associated to dropout. A possible explanation for this might be that the overall rate of individuals showing any form of depressive symptoms was high in both groups from the start, likely due to the comorbidity of depression and AUD (Subramaniam et al., 2017). Additionally, the measurement of depression used in this study was based on the frequency of depressive episodes, yet not their intensity. This means that depressed patients in this study were not necessarily as strongly affected by depression as the participants in the study of Markowitz and his colleagues (2015). Next to that, no difference between dropouts and completers were found for *anxiety* despite other studies finding anxiety to predict dropout (Krishnamurty et al., 2015). *Family* did not show meaningful differences between the groups either, despite Lowe and colleagues (2013) who found family related talk to be a major topic in alcohol related patient writings. Then again, if the findings report no difference between the groups and family related words being used more frequently than words of other dictionaries it could highlight the role family plays in AUD therapy for dropouts as well as completers, which might be worth investigating. *Insight* did not indicate meaningful differences between the groups as well, which opposes the findings that lack insight predicts dropout (Lincoln et al., 2014). Both *focus past* and *focus future* did not indicate meaningful differences between the groups, which is not in line with the findings of Alfonsson and colleagues (2016) that dropouts were less focused on the future. Reversely, focus past was theorized to be associated with ruminating in the past (Eisma et al., 2015), yet the findings of this study do not substantiate this theory. Though with the most used words for the dictionary *focus past,* exploring rumination might hold interesting findings for the role of the past on AUD therapy.

There are several additional possible explanations to these findings. First, other studies that investigated linguistic differences between groups compared groups that can be argued to differentiate more from each other than the ones of this study. Liehr and colleagues (2010) compared individuals participating in a mindfulness-based therapy to a control group not receiving mindfulness therapy. However, in this study, dropouts as well as completers participated in the same treatment program, answering the same questions and discussing the same material which might have led to similarities in the way they discussed this in therapy. Another explanation might be the choice of emails that were analyzed. By only analyzing the first four emails of both groups, meaningful differences between the groups might not have had enough time to develop based on the progression with the therapy like it was the case in the study of Arntz and colleagues (2012). To that end, analyzing the emails week by week would have been a better choice to pick up on changes in word choice in the course of therapy. Additionally, the first emails by the therapist included many questions about the general state of life like health, family situation and employment. This set of structured questions might have led the patients in their way of providing comparable answers. Another noteworthy finding is the difference in amount of words used by the two groups in their emails. Since the completers wrote much longer emails, a possible connection between email length and dropout could be that completers were more engaged in their therapy. By elaborating more on their situation, therapists in turn have more information to work with and can provide potentially better care creating a positive feedback cycle for the therapy. Yet verifying this requires more information that goes beyond just email lengths as participant satisfaction with therapy needs to be measured.

Text Mining as Method to Analyze Textual Patient Data

This study aimed to make a case for employing TM methods as means to effectively and efficiently analyze the large amount of data that online therapy provides. To this matter, online treatment programs like *Alcoholdebaas.nl* provide a well-suited surrounding. The therapist-client email correspondences provide a record of the therapy ready for TM analysis, provided that protection of individual privacy is safeguarded. Concerning the workload, the number of patient cases that can be analyzed is nearly irrelevant and only limited by computational power, whereas workload would scale linearly when human coders had been employed. Regarding speed and scale, TM and TA are ideally suited to keep up with the increase of online therapy that generates large amounts of textual data. This is of special value to online AUD therapy since interventions for substance abuse disorders form the biggest group of evidence-based online treatment programs. Therefore, it can be argued that AUD online therapy will produce a large share of data among online therapy programs and thus, will specially profit from utilizing automated analyses methods. This data will be very valuable for research to understand participants of online AUD therapies as the sample sizes will be higher than in traditional therapy settings, producing more robust research findings in online AUD and dropout research.

In terms of adequacy as a tool of measurement for this linguistic data, TM approaches excel when it comes to reliability, yet there are constrains to the validity of measurements. Reliability refers to the precision of a measurement in giving out results free of errors and yielding the same result every time the measurement is repeated with the data in the same setting (Parveen & Showkat, 2017). For computerized TM, the software strictly adheres to the rules it has been provided with, executes the same measurement repeatedly and exactly as it was programmed to. Though texts can have ambiguous meaning, metaphors and irony that are difficult to detect for computers (Spinczyk, Nabrdalik, & Rojewska, 2018), that can impact the reliability of measurements.

Concerning validity, defined as a methods ability to actually measure the constructs it intends to measure (Parveen & Showkat, 2017), for this studies proof-of-concept approach, achieving high levels of validity was less of a concern. Still, it has to be pointed out that the LIWC dictionaries are not clinically accurate measurement tools for sociodemographic as well as psychologic characteristics. Additionally, the selection of LIWC dictionaries was based on the scarce information that exists on AUD dropout, so with more research in the topic that can be used to adjust the LIWC dictionaries, validity can be expected to improve. For this study, the used LIWC dictionaries were considered to be approximations for the patient characteristics and the scarce findings from prior research indicated their possible connections to dropout. Still, one can argue that since only one linguistic difference between the groups was found, the validity of LIWC dictionaries as a measurement tool for linguistic patient characteristics was low. Then again, not all 73 categories were tested, and it is possible to create additional LIWC dictionaries that might better pick up on AUD related characteristics like the dictionary developed by Jensen & Hussong (2019) designed to capture alcohol related talk in students. Ultimately, TM approaches are not yet used extensively to investigate therapy dropout and therefore have less methodologic and theoretic background to build from. Still, the study exemplified the effectiveness and efficiency of the LIWC in dropout research that future research can draw from and improve on.

Strong Points and Limitations

A clear strong point is the relevancy of TM approaches to online AUD psychotherapy. As presented initially, there is a shift in therapy moving towards online treatment possibilities to take up the demand for therapy and AUD therapy is one of the largest among them. Analyzing the linguistic products of online therapy will most likely be a main path to understand how therapy is conducted online. Speed and automatization by using TM would make it a logical choice concerning effectiveness, efficiency and reliability. Another strong point is that by applying the TM methodology to the analysis of dropouts from psychotherapy new pathways can be created to analyze patient dropout. Obtaining data from dropouts that go beyond demographics and intake questionnaires are often difficult to obtain as dropouts are often lost to follow-up questions (Zandberg et al., 2016). Here, text mining can be a valuable alternative to supplement analysis methods on dropout research by analyzing the linguistic data they produced during their time in therapy. This study had data from 770 client cases, which is a lot more than others used in their studies to investigate AUD and dropout (Carolina et al., 2008; Jensen & Hussong, 2019). This makes the findings of patient characteristics more robust since outliers affect the results less than in studies with fewer participants. A further strongpoint is that the data obtained from Tactus was of participants that followed a professional addiction intervention delivered by experts. Therefore, the interactions between patients and therapists are more comparable to other therapeutic settings, like the classic face-to-face-therapy, than for example mining texts from unguided social media platforms like Facebook or Twitter. At the present there is not much research on patient linguistics in web-based AUD treatment programs that is focused on dropouts. Therefore, this study highlighted a gap in research and produced first findings in the subject by emphasizing the applicability of TM in dropout related research. In turn, creating groundwork for future research.

However, the study has limitations as well. While LIWC excels in counting words and sorting them, wordcounts alone miss out on the meaning of sentences. For instance, '*I love my family*' and '*I hate my family*' both contribute equally to the LIWC category family, yet the sentences have opposing meanings. The LIWC dictionary family cannot identify in what context family was spoken about. Yet this distinction might be relevant to further understand the relationship of LIWC categories to dropout. This problem may also be present with different tenses. Sentences like '*I used to be angry*' and '*I am angry*' both equally contribute to the dictionary of *negative emotions*, yet the first sentence does not indicate the current angriness of the author as it is a description of the past. This might unjustly inflate LIWC categories, while the author was describing a situation that might not affect him any longer. Related to that, the process of psychotherapy requires clients to discuss and elaborate on different topics, themes and emotions more than he or she normally would. Asking a patient to describe personal sadness magnifies this topic and the patient may spend more words describing this emotional state, in turn inflating the word counts. Generally, the linguistic results of this study concerning emotional states, but also other categories, should be interpreted cautiously and clients using e.g. more *sadness* related words should not considered to represent more depressed individuals. The original Dutch texts might have produced problems implications to this study as well. For instance, the Dutch word '*ben'* (like in '*ik ben'*, meaning '*I am*') was repeatedly mistaken for the name Ben and anonymized. Yet, it is difficult to identify all words that were censored instead of being matched to LIWC dictionaries to see the effect of this problem. While the translation was found to represent the

English LIWC categories fairly reliably, there might still have been words, especially slang words, that might not have been sorted into the correct LIWC dictionary. Yet another limitation is, that LIWC dictionaries are no ideal representation of e.g. sociodemographic characteristics. E.g. to measure education, no LIWC dictionary was deemed appropriate. Thus, LIWC is lacking in dictionaries that could capture the full range of sociodemographic characteristics and can only offer approximations.

Future Research and Closing Statement

Based on the findings of this study, several options to further the dropout research can be pointed out. Since it was found that the LIWC dictionaries lack in capturing the full extent of patient characteristics, it is proposed to further the creation of dictionaries to increase the applicability to LIWC for AUD dropout research. This is exemplified by the dictionary created by Jensen & Hussong (2019) and their success in capturing alcohol related talk using it. To capture the educational level of the author by linguistics alone for instance, one could aggregate the number of spelling mistakes and the complexity of the used vocabulary by the author to derive at a measurement that might reflect the educational level. Of course, it would require to be validated if spelling and choice of words reflect differences in educational level. Yet, creating such measurement tools could provide more substantiated LIWC dictionaries to capture differences in education since LIWC offers none currently. Another interesting analysis would be to include the emails of counsellors as well and to analyze if there are moderating effects of the language used by counsellors on the language used by clients. For instance, if a client uses a lot of negative words in an email and the counsellor uses more positive words in the reply, what would the effect be on the following client email? Such an analysis could give valuable insights into the linguistic processes of online therapy and how language influences dropout. Additionally, if no or only low moderating effects between counsellor language and client language are found, it might indicate that the working alliance is not functional since the two conversation partners are not responsive to texts of each other. Such client-counsellor relationship analyses might hold additional value in explaining client dropout by exploring the phenomenon even deeper than sorting dropouts and completers according to characteristics.

Still, the growing need for treatment possibilities makes online therapy a staple component of the current and future health professions. The adoption of TA approaches can be considered a necessity, when installing text-reliant online services. Since high dropout rates are a major implication for online therapy to live up to its potential, it is a challenge that requires to be addressed by researchers and intervention designers. The proposed method of using linguistics for analysis of dropouts is an ideal tool that could be integrated into online treatment infrastructure so it can monitor patient linguistics for signs of dropouts constantly. Identifying such signs could notify treatment providers that a patient needs extra attention to accomplish their treatment goals. While this study finding a first meaningful linguistic characteristics associated with dropout, it is pointed out that future research build on that and follow in this path. If further characteristics are identified, TM and TA approaches can realize their potential as a cost efficient and effective tool to improve online therapy administration and therefore save resources for the health care system while ultimately aiding patients in achieving the needed symptom relief.

Acknowledgements

For this M.Sc. thesis it was tapped into information and techniques of multiple fields such as computer sciences, data sciences and psychology which required multidisciplinary collaboration. Dr. E.T. Sang contributed with data pre-processing, data handling and setting up the various software utilized in this study. W.A.C. Smink, M.Sc. and Dr. A.M. Sools contributed with setting up the thesis and providing feedback throughout its process.

Ethic Statement

The patient emails from *Alcoholdebaas.nl* program included sensible information of the patients and therefore required sensitive handling. Patients had given their consent to make their data available for future research projects and the *What Works When for Whom* project as the superordinate project to this thesis had received ethical improvement from the University of Twente in the Netherlands. As already stated, the patient data had been anonymized and was securely stored on an encrypted and password protected device only the researchers had access to, to maximize privacy protection.

Funding

The project *What Works When for Whom* received its support from the Accelerating Scientific Discovery Call grant by the eScience Centre Amsterdam in the Netherlands (027.015.G04). Said eScience Centre receives its funding from the Netherlands Organisation for Scientific

Research (abbreviated NOW, *Dutch*: Nederlands organisatie for Wetenschappelijk Onderzoek). However, this thesis research received no funding.

References

Abbe A, Grouin C, Zweigenbaum P, Falissard B. Text mining applications in psychiatry: a systematic lit- erature review. International journal of methods in psychiatric research. 2016 Jun; 25(2):86–100. https://doi.org/10.1002/mpr.1481

Alcoholdebaas.nl (n.d.) Internetbehandeling. https://alcoholdebaas.nl/index.html

Alfonsson, S., Olsson, E., & Hursti, T. (2016). Motivation and treatment credibility predicts dropout, treatment adherence, and clinical outcomes in an internet-based cognitive behavioral relaxation program: A randomized controlled trial. *Journal of Medical Internet Research*, *18*(3). https://doi.org/10.2196/jmir.5352

Anderson, R. E. E., Spence, S. H., Donovan, C. L., March, S., Prosser, S., & Kenardy, J. (2012). Working alliance in online cognitive behavior therapy for anxiety disorders in youth: Comparison with clinic delivery and its role in predicting outcome. *Journal of Medical Internet Research*, *14*(3). https://doi.org/10.2196/jmir.1848

Arntz, A., Hawke, L. D., Bamelis, L., Spinhoven, P., & Molendijk, M. L. (2012). Changes in natural language use as an indicator of psychotherapeutic change in personality disorders. Behaviour research and therapy, 50(3), 191-202.

Back, S. E., Dansky, B. S., Carroll, K. M., Foa, E. B., & Brady, K. T. (2001). Exposure therapy in the treatment of PTSD among cocaine-dependent individuals: Description of procedures. *Journal of Substance Abuse Treatment*, *21*(1), 35–45. https://doi.org/10.1016/S0740-5472(01)00181-7

Bagnardi, V., Rota, M., Botteri, E., Tramacere, I., Islami, F., Fedirko, V., … La Vecchia, C. (2013). Light alcohol drinking and cancer: A meta-analysis. *Annals of Oncology*, *24*(2), 301–308. https://doi.org/10.1093/annonc/mds337

Barrett, M. S., Chua, W. J., Crits-Christoph, P., Gibbons, M. B., & Thompson, D. (2008). Early Withdrawal From Mental Health Treatment: Implications for Psychotherapy Practice. *Psychotherapy*, *45*(2), 247–267. https://doi.org/10.1037/0033-3204.45.2.247

Baumann, M., Spitz, E., Guillemin, F., Ravaud, J. F., Choquet, M., Falissard, B., … Mur, J. M. (2007). Associations of social and material deprivation with tobacco, alcohol, and psychotropic drug use, and gender: A population-based study. *International Journal of Health Geographics*, *6*, 1–12. https://doi.org/10.1186/1476-072X-6-50

Belleau, E. L., Chin, E. G., Wanklyn, S. G., Zambrano-Vazquez, L., Schumacher, J. A., & Coffey,

S. F. (2017). Pre-treatment predictors of dropout from prolonged exposure therapy in patients with chronic posttraumatic stress disorder and comorbid substance use disorders. *Behaviour Research and Therapy*, *91*, 43–50. https://doi.org/10.1016/j.brat.2017.01.011

Berger, M., Wagner, T. H., & Baker, L. C. (2005). Internet use and stigmatized illness. *Social Science and Medicine*, *61*(8), 1821–1827. https://doi.org/10.1016/j.socscimed.2005.03.025

Bliuc, A. M., Doan, T. N., & Best, D. (2019). Sober social networks: The role of online support groups in recovery from alcohol addiction. *Journal of Community and Applied Social Psychology*, *29*(2), 121–132. https://doi.org/10.1002/casp.2388

Brewer, J. A., & Potenza, M. N. (2008). The neurobiology and genetics of impulse control disorders: relationships to drug addictions. *Biochemical pharmacology*, *75*(1), 63-75.

Brouwer, W., Oenema, A., Crutzen, R., de Nooijer, J., de Vries, N. K., & Brug, J. (2009). What makes people decide to visit and use an internet-delivered behavior-change intervention?: A qualitative study among adults. *Health Education*, *109*(6), 460–473. https://doi.org/10.1108/09654280911001149

Can, D., Marín, R. A., Georgiou, P. G., Imel, Z. E., Atkins, D. C., & Narayanan, S. S. (2016). "It sounds like ...": A natural language processing approach to detecting counselor reflections in motivational interviewing. *Journal of Counseling Psychology*, *63*(3), 343–350. https://doi.org/10.1037/cou0000111

Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cognitive Behaviour Therapy*, *47*(1), 1–18. https://doi.org/10.1080/16506073.2017.1401115

Carolina, S., Informa, R., Number, W. R., House, M., Street, M., Policy, S., … Out, D. (2008). *Factors Associated with Dropping Out from Outpatient Treatment of Alcohol- Other Drug Abuse Factors Associated with Dropping Out from Outpatient Treatment of Alcohol-Other Drug Abuse*. (August 2014), 37–41. https://doi.org/10.1300/J020v20n02

Chapman, C., Slade, T., Hunt, C., & Teesson, M. (2015). Delay to first treatment contact for alcohol use disorder. *Drug and Alcohol Dependence*, *147*, 116–121. https://doi.org/10.1016/j.drugalcdep.2014.11.029

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

Cloud, R. N., & Peacock, P. L. (2001). Internet screening and interventions for problem drinking: Results from the www.carebetter.com pilot study. *Alcoholism Treatment Quarterly*, *19*(2), 23–44. https://doi.org/10.1300/J020v19n02_02

Clough, B. A., March, S., Leane, S., & Ireland, M. J. (2019). What prevents doctors from seeking help for stress and burnout? A mixed-methods investigation among metropolitan and regional-based australian doctors. *Journal of Clinical Psychology*, *75*(3), 418–432. https://doi.org/10.1002/jclp.22707

Coa, K., & Patrick, H. (2016). Baseline Motivation Type as a Predictor of Dropout in a Healthy Eating Text Messaging Program. *JMIR MHealth and UHealth*, *4*(3), e114. https://doi.org/10.2196/mhealth.5992

Cook, J. E., & Doyle, C. (2002). Working alliance in online therapy as compared to face-to-face therapy: Preliminary results. *Cyberpsychology and Behavior*, *5*(2), 95–105. https://doi.org/10.1089/109493102753770480

Cook, S., Heather, N., & McCambridge, J. (2015). The Role of the Working Alliance in Treatment for Alcohol Problems. *Psychology of Addictive Behaviors*, *29*(2), 371–381. https://doi.org/10.1037/adb0000058

Corrêa Filho, J. M., & Baltieri, D. A. (2012). Preditores psicossociais e clínicos de retenção ao tratamento para alcoolismo. *Revista Brasileira de Psiquiatria*, *34*(4), 413–421. https://doi.org/10.1016/j.rbp.2012.03.003

Crombie, I. K., Irvine, L., Williams, B., Sniehotta, F. F., Petrie, D., Jones, C., … Achison, M. (2018). Texting to Reduce Alcohol Misuse (TRAM): main findings from a randomized controlled trial of a text message intervention to reduce binge drinking among disadvantaged men. *Addiction*, *113*(9), 1609–1618. https://doi.org/10.1111/add.14229

Cunningham, J. A., Sobell, L. C., Sobell, M. B., & Gaskin, J. (1994). Alcohol and drug abusers' reasons for seeking treatment. *Addictive Behaviors*, *19*(6), 691–696. https://doi.org/10.1016/0306-4603(94)90023-X

Darke, S., Campbell, G., & Popple, G. (2012). Retention, early dropout and treatment completion among therapeutic community admissions. *Drug and Alcohol Review*, *31*(1), 64–71. https://doi.org/10.1111/j.1465-3362.2011.00298.x

Dauber, H., Pogarell, O., Kraus, L., & Braun, B. (2018). Older adults in treatment for alcohol use disorders: Service utilisation, patient characteristics and treatment outcomes. *Substance Abuse: Treatment, Prevention, and Policy*, *13*(1), 1–10. https://doi.org/10.1186/s13011-018-0176-z

Deane, F. P., Wootton, D. J., Hsu, C. I., & Kelly, P. J. (2012). Predicting dropout in the first 3 months of 12-step residential drug and alcohol treatment in an Australian sample. *Journal of Studies on Alcohol and Drugs*, *73*(2), 216–225. https://doi.org/10.15288/jsad.2012.73.216

Degenhardt, L., Charlson, F., Ferrari, A., Santomauro, D., Erskine, H., Mantilla-Herrara, A., … Vos, T. (2018). The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Psychiatry*, *5*(12), 987–1012. https://doi.org/10.1016/S2215-0366(18)30337-7

Degni, F., Vaherkylä, M., & Hurme, S. (2017). Health Behavior and Self-Assessed Health Among Some Long-Term Unemployed Living in Turku, Finland. *The Open Public Health Journal*, *10*(1), 232–243. https://doi.org/10.2174/1874944501710010232

Demsar, et al. (2013). Orange: data mining toolbox in Python. *International Journal of Conservation Science*, *7*(SpecialIssue1), 295–300.

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*, 37-46.

Dunlop, W. L., & Tracy, J. L. (2013). Sobering stories: Narratives of self-redemption predict behavioral change and improved health among recovering alcoholics. *Journal of Personality and Social Psychology*, *104*(3), 576–590. https://doi.org/10.1037/a0031185

Eisma, M. C., Boelen, P. A., van den Bout, J., Stroebe, W., Schut, H. A. W., Lancee, J., & Stroebe, M. S. (2015). Internet-Based Exposure and Behavioral Activation for Complicated Grief and Rumination: A Randomized Controlled Trial. *Behavior Therapy*, *46*(6), 729–748. https://doi.org/10.1016/j.beth.2015.05.007

Elbreder, M. F., de Souza e Silva, R., Pillon, S. C., & Laranjeira, R. (2011). Alcohol dependence: Analysis of factors associated with retention of patients in outpatient treatment. *Alcohol and Alcoholism*, *46*(1), 74–76. https://doi.org/10.1093/alcalc/agq078

Elisabeth, A., Stacey, L., & Maintaining, J. D. (2009). QUT Digital Repository : This is the author version published as : This is the accepted version of this article . To be published as : This is the author version published as : *Health (San Francisco)*, *13*(December), 913–917.

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*(May), 29–38. https://doi.org/10.1016/j.invent.2017.09.001

Fotheringham, M. J., Owies, D., Leslie, E., & Owen, N. (2000). Interactive health communication in preventive medicine: Internet-based strategies in teaching and research. *American Journal of Preventive Medicine*, *19*(2), 113–120. https://doi.org/10.1016/S0749-3797(00)00188-4

Fox, S., & Duggan, M. (2013). *Health Online 2013*. Retrieved from https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf (Archived by WebCite at https://www.webcitation.org/74sbsnhYG)

Geraghty, A. W. A., Wood, A. M., & Hyland, M. E. (2010). Attrition from self-directed interventions: Investigating the relationship between psychological predictors, intervention content and dropout from a body dissatisfaction intervention. *Social Science and Medicine*, *71*(1), 30–37. https://doi.org/10.1016/j.socscimed.2010.03.007

Gersh, E., Hallford, D. J., Rice, S. M., Kazantzis, N., Gersh, H., Gersh, B., & McCarty, C. A. (2017). Systematic review and meta-analysis of dropout rates in individual psychotherapy for generalized anxiety disorder. *Journal of Anxiety Disorders*, *52*(December), 25–33. https://doi.org/10.1016/j.janxdis.2017.10.001

Goldstein, B., Bradley, B., Ressler, K. J., & Powers, A. (2017). Associations Between Posttraumatic Stress Disorder, Emotion Dysregulation, and Alcohol Dependence Symptoms Among Inner City Females. *Journal of Clinical Psychology*, *73*(3), 319–330. https://doi.org/10.1002/jclp.22332

Graff-Iversen, S., Jansen, M. D., Hoff, D. A., Høiseth, G., Knudsen, G. P., Magnus, P., … Tambs, K. (2013). Divergent associations of drinking frequency and binge consumption of alcohol with mortality within the same cohort. *Journal of Epidemiology and Community Health*, *67*(4), 350–357. https://doi.org/10.1136/jech-2012-201564

György Purebl, Ionela Petrea, Laura Shields, Mónika Ditta Tóth, András Székely, Tamás Kurimay, David McDaid, Ella Arensman, Isabela Granic, K. M. A. (2017). *Joint Action on*

Mental Health and Wellbeing: Situation analysis and recommendations for action. 210. https://doi.org/10.1155/2015/281616

Hall, A. K., Bernhardt, J. M., Dodd, V., & Vollrath, M. W. (2015). The Digital Health Divide: Evaluating Online Health Information Access and Use Among Older Adults. *Health Education and Behavior*, *42*(2), 202–209. https://doi.org/10.1177/1090198114547815

Hammer, J. H., Parent, M. C., Spiker, D. A., & World Health Organization. (2018). Global status report on alcohol and health 2018. In *Global status report on alcohol* (Vol. 65). https://doi.org/10.1037/cou0000248

Hanley, T., Ersahin, Z., Sefi, A., & Hebron, J. (2017). Comparing Online and Face-to-Face Student Counselling: What Therapeutic Goals Are Identified and What Are the Implications for Educational Providers? *Journal of Psychologists and Counsellors in Schools*, *27*(1), 37–54. https://doi.org/10.1017/jgc.2016.20

Hans, E., & Hiller, W. (2013). Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: A meta-analysis of nonrandomized effectiveness studies. *Journal of Consulting and Clinical Psychology*, *81*(1), 75–88. https://doi.org/10.1037/a0031080

Hilbert, M., & López, P. (2014). *The World ' s Technological Capacity*. *60*(2011), 60–66. https://doi.org/10.1126/science.1200970

Hoogendoorn, M., Berger, T., Schulz, A., Stolz, T., & Szolovits, P. (2017). Predicting Social Anxiety Treatment Outcome Based on Therapeutic Email Conversations. *IEEE Journal of Biomedical and Health Informatics*, *21*(5), 1449–1459. https://doi.org/10.1109/JBHI.2016.2601123

Hotho, A., Nürnberger, A., & Paaß, G. (2005, May). A brief survey of text mining. In Ldv Forum (Vol. 20, No. 1, pp. 19-62).

Hussain, D., Ross, P., & Bednar, P. (2018). The perception of the benefits and drawbacks of internet usage by the elderly people. *Lecture Notes in Information Systems and Organisation*, *23*, 199–212. https://doi.org/10.1007/978-3-319-62051-0_17

Imel, Z. E., Steyvers, M., & Atkins, D. C. (2015). Computational psychotherapy research: Scaling up the evaluation of patient-provider interactions. *Psychotherapy*, *52*(1), 19–30. https://doi.org/10.1037/a0036841

Jensen, M., & Hussong, A. M. (2019). Text message content as a window into college student drinking: Development and initial validation of a dictionary of "alcohol-talk."

International Journal of Behavioral Development. https://doi.org/10.1177/0165025419889175

Johnson, J. (2020, February 10). *Internet Usage in Europe - Statistics & Facts.* Statista. https://www.statista.com/topics/3853/internet-usage-in-europe/

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

Kavanagh, D. J., Sitharthan, G., Young, R. M., Sitharthan, T., Saunders, J. B., Shockley, N., & Giannopoulos, V. (2006). Addition of cue exposure to cognitive-behaviour therapy for alcohol misuse: A randomized trial with dysphoric drinkers. *Addiction*, *101*(8), 1106–1116. https://doi.org/10.1111/j.1360-0443.2006.01488.x

Kay Lambkin, F. J. (2014). Lifetime Affective and Alcohol Use Disorder: Impact of Comorbidity on Current Functioning and Service Use in a Rural Population. *Journal of Addiction Research and Therapy*, *s10*. https://doi.org/10.4172/2155-6105.s10-009

Krishnamurthy, P., Khare, A., Klenck, S. C., & Norton, P. J. (2015). Survival modeling of discontinuation from psychotherapy: A consumer decision-making perspective. *Journal of Clinical Psychology*, *71*(3), 199–207. https://doi.org/10.1002/jclp.22122

Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy*, *50*(1), 42–51. https://doi.org/10.1037/a0030682

Lejuez, C. W., Zvolensky, M. J., Daughters, S. B., Bornovalova, M. A., Paulson, A., Tull, M. T., ... & Otto, M. W. (2008). Anxiety sensitivity: A unique predictor of dropout among inner-city heroin and crack/cocaine users in residential substance use treatment. *Behaviour research and therapy, 46*(7), 811-818.

Levy, H., Janke, A. T., & Langa, K. M. (2015). Health Literacy and the Digital Divide Among Older Americans. *Journal of General Internal Medicine*, *30*(3), 284–289. https://doi.org/10.1007/s11606-014-3069-5

Li,Lau, P., 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

Liehr, P., Marcus, M. T., Carroll, D., Granmayeh, L. K., Cron, S. G., & Pennebaker, J. W. (2010). Linguistic analysis to assess the effect of a mindfulness intervention on self-change for adults in substance use recovery. *Substance Abuse*, *31*(2), 79–85. https://doi.org/10.1080/08897071003641271

Lincoln, T. M., Rief, W., Westermann, S., Ziegler, M., Kesting, M. L., Heibach, E., & Mehl, S. (2014). Who stays, who benefits? Predicting dropout and change in cognitive behaviour therapy for psychosis. *Psychiatry Research*, *216*(2), 198–205. https://doi.org/10.1016/j.psychres.2014.02.012

Lowe, R. D., Heim, D., Chung, C. K., Duffy, J. C., Davies, J. B., & Pennebaker, J. W. (2013). In verbis, vinum? Relating themes in an open-ended writing task to alcohol behaviors. *Appetite*, *68*, 8–13. https://doi.org/10.1016/j.appet.2013.04.008

Luedke, A. J., Peluso, P. R., Diaz, P., Freund, R., & Baker, A. (2017). Predicting Dropout in Counseling Using Affect Coding of the Therapeutic Relationship: An Empirical Analysis. *Journal of Counseling and Development*, *95*(2), 125–134. https://doi.org/10.1002/jcad.12125

Lyons, E. J., Mehl, M. R., & Pennebaker, J. W. (2006). Pro-anorexics and recovering anorexics differ in their linguistic Internet self-presentation. *Journal of Psychosomatic Research*, *60*(3), 253–256. https://doi.org/10.1016/j.jpsychores.2005.07.017

Magill, M., Ray, L., Kiluk, B., Hoadley, A., Bernstein, M., Scott Tonigan, J., & Carroll, K. (2019). A meta-analysis of cognitive-behavioral therapy for alcohol or other drug use disorders: Treatment efficacy by contrast condition. *Journal of Consulting and Clinical Psychology*, *87*(12), 1093–1105. https://doi.org/10.1037/ccp0000447

Markowitz, J. C., Petkova, E., Neria, Y., Van Meter, P. E., Zhao, Y., Hembree, E., … Marshall, R. D. (2015). Is exposure necessary? A randomized clinical trial of interpersonal psychotherapy for PTSD. *American Journal of Psychiatry*, *172*(5), 430–440. https://doi.org/10.1176/appi.ajp.2014.14070908

McKellar, J., Kelly, J., Harris, A., & Moos, R. (2006). Pretreatment and during treatment risk factors for dropout among patients with substance use disorders. *Addictive Behaviors*, *31*(3), 450–460. https://doi.org/10.1016/j.addbeh.2005.05.024

McMurran, M., Huband, N., & Overton, E. (2010). Non-completion of personality disorder treatments: A systematic review of correlates, consequences, and interventions. *Clinical Psychology Review*, *30*(3), 277–287. https://doi.org/10.1016/j.cpr.2009.12.002

Miller, W. R., & Rollnick, S. (2012). Meeting in the middle: motivational interviewing and self-determination theory. *International Journal of Behavioral Nutrition and Physical*

Activity, 9(1), 25.

Mohr, D. C., Cuijpers, P., & Lehman, K. (2011). Supportive accountability: A model for providing human support to enhance adherence to eHealth interventions. *Journal of Medical Internet Research*, *13*(1). https://doi.org/10.2196/jmir.1602

Moritz, S., Schröder, J., Meyer, B., & Hauschildt, M. (2013). The more it is needed, the less it is wanted: Attitudes toward face-to-face intervention among depressed patients undergoing online treatment. *Depression and Anxiety*, *30*(2), 157–167. https://doi.org/10.1002/da.21988

O'Neil, K. M., Penrod, S. D., & Bornstein, B. H. (2003). Web-based research: Methodological variables' effects on dropout and sample characteristics. *Behavior Research Methods, Instruments, and Computers*, *35*(2), 217–226. https://doi.org/10.3758/BF03202544

Oldham, M., Kellett, S., Miles, E., & Sheeran, P. (2012). Interventions to increase attendance at psychotherapy: A meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology*, *80*(5), 928–939. https://doi.org/10.1037/a0029630

Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H. U., & Jönsson, B. (2012). The economic cost of brain disorders in Europe. *European Journal of Neurology*, *19*(1), 155–162. https://doi.org/10.1111/j.1468-1331.2011.03590.x

Parveen, H., Showkat, N. (2017, July) Validity, Reliability, Generalizability. Retrieved from: https://www.researchgate.net/publication/319128 421_Validity_Reliability_Generalizability

Pedersen, E. R., Marshall, G. N., & Schell, T. L. (2016). Study protocol for a web-based personalized normative feedback alcohol intervention for young adult veterans. *Addiction Science & Clinical Practice*, *11*(1), 6. https://doi.org/10.1186/s13722-016-0055-8

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Postel, M.G. (2011). *Well connected. Web-based treatment for problem drinkers.* Oldenzaal, the Netherlands: Nijmegen Institute for Scientist-Practitioners in Addiction, Radboud University Nijmegen.

Postel, Marloes 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), 1–14. https://doi.org/10.2196/jmir.1642

Postel, Marloes 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

Radtke, T., Ostergaard, M., Cooke, R., & Scholz, U. (2017). Web-based alcohol intervention: Study of systematic attrition of heavy drinkers. *Journal of Medical Internet Research*, *19*(6), 1–12. https://doi.org/10.2196/jmir.6780

Ray, L. A., Hutchison, K. E., & Bryan, A. (2006). Psychosocial predictors of treatment outcome, dropout, and change processes in a pharmacological clinical trial for alcohol dependence. *Addictive Disorders and Their Treatment*, *5*(4), 179–190. https://doi.org/10.1097/01.adt.0000210701.63165.5a

Rehm, J., Anderson, P., Barry, J., Dimitrov, P., Elekes, Z., Feijão, F., … Gmel, G. (2015). Prevalence of and potential influencing factors for alcohol dependence in europe. *European Addiction Research*, *21*(1), 6–18. https://doi.org/10.1159/000365284

Reinwand, D. A., Crutzen, R., Elfeddali, I., Schneider, F., Schulz, D. N., Smit, E. S., … De Vries, H. (2015). Impact of educational level on study attrition and evaluation of web-based computer-tailored interventions: Results from seven randomized controlled trials. *Journal of Medical Internet Research*, *17*(10). https://doi.org/10.2196/jmir.4941

Reynolds, D. J., Stiles, W. B., Bailer, A. J., & Hughes, M. R. (2013). Impact of exchanges and client-therapist alliance in online-text psychotherapy. *Cyberpsychology, Behavior, and Social Networking*, *16*(5), 370–377. https://doi.org/10.1089/cyber.2012.0195

Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review*, *32*(4), 329–342. https://doi.org/10.1016/j.cpr.2012.02.004

Rizvi, S. L., Vogt, D. S., & Resick, P. A. (2009). Cognitive and affective predictors of treatment outcome in cognitive processing therapy and prolonged exposure for posttraumatic stress disorder. *Behaviour Research and Therapy*, *47*(9), 737–743. https://doi.org/10.1016/j.brat.2009.06.003

Roberts, A., Murphy, R., Turner, J., & Sharman, S. (2019). Predictors of Dropout in Disordered Gamblers in UK Residential Treatment. *Journal of Gambling Studies*, *36*(1), 373–386. https://doi.org/10.1007/s10899-019-09876-7

Rochlen, A. B., Zack, J. S., & Speyer, C. (2004). Online Therapy: Review of Relevant Definitions, Debates, and Current Empirical Support. *Journal of Clinical Psychology*, *60*(3), 269–283. https://doi.org/10.1002/jclp.10263

Rogers, M. A. M., Lemmen, K., Kramer, R., Mann, J., & Chopra, V. (2017). Internet-delivered health interventions that work: Systematic review of meta-analyses and evaluation of website availability. *Journal of Medical Internet Research*, *19*(3), 1–28. https://doi.org/10.2196/jmir.7111

Rooke, S., Thorsteinsson, E., Karpin, A., Copeland, J., & Allsop, D. (2010). Computer-delivered interventions for alcohol and tobacco use: A meta-analysis. *Addiction*, *105*(8), 1381–1390. https://doi.org/10.1111/j.1360-0443.2010.02975.x

Ryan, S. A., & Kokotailo, P. (2019). Alcohol Use by Youth. *Pediatrics*, *144*(1), e20191357. https://doi.org/10.1542/peds.2019-1357

Sansone, A., Di Dato, C., de Angelis, C., Menafra, D., Pozza, C., Pivonello, R., … Gianfrilli, D. (2018). Smoke, alcohol and drug addiction and male fertility. *Reproductive Biology and Endocrinology*, *16*(1), 1–11. https://doi.org/10.1186/s12958-018-0320-7

Schomerus, G., Lucht, M., Holzinger, A., Matschinger, H., Carta, M. G., & Angermeyer, M. C. (2011). The stigma of alcohol dependence compared with other mental disorders: A review of population studies. *Alcohol and Alcoholism*, *46*(2), 105–112. https://doi.org/10.1093/alcalc/agq089

Schwarzinger, M., Pollock, B. G., Hasan, O. S. M., Dufouil, C., Rehm, J., Baillot, S., … Luchini, S. (2018). Contribution of alcohol use disorders to the burden of dementia in France 2008–13: a nationwide retrospective cohort study. *The Lancet Public Health*, *3*(3), e124–e132. https://doi.org/10.1016/S2468-2667(18)30022-7

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

Smink, W., Sools, A. M., Van Der Zwaan, J. M., Wiegersma, S., Veldkamp, B. P., & Westerhof, G. J. (2019). Towards text mining therapeutic change: A systematic review of text-based methods for Therapeutic Change Process Research. *PLoS ONE*, *14*(12), 1–21. https://doi.org/10.1371/journal.pone.0225703

Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. Y. (2008). Cheap and fast - But is it good?

Evaluating non-expert annotations for natural language tasks. *EMNLP 2008 - 2008 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference: A Meeting of SIGDAT, a Special Interest Group of the ACL*, (October), 254–263.

Spinczyk, D., Nabrdalik, K., & Rojewska, K. (2018). Computer aided sentiment analysis of anorexia nervosa patients' vocabulary. *BioMedical Engineering Online*, *17*(1). https://doi.org/10.1186/s12938-018-0451-2

Stelzhammer, V., Ozcan, S., Michael, G., Steeb, H., Hodes, G. E., Guest, C., … Bahn, S. (2015). Author ' s Accepted Manuscript Author ' s Accepted Manuscript. *Diagnostics in Neuropsychiatry*, 88–93. https://doi.org/10.1016/j.dineu.2015.08.001

Stolle, M., Sack, P. M., & Thomasius, R. (2009). Rauschtrinken im kindes- und jugendalter: Epidemiologie, auswirkungen und intervention. *Deutsches Arzteblatt*, *106*(19), 323–328. https://doi.org/10.3238/arztebl.2009.0323

Subramaniam, M., Mahesh, M. V., Peh, C. X., Tan, J., Fauziana, R., Satghare, P., … Chong, S. A. (2017). Hazardous alcohol use among patients with schizophrenia and depression. *Alcohol*, *65*, 63–69. https://doi.org/10.1016/j.alcohol.2017.07.008

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *80*(4), 547–559. https://doi.org/10.1037/a0028226

Swift, J. K., & Greenberg, R. P. (2014). A Treatment by disorder meta-analysis of dropout from Psychotherapy. *Journal of Psychotherapy Integration*, *24*(3), 193–207. https://doi.org/10.1037/a0037512

Tan, S. S. L., & Goonawardene, N. (2017). Internet health information seeking and the patient-physician relationship: A systematic review. *Journal of Medical Internet Research*, *19*(1). https://doi.org/10.2196/jmir.5729

Tanana, M., Hallgren, K. A., Imel, Z. E., Atkins, D. C., & Srikumar, V. (2016). A Comparison of Natural Language Processing Methods for Automated Coding of Motivational Interviewing. *Journal of Substance Abuse Treatment*, *65*, 43–50. https://doi.org/10.1016/j.jsat.2016.01.006

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

Teixidó-Compañó, E., Espelt, A., Sordo, L., Bravo, M. J., Sarasa-Renedo, A., Indave, B. I., … Brugal, M. T. (2018). Diferencias entre hombres y mujeres en el consumo de sustancias: el rol del nivel educativo y la situación laboral. *Gaceta Sanitaria*, *32*(1), 41–47. https://doi.org/10.1016/j.gaceta.2016.12.017

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*(April 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.

Van Ballegooijen, W., Cuijpers, P., Van Straten, A., Karyotaki, E., Andersson, G., Smit, J. H., & Riper, H. (2014). Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: A meta-analysis. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0100674

Van Beugen, S., Ferwerda, M., Hoeve, D., Rovers, M. M., Spillekom-Van Koulil, S., Van Middendorp, H., & Evers, A. W. M. (2014). Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: A meta-analytic review. *Journal of Medical Internet Research*, *16*(3), 1–15. https://doi.org/10.2196/jmir.2777

van der Sloot, K., Hendrickx, I., van Gompel, M., van den Bosch, A., & Daelemans, W. (2018). *Frog, a natural language processing suite for Dutch: Reference guide*. (June). Retrieved from https://frognlp.readthedocs.io/en/latest/ (last accessed: 7 February 2020)

Vernon, M. L. (2010). A review of computer-based alcohol problem services designedfor the general public. *Journal of Substance Abuse Treatment*, *38*(3), 203–211. https://doi.org/10.1016/j.jsat.2009.11.001

Vuoristo-Myllys, S., Psych, L. I. C. A., Lahti, J., Ph, D., Alho, H., & Ph, D. (2013). *Predictors of Dropout in an Outpatient Treatment for.pdf*. (November), 894–901.

Watson, H. J., Fursland, A., & Byrne, S. (2013). Treatment engagement in eating disorders: Who exits before treatment? *International Journal of Eating Disorders*, *46*(6), 553–559. https://doi.org/10.1002/eat.22085

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

Wiener, C. D., Moreira, F. P., Zago, A., Souza, L. M., Branco, J. C., Oliveira, J. F. De, … Oses, J. P. (2018). *Mood disorder , anxiety , and suicide risk among subjects with alcohol abuse and / or dependence : a population-based study*. 1–5. https://doi.org/10.1590/1516-4446-2016-2170

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, J., & Li, M. D. (2016). Converging findings from linkage and association analyses on susceptibility genes for smoking and other addictions. *Molecular psychiatry*, *21*(8), 992-1008.

Yeung, W. F., Chung, K. F., Ho, F. Y. Y., & Ho, L. M. (2015). Predictors of dropout from internet-based self-help cognitive behavioral therapy for insomnia. *Behaviour Research and Therapy*, *73*(July 2016), 19–24. https://doi.org/10.1016/j.brat.2015.07.008

Zachariae, R., Lyby, M. S., Ritterband, L. M., & O'Toole, M. S. (2016). Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - A systematic review and meta-analysis of randomized controlled trials. *Sleep Medicine Reviews*, *30*, 1–10. https://doi.org/10.1016/j.smrv.2015.10.004

Zandberg, L. J., Rosenfield, D., Alpert, E., McLean, C. P., & Foa, E. B. (2016). Predictors of dropout in concurrent treatment of posttraumatic stress disorder and alcohol dependence: Rate of improvement matters. *Behaviour Research and Therapy*, *80*, 1–9. https://doi.org/10.1016/j.brat.2016.02.005

Appendices

Appendix A

Overview of the Content of the Intervention with Typical Quotes from each Week of Treatment.

Treatment Procedure		Content	Counsellor	Client
Beginning		Welcoming	Dear PER, welcome. My name is PER and I am counsellor for LOC. In this message I will talk about your registration and discuss the program in more detail (…).	Dear PER, I checked this only until this morning. I still have to adapt to how this is working. There is something wrong with my computer. It says you sent me another message earlier and some homework assignments, so I am looking for it right now.
Part 1				
	1.	Advantages and disadvantages of drinking	In the list of medical conditions, you mention that you sometimes feel overly tired. Did you already talk with your general practitioner about that? This problem can be related to alcohol consumption. Less alcohol gives you more energy. What do you think about that?	Alcohol helps me tremendously to relax after a stressful day. Even stronger, I find it difficult to relax without alcohol. (…). Distorted sleeping rhythm, if a bottle of wine or beer is on the table, I am no longer willing to go to bed. I keep on drinking until I am dead tired.
	2.	When and how much do you drink?	I see that your consumption is equally high in weeks that you work and in	My girlfriend is anti-alcohol and anti-smoking. I do not smoke but every now and then I like

			weeks that you stay at home and your consumption was lower when you were on vacations. Just as you had predicted.	having a beer. Now, during the weekends when we are together, I do not drink. However, during the week I am alone. (…). When I come home and I do not have to work the following day or have a nightshift, I grab a beer. Yet, that beer becomes easily NUM and NUM beers.
	3.	Analysis of drinking situations	That was an awful situation. You said that deep inside you something began to tremble. In your moment description I read that you drank a glass of wine to calm down. Did the glass of wine have the effect you had anticipated? What did you do after you drank the glass of wine? What did this do with your feelings?	My stumbling block is that I do not feel well when I come home from work. The work itself is fun, but my director is just an idiot. When I come home stressed and do not feel well, I will drink some glasses of wine. It is like a break for me, just like NUM years ago when I was still smoking.
	4.	Recognize drinking situations	Based on the different situations that you described it seems that you often feel like having a drink when you are: out of house, at the golf club, at home with your partner, in a restaurant or café, at a party, at weekends. These we call your risk-situations. In these situations, you seem to have a higher urge to drink.	I had a partner that liked to drink. Now, there are no days on which I do not drink. Currently I drink some more because the vacation period is over and there are other circumstances like some nice parties and gatherings. I almost always drink NUM or NUM beers after work. Additionally, I always have a glass of red wine when eating. Mainly around NUM o'clock or between Num and NUM in the evening. What I drink above of that is mainly due to being in social gatherings.
Part 2				
	5.	Set a drinking goal	The choice to drink less is a good choice. I think that this is achievable (…). You can	I do not want to reach the point where a doctor tells me that due to a cirrhosis of the liver I may

		always drink a little. This means that you don't have to get nervous when you are offered a drink somewhere. Also, you avoid questions by others that you might consider uncomfortable.	never drink again. (…). In short, I do not want to set myself up for a severe alcohol addiction and detox clinics. I want to try to drink on a 'not harmful to health' level for half a year.
6.	Identifying unhelpful & helpful thoughts	You are going to investigate your own unhelpful risky thoughts and try to change them into more positive helpful thoughts. Helpful thoughts are thoughts that aid you in achieving your goals and that support you in your plans. Unhelpful thoughts tempt you to drink more than you had intended. (…). One risky thought you mentioned earlier was 'Now I really need wine to get me through this	The function of alcohol after parties, or other events, was like a reward. Or if I had an annoying meeting then I had at least something to look forward to at home. (…). You asked whether I wanted to think of helpful thoughts. Last week I noticed that alcohol increases my panic between NUM and NUM o'clock, while I actually drink to reduce the panic. The best way of not having a panic is therefore to not drink alcohol, but my fear of having a panic actually makes me want to drink. I know that there is some irrational & reversed logic in there (…)
7.	Formulate helpful behaviors for moments of craving	Just like your thoughts, behaviors can also be distinguished in helpful and unhelpful behaviors. An unhelpful behavior can be that you accepted the glass of wine that was offered to you at a family dinner. Such a risk behavior does not help you in moments you actually do not want to drink. It is better to do other things instead. Searching a distraction or doing something nice can help you overcome your craving. It can be difficult	I bought a training DVD and restored my stepping-machine. When I feel stressed, I will use the stepping-machine for NUM minutes and for NUM days a week I train with the DVD.

		to come up with alternatives just when the craving arises, so it can be useful to think of some activities beforehand and make a list of them. You do not have to choose complicated activities and even simple ones like vacuum-cleaning or taking a shower can be very effective.	
8.	Decision-making	For the task ‘decisions-making’ you create a schema of Decision-making moments that you go through before you decide to drink. For each moment in your decision-making there is the chance to step out. Read carefully and try to imagine what you can or could do at the different moments of decision-making.	NUM days ago, I wanted to drink a glass of wine. It was after lunch and I really felt that I needed to sit down with a glass of wine to relax. Yet, I felt that I had made so much progress in the last weeks and I did not want to risk it. I decided to try and make a cup of coffee instead and drink it. It helped and I did not need to have the wine.
9.	Formulating an action plan*	We now reached step NUM, creating an action plan, so that you have something that can serve as a guide and to prevent a relapse. You have read that there are moments when you can take other actions instead of choosing to drink. For the task ‘actionplan’ you will make a short, condensed overview of your personal motivation as well as your helpful thoughts and behaviors. It is good to have this all together in one place.	I am curious how we will proceed with the action plan and how this can increase my efforts to reach my goal. I am also interested in where you see my personal pitfalls and what, according to you, the mechanisms are that make me drink more than I plan.

	10. Wrapping up	Dear PER, thank you for the message. I like reading that the moments in which you want to drink but do not become more and more normal to you. Also, that you perceive the advantages of not drinking as greater to the advantages of drinking. We will wrap up the treatment here and therefore I want to give you a summary of the treatment. (…). For the next half year, you can always come back online to check your files and to re-read what we did. This can help you in solidifying the changes that you achieved. (…). For now, I wish you much success and all the best!	Dear PER, (…). Thank you for the conversations we had the past weeks. The fun thing is, that what you wrote in 'Subject' (Subject of the e-mails) always described the essence well. Especially in your last mail. 'Taking the matters in your own hand'. You are exactly right, and I think that this is the process that I have started. Thank you, I am on a good path and I think it will become better and better. I want to keep on improving. Thank you again!
Aftercare & Conclusion		Dear PER, it is nice to read that you are feeling well, and things have turned out just as you had hoped for.	I am feeling much better than I had expected. The live I had seems to be so far away now, instead of the actual NUM months. I never want to go back.

Note: The original content was in Dutch and had been translated by the researcher.
* The actual action plan was filled in online by the clients and was therefore not included in the emails. The quote used for the clients is only a discussion of the topic.

Appendix B

Frequency List for LIWC Positive Emotions

Dropout			Completer		
English	Dutch	N	English	Dutch	N
good	goed	1379	good	goed	1741
gladly	graag	480	better	beter	582
better	beter	471	true	waar	574
true	waar	470	gladly	graag	554
best	beste	415	best	beste	458
surely	zeker	330	surely	zeker	426
hope	hoop	321	well	lekker	388
well	lekker	307	advantages	voordelen	343
advantages	voordelen	267	hope	hoop	323
fun	leuk	198	fun	leuk	256
happy	gelukkig	159	energy	energie	217
energy	energie	153	relaxed	ontspannen	209
free	vrij	151	happy	gelukkig	208
relaxed	ontspannen	143	give	geeft	188
super	prima	143	free	vrij	187

Appendix C

Frequency List for LIWC Negative Emotions

Dropout			Completer		
English	Dutch	N	English	Dutch	N
alone	alleen	769	alone	alleen	917
disadvantages	nadelen	327	disadvantages	nadelen	405
last	last	302	last	last	384
problems	problemen	294	difficult	moeilijk	338
problem	probleem	272	problems	problemen	310
difficult	moeilijk	255	problem	probleem	282
only	slecht	218	only	slecht	260
complaints	klachten	204	complaints	klachten	233
pressure	druk	200	pressure	druk	207
unfortunately	helaas	147	pain	pijn	203
frightened	bang	129	frightened	bang	155
difficult	lastig	113	stress	stress	134
pain	pijn	103	sorrows	zorgen	130
sorrows	zorgen	103	sick	ziek	128
lost	kwijt	100	unfortunately	helaas	123

Appendix D

Frequency List for LIWC Anxiety

Dropout			Completer		
English	Dutch	N	English	Dutch	N
pressure	druk	200	pressure	druk	207
afraid	bang	129	afraid	bang	155
to worry	zorgen	103	stress	stress	134
stress	stress	89	worry	zorgen	130
living	woon	73	anxiety	angst	109
guilty	schuldig	55	guilty	schuldig	85
anxiety	angst	53	angry	boos	77
angry	boos	53	living	woon	70
lived	woonde	38	suspense	spanning	64
awful	vreselijk	37	suspenses	spanningen	52
drag	rem	33	frightened	angstig	50
restlessness	onrust	31	restlessness	onrust	47
suspense	spanning	29	awful	vreselijk	45
frightened	angstig	29	shame	schaamte	40
insecure	onzeker	29	tense	gespannen	40

Appendix E

Frequency List for LIWC Sadness

Dropout			Completer		
English	Dutch	N	English	Dutch	N
alone	alleen	769	alone	alleen	917
sadly	helaas	147	sadly	helaas	123
lost	kwijt	100	lost	kwijt	116
depressive	depressief	71	depression	depressie	98
unfortunately	jammer	52	depressive	depressief	93
depression	depressie	49	tiredness	vermoeidheid	88
tiredness	vermoeidheid	49	empty	leeg	66
dull	somber	41	grief	verdriet	62
empty	leeg	36	lonely	eenzaam	54
sorry	sorry	36	loss	verlies	46
grief	verdriet	35	energy	somber	45
lonely	eenzaam	29	unfortunately	jammer	44
loneliness	eenzaamheid	27	grieving	verdrietig	35
anxieties	angsten	24	anxieties	angsten	28
low	laag	24	miserable	ellende	28

Appendix F

Frequency List for LIWC Family

Dropout			Completer		
English	Dutch	N	English	Dutch	N
man/husband	man	300	man/husband	man	443
mother	moeder	274	mother	moeder	337
family	familie	257	family	familie	310
father	vader	215	father	vader	290
daughter	dochter	202	woman/wife	vrouw	236
wife	vrouw	201	daughter	dochter	216
parents	ouders	178	parents	ouders	194
son	zoon	108	son	zoon	187
relatives	gezin	101	brother	broer	146
brother	broer	73	parent	gezin	144
sister	zus	71	sister	zus	115
married	getrouwd	53	marriage	huwelijk	48
marriage	huwelijk	50	married	getrouwd	42
son	zoontje	50	sisters	zussen	35
parent	ouder	39	brothers	broers	33

Appendix G

Frequency List for LIWC Insight

Dropout			Completer		
English	Dutch	N	English	Dutch	N
knows	weet	837	knows	weet	1064
for real	echt	701	find	vind	886
find	vind	698	think	denk	861
think	denk	678	for real	echt	828
feel	voel	522	feel	voel	697
asks	vragen	495	ask	vragen	594
sick	wordt	424	become	wordt	531
see	zie	334	becoming	worden	438
ask	vraag	314	feeling	gevoel	423
become	worden	306	meaning	zin	421
meaning	zin	295	become	werd	419
feeling	gevoel	285	see	zie	413
relationship	relatie	268	ask	vraag	407
become	werd	253	relationship	relatie	276
became	geworden	204	found	vond	253

Appendix H

Frequency List for LIWC Focus Past

Dropout			Completer		
English	Dutch	N	English	Dutch	N
was	was	1596	was	was	2367
although	al	1456	although	al	1886
had	had	901	had	had	1240
go	ga	755	go	ga	959
find	vind	698	find	vind	886
goes	gaat	697	goes	gaat	787
drank	gedronken	597	drank	gedronken	719
had	gehad	540	had	gehad	588
was	geweest	477	do	doe	566
do	doe	464	was	geweest	545
went	ging	336	drank	dronk	470
eating	eten	334	went	ging	464
drank	dronk	326	comes	komt	461
comes	komt	318	can	kon	444
asks	vraag	314	come	komen	441

Appendix I

Frequency List for LIWC Focus Future

Dropout			Completer		
English	Dutch	N	English	Dutch	N
then	dan	2352	then	dan	3047
going	gaan	984	going	gaan	1262
will	wil	876	will	wil	1020
have to	moet	819	have to	moet	980
go	ga	755	go	ga	959
find	vind	698	find	vind	886
goes	gaat	697	goes	zou	840
Has to	zal	647	Has to	zal	819
should	zou	588	should	gaat	787
sometimes	soms	567	sometimes	soms	700
further	verder	505	further	verder	594
hope	hoop	321	hope	hoop	323
afterwards	daarna	220	afterwards	daarna	300
must	moeten	188	found	vond	253
possible	mogelijk	168	must	moeten	243

Apendix J

Frequency List for LIWC Work

Dropout			Completer		
English	Dutch	N	English	Dutch	N
work	werk	1094	work	werk	1198
busy	bezig	279	busy	bezig	344
working	werken	213	working	werken	318
began	begonnen	201	began	begonnen	216
pressure	druk	200	reading	lezen	210
reading	lezen	158	pressure	druk	207
job	baan	145	job	baan	144
begin	beginnen	121	colleague	collega	143
task	opdracht	100	begin	beginnen	141
works	werkt	100	began	begon	132
teaching /learning	leren	93	task	opdracht	129
began	begon	90	works	werkt	124
program	programma	81	read	lees	120
read	lees	80	program	programma	97
colleagu e	collega	79	teaching/ learning	leren	87