

# TeleScreen as a novel internet-based tool for classifying mental disorders presented in primary care

Citation for published version (APA):

Dijksman, I. (2018). *TeleScreen as a novel internet-based tool for classifying mental disorders presented in primary care*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20181011id>

## Document status and date:

Published: 01/01/2018

## DOI:

[10.26481/dis.20181011id](https://doi.org/10.26481/dis.20181011id)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

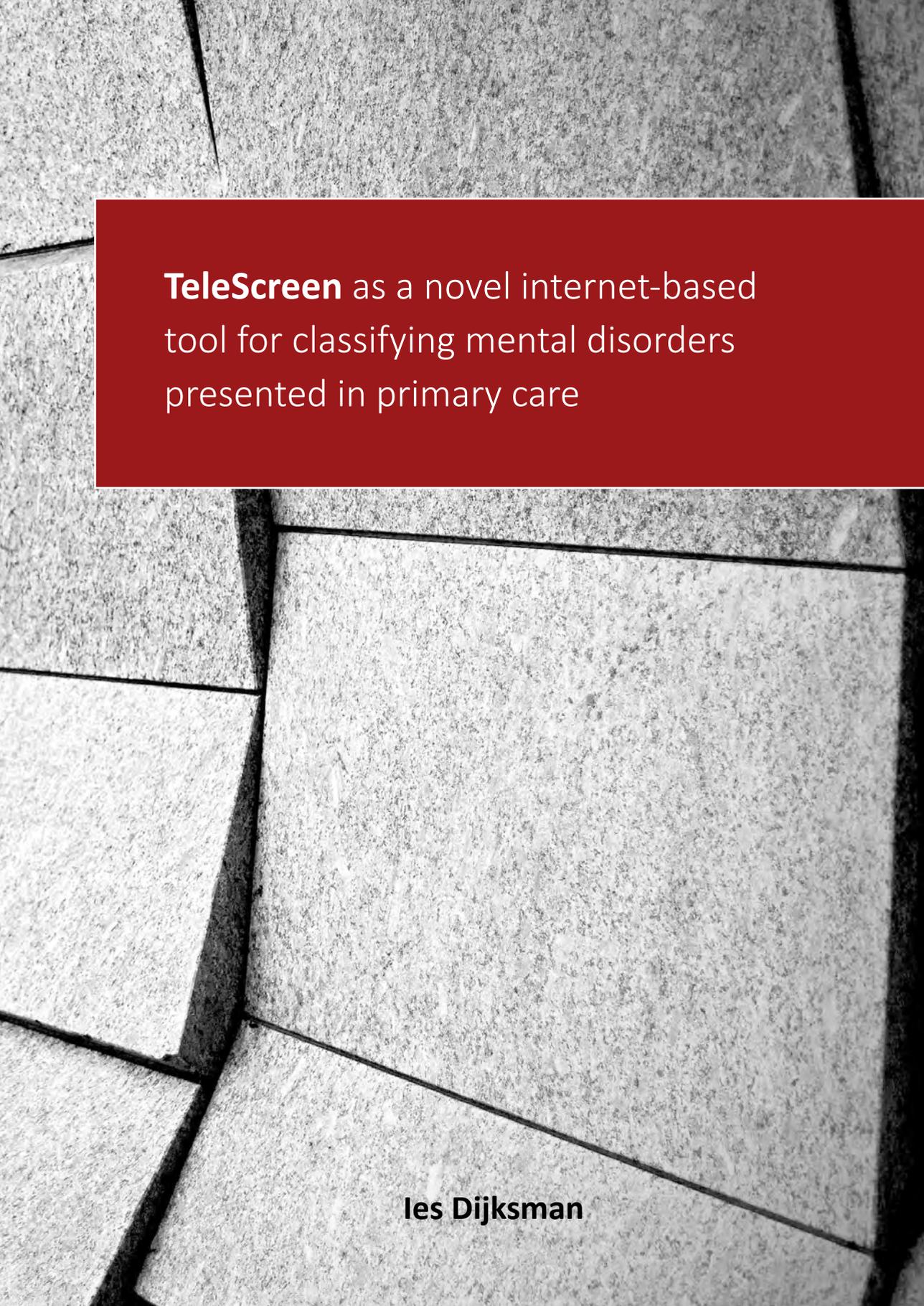
[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

The background of the entire page is a close-up, high-angle photograph of a stone wall. The stones are rectangular and have a rough, textured surface. The mortar joints are dark and clearly visible, creating a grid-like pattern. A solid red rectangular box is positioned in the upper left quadrant of the page, containing white text.

**TeleScreen** as a novel internet-based  
tool for classifying mental disorders  
presented in primary care

Ies Dijkman

The research presented in this thesis was conducted at the:

School for Public Health and Primary Care (CAPHRI), Department of Family Medicine,  
Maastricht University.

The research presented in this dissertation was funded by Stichting Robuust and  
TelePsy.

The printing of this thesis was financially supported by TelePsy.

All rights are reserved. No part of this book may be reproduced or transmitted in any  
form or by any means, without the written permission from the author or, where appro-  
priate, the publisher of the article.

© copyright Ies Dijkman, Maastricht 2018

Printing: Datawyse | Universitaire Pers Maastricht

ISBN 978-94-6380-009-9



# TeleScreen as a novel internet-based tool for classifying mental disorders presented in primary care

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,  
op gezag van de Rector Magnificus, Prof. dr. Rianne M. Letschert,  
volgens het besluit van het College van Decanen,  
in het openbaar te verdedigen  
op donderdag 11 oktober om 10:00 uur

door

Ies Dijkman  
Geboren op 21 november 1985 te Nijmegen

**Promotor**

Prof. dr. G.J. Dinant

**Copromotor**

Dr. M. Spigt

**Beoordelingscommissie:**

Prof. dr. B.P.F. Rutten (voorzitter)

Prof. dr. J.W.M. Muris

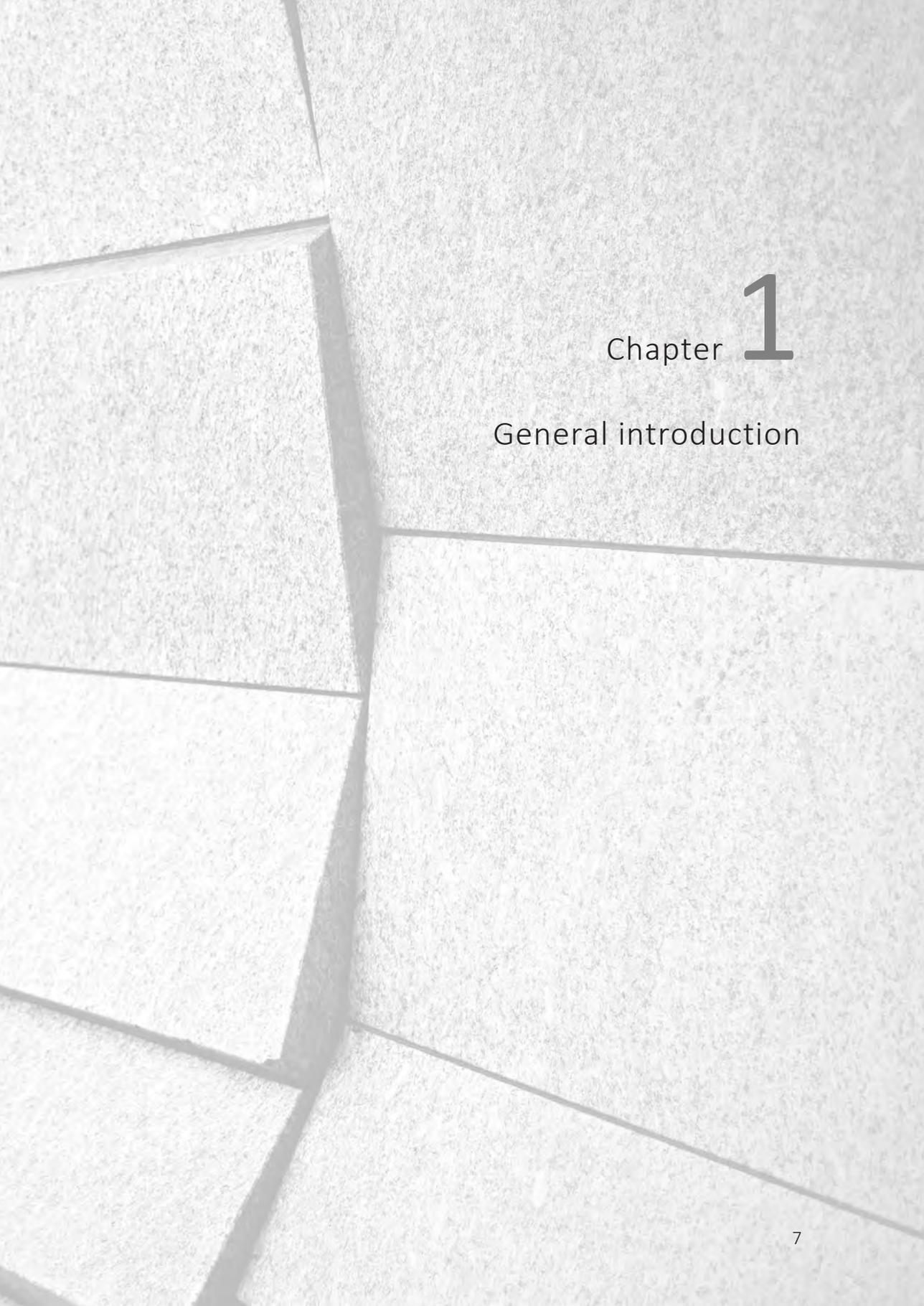
Prof. dr. H.C.W. de Vet (Vrije Universiteit Amsterdam)

Dr. V. Thewissen (Open Universiteit Heerlen)

# Contents

Chapter 1	General introduction	7
Chapter 2	eDiagnostics: a promising step towards primary mental health care Published as: Dijkman, I., Dinant, G. J., & Spigt, M. G. (2013). eDiagnostics: a promising step towards primary mental health care. <i>Fam Pract</i> , 30(6), 695-704.	19
Chapter 3A	The concurrent validity of a new eDiagnostic system for mental disorders in primary care Published as: Dijkman, I., Dinant, G. J., & Spigt, M. (2016). The concurrent validity of a new eDiagnostic system for mental disorders in primary care. <i>Fam Pract</i> , 33(6), 607-616.	37
Chapter 3B	eDiagnostics for mental disorders Published as: Dijkman, I., Dinant, G. J., & Spigt, M. (2017). E-diagnostiek voor psychische stoornissen. <i>Huisarts Wet</i> , 60(9), 432-435.	63
Chapter 4	The concurrent validity of a web-based self-report assessment for personality disorders Submitted	75
Chapter 5	The perception and needs of psychologists towards blended care Published as: Dijkman I, Dinant G-J, Spigt M. (2017). The Perception and Needs of Psychologists Toward Blended Care. <i>Telemed J E Health</i> , 23(12), 1-13.	91
	General discussion	113
	Summary	131
	Samenvatting	139
	Valorisation	147
	Dankwoord	157
	Curriculum Vitae	163





Chapter **1**

General introduction



## Introduction

The general practitioner (GP) is often the first person to contact in case of mental health problems. The Dutch healthcare authority stated that yearly more than 20% of all adults have mental health problems and about three quarter of these persons visit their GP (1). In the Netherlands, one speaks of a mental disorder when a person fulfils the required symptoms based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) (2) or the renewed fifth edition (DSM-5) (3). Major depressive and anxiety disorder(s) are the most commonly observed mental disorders (4). The long-term prognosis for these patients is poor; many patients do not recover, or sustain recurrent episodes (5). Mental disorders are associated with a high burden of disease (6), reduced quality of live (7), high societal costs (8, 9) and work disability (7, 10). Effective treatments for mental disorders exist, but they are not always accessible (11-16). Mental disorders can be missed or misinterpreted, for example if patients do not ask for help, if they normalize complaints, if symptoms are attributed to somatic complaints, or if patients do not recognize their problems (17-19). Furthermore, many people feel ashamed of having mental problems, so not everyone seeks help (20).

The GP plays a central role in managing patients with mental health problems. The GP is expected to identify mental health problems at an early stage, identify persons who are at risk of developing a mental disorder and prevent the onset of mental disorders. At risk for developing mental disorders are for example persons with a history of mental disorders, persons who often visit their GP, persons who present with unexplained physical complaints, persons with chronic pain, or persons who frequently ask for sleeping medication or sedatives (21-23). The GP does not only play an important role in the identification of mental disorders, but also in the decision to refer the patient for (specialized) mental healthcare treatment. In the Netherlands, a referral letter from the GP is required to access (reimbursed) mental healthcare, so the GP acts as a gatekeeper to (specialized) mental healthcare. This referral system is also implemented in other countries, such as the United Kingdom, Denmark, Norway and Spain, but it differs from a system in which patients have direct access to specialist care, which is the case in Belgium, France, Germany, Finland and Sweden (24, 25).

## The Dutch healthcare setting

In 2014 the Dutch government stimulated a shift from specialized mental healthcare to primary care. Since then, only patients with a suspicion of a mental disorder based on DSM-IV-TR (or DSM-5) are allowed to be referred to (reimbursed) mental healthcare; that is a primary care psychologist or specialized mental healthcare trajectory (26). This means that a referral from the GP to a mental healthcare professional should always be accompanied with a suspected diagnosis from DSM-IV-TR (or DSM-5). Patients without a clear

suspicion of a DSM-IV-TR (or DSM-5) disorder must be treated in general practice, by the GP or a practice nurse (PN). A PN received higher vocational training in nursing or psychology, but works under the supervision of the GP. One of the PNs' tasks is to clarify the problem presented by a patient by performing the first diagnostic assessment and prepare and discuss referral steps, when deemed necessary.

In the Netherlands, the majority of patients with a mental complaint are treated in general practice, most often by their GP. When patients are referred, the majority enters secondary mental healthcare (27). The incidence and prevalence of mental disorders did not reduce over time (4). Between 2010 and 2014 both PNs and GPs recorded increased numbers of patients presenting with mental health problems, as well as increased numbers of consultations per patient (28). Both GPs and PNs can allow relatively short consultation times to clarify presented mental health problems and patients with mental health problems heavily impact on the GP's workload (29). Furthermore, GPs often feel uncertain and are unfamiliar with psychological interventions, which may result in incorrect treatment, referrals and long waiting lists (30).

## **eMental health**

eMental health which involves the use of online information and communication technology to support mental healthcare (31, 32), is seen as one of the solutions to address the above mentioned mental healthcare challenges. Specific advantages are mentioned in the scientific literature. For example, eMental health has the potential to reach more people, because an unlimited number of patients can use an intervention once it is online (33). Also, patients and healthcare professionals are not restricted to a treatment location or time. As long as there is an Internet connection, patients can work on their own problems 24 hours per day (33). Another positive aspect of eMental health is that patients become more involved in their own treatment and there is more time for self-reflection (33). In this way, eMental health stimulates self-management and a more independent patient-doctor or patient-therapist relationship (34). Furthermore, because psychological mechanisms, such as affect and (negative) thoughts can be assessed real-time real-live, a patient is less dependent on retrospective information and answers on questionnaires may be more accurate (35, 36). Finally, people may express their feelings more openly, because of the perceived anonymity of the computer (37-39). For example suicidal behaviour can be better predicted using computer delivered tests (40). For these reasons Dutch healthcare policy makers actively encourage eMental health and it is expected that eMental health will play an increasingly important role in mental healthcare in the coming years (33).

## Screening questionnaires

One relevant example of eMental health is the use of online screening questionnaires. Online screening questionnaires have been promoted as a means to help identifying mental disorders, before patients visit their GP or PN and, subsequently, to ensure that patients receive appropriate treatment. However, despite high expectations and advantages, research on the effectiveness and usefulness of online screening questionnaires is scarce and the added value has therefore been intensively debated over the years (41).

General arguments against screening also apply to online screening and include: the fear of treating patients who are incorrectly labelled as having a mental disorder (the so called false positives) and fear of treating patients with mild disorders that would have resolved without intervention, fear of high costs for persons who might gain little health improvement, or fear of using the already scarce recourses for patients with mild symptoms while persons with more serious symptoms are on waiting lists (41). In addition, there is no clear evidence from randomized controlled trials proving that screening for depression for example, is cost-effective (23).

Gradually, alternative ways of using online questionnaires to identify mental disorders have been proposed, for example for monitoring patients who have been clinically diagnosed with a mental disorder and received treatment, or patients who are at high risk of developing a mental disorder (23). An alternative approach is the use of two verbal questions during one consultation to assess, for example, the major criteria for depression, while, after an affirmative answer has been reached, a more in-depth assessment may follow (42, 43).

GPs generally do not adopt online questionnaires in their routine care. They often question the validity and utility of questionnaires and report being sceptical about underlying motives for such questionnaires (44), for example seeing them as “academically” oriented and recommended by the government in order to save money through decreasing prescriptions for medicines. Other reasons not to adopt such questionnaires are the fear of reducing the human element of a live consultation, the fact that filling in questionnaires is time-consuming, and the fact that questionnaires are limited to identify one specific mental disorder, for example general distress only, or they provide a numeric score instead of the specific diagnosis that is needed for a referral decision (44-48). Other GPs are more positive about the use of online questionnaires, as long as they are easy to use and do not affect the GP’s workload (30).

Opinions of patients with regard to screening is often more positive. Patients stress the need for reaching a diagnosis, and additional information about the nature and course of their problems, plus understanding what to expect from the GP (30). Patients consider screening tools to be acceptable: they have the feeling that their symptoms are taken seriously and they see questionnaires as an efficient and structured supplement to the GP’s judgment (44, 49). Some, however, believe that screening questionnaires are stigmatizing and they feel aversion of being diagnosed (50). Therefore, it is recommended

that patients' views towards diagnostic steps are taken into consideration, before a diagnosis is made and a treatment trajectory is proposed (50). Both GPs and patients consider that questionnaires should be seen as one aspect of holistic care (44).

## **Diagnosis versus classification**

Initiated by frequently received questions about the DSM-5, the Dutch Association of Psychiatry described the difference between a diagnosis and a classification again (51). According to this standpoint a diagnosis implies the description of a mental disorder based on experienced complaints, problems and symptoms as observed by the clinician. It contains more elements than a description of symptoms. The stage of a mental disorder (e.g. first or recurrent episode), complicating factors, comorbidity, factors that contributed to the development of the mental disorder, factors that influence the course of a disorder, disturbed functions and subsequent limitations in daily functioning in different areas of life as well as factors involved in the success or failure of treatment are included in a diagnosis. A diagnosis could be adjusted or could change over time. Classifying is seen as the assignment of a mental disorder to a described category. The DSM-IV-TR or the DSM-5 is a classification system allowing us to reliably rank patients' symptoms (2, 3, 51). It provides categories and criteria for allocating symptoms to specific categories. Therefore, when discussing online questionnaires to assess mental disorders the term provisional DSM-IV-TR or DSM-5 classification is used in this dissertation instead of a DSM-IV-TR or DSM-5 diagnosis.

## **Choice of instruments in (Dutch) general practices**

PRIME-MD (primary care evaluation of disorders), developed at Columbia University, was one of the first online questionnaires for assessing specific mental disorders in primary care (52). Later, several online questionnaires, based on PRIME-MD, were developed. They range from questionnaires to assess general distress to those identifying specific mental disorders. In the Netherlands, the Four-Dimensional Symptom Questionnaire (4-DSQ) is often used in general practice. It is a validated self-report questionnaire to differentiate between general distress, depression, anxiety and somatization (53). To trace specific disorders the Patient Health Questionnaire (PHQ) (54, 55), Web Screening Questionnaire (WSQ) (56), or the Kessler 10 (K10) (57) can also be used. The PHQ is based on PRIME-MD, but requires less administration time. It assesses a depressive disorder, probable alcohol abuse or dependence, and somatoform and binge eating disorders (54). K10 identifies depressive and anxiety disorders in primary care. WSQ assesses depressive, anxiety disorders and alcohol abuse or dependence.

## TeleScreen

The TeleScreen instrument, which is at the center of the studies presented in this thesis, has been developed by TelePsy as a comprehensive screening system, which covers most diagnoses based on DSM-IV-TR (or renewed DSM-5). The system provides provisional classifications instead of established diagnoses. Strictly spoken, TeleScreen is not used as a screening system, meaning that it is not used in patients without known complaints. TeleScreen may be seen as a triage instrument. Triage is a Latin word that means ‘to sort’ (58), originally used to sort out injured soldiers during Napoleonic wars to determine priority and emergency of care (59). Nowadays triage is often used in emergency departments or in general practice for the assessment and management of patients with physical or mental complaints (60).

With TeleScreen extensive triage at the very start of a patient’s trajectory is possible. GPs or PNs can use TeleScreen for patients with a suspected mental disorder. Patients answer questions online, after which a psychologist has a telephone consultation with the patient to check the classifications generated by the system. After this telephone call with the patient, the psychologist provides an advice to the GP, within a week, summarizing which mental health problems are probably present and in which echelon the patient is preferably to be treated. Together with the provisional DSM-IV-TR (or nowadays the DSM-5) classification and echelon advice, the context and background of complaints are described in a short summary, also taking into account expressed requests for help, treatment history of the patient, and suicide risk, if applicable. As a next step, an established diagnosis can be reached in a more specialized mental healthcare setting, where in depth evaluation of mental complaints can be carried out.

## Research questions

The main objective of our research was to evaluate experiences of TeleScreen users and examine the validity of TeleScreen. The following research questions are addressed in this thesis:

1. What are experiences of patients, general practitioners and practice nurses with regard to TeleScreen?
2. What is the concurrent validity of TeleScreen when compared with the psychologists’ judgments, in patients, presenting in primary care, with a possible mental disorder?
3. What is the concurrent validity of TeleScreen, when compared with the results of the Structural Clinical Interview for DSM-IV-TR Axis-I (SCID-I) and Axis-II disorders (SCID-II), for patients with personality disorders?

4. What are the perceptions of psychologists, and their wishes and barriers towards blended care?

### *Outline of the thesis*

Chapter 2 describes an online survey evaluating experiences of TeleScreen service users (patients, GPs and PNs). We also monitored provisional classifications resulting from the system. We compared the echelon advice of GPs and PNs prior to results from TeleScreen (i.e. echelon advice resulting from the system). Chapter 3a, 3b and chapter 4 describe the concurrent validity of TeleScreen. In Chapter 3a TeleScreen is compared with the psychologists' judgment, among primary care patients with a suspicion of a mental disorder. We used routinely collected data of 675 primary care patients in the Netherlands. Furthermore, 100 patient records were randomly selected to investigate the inter-rater reliability among psychologists. Chapter 3b is based on chapter 3a. In Chapter 3b data are updated. We examined the validity of TeleScreen by comparing routinely collected results of 496 patients from Limburg GP practices with the judgment of a psychologist. In Chapter 4, TeleScreen was compared with results of SCID-I and SCID-II, in a sample of patients with personality disorders. In total 89 patients from a Dutch treatment center for personality disorders were recruited. Chapter 5 describes perceptions of psychologists and barriers with regard to blended care. Specific wishes for the way blended care should be designed are summarized. For this study, opinions of psychologists of the Dutch Association of Psychologists were collected through an e-mailed survey.

## References

1. De Nederlandse Zorgautoriteit. *Advies bekostiging POH-GGZ*. [http://www.nza.nl/104107/139830/740611/Advies\\_Bekostiging\\_POH-GGZ](http://www.nza.nl/104107/139830/740611/Advies_Bekostiging_POH-GGZ). Archived at: <http://www.webcitation.org/6waYTMSoE>, 2013.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders 4th edition, Text Revision (DSM-IV-TR)*. Washington, DC: American Psychiatric Association, 2000.
3. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing, 2013.
4. de Graaf R, ten Have M, van Gool C, van Dorsselaer S. Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc Psychiatry Psychiatr Epidemiol* 2012; 47: 203-13.
5. Wildeboer RM, van der Hoek L, Verhaak PF. Use of general practice services 5 years after an episode of mental illness: case-control study using electronic records. *Br J Gen Pract* 2016; 66: e347-53.
6. Whiteford H, Ferrari A, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the global burden of disease study 2010. *PLoS one* 2015; 10: e0116820.
7. Alonso J, Angermeyer MC, Bernert S *et al*. Disability and quality of life impact of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta psychiatr Scand* 2004; 420: 38-46.
8. Smit F, Cuijpers P, Oostenbrink J *et al*. Costs of nine common mental disorders: implications for curative and preventive psychiatry. *J Ment Health Policy and Econ* 2006; 9: 193-200.
9. Kessler RC, Aguilar-Gaxiola S, Alonso J *et al*. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) Surveys. *Epidemiol e Psichiatr Soc* 2009; 18: 23-33.
10. Wittchen HU, Jacobi F. Size and burden of mental disorders in Europe—a critical review and appraisal of 27 studies. *Eur Neuropsychopharmacol* 2005; 15: 357-76.
11. Lecrubier Y. Widespread underrecognition and undertreatment of anxiety and mood disorders: results from 3 European studies. *J Clin Psychiatry* 2007; 68: 36-41.
12. Harvey AG, Gumpert NB. Evidence-based psychological treatments for mental disorders: modifiable barriers to access and possible solutions. *Behav Res Ther* 2015; 68: 1-12.
13. Wang PS, Aguilar-Gaxiola S, Alonso J *et al*. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. *Lancet* 2007; 370: 841-50.
14. Fernández A, Haro JM, Martínez-Alonso M *et al*. Treatment adequacy for anxiety and depressive disorders in six European countries. *Br J Psychiatry* 2007; 190: 172-3.
15. Bijl RV, Ravelli A. Psychiatric morbidity, service use, and need for care in the general population: results of The Netherlands Mental Health Survey and Incidence Study. *Am J Public Health* 2000; 90: 602-607.
16. Alonso J, Codony M, Kovess V *et al*. Population level of unmet need for mental healthcare in Europe. *Br J Psychiatry* 2007; 190: 299-306.
17. Verhaak PF, Schellevis FG, Nuijen J, Volkerts AC. Patients with a psychiatric disorder in general practice: determinants of general practitioners' psychological diagnosis. *Gen Hosp Psychiatry* 2006; 28: 125-32.
18. Kessler D, Lloyd K, Lewis G, Gray DP. Cross sectional study of symptom attribution and recognition of depression and anxiety in primary care. *BMJ* 1999; 318:436-9.
19. Jorm AF. Mental health literacy. Public knowledge and beliefs about mental disorders. *Br J Psychiatry* 2000; 177: 396-401.
20. Postel MG, de Haan HA, De Jong CA. E-therapy for mental health problems: a systematic review. *Telemed J E Health* 2008; 14: 707-14.
21. Hassink-Franke L, Terluin B, Van Heest FB *et al*. NHG-Standaard Angst (tweede herziening) [NHG guideline Anxiety second revision]. *Huisarts Wet* 2012; 2: 68-77.
22. Van Weel-Baumgarten EM, Van Gelderen MG, Grundmeijer HGLM *et al*. NHG Standaard Depressie (tweede herziening) [NHG guideline Depression second revision]. *Huisarts Wet* 2012; 55: 252-9.

23. Palmer SC, Coyne JC. Screening for depression in medical care: pitfalls, alternatives, and revised priorities. *J Psychosom Res* 2003 54: 279-87.
24. Kulu-Glasgow I, Delnoij D, de Bakker D. Self-referral in a gatekeeping system: patients' reasons for skipping the general-practitioner. *Health Policy* 1998; 45: 221-38.
25. Verhaak PF, van den Brink-Muinen A, Bensing JM, Gask L. Demand and supply for psychological help in general practice in different European countries: access to primary mental health care in six European countries. *Eur J Public Health* 2004; 14: 134-40.
26. Bakker P, Jansen P. *Generalistische Basis GGZ. Verwijsmodel en productbeschrijvingen*. Enschede: Bureau HHM, 2013.
27. Verhaak PF, van Dijk CE, Nuijen J, Verheij RA, Schellevis FG. Mental health care as delivered by Dutch general practitioners between 2004 and 2008. *Scand J Prim Health Care* 2012; 30: 156-62.
28. Magnee T, de Beurs DP, de Bakker DH, Verhaak PF. Consultations in general practices with and without mental health nurses: an observational study from 2010 to 2014. *BMJ Open* 2016; 18; 6:e011579.
29. Zantinge EM, Verhaak PF, Kerssens JJ, Bensing JM. The workload of GPs: consultations of patients with psychological and somatic problems compared. *Br J Gen Pract* 2005; 55: 609-14.
30. Gidding L, Spigt M, Brouwer E *et al*. Experiences of general practitioners, practice nurses, psychologists and patients with stepped collaborative care for depression: a focus group study in a large primary healthcare organization. *Eur J Pers Cent Healthc* 2014; 2: 170-8.
31. Eysenbach G. What is e-health? *J Med Internet Res* 2001; 3: E20.
32. Riper H, Andersson G, Christensen H *et al*. Theme issue on e-mental health: a growing field in internet research. *J Med Internet Res* 2010; 12: e74.
33. Schalken F, Blankers M, Donker T, Duinstra U, Van Gemert M, van Hoogenhuyze C. *Handboek online hulpverlening: met internet zorg en welzijn verbeteren*. Bohn Stafleu van Loghum: Houten, 2013.
34. van der Vaart R, Witting M, Riper H *et al*. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. *BMC Psychiatry* 2014; 14; 14: 355.
35. Van Os J. De DSM-5 voorbij! Persoonlijke diagnostiek in een nieuwe GGZ. *Tijdschrift voor Psychiatrie* 2014; 56: 356-7.
36. Kelly J, Gooding P, Pratt D *et al*. Intelligent real-time therapy: Harnessing the power of machine learning to optimise the delivery of momentary cognitive-behavioural interventions. *J Ment Health* 2012; 21: 404-14.
37. Joinson A. Social desirability, anonymity, and Internet-based questionnaires. Behavior research methods, instruments, & computers. *Behav Res Methods* 1999; 31: 433-8.
38. Richman WL, Kiesler S, Weisband S, Drasgow F. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *J Appl Psychol* 1999; 84: 754-75.
39. Newman MG. Technology in psychotherapy: an introduction. *J Clin Psychol* 2004; 60: 141-5.
40. Levine S, Ancill RJ, Roberts AP. Assessment of suicide risk by computer-delivered self-rating questionnaire: preliminary findings. *Acta Psychiatr Scand* 1989; 80: 216-20.
41. Thombs BD, Coyne JC, Cuijpers P *et al*. Rethinking recommendations for screening for depression in primary care. *CMAJ* 2012; 184: 413-8.
42. Hackley B, Sharma C, Kedzior A, Sreenivasan S. Managing Mental Health Conditions in Primary Care Settings. *J Midwifery Womens Health* 2010; 55: 9-19.
43. Arroll B, Khin N, Kerse N. Screening for depression in primary care with two verbally asked questions: cross sectional study. *BMJ* 2003; 15; 327: 1144-6.
44. Dowrick C, Leydon GM, McBride A *et al*. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ* 2009; 338: b663.
45. Roche AM, Freeman T. Brief interventions: good in theory but weak in practice. *Drug Alcohol Rev* 2004; 23: 11-8.
46. Sinnema H, Franx G, Spijker J *et al*. Delivering stepped care for depression in general practice: results of a survey amongst general practitioners in the Netherlands. *Eur J Gen Pract* 2013; 19: 221-9.

47. Fleury MJ, Imboua A, Aube D, Farand L, Lambert Y. General practitioners' management of mental disorders: a rewarding practice with considerable obstacles. *BMC Fam Pract* 2012; 13: 19.
48. Shedler J. The Shedler QPD Panel (Quick PsychoDiagnostics Panel): a psychiatric "lab test" for primary care. In: Maruish M (ed). *Handbook of Psychological Assessment in Primary Care Settings*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
49. Wood F, Pill R, Prior L, Lewis G. Patients' opinions of the use of psychiatric case-finding questionnaires in general practice. *Health expect* 2002; 5: 282-8.
50. Wittkampf KA, van Zwieten M, Smits FT *et al*. Patients' view on screening for depression in general practice. *Fam Pract* 2008; 25: 438-44.
51. Nederlandse Vereniging voor Psychiatrie. *Vijftien veel gestelde vragen over DSM-5*. [https://www.boompsychologie.nl/actueel-item/80-1007\\_NVvP-beantwoordt-15-veelgestelde-vragen-over-de-DSM-5](https://www.boompsychologie.nl/actueel-item/80-1007_NVvP-beantwoordt-15-veelgestelde-vragen-over-de-DSM-5). Archived at: <http://www.webcitation.org/6wpYQK8Wg>, 2014
52. Spitzer RL, Williams JB, Kroenke K *et al*. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *JAMA* 1994; 272: 1749-56.
53. Terluin B, van Marwijk H, Adèr H *et al*. The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. *BMC Psychiatry* 2006; 6: 34.
54. Spitzer RL, Kroenke K, Williams JW. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999; 282: 1737-44.
55. Mitchell A, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009; 374: 609-19.
56. Donker T, van Straten A, Marks I, Cuijpers P. A brief Web-based screening questionnaire for common mental disorders: development and validation. *J Med Internet Res* 2009; 11: e19.
57. Donker T, Comijs H, Cuijpers P *et al*. The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Res* 2010; 30; 176: 45-50.
58. Brentnall EW. A history of triage in civilian hospitals in Australia. *Emerg Med Australas* 1997;9:50-4.
59. FitzGerald GJ. The national triage scale. *Emerg Med Australas* 1996; 8: 205-6.
60. Broadbent M, Moxham L, Dwyer T. The development and use of mental health triage scales in Australia. *Int J Ment Health Nurs* 2007; 16: 413-21.



## Chapter 2

### eDiagnostics: a promising step towards primary mental health care

Iles Dijkman  
Geert-Jan Dinant  
Mark G. Spigt

*Fam Pract* 2013;30(6):695-704

## Abstract

**Introduction.** There is a growing interest in eHealth applications in daily health care. Considering that a psychological examination, to a large extent, consists of filling out questionnaires, the use of the Internet seems logical. We evaluated an eDiagnostic system for mental health disorders that has recently been introduced in primary care in the Netherlands.

**Method.** We monitored the diagnoses produced by the system. Evaluation questionnaires from both GPs/practice nurses (PNs) and patients were collected. In addition, we compared the advice produced by the GPs/PNs and the advice produced by the system.

**Results.** The most prevalent disorders were mood, anxiety, and somatoform disorders ( $n = 353$ ). Patients ( $n = 242$ ; 69% response rate) were moderately enthusiastic about the eHealth approach and GPs/PNs ( $n = 49$ , 72% response rate per practice) were very enthusiastic. Patients showed no clear preference for a face-to-face consultation with a psychologist over an eDiagnostic system. GPs/PNs felt strengthened in their control function. In most cases, the system gave a different echelon advice (i.e. referral to primary or secondary mental health care) than the GPs/PNs ( $\kappa = 0.13$ ,  $P = .003$ ). Nevertheless, GPs/PNs accept the results of the examination and the advice given.

**Conclusions.** Using the Internet to diagnose mental health problems in primary care seems very promising. This system of using eDiagnostics before referral to a mental health institution may change the management of mental health care. Further research should investigate whether this tool is valid, reliable, and (cost) effective.

## Introduction

Mental disorders are highly prevalent worldwide<sup>1</sup>. Anxiety and mood disorders are most prevalent<sup>1</sup>. Mental health disorders are associated with reduced quality of life and work disability<sup>2,3</sup>, impaired social functioning, and high societal costs<sup>4</sup>. Symptoms are often missed<sup>5-7</sup> or incorrectly labelled by GPs<sup>8,9</sup>, which could lead to a delay in receiving the appropriate help<sup>10</sup>. Early detection is necessary to prevent mild cases from developing into severe ones, prevent the onset of secondary disorders<sup>11</sup>, and reduce the above-mentioned risks<sup>2-4</sup>. In the Netherlands, patients usually visit their GPs, who act as a gatekeeper, first. However, it is questionable whether GPs are capable of making a good initial judgement, since research has shown that 40% of diagnoses are not recognized by GPs (e.g. because patients do not ask for help, lack insight into their problems, attribute symptoms to somatic complaints, or normalize their complaints)<sup>5,7,12</sup>.

Early detection of mental disorders might be achieved by eHealth/eMental-health, which is the use of information and communication technology (ICT) to support (mental) health care<sup>13,14</sup>. This method could bridge the gap between primary and specialist health care, as it allows an intense exchange of information without costing more consultation time. There is increased international interest in eHealth, but the cost-effectiveness of most eHealth programmes requires further investigation<sup>15</sup>. Within the field of mental health care, eHealth has mainly been successful in the field of teletherapy<sup>16-18</sup>. Only a few online screening questionnaires have been developed to triage mental health disorders in primary care<sup>19-22</sup>. Those screening devices focused specifically on more commonly occurring mental disorders. No research has been dedicated to an online psychodiagnostic system to triage most disorders, included in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)<sup>36</sup>. Considering that a psychological examination, to a large extent, consists of filling out questionnaires, the use of the Internet in daily health care seems logical.

The great advantage of an Internet-based examination is that patients can complete diagnostic tests at home<sup>23</sup>. In addition, people tend to reveal themselves faster via the Internet, and, due to the perceived anonymity, responses are less influenced by social desirability<sup>24-26</sup>. Furthermore, suicidality can be better predicted using computer-delivered tests<sup>27</sup>. This could lead to a more accurate information collection process compared to conventional clinical interviews<sup>26</sup>.

On the other hand, the introduction of Internet-based applications in everyday care might meet unforeseen obstacles. For example, there are no visual cues to guide the physician<sup>28</sup>. In addition, the use of the Internet may not be suitable for crisis-prone patients or patients uncomfortable with expressing their problems in writing<sup>28</sup>. Research has shown resistance (mainly from therapists) to using a computer for therapeutic purposes,<sup>29</sup> and new implementations are generally not easily adopted by GPs<sup>30-33</sup>. Therefore, it is uncertain whether GPs and patients are ready and willing to use eDiagnostics for mental health problems.

We evaluated an online diagnostic system that assesses the nature and severity of possible mental disorders to support GPs in their control function. This system has recently been introduced in the Netherlands. This evaluation study was conducted using online questionnaires to examine whether patients, GPs, and practice nurses (PNs) would accept such a system.

## Methods

### *Design*

We performed surveys on users (GPs, PNs, and patients) of an eDiagnostic system called TelePsy. This system was initiated in September 2011. In this survey, all users were evaluated during the period of February to June 2012, when the system had reached a stable phase.

### *Setting*

In the Netherlands, patients first go to their GP if they need medical attention. The GP decides whether and what type of care is appropriate. The PN supports the GP in patients with mental health problems. A referral letter is necessary to gain access to specialist care, so the GP acts as a gatekeeper. This referral system is also implemented in other countries (for example, the UK, Denmark, Norway and Spain) and differs from a system in which patients have direct access to specialist care (e.g. in Belgium, France, Germany, Finland and Sweden)<sup>34,35</sup>.

### *The eDiagnostic system*

TelePsy is a health care institution that helps caregivers triage psychological and psychosocial problems. GPs can refer patients whom they suspect to have psychological problems, except for crisis-prone patients, illiterates, and patients who are unable to speak Dutch. TelePsy does not treat patients but helps to identify symptoms and possible causes of psychological problems. The procedure (shown in Fig. 1) starts after patients are referred to TelePsy by their GP/PN. Queries submitted by GPs frequently include whether the patient has a mood or anxiety disorder, whether there is an indication for a personality disorder, and whether patients should be referred to primary or secondary mental health care. Independent of the GP's ideas, referral results in the start of the e-tool, the TeleScreen. First, patients receive an e-mail with a secured electronic link to the online questionnaires. Patients give permission to share results with the GP/PN. Second, to screen axis 3–5 of the DSM-IV-TR, patients answer questions to the regarding demographic questionnaire (DS 3–5). Additionally, the diagnostic and statistic screening questionnaire, also based on the DSM-IV-TR<sup>36</sup>, for most axis I disorders is administered (DS 1).

If there is an indication for a disorder, specialized validated questionnaires are administered to rate the severity of the disorder. The Psy Questionnaire Netherlands (PQN) is administered to collect data on personality traits and coping strategies. Finally, patients are asked about their preferences in relation to therapy, using the Patient Preferences questionnaire (PP). A full overview of all questionnaires and validation information is included as Appendix 1. Third, the TelePsy psychologist reviews the answers and makes a working hypothesis based on the generated profile. Fourth, the psychologist contacts the patient by telephone. During this consultation, the psychologist checks the software-generated profile and builds a more specific understanding of the nature and scope of the patient's complaints. An optional step is to select an additional questionnaire to clarify complaints. Fifth, based on both the telephone consultation and the software-generated profile, the psychologist prepares a report for the GP. The report contains information about the preliminary DSM-IV-TR<sup>36</sup> classification and provides an echelon advice (i.e. no referral or referral to primary, secondary, or tertiary mental health care). The echelon advice is based on the guidelines of the Dutch Healthcare Authority ([www.nza.nl](http://www.nza.nl)) and Trimbos Institute (Netherlands Institute of Mental Health and Addiction, see also: [www.ggzrichtlijnen.nl](http://www.ggzrichtlijnen.nl)). Appendix 2 provides an overview of the algorithms in the system. A governmental website in the Netherlands ([www.kiesbeter.nl](http://www.kiesbeter.nl)) that contains independent information about local health-care services is used and (when possible) matched to patient preferences (i.e. preferences for number of kilometres to travel). Here, the role of TelePsy ends. Sixth, the patient and the GP decide which steps are to be taken. After a week, patients receive an evaluation questionnaire to evaluate the process. GPs/PNs receive a single evaluation questionnaire.

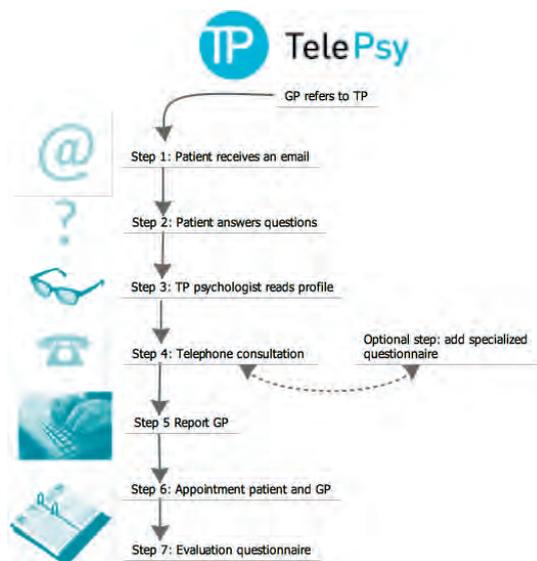


Figure 1. TelePsy Procedure

### *Study subjects*

*GPs and PNs.* In collaboration with several GPs, PNs, and mental health professionals, TelePsy started the development of the TeleScreen in 2010. From 2010, practices were recruited through information meetings. All users (99 GPs and PNs) were contacted through an electronic link connected to an e-mail address. In order to increase the response rate, we contacted the GPs/PNs who did not respond.

*Patients.* All patients were 18 years or older. Only patients who completed the entire procedure were included. A week after the dossier was closed, patients were invited to complete the evaluation questionnaire.

### *Questionnaires*

We used two different evaluation questionnaires, but all questions were scored on a scale from 1 to 5 (i.e. 1 = 'extremely satisfied' to 5 = 'extremely unsatisfied'). The patient questionnaire consisted of questions about the e-tool, contact with the psychologist, and an overall impression of TelePsy. The questionnaire for the GPs/PNs consisted of questions about the echelon advice, the use of the eDiagnostic system, the advice given, and their general impression (i.e. 'Do you think TelePsy will benefit patients?').

## **Results**

### *Participants*

*Patients.* A total of 353 patients completed the entire TelePsy procedure and received an invitation for participation. Of these patients, 242 (99 men and 143 woman) responded (69% response rate). The age of those who responded varied from 18 to 74 years, with a mean age of 42 (SD = 13.34). The characteristics of the patients are shown in Table 1. The most prevalent diagnoses were mood disorders (36%), anxiety disorders (18%), and somatoform disorders (8%). In total, 26% of the patients reported one or more co-morbid disorders (see Table 2). The two main co-morbid disorders were mood and anxiety disorders. Overall, 15% of the patients (64 out of 417 patients) did not complete the entire TelePsy procedure (no-show) and were therefore excluded from the evaluation study. Thirty per cent of the no-shows were patients who never started the test and 40% completed the questionnaire but missed the telephone consultation. Since these patients could not evaluate the entire process, they were excluded from the evaluation study. The reason for their non-compliance remained unknown. Improvements to the process of setting appointments with patients were completed after two months; after this improvement, the no-show rate decreased considerably from 15% to 6%. Due to technical problems a small number of patients were not invited for the evaluation study.

**Table 1.** Patient characteristics: DSM-IV-TR Classification system \*

	Non- Responders ( <i>n</i> = 111)	Responders ( <i>n</i> = 242)	Total ( <i>n</i> = 353)
	Diagnoses ( <i>n</i> = 150)	Diagnoses ( <i>n</i> = 329)	Total ( <i>n</i> = 479)
No diagnosis <i>N</i> (%)	9 (2)	14 (3)	23 (5)
Postponed diagnosis <i>N</i> (%)	5 (1)	7 (2)	12 (3)
Mood disorder <i>N</i> (%)	46 (10)	124 (26)	170 (36)
Depressive disorder	26%	64%	
Dysthymic disorder	0%	1%	
Bipolar disorder	1%	0%	
Depressive disorder not otherwise specified	1%	9%	
Anxiety disorder <i>N</i> (%)	24 (5)	64 (13)	88 (18)
Panic disorder	7%	27%	
Social phobia	7%	11%	
Post-traumatic stress disorder	5%	13%	
Obsessive-compulsive disorder	7%	8%	
Other	2%	10%	
Somatiform disorder <i>N</i> (%)	12 (3)	28 (6)	40 (8)
Somatization disorder	5%	3%	
Body dismorphic disorder	8%	10%	
Undifferentiated somatiform disorder	18%	53%	
Other	0%	5%	
Suspicion of a personality disorder <i>N</i> (%) <sup>a</sup>	10 (2)	16 (3)	26 (5)
Substance abuse disorder <i>N</i> (%)	11 (2)	11(2)	22 (5)
Alcohol	23%	36%	
Cannabis	5%	9%	
Cocaine	5%	0%	
Other	14%	9%	
Developmental disorder <i>N</i> (%)	10 (2)	12 (3)	22 (5)
Attention-deficit/hyperactivity disorder	23%	18%	
Autistic spectrum disorder	23%	36%	
Adjustment disorder <i>N</i> (%)	6 (1)	13 (3)	19 (4)
Psychotic disorder <i>N</i> (%)	3 (1)	4(1)	7 (2)
Eating disorder <i>N</i> (%)	2 (0)	9 (2)	11 (2)
Sleeping disorder <i>N</i> (%)	2 (0.5)	2 (0.5)	4 (1)
Other <i>N</i> (%) <sup>b</sup>	10 (2)	27 (6)	37 (8)

\* Note: these classifications are not DSM-IV-TR diagnoses; they are indicative and must be further examined.

<sup>a</sup> We recommended further research into underlying dysfunctional personality traits.

<sup>b</sup> Including dissociative disorder, depersonalization disorder, tic disorder, impulse control disorder, disruptive behaviour disorder, pathological gambling, identity disorder, sexual and gender identity disorder and V-codes.

**Table 2.** Co-morbid axis I diagnoses

	Non-responders diagnoses (n = 150)	Responders diagnoses (n = 329)	Total diagnoses (n = 479)
No Co-morbid axis I disorder	23%	51%	74%
Co-morbid axis I disorder	8%	18%	26%
1	6%	14%	-
2	2%	4%	-

*GPs and PN.* A total of 99 GPs/PNs were invited to participate and 49 responded (49% response rate). Among these respondents were 34 GPs and 15 PNs (18 men and 31 women). The age of the GPs/PNs varied from 26 to 58 years, with a mean age of 44 years. Of note, GPs and PNs within the same practice both received invitations. Some GPs may have decided to delegate the survey to the PN in the practice rather than complete the survey themselves; this was counted as a non-response. The response rate per practice was 72% (43 out of 60 practices).

### *Patient satisfaction*

Figure 2 shows the mean scores and standard deviation per question. Generally, patients were quite satisfied with the eDiagnostic computer system. As shown in Fig. 2, the average score on most questions was rated between 'neutral and satisfied' except for the questions on 'computer vs. psychologist' and 'emotional impact'. In general, patients reported being neutral on whether they would prefer a psychologist or an electronic questionnaire, but there seemed to be a slight preference for a psychologist (mean = 2.81, SD = .88). Examples of patients' elaborations on the item 'computer vs. psychologist' were less time pressure in the absence of a person, a user-friendly computer system, easy access, and less feelings of shame. On the other hand, some felt that they were less able to adequately express themselves via the system than with a face-to-face psychologist, needed help operating their computers, or had difficulty navigating through the online questionnaires. Overall, patients reported being neutral on the emotional impact of the process, but with quite a large standard deviation (mean = 2.88, SD = .93). The satisfaction rate of the patients, graded on a scale of 1 to 10, was 7.45 (SD = 1.11; data not shown in table). Of the 242 patients, 9 patients rated their satisfaction as  $\leq 5$ . One of these patients had a principal objection against an eDiagnostic system for psychological complaints, three patients did not agree with the results or advice of the psychological examination, and two patients were dissatisfied with the psychologist. The reasons for the remaining three low grades remained unknown.

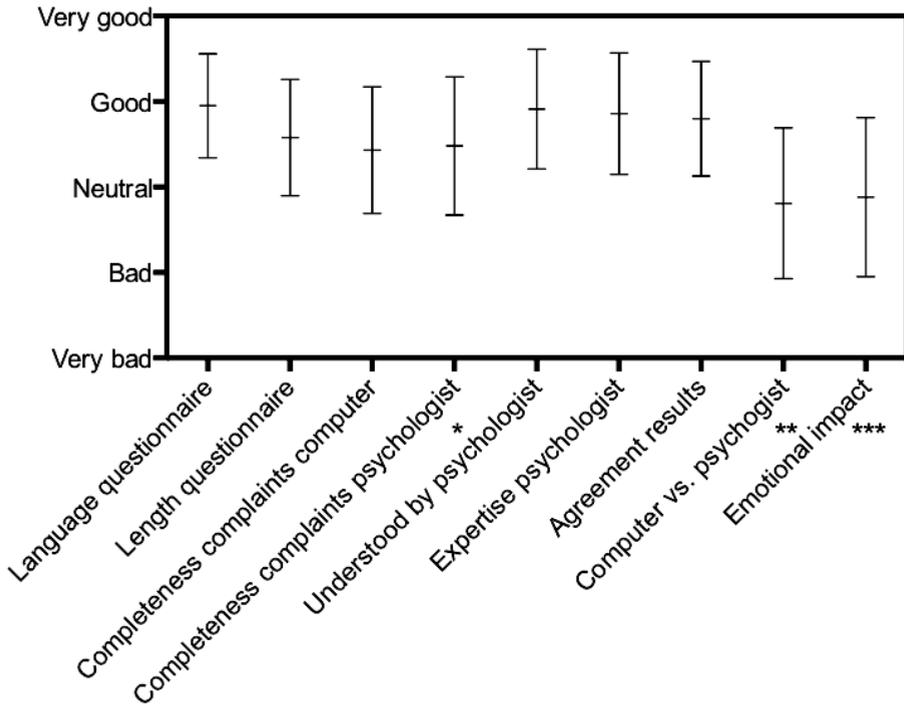


Figure 2. Patient satisfaction (n = 242)

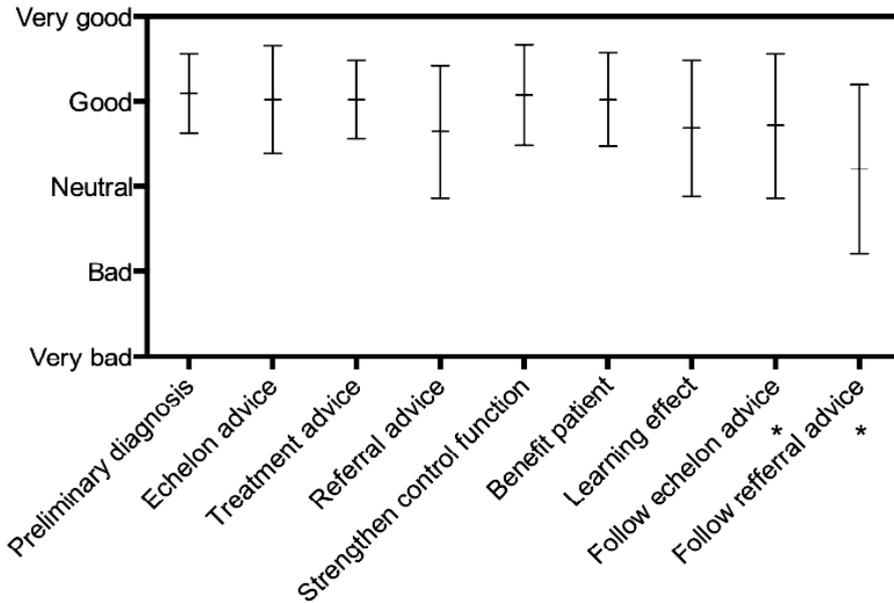
\* Only asked to those scoring negative or neutral on the preceding question (n = 127).

\*\* Rated on a scale from 1 to 5, from 'rather a psychologist' to 'rather a computer'.

\*\*\* Rated on a scale from 1 to 5, from 'very high emotional impact' to 'very low emotional impact'.

### GP/PN satisfaction

As shown in Fig. 3, GPs/PNs were very satisfied with the eDiagnostic computer system for their patients. They agreed with the preliminary diagnosis and echelon advice, felt strengthened in their control function, and thought that the system would benefit the patients. A slightly lower score was observed on the three questions, 'In general, did you follow the referral advice?', 'Did the reports have a learning effect?', and 'In general, did you follow the echelon advice?'. The neutral score on the question: 'Did you follow the referral advice?' stood out (see Fig. 3). Finally, the mean satisfaction rate of GPs/PNs, expressed in a grade from 1 to 10, was 8.07 (SD = .87, data not shown in table).



**Figure 3.** GP/PN satisfaction ( $n = 44^a$ )

<sup>a</sup>This number is lower due to the fact that some did not fill out the questions.

\* Rated on a scale from 1 to 5, from 'always' to 'never'.

### Agreement between GPs/PNs and the eHealth system

A total of 215 cases could be retrieved from which we had both a reaction from the GP/PN and an eHealth echelon advice. As seen in Table 3, there was no agreement between the GP/PN and the eHealth system ( $\kappa = 0.13, P = .003$ ). It is noteworthy that the eHealth system advised to refer 42 times (20%) to the PN, 49 times (23%) to primary care, and 112 (52%) times to secondary health care. The GPs/PNs would have referred 41 times (19%) to the PN, 84 times to primary care (39%) and 74 (34%) times to secondary health care. When the scores were categorized into three categories: primary, secondary health care, and other, Cohen's kappa was 0.12 ( $P = 0.03$ ; Table 4).

**Table 3.** Agreement between GPs/NPs ( $n = 215$ ) and eDiagnostic system ( $\kappa = 0.13, P = 0.003$ )

	eDiagnostic system					Total
	Practice Nurse	Primary health care	Secondary health care	Tertiary health care	Other	
Practice Nurse	6%	3%	9%	0%	1%	41 (19%)
Primary health care	8%	11%	19%	0%	1%	84 (39%)
Secondary health care	4%	7%	22%	0%	1%	74 (34%)
Tertiary health care	0%	0%	0%	0%	0%	1 (0%)
Other	2%	2%	2%	0%	1%	15 (7%)
Total	42 (20%)	49 (23%)	112 (52%)	1(0%)	11(5%)	215 (100%)

**Table 4.** Agreement between GPs/NPs ( $n = 215$ ) and eDiagnostic system ( $\kappa = 0.12$ ,  $P = 0.03$ )

		eDiagnostic system			Total
		Primary health care	Secondary health care	Other	
GPs/NPs	Primary health care	29%	29%	3%	132 (61%)
	Secondary health care	12%	22%	1%	74 (35%)
	Other	3%	1%	0%	9 (4%)
Total		94 (44%)	112 (52%)	9 (4%)	215 (100%)

## Discussion

### *Summary of the most important results*

The patients showed no clear preference for a face-to-face consultation with a psychologist over an eDiagnostic system for mental health problems. The patients were quite satisfied; they felt understood, agreed with the results of the examination, and were satisfied with the degree to which their complaints were taken into account by the computer or by the follow-up telephone consultation with the psychologist. GPs/PNs were very satisfied with the eDiagnostic system. They felt strengthened in their control function and thought the system would benefit their patients. In most cases, the system gave a different echelon advice than the GPs/PNs would have expected. The GPs/PNs reported agreement with the results of the examination and the advice given, but quite often they did not follow the specific referral advice.

### *Possible explanations of the evaluation results*

A literature study by Emmelkamp (2005)<sup>29</sup> showed resistance from mainly therapists, not patients, in using a computer for therapeutic purposes. This study is, to our knowledge, the first to investigate patients' attitudes toward eDiagnostics. Our results are in line with the above-mentioned results for e-therapy, showing no clear resistance to a computer-based system.

Since many studies report difficulties concerning the adoption of new interventions by GPs, we expected a restrained attitude towards the implementation of the eDiagnostic system<sup>30-33</sup>. Research shows that the implementation of innovative plans is complex. The nature of the innovation, GPs and patients' characteristics, and economic, organisational, social, and political factors are involved in the process<sup>31-33</sup>. Taking the literature into account, the positive evaluations of this system by GPs/PNs are surprising. The eDiagnostic system is not time-consuming for GPs/PNs. They do not need to interpret the questionnaires themselves or change their behaviour, and they receive additional information about their patients, which increases their control function. The agreement of GPs with the results of the examination and the advice given is surprising. It is conceivable that GPs, when they receive the full diagnostic report, also follow the line of reasoning for the

echelon advice given by TelePsy, but the exact reasons why they agree so easily with an unexpected advice remains unknown. The only score that seems to be discordant with the positive trend is the neutral reaction of GPs/PNs on the follow-up of the referral advice. This low score may be explained by GPs'/PNs' preference to use their own established referral channels.

### *Echelon results in relation to the literature*

The results show surprisingly high disagreement between GPs/PNs and the system concerning the echelon advice. It is not possible to make a judgment on the validity of the eDiagnostic system based on the results of this study. Research showed that the disadvantage of screening tools is the possibility of false positives, where healthy patients could be labelled with one or more diagnoses<sup>37</sup>. Considering this, the preliminary DSM-IV-TR classification and the subsequent echelon advice of the eDiagnostic system might represent a slight overdiagnosis. The advantage of the GPs/PNs over an eDiagnostic system is their history with the patient. Research has shown that an increase in office visits and contact with the patient improves the detection of a mental disorder by GPs<sup>5,38,39</sup>. The eDiagnostic system, however, is based on information filled out at a specific moment and lacks the extra information that the GPs/PNs might have. This might hinder the eDiagnostic system's judgment regarding the echelon advice. On the other hand, GPs/PNs might not be able to extensively explore every complaint due to a short consultation time. This might positively influence the judgment given by the eDiagnostic system. Finally, GPs/PNs are dependent on the presentation of complaints by patients. If patients lack insight, interpret their complaints as somatic, or normalize their problems, potential diagnoses are missed<sup>5,7,12</sup>. This may impede GPs'/PNs' judgment regarding the echelon advice. In the Netherlands, the policy is to treat more patients in primary instead of secondary mental health care. Although the results of this study show more secondary than primary health care referrals generated by the eDiagnostic system, this should not be interpreted as representative of demographic referral statistics. Since GPs decide which patients undergo the TeleScreen, the results discussed here are not based on a random demographic patient selection.

### *Limitations*

These high positive evaluations should be considered against the background of some limitations. It is conceivable that the GP/PN group is biased. First, the GPs/PNs in our study could be considered as early adopters, who are more prepared to change, open to new ideas, and willing to take risks<sup>32</sup>. The current positive evaluations may over-represent the true satisfaction scores of GPs/PNs. Second, only patients who completed the process were included. It is possible that patients who did not complete the procedure (15% no-shows) were those with a negative opinion towards an eDiagnostic system. Finally, the

eDiagnostic system is not suitable for non-Dutch speakers, crisis-prone patients, illiterates, and patients who do not have access to the Internet.

*Conclusions, clinical implications, and future research*

To our knowledge, this is the first study to evaluate an eDiagnostic system for mental health problems implemented in routine primary care. It can be concluded that patients are quite satisfied with an eDiagnostic system. GPs/PNs are satisfied with such a system for their patients. In most cases, the system gave a different echelon advice from what the GPs/PNs would have given. Nevertheless, the GPs/PNs agreed with the results of the examination and the advice given. This system may change the management of mental health care by using eDiagnostics *prior* to a referral to a mental health institution. We evaluated the system in a gatekeeper-dependent system. However, we believe that the system could also be useful in countries in which direct access to specialist care is possible. Among other benefits, eDiagnostics could lead to a more (cost) efficient process, a reduction in undetected disorders, a reduction in the burden on the patient, and an improved control function of the GP. The question remains whether it is possible to make a correct estimation of the nature and severity of symptoms through the Internet and a telephone consultation. Now we know that patients and GPs/PNs are satisfied with the system; further research should investigate if this system is valid, reliable, and cost-efficient.

## References

1. Kessler RC, Angermeyer M, Anthony JC *et al.* Prevalence, Severity, and Unmet Need for Treatment of Mental Disorders in the World Health Organization World Mental Health Surveys. *JAMA* 2004; 291: 2581-2590.
2. Kessler RC, Aguilar-Gaxiola S, Alonso J *et al.* The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiol Psychiatr Soc* 2009; 18: 23-33.
3. Alonso J, Angermeyer M, Bernert S *et al.* Disability and quality of life impact of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMed) project. *Acta Psychiatr Scand* 2004; 109: 38-46.
4. Wittchen HU, Jacobi F. Size and burden of mental disorders in Europe—a critical review and appraisal of 27 studies. *Eur Neuropsychopharmacol* 2005; 15: 357-76.
5. Verhaak PF, Schellevis FG, Nuijen J, Volkerts AC. Patients with a psychiatric disorder in general practice: determinants of general practitioners' psychological diagnosis. *Gen Hosp Psychiatry* 2006; 28: 125-32.
6. van Beljouw I, Verhaak P, Prins M *et al.* Reasons and determinants for not receiving treatment for common mental disorders. *Psychiatr Serv* 2010; 61: 250-7.
7. Kessler D, Lloyd K, Lewis G, Gray DP. Cross sectional study of symptom attribution and recognition of depression and anxiety in primary care. *BMJ* 1999; 318: 436-9.
8. Janssen EH, van de Ven PM, Terluin B *et al.* Recognition of anxiety disorders by family physicians after rigorous medical record case extraction: results of the Netherlands Study of Depression and Anxiety. *Gen Hosp Psychiatry* 2012; 34: 460-7.
9. Joling KJ, van Marwijk HW, Piek E *et al.* Do GPs' medical records demonstrate a good recognition of depression? A new perspective on case extraction. *J Affect Disord* 2011; 133: 522-7
10. Wang PS, Berglund P, Olfson M *et al.* Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005; 62: 603-13.
11. Kessler RC, Merikangas KR, Berglund P *et al.* Mild disorders should not be eliminated from the DSM-V. *Arch Gen Psychiatry* 2003; 60: 1117-22
12. Jorm AF. Mental health literacy. Public knowledge and beliefs about mental disorders. *Br J Psychiatry* 2000; 177: 396-401
13. Eysenbach G. What is e-health? *J Med Internet Res* 2001; 3: E20.
14. Riper H, Andersson G, Christensen H *et al.* Theme issue on e-mental health: a growing field in internet research. *J Med Internet Res* 2010; 12: e74
15. Black AD, Car J, Pagliari C *et al.* The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Med* 2011; 8: e1000387.
16. Spek V, Cuijpers P, Nyklíček I *et al.* Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychol Med* 2007; 37: 319-28.
17. Cuijpers P, Marks IM, van Straten A *et al.* Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. *Cogn Behav Ther* 2009; 38: 66-82.
18. Riper H, van Straten A, Keuken M *et al.* Curbing problem drinking with personalized-feedback interventions: a meta-analysis. *Am J Prev Med* 2009; 36: 247-55.
19. Gega L, Kenwright M, Mataix-Cols D, Cameron R, Marks IM. Screening people with anxiety/depression for suitability for guided self-help. *Cogn Behav Ther* 2005; 34: 16-21.
20. Farvolden P, McBride C, Bagby RM, Ravitz P. A Web-based screening instrument for depression and anxiety disorders in primary care. *J Med Internet Res* 2003; 5: e23.
21. Lin CC, Bai YM, Liu CY *et al.* Web-based tools can be used reliably to detect patients with major depressive disorder and subsyndromal depressive symptoms. *BMC Psychiatry* 2007; 7: 12.
22. Donker T, van Straten A, Marks I, Cuijpers P. A brief Web-based screening questionnaire for common mental disorders: development and validation. *J Med Internet Res* 2009; 11: e19.
23. Buchanan T. Online assessment: desirable or dangerous? *Prof Psychol Res Pract* 2002; 33: 148
24. Joinson A. Social desirability, anonymity, and Internet-based questionnaires. *Behav Res Methods Instrum Comput* 1999; 31: 433-8
25. Richman WL, Kiesler S, Weisband S, Drasgow F. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *J Appl Psychol* 1999; 84: 754.

26. Newman MG. Technology in psychotherapy: an introduction. *J Clin Psychol* 2004; 60: 141–5.
27. Levine S, Ancill RJ, Roberts AP. Assessment of suicide risk by computer-delivered self-rating questionnaire: preliminary findings. *Acta Psychiatr Scand* 1989; 80: 216–20.
28. Rochlen AB, Zack JS, Speyer C. Online therapy: review of relevant definitions, debates, and current empirical support. *J Clin Psychol* 2004; 60: 269–83.
29. Emmelkamp PMG. Technological innovations in clinical assessment and psychotherapy. *Psychother Psychosom* 2005; 74: 336–43.
30. Roche AM, Freeman T. Brief interventions: good in theory but weak in practice. *Drug Alcohol Rev* 2004; 23: 11–8.
31. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. *Med J Aus* 2004; 180: 57.
32. Grol R. Implementing guidelines in general practice care. *Qual Health Care* 1992; 1: 184–91.
33. Vogt F, Armstrong D, Marteau TM. General practitioners' perceptions of the effectiveness of medical interventions: an exploration of underlying constructs. *Implement Sci* 2010; 5: 17.
34. Kulu-Glasgow I, Delnoij D, de Bakker D. Self-referral in a gatekeeping system: patients' reasons for skipping the general practitioner. *Health Policy* 1998; 45: 221–38.
35. Verhaak PF, van den Brink-Muinen A, Bensing JM, Gask L. Demand and supply for psychological help in general practice in different European countries: access to primary mental health care in six European countries. *Eur J Public Health* 2004; 14: 134–40.
36. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders 4th edition, Text Revision (DSM-IV-TR)*. Washington, DC: American Psychiatric Association, 2000.
37. Grimes DA, Schulz KF. Uses and abuses of screening tests. *Lancet* 2002; 359: 881–4.
38. Bushnell J; MaGPIe Research Group. Frequency of consultations and general practitioner recognition of psychological symptoms. *Br J Gen Pract* 2004; 54: 838–43.
39. Klinkman MS, Coyne JC, Gallo S, Schwenk TL. False positives, false negatives, and the validity of the diagnosis of major depression in primary care. *Arch Fam Med* 1998; 7: 451–61.
40. Draper H, Wilson S. Research ethics approval: comprehensive mechanisms are essential but not available. *Fam Pract* 2007; 24: 527–8.
41. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV-Axis I disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/ PSY SCREEN)*. New York: Biometrics Research, New York State Psychiatric Institute, 2002.
42. Groenestijn M, Akkerhuis G, Kupka R, Schneider N, Nolen W. *Gestructureerd klinisch interview voor de vaststelling van DSM-IV As-I stoornissen*. Lisse, The Netherlands: Swets en Zeitlinger, 1999.
43. Sheehan D, Lecrubier Y, Sheehan K *et al*. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatr* 1998; 59: 22–33.
44. First MB, Gibbon M, Spitzer RL, Williams JBW, Benjamin LS. *Structured Clinical Interview for DSM-IV Axis II Personality disorders (SCID-II)*. Washington, DC: American Psychiatric Press, 1997.
45. van der Does AJW. BDI-II-NL. Handleiding. *De Nederlandse versie van de Beck Depression Inventory-2nd edition*. Lisse, The Netherlands: Harcourt Test Publishers, 2002.
46. Evers A, van Vliet-Mulder JC, Groot CJ. *Documentatie van tests en testresearch in Nederland, aanvulling 2005/01 (COTAN)*. Amsterdam, The Netherlands: Boom test uitgevers, 2005.
47. Willems LFM, Tuender-de Haan HA, Defares PB. Een schaal om sociale angst te meten. *Ned Tijdschr Psychol* 1973; 28: 415–22.
48. Saunders JB, Aasland O, Babor T, Fuente J, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction* 1993; 88: 791–804.
49. Kraaimaat F. De Inventarisatielijst Dagelijkse Bezigheden (IDB). *Gedragstherapie* 1994; 27: 149–53.
50. Fairburn C, Beglin S. The assessment of eating disorders: interview versus questionnaire. *Int J Eating Disord* 1994; 16: 363–70.
51. Van Furth E. *Nederlandse vertaling van de EDE-Q*. Leidschendam, The Netherlands: Robert-Fleury Stichting, 2000.

## Appendix 1.

Name	Item information
DS 3-5 (Axis 3-5 of the DSM-IV-TR) <sup>36</sup>	Demographic items, possible somatic complaints, the patients' request for help, patients' earlier psychological treatment and psychosocial problems.
DS 1 <sup>a</sup> (Most Axis I disorders are requested)	This questionnaire is developed by TelePsy and has not been validated yet. The questionnaire is inspired by the Structured Clinical Interview DSM-IV-TR <sup>41,42</sup> and Mini-International Neuropsychiatric Interview-500 (MINI-500). <sup>43</sup> The questionnaire screens for most Axis-I disorders based on the DSM-IV-TR. <sup>36</sup>
Psy Questionnaire Netherlands–Personality (PQN-P)	Personality traits and coping strategies are requested.
Psy Questionnaire Netherlands–Coping (PQN-C)	
Patient Preferences	Patient preferences in relation to therapy are administered. Also, items regarding the motivation for therapy are included.
Specialized validated questionnaires <sup>b</sup>	<ul style="list-style-type: none"> <li>–Structured Clinical Interview DSM-IV-II<sup>44</sup> (SCID-II)</li> <li>–Beck Depression Inventory - second edition Dutch version (BDI-II-NL)<sup>45,46</sup></li> <li>–Sociale Angst Schaal<sup>47</sup> (Social Anxiety Scale)</li> <li>–Alcohol Use Disorder Identification test<sup>48</sup></li> <li>–Inventarisatielijst Dagelijkse Bezigheden<sup>49</sup> (Self-rating questionnaire for obsessive-compulsive behaviour).</li> <li>–Eating Disorder Examination–Questionnaire<sup>50,51</sup></li> <li>–Klachtenlijst Chronische Stress (Chronic Stress Questionnaire)</li> </ul>

<sup>a</sup>This self-developed questionnaire adapts to positive-answered items. There are standard questions and advanced questions. The advanced questions are only activated when the standard questions are answered positively.

<sup>b</sup>Only when cutoff levels are reached, specialized validated questionnaires are administered in order to rate the severity of the disorder.

## Appendix 2.

Referral (characteristics)	To (job description)	If present/necessary
No referral/self-help		No request for help/no symptoms/symptoms in remission
Primary mental health care	Practice nurse (support GP within the field of mental health care)	<i>Aimed at</i> clarifying the problem-screening diagnostics <i>And/or</i> Prepare and discuss referral steps (under responsibility of the GP) <i>And/or</i> Motivational consult <i>And/or</i> Psycho-education <i>And/or</i> Guidance or support self-management <i>And/or</i> Intervention aimed at improvement in functioning for patients with mental health problems (i.e. behavioural change or support in self-help) <i>And/or</i> Prevention (indicated prevention, health care-related prevention, relapse prevention)
	Social work (practical and social interventions)	Axis IV (psychosocial problems)
	Primary care psychologists	Maximum of two DSM-IV axis I disorder(s) <i>And</i> No comorbidity axis II disorder <i>And</i> Mild symptoms Global assessment of functioning (GAF) score 51> <i>And</i> First episode (not recurrent) <i>Or</i> Earlier treatment successfully completed <i>And</i> No significant disruptions or expected psychosis <i>And</i> Brief diagnosis and brief treatment is sufficient
Secondary mental health care Ambulant treatment		Complex-severe axis I disorder(s) <i>And/or</i> Axis II disorder <i>And/or</i> GAF score< 51 <i>And/or</i> History of treatment in primary mental health care unsuccessful.
Tertiary mental health care Treatment/guidance in psychiatric department of a general hospital (e.g. crisis)/residential care		Severe axis II disorders <i>And/or</i> Complex comorbidity (axis I and axis III) <i>And/ or</i> History of treatment in secondary mental health care unsuccessful <i>And/or</i> Other complicated factors (e.g. insufficient intellectual abilities).

For the full text of the guidelines, see [www.nza.nl](http://www.nza.nl) and [www.ggzrichtlijnen.nl](http://www.ggzrichtlijnen.nl).



# Chapter 3A

The concurrent validity of a new eDiagnostic system for mental disorders in primary care

Iles Dijkman  
Geert Jan Dinant  
Mark Spigt

*Fam Pract* 2016;33(6):607-616

## Abstract

**Background.** An eDiagnostic system was implemented to classify mental disorders, to support general practitioners.

**Objective.** Assessing the validity of the system, compared to the psychologists' judgment.

**Methods.** Concurrent validity, using routinely collected data of 675 primary care patients with a suspicion of a mental disorder in the Netherlands. Four psychologists classified the patients according to the DSM-IV-TR. Hundred records were randomly selected to investigate the inter-rater reliability among psychologists. To investigate the concurrent validity of the system the sensitivity, specificity, positive predictive values (PPVs), negative predictive values (NPVs) and Cohen's  $\kappa$ -values ( $\kappa$ -values) were calculated.

**Results.** Inter-rater agreement between psychologists were fair to good or excellent. The system could correctly estimate the echelon (sensitivity range: 0.85- 0.95, specificity range: 0.88 - 0.98) and correctly identify most Axis I classifications (sensitivity: 0.46-1.00, specificity: 0.75-0.99), except for Asperger's, sexual and adjustment disorders (sensitivity: 0.10-0.24, specificity: 0.97-0.99). It could determine the absence of a personality disorder (sensitivity: 0.81, specificity: 0.84, PPV: 0.77, NPV: 0.87 and  $\kappa$ -value: 0.65). The sensitivities and specificities for most specific personality disorders were good, but the PPVs for several specific Axis II classifications were low (PPV range: 0.06-0.77). The system was inaccurate in identifying the global assessment of functioning of patients (e.g.  $\kappa$ -values varied from 0.17-0.46).

**Conclusions.** Generally, the system can be seen as a valid instrument for most DSM-IV-TR classifications in primary care patients. It could assist healthcare professionals in the assessment and classification of mental disorders. Future research should include comparison to an independently administered structured clinical interview.

## Introduction

Mental disorders are highly prevalent worldwide, with the most common being mood and anxiety disorders [1]. They cause high societal costs [2, 3], are associated with work disability [4] [5], high burden of disease [6] and reduced quality of life [4]. General practitioners, who are crucial in the first steps of managing patients with mental disorders, often have to decide in short consultations if – and what type of – care is necessary. However, symptoms can be missed or misinterpreted if, e.g. patients do not ask for help, if complaints are normalized, if symptoms are attributed to somatic complaints, or if patients do not recognize their problems [7-9]. Consequently, many diagnoses are under-recognized [7, 8, 10, 11], over-recognized [12], or incorrectly labelled by general practitioners [13, 14]. Although effective treatments for mental disorders are available, many patients do not receive adequate treatment [10, 15-19], resulting in a poor prognosis including progression from mild to severe disorders or development of comorbid disorders [20-23]. Over-recognized symptoms could lead to misallocation of treatment resources [24-27].

To increase the identification of mental disorders, screening questionnaires have been recommended in general practice [28, 29]. However, general practitioners generally do not adopt these questionnaires in their routine care [30-32]. Questionnaires are considered time-consuming, and limited to detecting a specific mental disorder or general distress only, and providing a numeric score instead of a specific diagnosis needed for an optimal decision [31, 33]. Other reported arguments are the unfamiliarity with the questionnaires, the belief that the instruments are not necessary, lack of available instruments, lack of reimbursement and insufficient knowledge and skills to interpret the scores [31]. Specialists in mental healthcare often use questionnaires, but the process of diagnosing a patient in secondary care often takes several consultations and therefore can only be conducted in a referred population.

eHealth/eMental-health, which involves the use of information and communication technology to support mental healthcare [34, 35], offers new opportunities. The use of the Internet to support caregivers can confer great advantages [36, 37] as it allows an intense exchange of information between the patient and healthcare professional without costing additional time. Nevertheless, only few online screening questionnaires have been developed to identify mental disorders in primary care [38-42]. Most instruments focus only on depression or anxiety disorders [39-41], while an instrument covering all relevant classifications is needed. No studies have investigated an eDiagnostic system to triage most classifications according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) [43].

Recently, an eDiagnostic system of this nature has been implemented in the Netherlands. The system is more comprehensive than a screening system, but less extensive than a complete diagnostic system, since the system provides provisional classifications instead of firm diagnoses. With this system a full psychological examination is available

to all primary care patients of whom the general practitioner has a clinical suspicion of a mental disorder. Patients answer online questions, and are contacted by telephone by a psychologist to check the probable classifications generated by the system (total psychologist consultation time varied from 30 till 45 minutes). The caregiver, usually a general practitioner, receives a report with a preliminary DSM-IV-TR classification and echelon advice (e.g. no referral, referral to primary or secondary healthcare), based on the guidelines. Currently, this system is employed in the Netherlands; approximately 200 General practitioners used this system in the Netherlands in 2013. Access to the data generated by the system, and the subsequent evaluation of the psychologist made it possible for us to validate the process, by comparing the classifications generated by the eDiagnostic system with the classifications made by the psychologists.

## Method

### *Design*

We evaluated the concurrent validity of the eDiagnostic system (called TeleScreen) when compared to the psychologists' judgment, among primary care patients with a suspicion of a mental disorder. Routinely collected data of the patients in the Netherlands (in the province of Limburg) were gathered prospectively, from November 2013 until April 2014. In addition, 100 patient records were randomly selected from the data, to investigate the inter-rater reliability between the psychologists.

### *Setting*

In the Netherlands, if patients need medical attention they first visit their general practitioner. It is the role of the general practitioner to decide whether and what type of care is required. A practice nurse often supports the general practitioner, e.g. in diabetes care. Patients with mental health problems can now be supported by practice nurses specialized in mental healthcare. When general practitioners decide to refer the patients, patients can be referred to primary mental healthcare (e.g. short-term treatments) or secondary mental healthcare (e.g. specialist care for complex or severe disorders).

### *Procedure of the eDiagnostic system as used in this study*

The general practitioners or practice nurses decided which patients they would refer to the eDiagnostic system. Patients were chosen for referral by general practitioners or practice nurses if a clinical suspicion of mental disorders was present by the caregiver. Frequently asked questions by the caregivers were whether there was a suspicion of a mood or anxiety disorder, a personality or developmental disorder, and questions regarding the echelon. The eDiagnostic system is not suitable for crisis-prone patients, illiterates

and patients with insufficient understanding of the Dutch language. When the caregivers decided to refer the patient, they created an account and entered the patients' contact information.

Patients received an e-mail with a secured electronic link to the online questionnaires. They gave permission to share results with their caregiver and provided informed consent to participate in the scientific evaluation. Subsequently, patients answered the online questions to classify Axis I to Axis V classifications, according to the DSM-IV-TR [43]. When the patient met a DSM-IV-TR diagnosis, an indication was automatically generated by the system. The cut-off points were based on the DSM-IV-TR [43]. The eDiagnostic system was developed by TelePsy, a mental healthcare institution in the Netherlands and has not been validated before this time.

The diagnostic and statistical screening questionnaire for Axis-I disorders, called the DS 1, was administered first. The DS 1 was based on the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I) [44] and MINI-International Neuropsychiatric Interview-500 [45]. The DS 1 was adapted to positive answered items. So, if a person fails to meet the required number of symptoms of a disorder, based on the DSM-IV-TR, follow-up questions of that specific disorder were not administered. For example when a patient fails to meet at least one of the two main symptoms of a major depression (depressed mood or loss of interest or pleasure), follow-up questions were not queried by the system. Also, the diagnostic and statistical screening questionnaire for Axis II disorders (personality disorders) was administered (DS 2). This questionnaire was based on the Structured Clinical Interview for DSM-IV-TR Axis II Disorders (SCID-II) [46]. In addition, Axis 3-5 of the DSM-IV-TR was assessed by a demographic questionnaire, called the DS 3-5. Information about patients' somatic complaints, psychosocial problems and the patients' self-reported global assessment of functioning (GAF) was collected.

Subsequently, additional information was automatically queried by the system. This extra information did not have an influence on the generated classifications of the system (i.e. classifications generated by the DS 1 – DS 5), but supported psychologists in their diagnostic judgment. Patients were asked about their work, family, hobbies, request for help, previously received care and preferences in relation to therapy. Validated questionnaires were automatically administered by the system for some Axis I classifications, to rate the severity of the disorder and to form a better judgment of the patients' complaints (i.e. Beck Depression Inventory – II- NL [47], Alcohol Use Disorder Identification Test [48], see Supplementary Table S1). Finally, to support the psychologists in the identification of personality pathology, nine additional questions were administered to check whether the patient met the general diagnostic criteria of a personality disorder. These questions were based on the DSM-IV-TR [43] and were developed by TelePsy.

Within two days after test completion, the psychologists reviewed the answers and made a working hypothesis based on the classifications generated by the system. The generated classifications of the system were always reviewed by the psychologist, even if no classifications were indicated by the system. The psychologists could form their

diagnostic judgment on the eDiagnostic test (e.g. classifications generated by the DS 1 – DS 5). However, to form a better judgment of the patients' complaints the psychologists could also use the additional information of the patient regarding the request for help, history of complaints, the information provided by the validated questionnaires and the answers to the questions concerning the identification of personality pathology.

Finally, psychologists contacted all included patients by telephone, to evaluate the complaints, ask their own questions, critically check the generated classifications of the system and if necessary, clarify ambiguous test results. If necessary, the psychologists could contact their supervisor for support. The psychologists provided their own preliminary DSM-IV-TR classification and echelon advice after consultation. The echelon advice was based on the guidelines of the Dutch Healthcare Authority and the Dutch Institute of Mental Health and Addiction, an overview is shown in Dijkman *et al.* (2013) [49].

### *Participants*

Patients who completed the entire procedure and gave informed consent were included. All patients were 18 years or older. Patients who were unable to perform the diagnostic tests (i.e. patients with insufficient understanding of the Dutch language, a cognitive impairment or severe limited reality testing capabilities) were excluded.

### *Psychologists*

All raters had a Master of Science degree in Psychology and a basic certificate of psychodiagnostics. They were trained and under control of a supervisor, who was a registered healthcare psychologist and psychotherapist.

### *Statistical Analysis*

*Inter-rater reliability.* To investigate the inter-rater reliability between psychologists, 100 patients were randomly selected from the data. Cohen's kappa values ( $\kappa$ -values) were calculated with 95% confidence intervals (CI). Based on Fleiss [50],  $\kappa$ -values lower than 0.4 can be interpreted as poor, between 0.40 and 0.75 as fair to good and above 0.75 as excellent agreement. All patient records were observed by two different psychologists and four psychologists were involved in this evaluation. The psychologist did not know the test outcomes of the opposite assessor. Because inter-rater reliability will be influenced by the illness base rate,  $\kappa$ -values were calculated only when at least five classifications were present for each Axis I – Axis V classification and the echelon advice by both raters. Analysis was done using SPSS 21.0.

*Concurrent validity.* To investigate the concurrent validity between the system and the psychologists' judgment, we calculated sensitivity and specificity scores [51], positive predictive values (PPV), negative predictive values (NPV) and  $\kappa$ -values (95% CI). Acceptable values vary per situation, depending on the situation (tests aim, expected costs and

benefits) and clinical consequences of positive and negative misclassifications [52, 53]. As missed symptoms can affect the quality of life of patients or can even be fatal, the eDiagnostic system was developed to minimise false negatives. Therefore, the amount of false positives was expected to be higher than the number of false negatives.

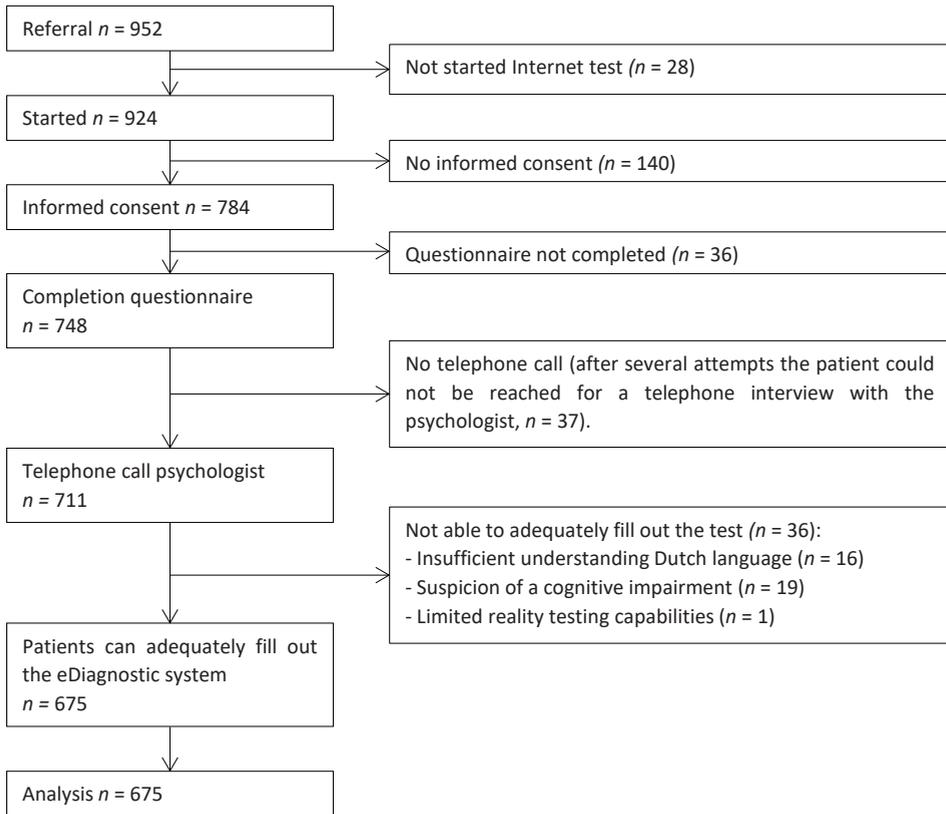
## Results

### *Participants*

A flowchart of the inclusion is shown in Figure 1. A total of 952 patients received an invitation for participation, but 277 patients dropped out or were excluded during the inclusion process (71% of the patients were included in the analysis). In total, 675 patients (276 men and 399 women) from 92 general practices were included. The age of the respondents varied from 18 to 83 years, with a mean age of 38 (SD = 13.10).

### *Psychologists*

Four mental health professionals (one man and three women) classified the patients according to the DSM-IV-TR. The age of the mental health professionals varied from 25 to 29 years, mean age was 27 years (SD = 1.79). The years of experience varied from 0.5 to 3.5 years (M = 2.5 years SD = 1.15).



**Figure 1:** Flowchart of the inclusion procedure of patients.

### *Inter-rater reliability*

The inter-rater reliability between the raters for most Axis I classifications was considered acceptable (see Table 1); most could be interpreted as fair to good, or excellent agreement, except for the depressive disorder NOS ( $\kappa$ -value = 0.35). Agreement between the raters for Axis II classifications was fair to good or excellent, with the exception of the obsessive-compulsive personality disorder ( $\kappa$ -value = 0.39) and the classification for Axis II deferred ( $\kappa$ -value = -0.13). The inter-rater reliability between Axis III to Axis V classifications and the agreement between the raters for the echelon advice was considered fair to good or excellent. The percentage agreement was high for most Axis I – Axis V classifications and the echelon advice, except for the diagnosis or condition on Axis II deferred (0.62). This demonstrates that low kappa values can be obtained despite high percentage agreement.

**Table 1.** Agreement between Axis I to Axis V classifications and the echelon advice of both raters ( $n = 100$ ).

	Cohen's $\kappa$ (95% CI)	Percentage agreement
Axis I classification (clinical disorders)		
Major depression	0.67 (0.52 to 0.82)	0.85
Depressive disorder NOS	0.35 (0.07 to 0.63)	0.88
Posttraumatic stress-disorder	0.62 (0.41 to 0.83)	0.89
Panic disorder	0.70 (0.52 to 0.88)	0.91
Social phobia	0.81 (0.60 to 1.02)	0.97
Alcohol abuse/dependence	0.79 (0.51 to 1.07)	0.98
Cannabis abuse/dependence	0.86 (0.68 to 1.04)	0.98
Attention-deficit/hyperactivity disorder	0.82 (0.58 to 1.06)	0.98
Autistic disorder	0.42 (0.11 to 0.73)	0.91
Axis II classification (personality disorders)		
No personality disorder	0.43 (0.23 to 0.63)	0.79
Avoidant personality disorder	0.41 (0.15 to 0.67)	0.87
Obsessive-compulsive personality disorder	0.39 (0.08 to 0.70)	0.90
Borderline personality disorder	0.78 (0.62 to 0.94)	0.93
Personality disorder NOS	0.46 (0.19 to 0.73)	0.92
Diagnosis or condition on Axis II deferred	-0.13 (-0.25 to -0.01)	0.62
Axis III classification (current somatic disorder) <sup>a</sup>		
No somatic complaints	0.56 (0.39 to 0.73)	0.78
Single somatic complaints (no effect for treatment)	0.51 (0.32 to 0.70)	0.78
Complex somatic complaints (effects treatment)	0.49 (0.22 to 0.76)	0.89
Axis IV classification (psychosocial and environmental problems)		
No Axis IV problems	0.52 (0.25 to 0.79)	0.91
Problems with primary support	0.66 (0.51 to 0.81)	0.84
Problems related to social environment	0.62 (0.47 to 0.77)	0.81
Educational problems	0.75 (0.54 to 0.96)	0.95
Occupational problems	0.65 (0.50 to 0.80)	0.83
Housing problems	1.00 (1.00 to 1.00)	1.00
Economic problems	0.79 (0.65 to 0.93)	0.92
Axis V classification (global assessment of functioning scale)		
41-50 serious impairments	0.75 (0.61 to 0.89)	0.90
51-60 moderate impairments	0.69 (0.54 to 0.84)	0.86
Echelon advice <sup>b</sup>		
Primary mental health care	0.62 (0.46 to 0.78)	0.85
Secondary mental health care	0.67 (0.52 to 0.82)	0.86

$\kappa$ -values were only calculated when at least a number of five classifications were present for each Axis I – Axis V classification and the echelon advice by both raters (i.e. referral to general practice is not shown in the Table because this echelon advice was assessed by the two raters in less than 5 times). <sup>a</sup> There were seven missing values, <sup>b</sup> There was one missing value.

*Concurrent validity*

*Axis I and Axis II classifications.* The sensitivity and specificity scores, the PPVs, NPVs and  $\kappa$ -values (95% CI) for Axis I and Axis II classifications of the eDiagnostic system vs. the judgments of the psychologists are presented in Table 2 and 3. In total, the scores for 25 classifications could be calculated. The most common classifications were mood and anxiety disorders. Moderate to good sensitivities were observed for most Axis I classifications, some sensitivities were low (e.g. Asperger's disorder, adjustment disorder and sexual disorders). The specificity of test scores was considered good for all Axis I classifications (0.75-0.99). PPVs varied considerably (0.06 - 0.82). For a number of classifications, the PPVs were moderate to good. Therefore, when the system detected an abnormality, the probability that there was a classification according to the psychologist was high and the number of false positives was rather low (see Table 2). However, for a number of classifications the PPVs were rather low (see Table 2). NPVs were high for all classifications (scores varied from 0.86 - 1.00), so when the system detected no abnormalities, it was almost certain there was no classification according to the psychologist. Fair to good or excellent agreement was found for 12 Axis I classifications (see: Table 2). Poor agreement was found for the other classifications. A more skewed distribution of the marginals could have contributed to the low scores [53, 54]. Therefore, the raw data provides us more information (see Table 2). For a number of classifications the number of false positives was rather high. The most notable pertained to generalized anxiety disorder and schizophrenia or a psychotic disorder.

In total, 13 Axis II classifications were observed during the study period, the scores for 9 classifications could be calculated. The system was able to determine the absence of a personality disorder (e.g. No Axis II classification: sensitivity: 0.81, specificity: 0.84, PPV: 0.77, NPV 0.87 and  $\kappa$ -value: 0.65). The sensitivities and specificities for most specific personality disorders were considered good (see Table 3). However, PPVs for several specific Axis II classifications were rather low. NPVs were high for all Axis II classifications (0.87-1.00). So when the system detected no abnormalities, it was almost certain there was no classification according to the psychologist. Fair to good agreement was found between the system and the psychologists' judgment in approximately half of the classifications ( $\kappa$ -values: 0.43-0.65). For the remaining classifications the agreement was poor (this could be due to the fact that the distribution of the marginal was highly skewed) [53, 54].

*Axis III - Axis V classifications and the echelon advice.* Diagnostic properties for Axis III (somatic disorders), Axis IV (psychosocial and environmental problems), Axis V (global assessment of functioning scale) classifications, and the echelon advice are shown in Table 4.

**Table 2.** Diagnostic properties of the eDiagnostic system compared with the judgment of the psychologist for Axis I classifications ( $n = 675$ ).

Axis I classifications (clinical syndromes)	Total Int <sup>a</sup>	Total PsA <sup>b</sup>	TP <sup>c</sup>	FP <sup>d</sup>	FN <sup>e</sup>	TN <sup>f</sup>	Sens <sup>g</sup>	Spec <sup>h</sup>	PPV	NPV	Cohen's $\kappa$ (95% CI)
No Axis I classification	50	15	7	43	8	617	0.47	0.93	0.14	0.99	0.19 (0.05-0.33)
Major depression	416	376	341	75	35	224	0.91	0.75	0.82	0.86	0.67 (0.61-0.73)
Dysthymia	118	35	32	86	3	554	0.91	0.87	0.27	0.99	0.37 (0.27-0.47)
Bipolar disorder	35	3	-	-	-	-	-	-	-	-	-
Depressive disorder NOS	77	97	58	19	39	559	0.60	0.97	0.75	0.93	0.62 (0.53-0.71)
Posttraumatic stress-disorder	97	98	69	28	29	549	0.70	0.95	0.71	0.95	0.66 (0.58-0.74)
Panic disorder	115	84	61	54	23	537	0.73	0.91	0.53	0.96	0.55 (0.46-0.64)
Social phobia	73	41	32	41	9	593	0.78	0.94	0.44	0.99	0.52 (0.41-0.63)
Obsessive-compulsive disorder	69	23	15	54	8	598	0.65	0.92	0.22	0.99	0.29 (0.17-0.41)
Specific phobia	30	12	8	22	4	641	0.67	0.97	0.27	0.99	0.37 (0.18-0.56)
Generalized anxiety disorder	166	38	31	135	7	502	0.82	0.79	0.19	0.99	0.23 (0.15-0.31)
Agoraphobia	56	2	-	-	-	-	-	-	-	-	-
Alcohol abuse/dependence	59	30	29	30	1	615	0.97	0.95	0.49	1.00	0.63 (0.51-0.75)
Cannabis abuse/dependence	32	22	21	11	1	642	0.95	0.98	0.66	1.00	0.77 (0.64-0.90)
Cocaine abuse/dependence	11	6	6	5	0	664	1.00	0.99	0.55	1.00	0.70 (0.45-0.95)
Other drug abuse/dependence	21	5	5	16	0	654	1.00	0.98	0.24	1.00	0.38 (0.14-0.62)
Hypochondriasis	18	12	8	10	4	653	0.67	0.98	0.44	0.99	0.52 (0.30-0.74)
Body dysmorphic disorder	40	7	7	33	0	635	1.00	0.95	0.18	1.00	0.29 (0.12-0.46)
Pain disorder	25	26	12	13	14	636	0.46	0.98	0.48	0.98	0.45 (0.28-0.62)
Undifferentiated somatoform disorder	0	18	-	-	-	-	-	-	-	-	-
Autistic disorder	76	7	7	69	0	599	1.00	0.90	0.09	1.00	0.15 (0.05-0.25)
Asperger's disorder	22	21	5	17	16	637	0.24	0.97	0.23	0.98	0.21 (0.04-0.38)
ADHD <sup>i</sup>	39	48	29	10	19	617	0.60	0.98	0.74	0.97	0.64 (0.52-0.76)
Eating disorders	53	27	23	30	4	618	0.85	0.95	0.43	0.99	0.55 (0.42-0.68)

Axis I classifications (clinical syndromes)	Total Int <sup>a</sup>	Total PsA <sup>b</sup>	TP <sup>c</sup>	FP <sup>d</sup>	FN <sup>e</sup>	TN <sup>f</sup>	Sens <sup>g</sup>	Spec <sup>h</sup>	PPV	NPV	Cohen's $\kappa$ (95% CI)
Adjustment disorders	8	25	3	5	22	645	0.12	0.99	0.38	0.97	0.17 (-0.01-0.35)
Sexual disorders	5	10	1	4	9	661	0.10	0.99	0.20	0.99	0.13 (-0.11-0.37)
Sleeping disorders	22	3	-	-	-	-	-	-	-	-	-
Schizophrenia / psychotic disorders	102	8	6	96	2	571	0.75	0.86	0.06	1.00	0.09 (0.02-0.16)
Impulse control disorders	45	5	3	42	2	628	0.60	0.94	0.07	1.00	0.11 (-0.01-0.23)
Diagnosis or condition on Axis I deferred	0	14	-	-	-	-	-	-	-	-	-

<sup>a</sup> Total number of classifications from the eDiagnostic system, <sup>b</sup> Total number of classifications from the judgment of the psychologist after telephone call, <sup>c</sup> true positive, <sup>d</sup> false positive <sup>e</sup> false negative, <sup>f</sup> true negative, <sup>g</sup> sensitivity, <sup>h</sup> specificity, <sup>i</sup> Attention-deficit/hyperactivity disorder. TP, FP, FN, TN, sensitivities, specificities, PPVs, NPVs and  $\kappa$ -values were calculated only when at least five numbers of classifications were present for both groups.

**Table 3.** Diagnostic properties of the eDiagnostic system compared with the judgment of the psychologist for Axis II classifications ( $n = 675$ ).

Axis II classifications (personality disorders)	Total Int <sup>a</sup>	Total PSA <sup>b</sup>	TP <sup>c</sup>	FP <sup>d</sup>	FN <sup>e</sup>	TN <sup>f</sup>	Sens <sup>g</sup>	Spec <sup>h</sup>	PPV	NPV	Cohen's $\kappa$ (95% CI)
No Axis II classification	281	267	217	64	50	344	0.81	0.84	0.77	0.87	0.65 (0.59-0.71)
Avoidant PD	164	88	82	82	6	505	0.93	0.86	0.50	0.99	0.58 (0.50-0.66)
Dependent PD	94	36	35	59	1	580	0.97	0.91	0.37	1.00	0.50 (0.39-0.61)
Obsessive-compulsive PD	167	62	58	109	4	504	0.94	0.82	0.35	0.99	0.43 (0.35-0.51)
Narcissistic PD	53	7	7	46	0	622	1.00	0.93	0.13	1.00	0.22 (0.08-0.36)
Antisocial PD	30	6	6	24	0	645	1.00	0.96	0.20	1.00	0.32 (0.12-0.52)
Borderline PD	170	82	79	91	3	502	0.96	0.85	0.46	0.99	0.55 (0.47-0.63)
Histrionic PD	30	0	-	-	-	-	-	-	-	-	-
Paranoid PD	133	8	8	125	0	542	1.00	0.81	0.06	1.00	0.09 (0.03-0.15)
Schizoid PD	59	0	-	-	-	-	-	-	-	-	-
Schizotypal PD	41	1	-	-	-	-	-	-	-	-	-
PD NOS	235	100	59	176	41	399	0.59	0.69	0.25	0.91	0.18 (0.11-0.25)
Diagnosis or condition on Axis II deferred	0	105	-	-	-	-	-	-	-	-	-

<sup>a</sup> Total number of classifications from the eDiagnostic system, <sup>b</sup> Total number of classifications from the judgment of the psychologist after telephone call, <sup>c</sup> true positive, <sup>d</sup> false positive <sup>e</sup> false negative, <sup>f</sup> true negative, <sup>g</sup> sensitivity, <sup>h</sup> specificity, <sup>i</sup> specificity, TP, FP, FN, TN, sensitivities, specificities, PPVs, NPVs and  $\kappa$  -values were calculated only when at least five numbers of classifications were present for both groups.

**Table 4.** Diagnostic properties of the eDiagnostic system compared with the judgment of the psychologist for the echelon advice and Axis III to Axis V classifications (*n* = 675).

Echelon advice	Total Int <sup>a</sup>	Total PsA <sup>b</sup>	TP <sup>c</sup>	FP <sup>d</sup>	FN <sup>e</sup>	TN <sup>f</sup>	Sens <sup>g</sup>	Spec <sup>h</sup>	PPV	NPV	Cohen's κ (95% CI)
No referral	1	8	-	-	-	-	-	-	-	-	-
General practice (general practitioner/practice nurse)	4	21	-	-	-	-	-	-	-	-	-
Social Work	38	4	-	-	-	-	-	-	-	-	-
Primary mental health care	173	189	161	12	28	474	0.85	0.98	0.93	0.94	0.85 (0.80-0.90)
Secondary mental health care	459	455	433	26	22	194	0.95	0.88	0.94	0.90	0.84 (0.79-0.89)
Axis III (current somatic disorder)											
Somatic disorder	322	334	296	26	38	315	0.89	0.92	0.92	0.89	0.81 (0.76-0.86)
Axis IV (psychosocial and environmental problems)											
No Axis IV problems	84	55	49	35	6	585	0.89	0.94	0.58	0.99	0.67 (0.58-0.76)
Problems primary support	323	371	317	6	54	298	0.85	0.98	0.98	0.85	0.82 (0.78-0.86)
Problems social environment	331	331	318	13	13	331	0.96	0.96	0.96	0.96	0.92 (0.89-0.95)
Occupational problems	306	314	295	11	19	350	0.94	0.97	0.96	0.95	0.91 (0.88-0.94)
Economic problems	242	222	222	20	0	433	1.00	0.96	0.92	1.00	0.93 (0.90-0.96)
Housing problems	72	69	69	3	0	603	1.00	1.00	0.96	1.00	0.98 (0.95-1.01)
Problems legal system/crime	14	15	13	1	2	659	0.87	1.00	0.93	1.00	0.89 (0.77-1.01)
Educational problems	95	92	91	4	1	579	0.99	0.99	0.96	1.00	0.97 (0.94-1.00)
Axis V (global assessment of functioning scale) <sup>i</sup>											
31-40 major impairments	33	4	-	-	-	-	-	-	-	-	-
41-50 serious impairments	92	199	81	11	118	464	0.41	0.98	0.88	0.80	0.46 (0.38-0.53)
51-60 moderate impairments	335	407	254	81	152	187	0.63	0.70	0.76	0.55	0.31 (0.24-0.38)
61-70 mild impairments	127	52	29	98	23	524	0.56	0.84	0.23	0.96	0.24 (0.15-0.33)
71-80 slight impairments	35	8	4	31	4	635	0.50	0.95	0.11	0.99	0.17 (0.02-0.32)
81-90 minimal impairments	43	4	-	-	-	-	-	-	-	-	-
91-100 no impairments	9	1	-	-	-	-	-	-	-	-	-

<sup>a</sup> Total amount of classifications from the eDiagnostic system, <sup>b</sup> Total amount of classifications from the judgment of the psychologist after telephone call, <sup>c</sup> true positive, <sup>d</sup> false positive <sup>e</sup> false negative, <sup>f</sup> true negative, <sup>g</sup> sensitivity, <sup>h</sup> specificity, <sup>i</sup> there was one missing value (*n* = 674). TP, FP, FN, TN, sensitivities, specificities, PPVs, NPVs and κ-values were calculated only when at least five numbers of classifications were present for both groups.

Results showed that the eDiagnostic system was able to correctly estimate the echelon advice. Good sensitivities (0.85-0.95), specificities (0.88-0.98), PPVs (0.93-0.94), NPVs (0.90-0.94), and excellent agreement was found for referral to primary and secondary care ( $\kappa$ -value = 0.85 and 0.84, respectively).

As shown in Table 4, the diagnostic properties for Axis III- and Axis IV classifications were high. Meaning that the system could accurately estimate the somatic complaints and psychosocial symptoms.

Overall, the system was less accurate in identifying the global assessment of functioning (GAF) of patients according to the psychologist. The sensitivities were moderate (ranging from 0.41-0.63). The specificities and NPVs were considered moderate to good. PPVs were high for the estimated 'moderate impairments' (PPV: 0.76) and 'serious impairments' (PPV: 0.88), but low for the remaining GAF-scores (see Table 4). Fair to good agreement was found for 'serious impairments' ( $\kappa$ -value = 0.46), the remaining  $\kappa$ -values showed poor agreement.

## Discussion

### *Summary of the most important findings*

The present study aimed to investigate the concurrent validity of a new eDiagnostic system, recently introduced in the Netherlands. Our study showed that the system could be used to correctly estimate the echelon advice (i.e. referral to primary or secondary care) and correctly identify and exclude most Axis I classifications, such as mood, anxiety and substance-related disorders. The eDiagnostic system was less accurate in identifying patients with Asperger's disorder, sexual disorder and adjustment disorder. For several classifications the sensitivities and specificities were good, but the PPVs were rather low (e.g. generalized anxiety disorder and schizophrenia or a psychotic disorder). The system was able to determine the absence of a personality disorder and the sensitivities and specificities for most specific personality disorders were good. Still, the PPVs for several specific Axis II classifications were rather low. Finally, the system was inaccurate in identifying the GAF of the patient according to the psychologist.

### *Results in relation to literature*

Only few online screening questionnaires have been developed to identify Axis I disorders in primary care [38-42]. These questionnaires were mainly focused on depression or anxiety disorders [39-41]. Overall, the sensitivities, specificities and NPVs of the specific depression and anxiety disorders of the eDiagnostic system were comparable to the findings in other studies [39-41]. The observed PPVs in our study were lower than the ones of Farvolden et al. (2003) [41]; though they were slightly higher than Donker et al. (2010) [39] and Lin et al. (2007) [40]. Of note is that both PPVs and NPVs are influenced by the

prevalence of the disease in the tested population and should therefore be regarded with caution. A lower prevalence will decrease the PPV and increase the NPV. Still, other reasons might further explain the considerable amount of false positives. For example, the system is built to minimise the chance of missed symptoms for a broad range of mental disorders, which a priori leads to the probability of false positives. In addition, the system is not able to interpret the results or deal with diagnostic issues, such as differential diagnosis or overlapping symptoms. Furthermore, literature shows that computer administered instruments reduce social desirability when compared to (face-to-face) interviews, particularly when respondents reveal highly sensitive personal behaviour [55]. This might further explain the disagreement between the eDiagnostic system and the psychologists' judgment.

Few interviewer-administered screening questionnaires exist assessing the general diagnostic criteria of a personality disorder [56-58]. Also a number of self-report questionnaires for screening a personality disorder exist [59-63]. The self-report questionnaires for screening a personality disorder have the tendency to over-diagnose symptoms when compared to a structured clinical interview [59-63]. Our results showed that the eDiagnostic system could accurately distinguish a personality disorder from no personality disorder, but a high number of false positives was found for several specific personality disorders. This is consistent with previous research in which the self-report questionnaires could be used to screen patients for personality pathology on the overall level, but the questionnaire could not be used as a substitute for a clinical interview due to a high number of false positives [59, 60].

Problems have been raised concerning the reliability and validity of GAF ratings by clinicians [64-69]. The reliability of GAF ratings could be good especially in research settings, but was insufficient in routine clinical practice [64-66]. Research showed problems regarding the validity of GAF ratings [65-69]. Prior training and test awareness increased both the reliability and validity of the ratings [64, 65, 70, 71]. In the DSM-5, the GAF ratings have been eliminated, among others due to lack of conceptual clarity and questionable psychometric properties in practice [72]. Taking this into consideration, it is not surprising that the agreement between the patients' GAF ratings and clinicians' GAF ratings was poor. Patients were not trained to rate their own GAF scores, while the psychologists were trained to interpret the scores.

### *Implications for medical practice*

The eDiagnostic system could adequately identify and exclude a number of mood, anxiety, substance-related disorders, determine the absence of personality pathology and estimate the echelon advice, without the intervention of a psychologist. Therefore, the system appears to be a valid source of information for the general practitioner and mental health professional in the assessment and classification of mental disorders in primary care patients with a suspicion of a mental disorder. The system seems therefore more

comprehensive than a screening system, but it cannot, at this point of time, replace clinical assessment. The eDiagnostic system could be used in several ways. First, it could be used including a consultation of a psychologist as it is used in this study. Second, we believe that it would be possible for the general practitioner to interpret and check the classifications generated by the system, without the judgment of a psychologist. According to the Dutch College of General Practitioners' (NHG) practice guidelines, it is the role of the general practitioner to assess whether there is anxiety or an anxiety disorder, and general practitioners should try to name the specific anxiety disorder [29]. Also it is the role of the general practitioner to assess hypochondriasis, depressive symptoms, a major depression, dysthymia, problematic alcohol use and sleeping problems [28, 29, 73, 74]. Results showed that the system could assist general practitioners in the assessment and classification of these problems. In using our research findings, an abridged version of the eDiagnostic system, called the QuickScreen, has been compiled for general practitioners to support them in the assessment of psychological problems mentioned in the guidelines. A third option would be to use the system after patients are referred to mental healthcare institutions, before the start of a healthcare trajectory (e.g. when patients are on the waiting list). This allows the caregivers to quickly decide if and what kind of healthcare trajectory is necessary. Finally, we believe that the system could be valuable during the intake consultation with a psychologist or psychiatrist after patients are referred to a mental healthcare institution. If patients were to complete the questions in advance, this could save time for the healthcare professional as they could receive and analyse patient information before consultation. It is not recommended to directly send the outcomes to the patients without the intervention of a healthcare professional. All caregivers should be aware that the system could generate false positives and false negatives and that test scores remain dependent on the introspective ability of the patient (for instance in patients with autism spectrum disorders or psychotic disorders) [8, 9]. The level of education, diagnostic category and level of suffering among patients may also have a negative effect on the validity of the test results. For now, the system showed to be a valuable assistant to quickly classify mental disorders. It is possible to identify and exclude a number of classifications, however the system cannot totally replace clinical judgment from a professional for all classifications. After revising some questions, future research should determine whether the computer would be able to replace a healthcare professional, if desired.

## Limitations

Our results should be considered against the background of some limitations. First, the judgment of the psychologist was partly based on the classifications generated by the eDiagnostic system (DS 1 – DS 5) and results could be prone to incorporation bias [75]. This may have positively influenced the results, as the psychologists may have tended to

seek and overvalue data consistent with the test and disregard data inconsistent with the results. Second, as discussed earlier, results may be influenced by the patients' introspective ability which in turn may be influenced by patient characteristics, such as diagnostic category and level of suffering. This may lead to poor agreement between the eDiagnostic system and psychologists for patients with certain classifications (e.g. Asperger's disorder). Also, it is known that people reveal themselves faster via the Internet, due to reduced social desirability (especially in highly sensitive personal behavior) [55]. It is possible that the eDiagnostic system identified those psychological troubles better than the psychologists did, due to social desirable answers. Third, the study was partly financed by the developers of the eDiagnostic system. However, they played no role in the study design, data collection, analysis, interpretation of data, in the writing of the report and in the decision to submit the article for publication. Fourth, the study was conducted in a real patient setting (patients of 92 practices), which entails certain limitations. The most obvious way to investigate the eDiagnostic system would be to blindly compare the system to an independently administered structured interview (i.e. comparison to a SCID-I and SCID-II). However, doing interviews with patients would be logistically very complex (it takes about 120-180 minutes to administer a SCID-I and SCID-II). This would be a big burden for the patient and it would delay the referral of the patients, which would have been ethically challenging, especially in this patient group. Therefore, we decided to use the routinely collected data as efficiently as possible. One may wonder whether the judgment of the psychologist after telephone call was a valid criterion and method of measurement. The high inter-rater agreement between the psychologists pleads for the psychologists in this study. When the system would be compared to a structured clinical interview, we expect that our overall conclusions would be maintained. However, since there would be more time to critically clarify all classifications using a structured clinical interview, and also taking into account the possible incorporation bias, we expect more false positive results in such an evaluation. We expect that the false negative results would remain the same, since the system is built to decrease the chance of missed symptoms. Future research should be done with an independently administered structured clinical interview.

## Conclusions, clinical implications and future research

This study breaks ground in the management of mental healthcare, using eDiagnostics *prior* to a referral to a mental health institution instead of afterwards. To our knowledge, this is the first study to validate an eDiagnostic system for most mental disorders in primary care. Generally, the eDiagnostic system can be seen as a valid instrument for most classifications for primary care patients. The system cannot, at this point of time, replace clinical assessment. For several classifications the system generates too many false positives and therefore a psychologist is currently used to reduce the amount of false

positives. The system could assist general practitioners and practice nurses in the assessment and classification of mental disorders. In addition, it can be useful for already referred patients (e.g. when patients are on the waiting list of a healthcare institution or during an intake procedure). Future research should include comparison to an independently administered structured clinical interview.

## References

1. Demyttenaere K, Bruffaerts R, Posada-Villa J *et al.* Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA* 2004; 291: 2581-90.
2. Kessler RC, Aguilar-Gaxiola S, Alonso J *et al.* The global burden of mental disorders: an update from the WHO World Mental Health (WMH) Surveys. *Epidemiol Psychiatr Soc* 2009; 18: 23-33.
3. Smit F, Cuijpers P, Oostenbrink J *et al.* Costs of nine common mental disorders: implications for curative and preventive psychiatry. *J Mental Health Policy Econ* 2006; 9: 193-200.
4. Alonso J, Angermeyer MC, Bernert S *et al.* Disability and quality of life impact of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatr Scand* 2004; 420: 38-46.
5. Wittchen HU, Jacobi F. Size and burden of mental disorders in Europe—a critical review and appraisal of 27 studies. *Eur Neuropsychopharmacol* 2005; 15: 357-76.
6. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. *PLoS One* 2015; 10: e0116820.
7. Verhaak PF, Schellevis FG, Nuijen J, Volkers AC. Patients with a psychiatric disorder in general practice: determinants of general practitioners' psychological diagnosis. *Gen Hosp Psychiatry* 2006; 28: 125-32.
8. Kessler D, Lloyd K, Lewis G, Gray DP. Cross sectional study of symptom attribution and recognition of depression and anxiety in primary care. *BMJ* 1999; 318: 436-39.
9. Jorm AF. Mental health literacy. Public knowledge and beliefs about mental disorders. *Br J Psychiatry* 2000; 177: 396-401.
10. Lecrubier Y. Widespread underrecognition and undertreatment of anxiety and mood disorders: results from 3 European studies. *J Clin Psychiatry* 2007; 68: 36-41.
11. Carey M, Jones K, Meadows G *et al.* Accuracy of general practitioner unassisted detection of depression. *Aust N Z J Psychiatry* 2014; 48: 571-78.
12. Mitchell A, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009; 374: 609-19.
13. Janssen EHC, van de Ven PM, Terluin B *et al.* Recognition of anxiety disorders by family physicians after rigorous medical record case extraction: Results of the Netherlands Study of Depression and Anxiety. *Gen Hosp Psychiatry* 2012; 34: 460-7.
14. Joling KJ, van Marwijk HWJ, Piek E *et al.* Do GPs' medical records demonstrate a good recognition of depression? A new perspective on case extraction. *J Affect Disord* 2011; 133: 522-7.
15. Harvey AG, Gumpert NB. Evidence-based psychological treatments for mental disorders: Modifiable barriers to access and possible solutions. *Behav Res Ther* 2015; 68: 1-12.
16. Wang PS, Aguilar-Gaxiola S, Alonso J *et al.* Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. *Lancet* 2007; 370: 841-50.
17. Fernandez A, Haro JM, Martinez-Alonso M *et al.* Treatment adequacy for anxiety and depressive disorders in six European countries. *Br J Psychiatry* 2007; 190: 172-3.
18. Bijl RV, Ravelli A. Psychiatric morbidity, service use, and need for care in the general population: results of The Netherlands Mental Health Survey and Incidence Study. *Am J Public Health* 2000; 90: 602-7.
19. Alonso J, Codony M, Kovess V *et al.* Population level of unmet need for mental healthcare in Europe. *Br J Psychiatry* 2007; 190: 299-306.
20. Rost K, Zhang M, Fortney J, Smith J, Coyne J, Smith GR Jr. Persistently poor outcomes of undetected major depression in primary care. *Gen Hosp Psychiatry* 1998; 20: 12-20.
21. Essau CA. Course and outcome of somatoform disorders in non-referred adolescents. *Psychosomatics* 2007; 48: 502-9.
22. Essau CA. Course and outcome of major depressive disorder in non-referred adolescents. *J Affect Disord* 2007; 99: 191-201.

23. Kessler RC, Merikangas KR, Berglund P, Eaton WW, Koretz DS, Walters EE. Mild disorders should not be eliminated from the DSM-V. *Arch Gen Psychiatry* 2003; 60: 1117-22.
24. Druss BG, Wang PS, Sampson NA *et al.* Understanding mental health treatment in persons without mental diagnoses: results from the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2007; 64: 1196-203.
25. Kessler RC, Demler O, Frank RG *et al.* Prevalence and treatment of mental disorders, 1990 to 2003. *N Engl J Med* 2005; 352: 2515-23.
26. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005; 62: 629-40.
27. Bruffaerts R, Posada-Villa J, Al-Hamzawi AO *et al.* Proportion of patients without mental disorders being treated in mental health services worldwide. *Br J Psychiatry* 2015; 206: 101-9.
28. Van Weel-Baumgarten EM, Van Gelderen MG, Grundmeijer HGLM *et al.* NHG Standaard Depressie (tweede herziening) [NHG guideline Depression second revision]. *Huisarts Wet* 2012; 55: 252-9.
29. Hassink-Franke L, Terluin B, Van Heest FB, Hekman J, Van Marwijk H, van Avendonk M. NHG-Standaard Angst (tweede herziening) [NHG guideline Anxiety second revision]. *Huisarts Wet* 2012; 55: 68-77.
30. Roche AM, Freeman T. Brief interventions: good in theory but weak in practice. *Drug Alcohol Rev* 2004; 23: 11-8.
31. Sinnema H, Franx G, Spijker J *et al.* Delivering stepped care for depression in general practice: results of a survey amongst general practitioners in the Netherlands. *Eur J Gen Pract* 2013; 19: 221-9.
32. Fleury MJ, Imboua A, Aubé D, Farand L, Lambert Y. General practitioners' management of mental disorders: a rewarding practice with considerable obstacles. *BMC Fam Pract* 2012; 13: 1-12.
33. Shedler J. *The Shedler QPD Panel (Quick PsychoDiagnostics Panel): a psychiatric "lab test" for primary care.* In: Maruish M, editor. *Handbook of psychological assessment in primary care settings.* Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
34. Eysenbach G. What is e-health? *J Med Internet Res* 2001; 3: e20.
35. Riper H, Andersson G, Christensen H *et al.* Theme issue on e-mental health: a growing field in internet research. *J Med Internet Res* 2010; 12: e74.
36. Buchanan T. Online assessment: desirable or dangerous? *Prof Psychol Res Pr* 2002; 33: 148-54.
37. Emmelkamp PM. Technological innovations in clinical assessment and psychotherapy. *Psychother and Psychosom* 2005; 74: 336-43.
38. Donker T, van Straten A, Marks I, Cuijpers P. A brief web-based screening questionnaire for common mental disorders: development and validation. *J Med Internet Res* 2009; 11: e19.
39. Donker T, Comijs H, Cuijpers P *et al.* The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Res* 2010; 176: 45-50.
40. Lin C, Bai Y, Liu C *et al.* Web-based tools can be used reliably to detect patients with major depressive disorder and subsyndromal depressive symptoms. *BMC Psychiatry* 2007; 7: 12.
41. Farvolden P, McBride C, Bagby RM, Ravitz P. A Web-based screening instrument for depression and anxiety disorders in primary care. *J Med Internet Res* 2003; 5: e23.
42. Spitzer RL, Williams JB, Kroenke K *et al.* Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *JAMA* 1994; 272: 1749-56.
43. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders 4th edition, Text Revision (DSM-IV-TR).* Washington, DC: American Psychiatric Association, 2000.
44. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV-TR Axis I disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/ PSY SCREEN).* New York: Biometrics Research, New York State Psychiatric Institute, 2002.
45. Sheehan DV, Lecrubier Y, Sheehan KH *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998; 59 Suppl 20: 22-33; quiz 34-57.
46. First MB, Gibbon M, Spitzer RL, Williams JBW, Benjamin LS. *Structured Clinical Interview for DSM-IV Axis II Personality disorders (SCID-II).* Washington, DC: American Psychiatric Press, 1997.

47. van der Does A. *BDI-II-NL. Handleiding. De Nederlandse versie van de Beck Depression Inventory-2nd edition*. Lisse: Harcourt Test Publishers, 2002.
48. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addiction* 1993; 88: 791-804.
49. Dijkman I, Dinant GJ, Spigt M. eDiagnostics: a promising step towards primary mental health care. *Fam Pract* 2013; 30: 695-704.
50. Fleiss JL. *Statistical Methods for Rates and Proportions. 2nd edition*. New York: Wiley, 1981.
51. Mokkink LB, Terwee CB, Patrick DL *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19: 539-49.
52. Smits N, Smit F, Cuijpers P, De Graaf R. Using decision theory to derive optimal cut-off scores of screening instruments: an illustration explicating costs and benefits of mental health screening. *Int J Methods Psychiatr Res* 2007; 16: 219-29.
53. De Vet HCW, Terwee CB, Mokkink LB, Knol D. *Measurement in Medicine: a Practical Guide*. Cambridge: Cambridge University Press, 2011.
54. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; 43: 543-9.
55. Richman WL, Kiesler S, Weisband S, Drasgow F. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *J Appl Psychol* 1999, 84: 754-775.
56. Langbehn DR, Pfohl BM, Reynolds S *et al*. The Iowa Personality Disorder Screen: development and preliminary validation of a brief screening interview. *J Person Disord* 1999; 13: 75-89.
57. Van Horn E, Manley C, Leddy D, Cicchetti D, Tyrer P. Problems in developing an instrument for the rapid assessment of personality status. *Eur Psychiatry* 2000; 15: 29-33.
58. Moran P, Leese M, Lee T *et al*. Standardised Assessment of Personality - Abbreviated Scale (SAPAS): preliminary validation of a brief screen for personality disorder. *Br J Psychiatry* 2003; 183: 228-32.
59. Ottosson H, Bodlund O, Ekselius L *et al*. DSM-IV and ICD-10 personality disorders: a comparison of a self-report questionnaire (DIP-Q) with a structured interview. *Eur Psychiatry* 1998; 13: 246-53.
60. Hyler SE, Skodol AE, Oldham JM, Kellman HD, Doidge N. Validity of the Personality Diagnostic Questionnaire-Revised: a replication in an outpatient sample. *Compr Psychiatry* 1992; 33: 73-7.
61. Ekselius L, Lindstrom E, von Knorring L, Bodlund O, Kullgren G. SCID II interviews and the SCID Screen questionnaire as diagnostic tools for personality disorders in DSM-III-R. *Acta Psychiatr Scand* 1994; 90: 120-3.
62. Hyler SE, Skodol AE, Kellman HD, Oldham JM, Rosnick L. Validity of the Personality Diagnostic Questionnaire--revised: comparison with two structured interviews. *Am J Psychiatry* 1990; 147: 1043-8.
63. Lenzenweger MF, Loranger AW, Korfine L, Neff C. Detecting personality disorders in a nonclinical population. Application of a 2-stage procedure for case identification. *Arch Gen Psychiatry* 1997; 54: 345-51.
64. Vatnaland T, Vatnaland J, Friis S, Opjordsmoen S. Are GAF scores reliable in routine clinical use? *Acta Psychiatr Scand* 2007; 115: 326-30.
65. Hilsenroth MJ, Ackerman SJ, Blagys MD *et al*. Reliability and validity of DSM-IV axis V. *Am J Psychiatry* 2000; 157: 1858-63.
66. Goldman HH, Skodol AE, Lave TR. Revising axis V for DSM-IV: a review of measures of social functioning. *Am J Psychiatry* 1992; 149: 1148-56.
67. Moos RH, McCoy L, Moos BS. Global assessment of functioning (GAF) ratings: determinants and role as predictors of one-year treatment outcomes. *J Clin Psychol* 2000; 56: 449-61.
68. Hay P, Katsikitis M, Begg J, Da Costa J, Blumenfeld N. A two-year follow-up study and prospective evaluation of the DSM-IV axis V. *Psychiatr Serv* 2003; 54: 1028-30.
69. Smith GN, Ehmann TS, Flynn SW *et al*. The assessment of symptom severity and functional impairment with DSM-IV axis V. *Psychiatr Serv* 2011; 62: 411-17.
70. Bates LW, Lyons JA, Shaw JB. Effects of brief training on application of the Global Assessment of Functioning Scale. *Psychol Rep* 2002; 91: 999- 1006.

71. Dworkin RJ, Friedman LC, Telschow RL, Grant KD, Moffic HS, Sloan VJ. The longitudinal use of the Global Assessment Scale in multiple-rater situations. *Community Ment Health J* 1990; 26: 335-44.
72. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (5th ed)*. Arlington, VA: American Psychiatric Publishing, 2013
73. Boomsma LJ, Drost IM, Larsen IM *et al*. NHG-Standaard Problematisch alcoholgebruik (derde herziening) [NHG guideline problematic alcohol consumption third revision]. *Huisarts Wet* 2014; 57: 638-46.
74. NHG-werkgroep Slaapproblemen en slaapmiddelen. NHG-Standaard Slaapproblemen en slaapmiddelen (tweede herziening) [NHG guideline Sleep problems and sleeping pills]. *Huisarts Wet* 2014; 57: 352-61.
75. Worster A, Carpenter C. Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias. *CJEM* 2008; 10: 174-5.
76. Draper H, Wilson S. Research ethics approval: comprehensive mechanisms are essential but not available. *Fam Pract* 2007; 24: 527-8
77. Evers A, Van Vliet-Mulder J, Groot CJ. *Documentatie van tests en testresearch in Nederland*. Assen: Van Gorcum, 2000.
78. Willems LFM, Tuender-de Haan HA, Defares PB. Een schaal om sociale angst te meten. *Ned Tijdschr Psychol* 1973, 28: 415-22.
79. Kraaimaat FW, van Dam-Baggen CMJ. Ontwikkeling van een zelfbeoordelingsvragenlijst voor obsessief-compulsief gedrag. *Ned Tijdschr Psychol* 1976; 31: 201-11.
80. Fairburn CG, Beglin SJ. Assessment of eating disorders: interview or self-report questionnaire? *Int J Eat Disord* 1994; 16: 363-70.
81. van Furth E. *Nederlandse vertaling van de EDE-Q*. Leidschendam: Robert-Fleury Stichting, 2000.
82. Berg KC, Peterson CB, Frazier P, Crow SJ: Psychometric evaluation of the eating disorder examination and eating disorder examination-questionnaire: a systematic review of the literature. *Int J Eat Disord* 2012; 45: 428-38.
83. Kloens GJ, Barelds DPH, Luteijn F, Schaap CPDR. *Psychometrische eigenschappen van enige vragenlijsten in de eerste lijn (proefschrift)*. <http://www.innovatieduo.nl/nieuws-en-publicaties.aspx>. Archived at: <http://www.webcitation.org/query?id=1431170761366535>, 2005.
84. Chambless DL, Caputo GC, Bright P, Gallagher R. Assessment of fear of fear in agoraphobics: the body sensations questionnaire and the agoraphobic cognitions questionnaire. *J Consult Clin Psychol* 1984; 52: 1090-7.
85. Wiegel M, Meston C, Rosen R. The female sexual function index (FSFI): cross-validation and development of clinical cutoff scores. *J Sex Marital Ther* 2005; 31: 1-20.
86. ter Kuile M, Brauer M, Laan E. The Female Sexual Function Index (FSFI) and the Female Sexual Distress Scale (FSDS): psychometric properties within a Dutch population. *J Sex Marital Ther* 2006; 32: 289-304.
87. Rosen R, Brown C, Heiman J *et al*. The Female Sexual Function Index (FSFI): a multidimensional self-report instrument for the assessment of female sexual function. *J Sex Marital Ther* 2000; 26: 191-208.
88. Rosen RC, Riley A, Wagner G *et al*. The International Index of Erectile Function (IIEF): a multidimensional scale for assessment of erectile dysfunction. *Urology* 1997; 49: 822-30.
89. Arrindell WA, Ettema JHM. *SCL-90: handleiding bij een multidimensionele psychopathologie indicator*. Lisse: Swets & Zeitlinger, 1986.

*Supplementary Table S1. Validation information additional questionnaires.*

Name test	Information questionnaire
Beck Depression Inventory - second edition Dutch version (BDI-II-NL) [47, 77]	The Dutch version of the BDI-II-NL consists of 21 items, scored on a range from 0 to 3 (the maximum score = 63) [47]. The BDI-II is a reliable and valid instrument to determine the severity of a depression for adults and adolescents of 13 years and older [47, 77].
Sociale Angst Schaal [78] (Social Anxiety Scale)	The Social Anxiety Scale contains 24 items, rated on a scale from 0-4 (maximum score = 96). Support was found for the reliability and validity for the Social Anxiety Scale [77, 78].
Alcohol Use Disorder Identification Test (AUDIT) [48]	The AUDIT identifies persons with hazardous alcohol consumption and dependence, it has been developed for patients in primary care. The screening questionnaire consists of 10 items that can be scored on a scale from 0-4 (maximum score = 40). High sensitivity and specificity scores were found for the detection of hazardous and harmful alcohol use among patients [48]. In the present study, the Dutch version of the AUDIT was used.
Inventarisatielijst Dagelijkse Bezigheden (IDB) [79] (Self-rating questionnaire for obsessive-compulsive behaviour).	The IDB has been developed to identify obsessive and compulsive behavior and it consists of 32 items, scored on a scale from 1 to 5 (1 = 'never'; 5 = 'very often') [79]. Minimum score = 32; maximum score = 160. The IDB is a reliable and valid instrument to identify obsessive and compulsive behavior [77].
Eating Disorder Examination-Questionnaire [80, 81]	The EDE-Q is a self-report questionnaire, assessing eating disorder symptoms. Items can be scored on a seven-point rating scale (score 0 = '0 days', 6 = '28 days'). The EDE-Q has four subscales: Restraint, Eating Concern, Shape Concern, and Weight Concern. In addition it consists of seven items that measure the frequency of key behaviours [80, 81]. Support was found for the reliability and validity of the EDE-Q for assessing eating disorder symptoms [82].
Psy Questionnaire Netherlands-Personality (PQN-P) and Psy Questionnaire Netherlands-Coping (PQN-C) [83]	The PQN assesses personality traits and coping strategies. The questionnaire is based on a doctoral thesis in 2005 and follow-up studies from 2005 to 2012 by drs. Gert Jan Kloens [83].
Agoraphobic cognitions questionnaire (ACQ) [84]	The ACQ measures thoughts one might have concerning negative consequences of experiencing anxiety. It consists of 14 questions on a five-point rating scale (1 = 'thought never occurs' to 5 = 'thought always occurs'). Support was found for the reliability and validity of the ACQ for assessing fear in agoraphobics [84].
Female Sexual Function Index (FSFI) [85-87]	The FSFI is a validated 19-item self-report questionnaire, rated on a six-point scale, concerning evaluation of female sexual function over the past 4 weeks. The questionnaire assesses six domains: sexual desire, arousal, lubrication, orgasm, satisfaction and pain [85-87].
International Index of Erectile Function (IIEF) [88]	The IIEF is a valid self-report questionnaire (consisting of 15 items) assessing different domains of male sexual function (erectile function, orgasmic function, sexual desire, intercourse satisfaction and overall satisfaction) [88].
Symptom Check List-90-Revised (SCL-90-R) [77, 89]	The SCL-90-R is a reliable and valid multidimensional psychopathology questionnaire that determines the extent to which a person suffered from various mental and/or physical symptoms over the past week. It consists of 90 items, on a five-point scale (i.e. 1 = 'not at all distressed' to 5 'extremely distressed'). It has nine primary symptom dimensions and three summary scores (i.e. global scores) [77, 89].





# Chapter 3B

## eDiagnostics for mental disorders

Iles Dijkman  
Geert Jan Dinant  
Mark Spigt

*Huisarts Wet 2017;60(9),432-435.*

This article is based on the following publication: Dijkman, I., Dinant, G. J., & Spigt, M. (2016). The concurrent validity of a new eDiagnostic system for mental disorders in primary care. *Fam Pract*, 33(6), 607-616. Publication in Huisarts and Wetenschap was done with permission of the publisher. Data has been updated (see Methods). To prevent repetition from the previous chapter (3a), we only described the updated results in this chapter.

## Abstract

**Background.** An eDiagnostic system was implemented to classify mental disorders, to support general practitioners.

**Method.** We examined the validity of the eDiagnostic system by comparing routinely collected results of 496 patients from Limburg GP practices with the judgment of a psychologist. We analyzed the concurrent validity of the eDiagnostic system via sensitivity, specificity, positive and negative predictive value and Cohen's kappa.

**Results.** The system could correctly recognize, exclude and predict various clinical disorders. It was able to adequately determine the absence of an Axis I disorder (sensitivity and specificity 75%) and it gave a correct advice for the echelon (sensitivity 85 to 92%, specificity 96 to 99%). However, for a number of clinical disorders, the system generated many false positives and false negatives and it was somewhat inaccurate in assessing the general assessment of functioning (Axis V of the DSM-IV-TR).

**Conclusion.** The eDiagnostic system is a valid screening instrument in general practice for most DSM-IV-TR classifications. It could assist GPs in the diagnostics and classification of mental disorders, but it cannot replace a clinical judgment because of the chance of a false positive or a false negative result. A version based on DSM-5 is currently being validated.

### **What is known?**

- Valid online questionnaires are available, but most have limited usefulness for the GP.
- An eDiagnostic system has been developed to classify mental disorders based on the DSM-IV-TR.

### **What's new?**

- The eDiagnostic system is a valid instrument to assess most DSM-IV-TR classifications in general practice.
- The eDiagnostic system can support GPs and practice nurses (PNs) in the diagnostics and classification of mental disorders and provide an adequate echelon advice.
- Using eDiagnostics before instead of after a referral is new.

### **Concurrent validity**

Concurrent validity is the level of agreement between a certain test and a reference test, possibly a gold standard. We compared the results of an eDiagnostic system with the judgment of a psychologist.

## Introduction

Recently, TelePsy ([www.telepsy.nl](http://www.telepsy.nl)) has developed an eDiagnostic system to classify mental disorders based on the DSM-IV-TR. A version based on the DSM-5 is currently being validated. The system, called TeleScreen, can be used by GPs and PNs as a source of information in the classification of mental disorders. The system is adaptive. If a person does not meet the required number of symptoms of a mental disorder, no additional questions are assessed. It is more comprehensive than a screening system, but less extensive than a complete diagnostic system, since it provides provisional classifications instead of established diagnoses. Patients answer online questions and are contacted by a healthcare professional to check the provisional classifications of the system. The system is used as follows: GPs or practice nurses (PNs) could interpret results or an (external) psychologist could check for the generated classifications.

Access to the data in the system and the additional judgment of the external psychologist made it possible to assess the validity of the system by comparing classifications of the system with classifications of the psychologist. The research question was: what is the concurrent validity of the eDiagnostic system when compared with the judgment of the psychologist for patients in primary care?

## Methods

To assess the validity of the system, a prospective study was conducted among patients in Limburg from June until November 2016. This paper contains new results when compared with earlier results (1).

In the opinion of the medical review committee of Maastricht University, this study did not involve a study with persons in the sense of the Medical Research Involving Subjects Act (WMO).

## Procedure

The procedure and the eDiagnostic system is broadly described in our original article (1). The GP and PN decided which patients they would refer to the eDiagnostic system. Frequently asked questions by the caregivers were whether there was a suspicion of a mood or anxiety disorder, a personality or developmental disorder and questions regarding the echelon (e.g. general practice, primary care psychologist or specialized mental healthcare). After referral, patients received an e-mail with a secured electronic link to the system. Subsequently, patients answered online questions to classify Axis I to Axis V classifications. When a patient fulfilled the required number of symptoms of a DSM-IV-

TR disorder, an indication was automatically generated by the system on every five Axes including an echelon advice.

After the patient completed the online questionnaire, a psychologist reviewed the answers and made a working hypothesis. Psychologists could form their diagnostic judgment on classifications generated by the system, supplemented with information from the system about, for example, request for help, treatment history, suicidal thoughts or the severity of a possible disorder as assessed by other validated questionnaires.

Finally, psychologists contacted all included patients by telephone, to evaluate complaints, ask their own questions, critically check generated classifications of the system and if necessary, clarify ambiguous test results. After the consultation, psychologists sent a report with their own preliminary DSM-IV-TR classification and echelon advice to the general practitioners' information system, based on the guidelines (2).

### *In- and exclusion criteria*

Patients older than 18 years, who completed the entire procedure and gave informed consent were included. Excluded were patients whose complaints were too complex to to perform adequate diagnostic investigation and where classification was not possible, patients with insufficient understanding of the Dutch language, patients with a suspicion of a neurological disease and patients with limited intellectual capacities based on the judgment of the psychologist.

### *Assessors*

Four psychologists (one woman and three men, age 29-30 years) classified patients according to the DSM-IV-TR. Their working experience varied between six months till five years. All psychologists had a Master of Science degree in Psychology or Health sciences and had a basic certificate of psychodiagnostics. They were under control of a supervisor, who was a registered health care psychologist and psychotherapist.

## **Statistical analysis**

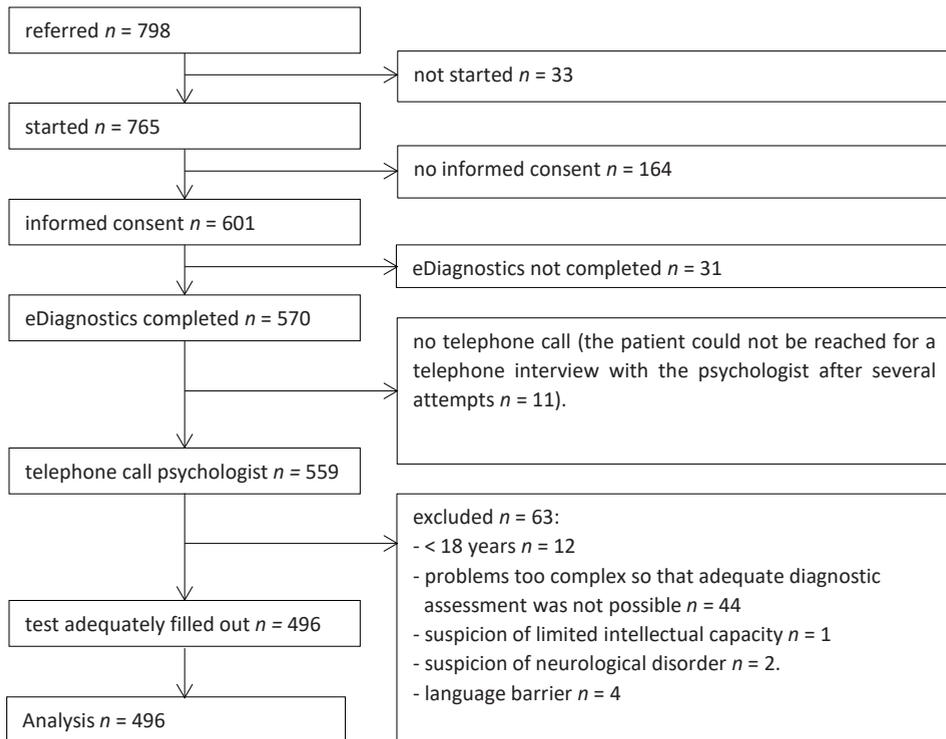
### *Concurrent validity*

The concurrent validity was examined using sensitivity, specificity (3), positive predictive values (PPVs), negative predictive values (NPVs) and kappa values. Acceptable values vary per situation, depending on the situation, such as the aim of the test and clinical consequences of positive and negative misclassifications (4,5).

## Results

### Participants

A total of 798 patients were referred to the eDiagnostic system, 601 patients gave permission for participation in the research. 559 patients completed the entire diagnostic process. In total, 496 patients were included and 63 patients were excluded (see figure 1). The average age of participants was 38 years (SD 13; ranging from 18-72).



**Figure 1.** Inclusion procedure

### Validity

Table 1 shows an overview of the validity of the examined eDiagnostic system. When looking at the main categories, moderate to good scores were found for most clinical and personality disorders, except for adjustment disorders. The agreement on specific disorders was moderate to good too, but for some disorders the number of false positives and false negatives was high. For example this concerned 'undifferentiated somatoform disorder', depressive disorder not otherwise specified' and 'social phobia'.

The validity of the eDiagnostic system for somatic disorders, psychosocial problems and environmental problems and the echelon advice was high.

The system was less accurate in identifying the general assessment of functioning (GAF). Patients were for instance less accurate in identifying serious impairments, according to the psychologists

**Table 1** Diagnostic properties of an eDiagnostic system for psychological complaints in primary care and in agreement with the judgment of a psychologist ( $n = 496$ )

Classification	Total Sys	Total Psy	TP	FP	FN	TN	Sens	Spec	PPV	NPV	kappa (95% CI)
Axis I (clinical disorders)											
Mood disorders	316	201	191	125	10	170	95%	58%	60%	94%	0.48 (0.45 – 0.52)
Major depression	225	165	145	80	20	251	88%	76%	64%	88%	0.58 (0.55 – 0.62)
Depressive disorder NOS	69	29	13	56	16	411	45%	88%	19%	96%	0.20 (0.14 – 0.26)
Dysthymia	10	5	0	10	5	481	-	-	-	-	-
Bipolar disorder	12	1	1	11	0	484	-	-	-	-	-
Anxiety disorders	186	151	115	71	36	274	76%	79%	62%	88%	0.52 (0.48 – 0.56)
Panic disorder	84	34	27	57	7	405	79%	88%	32%	98%	0.40 (0.34 – 0.46)
Posttraumatic stress-disorder	69	49	43	26	6	421	88%	94%	62%	99%	0.69 (0.64 – 0.74)
Generalized anxiety disorder	11	21	4	7	17	468	-	-	-	-	-
Obsessive-compulsive disorder	24	9	7	17	2	470	-	-	-	-	-
Social phobia	36	36	16	20	20	440	44%	96%	44%	96%	0.40 (0.32 – 0.48)
Agoraphobia	24	2	2	22	0	472	-	-	-	-	-
Specific phobia	32	7	7	25	0	464	-	-	-	-	-
Somatoform disorders	149	85	53	96	32	315	62%	77%	36%	91%	0.30 (0.25 – 0.35)
Hypochondriasis	12	11	8	4	3	481	-	-	-	-	-
Pain disorder	36	17	12	24	5	455	-	-	-	-	-
Undifferentiated somatoform disorder	79	41	21	58	20	397	51%	87%	27%	95%	0.27 (0.21 – 0.33)
Somatoform disorder NOS	13	17	1	12	16	467	-	-	-	-	-
Adjustment disorders	27	66	10	17	56	413	15%	96%	37%	88%	0.15 (0.09 – 0.21)
Eating disorders	24	11	9	15	2	470	-	-	-	-	-
Substance use disorders	60	41	35	25	6	430	85%	95%	58%	99%	0.66 (0.60 – 0.72)
Alcohol abuse/dependence	31	17	17	14	0	465	-	-	-	-	-
Cannabis abuse/dependence	25	23	16	9	7	464	-	-	-	-	-
Sexual disorders	60	18	15	45	3	433	-	-	-	-	-
Sleeping disorders	75	17	14	61	3	418	-	-	-	-	-
Pervasive developmental disorders	70	14	2	68	12	414	-	-	-	-	-
Attention deficit disorders	95	85	79	16	6	395	93%	96%	83%	99%	0.85 (0.82 – 0.88)
ADHD	38	47	33	5	14	444	70%	99%	87%	97%	0.76 (0.70 – 0.81)
ADD	34	34	26	8	8	454	76%	98%	76%	98%	0.75 (0.69 – 0.81)
No Axis I classification	43	50	20	23	30	423	40%	95%	47%	93%	0.37 (0.30 – 0.44)
Axis II (personality disorders)											
No Axis II classification	259	267	201	58	66	171	75%	75%	78%	72%	0.50 (0.46 – 0.54)
Axis III (somatic disorder)											
Somatic disorder	204	204	203	1	1	291	100%	100%	100%	100%	0.99 (0.99 – 1.00)

Classification	Total Sys	Total Psy	TP	FP	FN	TN	Sens	Spec	PPV	NPV	kappa (95% CI)
Axis IV (psychosocial and environmental problems)											
Problems with primary support	209	209	209	0	0	287	100%	100%	100%	100%	1.00 (1.00 – 1.00)
Problems related to social environment	249	249	247	2	2	245	99%	99%	99%	99%	0.98 (0.98 – 0.99)
Educational problems	71	67	67	4	0	425	100%	99%	94%	100%	0.97 (0.95 – 0.98)
Occupational problems	221	220	219	2	1	274	100%	99%	99%	100%	1.00 (0.98 – 1.00)
Housing problems	55	52	52	3	0	441	100%	99%	95%	100%	0.97 (0.95 – 0.99)
Economic problems	145	136	136	9	0	351	100%	98%	94%	100%	0.95 (0.94 – 0.97)
Axis V (general assessment of functioning, GAF-score) *											
Serious impairments	32	118	32	0	86	378	27%	100%	100%	81%	0.05 (0.04 – 0.06)
Moderate impairments	219	271	157	62	114	163	58%	72%	72%	59%	0.01 (0.09 – 0.11)
Slight impairments	211	95	84	127	11	274	88%	68%	40%	96%	0.10 (0.09 – 0.11)
Echelon advice											
General practice with PN	92	88	81	11	7	397	92%	97%	88%	98%	0.88 (0.85 – 0.91)
Primary care psychologist	162	127	125	37	2	332	98%	90%	77%	99%	0.81 (0.78 – 0.84)
Specialized mental healthcare	231	268	229	2	39	226	85%	99%	99%	85%	0.84 (0.81 – 0.86)

Tot Sys = total score eDiagnostic system; Tot Psy = total score psychologist; TP = true positive; FP = false positive; FN = false negative; TN = true negative; sens = sensitivity; spec = specificity; PPV = positive predictive value; NPV = negative predictive value;  $\kappa$  = Cohens kappa; 95%-BI = 95%-confidence interval; NOS = not otherwise specified; ADHD = Attention-deficit hyperactivity disorder Predominately Combined Presentation and Predominantly Hyperactive-Impulsive Presentation; ADD = Attention-deficit hyperactivity disorder Predominantly Inattentive Presentation. The figures are numbers of classifications, unless otherwise specified; Classifications are only calculated when they had a prevalence of more than 4% according to the psychologist (11). \* General Assessment of Functioning (GAF): score 41-50 = serious impairments; score 51-60 = moderate impairments; score 61-70 = slight impairments.

## Discussion

The examined eDiagnostic system could correctly determine various mood, anxiety, substance abusive disorders and the attention deficit/hyperactivity disorder as well as the absence of an Axis II disorder, without intervention of the psychologist. It was also able to give an adequate echelon advice. However, the number of false positives and false negatives for several clinical disorders was high. The system appeared to be a valid source of information for the classification of mental disorders in general practice.

## Limitations

First, the judgment of the psychologist was partly based on the classifications generated by the eDiagnostic system, therefore results could be prone to incorporation bias (6). This may have positively influenced the results. In addition, the study was conducted in a real patient setting, which entails certain limitations. The most obvious way to investigate the

eDiagnostic system would be to blindly compare the system to an independently administered structured interview (7,8). However, performing interviews with patients would be logistically very complex. For example it takes about 120–180 minutes to assess Axis I and Axis II with such an interview. This would be a big burden for the patient and it would delay the referral of patients. Therefore, we decided to use routinely collected data. One may wonder whether the judgment of the psychologist after telephone call was a valid criterion and method of measurement. The high inter-rater agreement between the psychologists as observed in our previous study pleads for the psychologists as a reference standard (1). When the system would be compared to a structured clinical interview, we expect that our overall conclusions would be maintained. However, since there would be more time to critically clarify all classifications using a structured clinical interview and also taking into account the possible incorporation bias, we expect more false positives in such an evaluation. We expect that the false negatives would remain the same, since the system is built to decrease the chance of missed symptoms.

A next limitation is that the principal investigator is Research and Development Manager at TelePsy. One may wonder whether this had an influence on the article. The results are however, not very sensitive for interpretation; all results are displayed in the tables so the reader could check the conclusions drawn.

A final limitation is that the instrument was based on the DSM-IV-TR. The eDiagnostic system based on DSM-5 is currently validated.

### *Results in relation to literature*

Only few online screening questionnaires have been developed to identify Axis I disorders in primary care. These questionnaires were mainly focused on mood or anxiety disorders. Overall, results were comparable to findings in other studies, although for certain classifications slightly higher or lower scores were found (9-12).

There are no eDiagnostic systems available to assess Axis II disorders. An earlier version of the eDiagnostic system classified specific personality disorders (1). Since too many false positives were found for specific Axis II disorders, it was decided not to classify specific personality disorders, but to group results on 'any personality disorder'. In agreement with earlier research, results showed that the system could adequately exclude an Axis II disorder (1).

We examined an eDiagnostic system and wrote a scientific report; the discussion with regard to the DSM and screening questionnaires was behind the scope of this article (13, 14).

### *Recommendations for practice*

The eDiagnostic system can be used in general practice, for instance as an aid to avoid missing mental disorders, as a guideline for a consultation, when in doubt about the nature and severity of complaints, prior to an external mental health referral, for checking a working hypothesis, when a process is stagnated or before a mental health trajectory

in general practice (e.g. PN or medication). The system is able to support the GP or PN in the classification of mental disorders. In order to do so, the GP could ask for the consultation of the external TelePsy psychologist, or results could be interpreted by GPs or PNs themselves. In using our research findings, an abridged version of the eDiagnostic system is available at TelePsy (QuickScreen) to be used in general practice.

Healthcare professionals have to realise that test results are dependent on the introspective ability of a patient. The level of education, diagnostic category and level of suffering could have a negative effect on the validity of test results. Because missed symptoms could affect the quality of life of patients, it is important to realise that the system is used to minimise the chance on false negatives. This a priori leads to the probability of false positives. In addition, the system is not able to interpret results or deal with diagnostic issues, such as differential diagnosis or overlapping symptoms.

The examined system could be used as an aid to check whether there is an indication for a DSM-IV-TR disorder and is able to support the GP in an echelon advice. It is possible to identify, exclude and predict a number of disorders, however no definite diagnoses could be made without clinical judgment.

## **Conclusion**

Research that examines the validity of an eDiagnostic system to assess a DSM-IV-TR classification is new. The examined system could be applied to examine whether a patient meets the criteria of a mental disorder and aid in the determination of an echelon advice, but it cannot replace clinical judgment.

## References

1. Dijkman I, Dinant GJ, Spigt M. The concurrent validity of a new eDiagnostic system for mental disorders in primary care. *Fam Pract* 2016; 33; 607-616.
2. Bakker P, Jansen P. *Generalistische Basis GGZ: Verwijzmodel en productbeschrijvingen*. Enschede: Bureau HHM, 2013.
3. Mokkink LB, Terwee CB, Patrick DL *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19: 539-49.
4. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: a practical guide*. Cambridge: Cambridge University Press, 2011.
5. Smits N, Smit F, Cuijpers P, De Graaf R. Using decision theory to derive optimal cut-off scores of screening instruments: an illustration explicating costs and benefits of mental health screening. *Int J Methods Psychiatr Res* 2007; 16: 219-29.
6. Worster A, Carpenter C. Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias. *CJEM* 2008; 10: 174-5.
7. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured clinical interview for DSM-IV Axis I disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/ PSY SCREEN)*. New York: Biometrics Research, 2002.
8. First M, Gibbon M, Spitzer R, Williams J, Benjamin L. *Structured Clinical Interview for DSM-IV Axis II Personality disorders (SCID-II)*. Washington, DC: American Psychiatric Press, 1997.
9. Donker T, Comijs H, Cuijpers P *et al*. The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Res* 2010 30; 176: 45-50.
10. Farvolden P, McBride C, Bagby RM, Ravitz P. A web-based screening instrument for depression and anxiety disorders in primary care. *J Med Internet Res* 2003; 5:e23.
11. Donker T, van Straten A, Marks I, Cuijpers P. A brief Web-based screening questionnaire for common mental disorders: development and validation. *J Med Internet Res* 2009; 11:e19.
12. Nguyen DP, Klein B, Meyer D, Austin DW, Abbott JA. The Diagnostic Validity and Reliability of an Internet-Based Clinical Assessment Program for Mental Disorders. *J Med Internet Res* 2015;21; 17: e218.
13. Lucassen P, Postma S, Olde Hartman T. Ggz-vragenlijsten staan persoonsgerichte zorg in de weg. *Huisarts Wet* 2017; 60: 112-4.
14. Terluin B, Oud MJT. Is de diagnose'depressie'nog wel bruikbaar? *Ned Tijdschr Geneesk* 2012; 156:A5684



# Chapter 4

## The concurrent validity of a web-based self-report assessment for personality disorders

Farid Chakhssi<sup>a</sup>  
Ies Dijkman<sup>a</sup>  
Mara L. Velmans  
J. Monique Zoet  
Jeanine M. Oostendorp  
Geert-Jan Dinant  
Mark Spigt

<sup>a</sup>These two authors contributed equally to this manuscript.

*Submitted*

## Abstract

**Background.** Incorporating web-based tools to assess DSM-IV-TR disorders, could help improve routine assessments of personality disorders. TeleScreen, a web-based self-report questionnaire for DSM-IV-TR disorders is used in the Netherlands. TeleScreen has not yet been compared to well-validated structured interviews for personality disorders, and their co-occurring clinical disorders.

**Objective.** To examine the concurrent validity of TeleScreen with Structured Clinical Interview for DSM-IV-TR Axis-I disorders (SCID-I) and personality disorders (SCID-II).

**Method.** Patients with personality disorders ( $n = 89$ ) were assessed with TeleScreen and independently interviewed with the SCID-I and SCID-II. The concurrent validity was examined using sensitivity, specificity, positive and negative predictive values (PPV and NPV) and Cohen's Kappa.

**Results.** TeleScreen showed moderate to good validity for any personality disorder, borderline personality disorder, avoidant personality disorder and obsessive-compulsive personality disorder (e.g. sensitivity: 0.70 – 0.97; specificity: 0.41-0.82; PPV: 0.56 – 0.88 and NPV: 0.78-0.88), with the exception of the dependent and paranoid personality disorder. Clinical disorders showed moderate to good values, except for social phobia, dysthymia and eating disorders (sensitivity: 0.20 – 0.29; specificity: 0.84-0.91; PPV: 0.08-0.54; NPV: 0.78 – 0.85).

**Conclusions.** Generally, TeleScreen appears to be a valid source of information for professionals in the assessment and classification of a DSM-IV-TR disorder for persons with personality disorders.

## Introduction

Personality disorders are among the most frequent psychiatric disorders treated by mental health professionals. About a quarter of patients in primary care and 45% in psychiatric outpatient settings meet the criteria for one or more personality disorders [1, 2]. Personality disorders are associated with a wide range of clinical disorders [3, 4], increased mortality [5], poor treatment outcome, extensive health care use [6, 7] high burden of disease [8] and high economic burden on society [9]. Patients with personality disorders that remain undetected in clinical practice might be given treatment that are ineffective or harmful and therefore, timely and accurate assessment of personality disorders is important [10].

To assess personality disorders and their co-occurring clinical disorders, the Structural Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) Axis-I disorders [SCID-I; 11] and for Axis-II disorders [SCID-II; 12] are widely regarded as valid and reliable measures [13, 14]. However, personality disorders are not routinely assessed in clinical practice [15-19]. One of the most important reasons is the absence of quick and valid assessment instruments for primary and specialized mental health care settings [10]. For example, semi-structured assessments for personality disorders and their co-occurring clinical disorders using the SCID-I and SCID-II usually take up to 2-6 hours [13]. Another important reason, is the presenting co-occurring clinical disorders that may mask the personality pathology. Patients with personality disorders are likely to have comorbid disorders, such as depressive or anxiety disorders, that may dominate the clinical picture [4, 10]. To improve the assessment of personality disorders in clinical practice, there is a need for novel methods that are less time consuming, easy to use for both patient and professional and that take into account the co-occurring clinical disorders.

Incorporating web-based tools, such as a web-based self-report version of the SCID-II [20] or a web-based self-report questionnaire for DSM-IV-TR clinical disorders and personality disorders [21], into routine intake procedures in mental health care could help improve routine assessments of personality disorders. Previous studies have indicated that clinicians preferred using web-based tools above semi-structured interviews, that it reduced time needed by clinicians substantially and that patients were satisfied with the web-based version [20, 21].

Previously we also examined the validity of a web-based self-report questionnaire for DSM-IV-TR clinical disorders and personality disorders, called TeleScreen [22]. In a sample of 675 adult primary care patients, we compared the results of TeleScreen, in terms of a DSM-IV-TR classification, with the clinical judgment of psychologists who had consulted the patients by phone. Results showed modest to high concurrent validity for the clinical disorders and for the personality disorders. Although these results were encouraging, additional research was deemed needed where TeleScreen would be compared to well-validated measures like independently administered semi-structured clinical interviews.

The current study compares the classifications of TeleScreen, with independent semi-structured assessments of trained raters using the SCID-I and SCID-II in a sample of patients with personality disorders.

## Method

### *Design*

We examined the concurrent validity of TeleScreen when compared to the SCID-I and SCID-II, among patients referred for specialized treatment of personality pathology. The first 31 patients were recorded on digital voice recorders, in order to investigate the interrater reliability between the raters of the SCID-I and SCID-II.

### *Participants*

Participants were recruited at two inpatient and one day-hospital unit from a Dutch psychotherapy center specializing in the treatment of personality disorders, called Scelta. All patients were referred for specialized treatment of personality disorder pathology. All patients that were consecutively admitted between June 2016 and July 2017 were invited to participate in the study. Excluded for this study were patients with insufficient understanding of the Dutch language, who showed signs of alcohol or drugs intoxication or showed acute psychotic symptoms during SCIDs assessments.

### *Raters*

The group of raters consisted of 10 first raters and 2 second raters. All raters were bachelor- or master-level psychologists with a mean age of 24.4 years ( $SD = 2.2$ ). The raters were trained in administrating the SCID-I and SCID-II by two of the co-authors (JMO & JMZ), who had extensive experience in SCIDs assessments (average years of experience is approximately 10 years). The 1-day training consisted of theory on SCID assessments, scoring case material of patients with different personality disorders previously assessed by the trainers, and role-playing interview situations. In addition, the raters accompanied the trainers for one time when performing a SCID-I and SCID-II assessment, after which the raters performed at least two SCID-I and SCID-II assessments accompanied by the trainers. Finally, raters received feedback on all consecutive SCIDs assessments from the SCID trainers.

### *Materials*

*SCID-I and SCID-II.* The Structured Clinical Interview for DSM-IV-TR Axis-I Disorders (SCID-I; [11, 23]) is a diagnostic interview designed to assess the most common clinical disorders as described in the fourth edition of the Diagnostic and Statistical Manual [24]. Test-retest and interrater reliability for the SCID-I is high [13]. The Structured Clinical Interview for DSM-IV-TR personality disorders (SCID-II; [12, 25]) consists of modules for all 11 personality disorders in which questions are grouped per personality disorder. The assessment of each personality disorder starts with an open question, after which further explanations are asked. For each disorder all criteria are addressed and scored on a three point scale (1= omission or the criterion is incorrect; 2 = criterion is doubtful; 3 = criterion is

present or correct). Previous research demonstrated that the SCID-II yields reliable personality disorder assessments [13].

*TeleScreen.* TeleScreen was previously described in Dijkman, Dinant and Spigt (2013) [21] and Dijkman, Dinant and Spigt (2016) [22]. TeleScreen is an automated web-based adaptive self-report questionnaire to assess the likelihood of DSM-IV-TR clinical disorders and personality disorders based on the SCID-I [11], MINI-International Neuropsychiatric Interview [26] and SCID-II [12]. For each clinical disorder, there are entrance questions and follow-up questions. To improve efficiency, follow-up questions are skipped according to the answers to the entrance questions. For example, when participants answer negatively on one of the core symptoms of major depressive disorder (i.e. depressed mood or a pronounced loss of interest or pleasure), the follow-up questions for major depressive disorder are automatically skipped and the participants are classified as not fulfilling the criteria for major depression. To assess personality disorders, TeleScreen presents a fixed set of 125 statements representing DSM-IV-TR personality disorder traits that can be scored with 'correct' or 'incorrect'. For example, participants are asked to answer correct or incorrect to the following statement: *"Throughout the largest extent of your life, but especially the last 5 years: as soon I sense that someone will abandon me, I will take extreme precautions to prevent that from happening. (For instance, by calling the person excessively, become overly nice, or beg the other person not to leave me)"* which directly corresponds with DSM-IV-TR borderline personality disorder criterion regarding frantic efforts to avoid real or imagined abandonment. For each personality disorder, all personality disorder traits are assessed. When a patient fulfils the required number of personality disorder traits based on the DSM-IV-TR, the presence of a personality disorder is indicated. In addition, patients' demographics, somatic complaints, psychosocial problems (e.g. problems with primary support group, occupational or economic problems) and global assessment of functioning (GAF-score) are assessed, by using both open and closed questions, to provide a DSM-IV-TR multi-axial classification.

### *Procedure*

Approval for the study was obtained from the local institutional research committee and the ethical board of the University of Twente (#16058). After participants provided written informed consent, the first 50 patients were administered the SCID-I and the SCID-II in two or three 2-hr sessions. Three weeks after SCIDs assessments participants received a link to TeleScreen via email. For the last 39 patients, the assessment of SCIDs and TeleScreen was reversed. These patients first received a link to TeleScreen, and three weeks later SCIDs assessments took place. In the first 31 cases, both SCIDs were administered face-to-face by the first raters and recorded on digital voice recorders. The second raters listened to recorded interviews, and assessed SCIDs criteria independent and blind for the first raters' scores.

### *Statistical Analysis*

*Inter-rater reliability.* Cohen's Kappa and their 95% confidence intervals (CI) were calculated to assess agreement between the raters of the categorical judgements (present/absent) of SCIDs clinical disorders and personality disorders. According to Fleiss [27], Cohen's Kappa below 0.40 indicate poor agreement, values between 0.40 and 0.75 as fair to good, and values above 0.75 as excellent.

*Concurrent validity.* The categorical presence (absent/present) of DSM-IV-TR clinical disorders and personality disorders according to SCIDs interviews were compared with categorical presence of the DSM-IV-TR clinical disorders and personality disorders based on TeleScreen. Sensitivity, specificity, positive predictive values (PPVs) and negative predictive values (NPVs) as well as the Cohen's Kappa were calculated to investigate the concurrent validity between SCIDs and TeleScreen. Acceptable values for sensitivity, specificity, PPVs and NPVs, differ depending on the tests' situation (tests aim, costs and benefits) and clinical consequences [28, 29]. TeleScreen was developed to avoid missing disorders, since missing disorders could have an effect on the quality of life of patients, or could even be fatal. As a consequence the number of false positives was expected to be higher than the number of false negatives. Also, since results could be influenced by the base rate of the disorders, analyses were only performed for categories of disorders when they had a prevalence rate of more than 4%, according to SCIDs [e.g. 30].

## **Results**

### *TeleScreen assessment*

In total, 95 patients were invited to participate, 5 patients refused and 1 patient was excluded because he met an exclusion criterion (i.e., active psychotic symptoms during SCIDs assessments). The remaining 89 patients that agreed to participate, included 72 (80.9%) women and 17 (19.1%) men, with an average age of 31.8 years ( $SD = 10.8$ ; range 18 – 59). In the current sample, the average length of TeleScreen assessments was 52 minutes ( $SD = 20$  minutes, range 22 – 120 minutes). The mean number of (login) sessions was 2.3 ( $SD = 1.8$ , range 1 – 11 sessions).

### *Interrater reliability*

The interrater reliability between the raters of SCIDs for clinical disorders and personality disorders are shown in Table 1. The contingency tables are shown in the appendix. For all classifications fair to good or excellent agreement was found, except for 'any clinical disorder', that showed poor agreement (see Table 1). This low value could, however, be due to the fact that the distribution of the marginals were highly skewed. The percentage agreement for all clinical disorders and personality disorders were high.

**Table 1** Agreement between the raters of SCIDs clinical disorders (n = 31) and personality disorders (n = 30)

DSM-IV-TR categories	Total classifications Rater 1	Total classifications Rater 2	Cohen's Kappa (95% CI)	Percentage agreement
Any clinical disorder	30	30	-0.33 (-0.38 - -0.28)	0.94
Any mood disorder	27	24	0.67 (0.34 - 1.01)	0.90
Major depression	26	22	0.64 (0.33 - 0.95)	0.87
Any anxiety disorder	20	19	0.66 (0.38 - 0.93)	0.84
Posttraumatic stress-disorder	15	15	0.74 (0.50 - 0.98)	0.87
Panic disorder	4	5	0.87 (0.62 - 1.12)	0.97
Any substance disorder	3	4	0.84 (0.53 - 1.14)	0.97
Eating disorder	11	9	0.71 (0.44 - 0.97)	0.87
Any PD	22	22	1.00 (1.00 - 1.00)	1.00
Cluster C	15	15	0.87 (0.69 - 1.05)	0.93
Avoidant PD	9	9	0.84 (0.63 - 1.05)	0.93
Dependent PD	-	-	-	-
Obsessive-compulsive PD	-	-	-	-
Obsessive-compulsive PD	8	8	1.00 (1.00 - 1.00)	1.00
Cluster B	9	8	0.92 (0.76 - 1.08)	0.97
Narcissistic PD	-	-	-	-
Antisocial PD	-	-	-	-
Histrionic PD	-	-	-	-
Borderline PD	9	8	0.92 (0.76-1.08)	0.97
Cluster A	8	5	0.71 (0.41 - 1.01)	0.90
Schizoid PD	-	-	-	-
Schizotypal PD	-	-	-	-
Paranoid PD	6	3	0.62 (0.24 - 1.00)	0.90

Note. = analyses were only performed for categories of disorders when there were at least three observations according to both raters. PD = Personality disorders.

### *Concurrent validity of the personality disorders*

The classifications according to TeleScreen and SCID-II are shown in Table 2. Most prevalent in the current sample were cluster B (dramatic, emotional or erratic personality disorders) and cluster C (anxious or fearful) personality disorders. Generally, moderate to good validity was found for the categories: any personality disorder, Cluster C- and Cluster B personality disorders, obsessive compulsive-, and borderline personality disorder. For example TeleScreen could correctly identify, exclude and predict whether a borderline personality disorder was present or absent. Also TeleScreen could adequately identify 'any personality disorder' and predict whether a personality disorder was present or absent according to the SCID, however the specificity of 'any personality disorder' was 0.41. The validity for the avoidant-, dependent-, paranoid personality disorder and cluster A personality disorders (the odd or eccentric personality disorders) varied considerably, some properties were high (e.g. specificity of 0.81 and NPV of 0.93 for the dependent

personality disorder) and some properties were rather low (e.g. sensitivity of 0.38 and PPV of 0.17 for the dependent personality disorder).

**Table 2:** Concurrent validity of TeleScreen compared with SCID-II for personality disorders (n = 89)

Personality Disorder	Total TS	Total SCID	TP	FP	FN	TN	Sens	Spec	PPV	NPV	Cohen's Kappa
Any PD	80	72	70	10	2	7	0.97	0.41	0.88	0.78	0.47 (0.22-0.72)
Cluster C <sup>a</sup>	65	51	48	17	3	20	0.94	0.54	0.74	0.87	0.51 (0.33 - 0.68)
Avoidant PD <sup>a</sup>	57	36	32	25	4	27	0.89	0.52	0.56	0.87	0.38 (0.21 - 0.54)
Dependent PD <sup>a</sup>	18	8	3	15	5	65	0.38	0.81	0.17	0.93	0.12 (-0.11-0.35)
Obsessive-compulsive PD <sup>a</sup>	28	23	16	12	7	53	0.70	0.82	0.57	0.88	0.48 (0.28 - 0.68)
Cluster B <sup>a</sup>	50	37	32	18	5	33	0.86	0.65	0.64	0.87	0.49 (0.32 - 0.66)
Narcissistic PD <sup>a</sup>	2	2	1	1	1	85	-	-	-	-	-
Antisocial PD <sup>a</sup>	13	1	1	12	0	75	-	-	-	-	-
Borderline PD <sup>a</sup>	46	36	31	15	5	37	0.86	0.71	0.67	0.88	0.55 (0.38 - 0.72)
Histrionic PD <sup>a</sup>	2	0	0	2	0	86	-	-	-	-	-
Cluster A <sup>a</sup>	45	6	3	42	3	40	0.50	0.49	0.07	0.93	-0.00 (-0.11 – 0.10)
Paranoid PD <sup>a</sup>	37	6	3	34	3	48	0.50	0.59	0.08	0.94	0.03 (-0.10 - 0.15)
Schizoid PD <sup>a</sup>	15	0	0	15	0	73	-	-	-	-	-
Schizotypal PD <sup>a</sup>	12	1	1	11	0	76	-	-	-	-	-

*Note.* = analyses were only performed for categories of disorders when they had a prevalence rate of more than 4% according to the raters; PD = Personality disorder, Total TS = Total number of TeleScreen classifications, Total SCID = Total number of SCID classifications, TP = True Positive, FP = False positive, FN = False negative, TN = True negative, Sens = sensitivity, Spec = specificity; PPV = Positive predictive value, NPV = Negative predictive value, <sup>a</sup> there was one missing value. PD = Personality disorders.

### *Concurrent validity of the clinical disorders*

The automated classifications of clinical disorders generated by TeleScreen and the DSM-IV-TR classifications by the raters based on the SCID-I, are shown in Table 3. Most common were the anxiety and mood disorders. Generally, the validity for the main categories: any clinical disorder, any anxiety and any substance-related disorders could be interpreted as moderate to good or excellent. TeleScreen could correctly identify and exclude whether there was any clinical, anxiety or substance-related disorder, and adequately predict whether those main categories were present or absent, according to the SCID. The validity for any mood- and any eating disorder varied (see Table 3). For example the sensitivity and positive predictive value for any mood disorder was rather high and the specificity and negative predictive value for any mood disorder was rather low.

**Table 3** Concurrent validity of TeleScreen compared with SCID-I for clinical disorders (n = 89)

Clinical disorder	Total TS	Total SCID	TP	FP	FN	TN	Sens	Spec	PPV	NPV	Cohen's Kappa
Any Axis I disorder	87	87	87	0	0	2	1.00	1.00	1.00	1.00	1.00 (1.00 - 1.00)
Any mood	76	67	61	15	6	7	0.91	0.32	0.80	0.54	0.27 (0.04 - 0.49)
Major Depression	63	60	48	15	12	14	0.80	0.48	0.76	0.54	0.29 (0.08 - 0.50)
Dysthymia	12	5	1	11	4	73	0.20	0.87	0.08	0.95	0.04 (-0.17 - 0.26)
Bipolar disorder	2	2	0	2	2	85	-	-	-	-	-
Any anxiety	62	68	58	4	10	17	0.85	0.81	0.94	0.63	0.60 (0.42 - 0.79)
Posttraumatic stress-disorder	41	50	35	6	15	33	0.70	0.85	0.85	0.69	0.53 (0.36 - 0.70)
Panic disorder	28	12	10	18	2	59	0.83	0.77	0.36	0.97	0.38 (0.18 - 0.58)
Social phobia	15	14	3	12	11	63	0.21	0.84	0.20	0.85	0.05 (-0.17 - 0.28)
Obsessive-compulsive disorder	13	9	6	7	3	73	0.67	0.91	0.46	0.96	0.48 (0.21 - 0.76)
Specific phobia	10	7	3	7	4	75	0.43	0.91	0.30	0.95	0.29 (-0.02 - 0.59)
Generalized anxiety disorder	8	2	0	8	2	79	-	-	-	-	-
Any Drugs	26	14	11	15	3	60	0.79	0.80	0.42	0.95	0.43 (0.23 - 0.64)
Alcohol abuse/dependence	12	9	7	5	2	75	0.78	0.94	0.58	0.97	0.62 (0.37 - 0.88)
Cannabis abuse/dependence	6	3	2	4	1	82	-	-	-	-	-
Cocaine abuse/dependence	0	0	0	0	0	89	-	-	-	-	-
Other drug abuse/dependence	13	4	2	11	2	74	0.50	0.87	0.15	0.97	0.18 (-0.09 - 0.44)
Somatoform/pain disorder	17	3	2	15	1	71	-	-	-	-	-
Eating disorder	13	24	7	6	17	59	0.29	0.91	0.54	0.78	0.23 (0.01 - 0.45)
Psychotic disorder	3	0	0	3	0	86	-	-	-	-	-

*Note.* = analyses were only performed for categories of disorders when they had a prevalence rate of more than 4% according to the raters; Total TS = Total number of TeleScreen classifications, Total SCID = Total number of SCID classifications, TP = True Positive, FP = False positive, FN = False negative, TN = True negative, Sens = sensitivity, Spec = specificity; PPV = Positive predictive value, NPV = Negative predictive value.

Most sensitivities and specificities for the specific clinical disorders were moderate to good, except for the social phobia (sensitivity: 0.21) and dysthymia (sensitivity: 0.20). The PPVs varied. Some clinical disorders could be well predicted by TeleScreen, according to the SCID (e.g. a major depression, posttraumatic stress-disorder). However, several of their PPVs were low. For example, when TeleScreen detected a dysthymia, social phobia or panic disorder, there was a chance that it was a false positive. When looking at the Cohen's Kappa, fair to good or excellent agreement between TeleScreen and SCID was found for six classifications, the seven remaining Cohen's Kappa could be interpreted as poor.

## Discussion

We investigated the concurrent validity of TeleScreen, a web-based adaptive self-report version for DSM-IV-TR clinical disorders and personality disorders, when compared to independent semi-structured assessments using the SCID-I and SCID-II in a sample of patients with personality disorders. The mean time to fill out TeleScreen was 52 minutes.

TeleScreen showed moderate to good validity for any personality disorder, borderline, avoidant and obsessive-compulsive personality disorder, except for the dependent- and paranoid personality disorder. The concurrent validity of the clinical disorders reached moderate to good values, with the exception of social phobia, dysthymia and eating disorders.

In line with our previous research, the concurrent validity of personality disorders was moderate to good, although for some classifications slightly better (e.g. borderline personality disorder), and for some slightly lower values were found (e.g. dependent personality disorder) [22]. To improve TeleScreen, several questions could be revised. For example to increase the sensitivity of the dependent personality disorder, questions could be made more broad, so more patients would recognize the symptoms more easily. Though one should also be careful with adjusting too many questions, since the number of false positives will likely increase as a consequence. It is important to keep in mind that both PPV and NPV are influenced by prevalence rates. The number of dependent and paranoid personality disorders was small in the current sample, and may have had an influence on the PPV and NPV. When the prevalence of a disease changes in a different test population, both PPV and NPV change as well: a higher prevalence will increase the PPV and decrease the NPV and vice versa [28]. Because of the low prevalence rate several other personality disorders such as the narcissistic-, antisocial-, schizoid and schizotypal personality disorders could not be examined. The current study should therefore be replicated in a larger sample to examine the concurrent validity for all personality disorders.

To assess clinical disorders, several online (screening) questionnaires have been developed and validated. Most questionnaires identified one clinical disorder or the most common clinical disorders [31-36]. The electronic Psychological Assessment and Screening System (e-PASS) [30], is an exception and has a considerably wider diagnostic reach, also taking into account less frequent clinical disorders. In general, it seems that the e-PASS showed better concurrent validity for social phobia, panic disorder, eating disorders and major depression, while TeleScreen showed better concurrent validity in assessing posttraumatic-stress disorder, obsessive compulsive disorders and alcohol abuse/dependence [30]. e-PASS, however, does not assess personality disorders and it functions as a diagnostic and referral tool, used as a starting point to access online treatment.

In line with previous research of TeleScreen in primary care, the concurrent validity of clinical disorders reached moderate to good values too [22], with the exception of the classifications for social phobia, dysthymia and eating disorders. Regarding eating disorders, Dijkman, Dinant and Spigt (2016) [22] also reported a somewhat lower PPV (PPV = .46). In the current sample of patients with personality disorders, where eating disorders are highly comorbid [37], patients may have not been aware of their symptoms or may have attributed these symptoms to personality pathology, and therefore did not endorse TeleScreen questions related to eating disorders. Again, the current study should be replicated in a larger samples using more disease-specific measures to examine the concurrent validity for clinical disorders.

Our study was conducted in a real patient setting, which contributes to the external validity, but this also ensued several limitations that should be taken into consideration. First, the number of included patients was relatively small and limited to treatment-seeking patients, who were assessed at a specialized treatment centre for personality disorders. Different prevalence rates and different PPVs of other personality disorders such as schizoid, schizotypal, antisocial or narcissistic personality disorder, may be expected in individuals who are recruited in a general population or in forensic psychiatry. Second, the time frame between the administration of SCIDs and TeleScreen was three weeks. It could be that the representation of the complaints (especially with regard to the clinical disorders) was changed. This time frame, however, was chosen to minimize the burden of consecutive assessments. Third, although, the raters were trained in SCIDs assessment and were supervised during the study, most of them were bachelor- or master level psychologists with little clinical experience. The inter-rater reliabilities were, however, fair to excellent. This is in line with a previous study that showed that clinically inexperienced students, even those that are untrained, are able to reliably assess patients' level of personality pathology from clinical interview material that converge with expert ratings [38]. Finally, the current study solely focused on categorical agreement between the two measures. Future studies should examine agreement between dimensional data of TeleScreen symptom counts and dimensional assessment of personality pathology, such as the borderline personality disorder severity index - 4<sup>th</sup> edition (BP DSI-IV; [39]).

The current study contributes further to the evidence that TeleScreen can routinely inform professionals when assessing clinical and personality disorders, in primary and specialized mental healthcare settings, at lower costs and in a user-friendly way. The present study focused on the classifications of TeleScreen, used as self-report, future research should examine whether TeleScreen augmented with the evaluation of a mental health professional could be comparable to the classifications of the SCID-I and SCID-II. Also, the introduction of the DSM-5 implies several changes when compared to the DSM-IV-TR. With regard to the clinical disorders, several changes have been made (e.g. the posttraumatic stress-disorder, dysthymia and eating disorders). There were, however, no criteria changes for personality disorders in the DSM-5. The results of the current study could therefore be applied to categorical DSM-5 personality disorders. The current study suggests to re-examine some questions in the TeleScreen for clinical and personality disorders that showed low validity.

Taken together, the web-based TeleScreen for DSM-IV-TR clinical and personality disorders showed promising results in identifying and excluding several specific clinical disorders and personality disorders. If developed further, TeleScreen may be a promising approach to help clinicians in the assessment and classification of mental disorders of their patients.

## References

1. Moran P, Jenkins R, Tylee A, Blizard R, Mann A. The prevalence of personality disorder among UK primary care attenders. *Acta Psychiatr Scand* 2000; 102:52-57.
2. Zimmerman M, Rothschild L, Chelminski I. The prevalence of DSM-IV personality disorders in psychiatric outpatients. *Am J Psychiatry* 2005;162:1911-1918.
3. Huang Y, Kotov R, de Girolamo G *et al.* DSM-IV personality disorders in the WHO World Mental Health Surveys. *Br J Psychiatry* 2009; 195: 46-53.
4. Lenzenweger MF, Lane MC, Loranger AW, Kessler RC. DSM-IV personality disorders in the National Comorbidity Survey Replication. *Biol Psychiatry* 2007; 62: 553-564.
5. Fok M, Hayes R, Chang C, Stewart R, Callard F, Moran P. Life expectancy at birth and all-cause mortality among people with personality disorder. *J Psychosom Res* 2012;73:104-107.
6. Bender DS, Dolan RT, Skodol AE *et al.* Treatment utilization by patients with personality disorders. *Am J Psychiatry* 2001;158:295-302.
7. Moran P, Rendu A, Jenkins R, Tylee A, Mann A. The impact of personality disorder in UK primary care: a 1-year follow-up of attenders. *Psychol Med* 2001; 31: 1447-1454.
8. Soeteman DI, Verheul R, Busschbach JJ. The burden of disease in personality disorders: diagnosis-specific quality of life. *J Pers Disord* 2008; 22: 259-268.
9. Soeteman DI, Hakkaart-van Roijen L, Verheul R, Busschbach JJ. The economic burden of personality disorders in mental health care. *J Clin Psychiatry* 2008; 69: 259-265.
10. Tyrer P, Reed GM, Crawford MJ. Classification, assessment, prevalence, and effect of personality disorder. *Lancet* 2015; 385: 717-726.
11. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV-TR Axis I disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/ PSY SCREEN)*. New York: Biometrics Research, New York State Psychiatric Institute, 2002.
12. First MB, Gibbon M, Spitzer RL, Williams JBW, Benjamin LS. *Structured Clinical Interview for DSM-IV Axis II Personality disorders (SCID-II)*. Washington, DC: American Psychiatric Press, 1997.
13. Lobbstaël J, Leurgans M, Arntz A. Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clin Psychol Psychother* 2011;18:75-79.
14. Widiger TA, Samuel DB. Evidence-based assessment of personality disorders. *Psychol Assess* 2005; 17: 278-287.
15. Newton-Howes G, Tyrer P, Anagnostakis K *et al.* The prevalence of personality disorder, its comorbidity with mental state disorders, and its clinical significance in community mental health teams. *Soc Psychiatr Epidemiol* 2010; 45: 453-460.
16. Beckwith H, Moran PF, Reilly J. Personality disorder prevalence in psychiatric outpatients: a systematic literature review. *Personal Ment Health* 2014; 8: 91-101.
17. Gunderson JG. Clinical practice. Borderline personality disorder. *N Engl J Med* 2011; 364: 2037-2042.
18. Westen D. Divergences between clinical and research methods for assessing personality disorders: Implications for research and the evolution of Axis II. *Am J Psychiatry* 1997; 154: 895-903.
19. Zimmerman M, Mattia JI. Differences between clinical and research practices in diagnosing borderline personality disorder. *Am J Psychiatry* 1999; 156: 1570-1574.
20. Brodey BB, First M, Linthicum J, Haman K, Sasiela JW, Ayer D. Validation of the NetSCID: an automated web-based adaptive version of the SCID. *Compr Psychiatry* 2016; 66: 67-70.
21. Dijkman I, Dinant GJ, Spigt M. eDiagnostics: a promising step towards primary mental health care. *Fam Pract* 2013; 30: 695-704.
22. Dijkman I, Dinant GJ, Spigt M. The concurrent validity of a new eDiagnostic system for mental disorders in primary care. *Fam Pract* 2016; 33: 607-616.
23. Groenestijn M, Akkerhuis G, Kupka R, Schneider N, Nolen W. *Gestructureerd klinisch interview voor de vaststelling van DSM-IV As-I stoornissen*. Lisse, The Netherlands: Swets en Zeitlinger, 1999.
24. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders 4th edition, Text Revision (DSM-IV-TR)*. Washington, DC: American Psychiatric Association, 2000.

25. Weertman A, Arntz A, Kerkhofs M. *Gestructureerd diagnostisch interview voor DSM-IV persoonlijkheidsstoornissen (SCID II)*. Lisse: Swets Test, 2000.
26. Lecrubier Y, Sheehan DV, Weiller E *et al*. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *European psychiatry* 1997; 12; 224-231.
27. Fleiss JL. *Statistical Methods for Rates and Proportions. 2nd edition*. New York: Wiley, 1981.
28. De Vet H, Terwee C, Mokkink L, Knol D. *Measurement in medicine: a practical guide*. New York: Cambridge University Press, 2011.
29. Smits N, Smit F, Cuijpers P, De Graaf R. Using decision theory to derive optimal cut-off scores of screening instruments: an illustration explicating costs and benefits of mental health screening. *Int J Methods Psychiatr Res* 2007; 16: 219-229.
30. Nguyen DP, Klein B, Meyer D, Austin DW, Abbott JA. The Diagnostic Validity and Reliability of an Internet-Based Clinical Assessment Program for Mental Disorders. *J Med Internet Res* 2015; 17: e218.
31. Donker T, van Straten A, Marks I, Cuijpers P. A brief Web-based screening questionnaire for common mental disorders: development and validation. *J Med Internet Res* 2009; 11: e19.
32. Donker T, Comijs H, Cuijpers P *et al*. The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Res* 2010 30; 176: 45-50.
33. Lin CC, Bai YM, Liu CY *et al*. Web-based tools can be used reliably to detect patients with major depressive disorder and subsyndromal depressive symptoms. *BMC Psychiatry* 2007; 7: 12.
34. Farvolden P, McBride C, Bagby RM, Ravitz P. A Web-based screening instrument for depression and anxiety disorders in primary care. *J Med Internet Res* 2003; 5: e23.
35. Spitzer RL, Williams JB, Kroenke K *et al*. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *JAMA* 1994; 272: 1749-56.
36. Spitzer RL, Kroenke K, Williams JBW *et al*. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 1999; 282: 1737-1744.
37. Martinussen M, Friborg O, Schmierer P *et al*. The comorbidity of personality disorders in eating disorders: a meta-analysis. *Eat Weight Disord* 2017; 22: 201-209.
38. Zimmermann J, Benecke C, Bender DS *et al*. Assessing DSM-5 level of personality functioning from videotaped clinical interviews: a pilot study with untrained and clinically inexperienced students. *J Pers Assess* 2014; 96: 397-409.
39. Arntz A, van den Hoorn M, Cornelis J., Verheul R, van den Bosch WM, de Bie AJ. Reliability and validity of the borderline personality disorder severity index. *Journal of Personality Disorders* 2003; 17: 45-59.

*Appendix A: Contingency table for SCIDs clinical disorders (n = 31)*

Clinical disorder	Total classifications		Matrix			
	Rater 1	Rater 2	Rater 1		Rater 2	
Any clinical disorder	30	30	Rater 1		Rater 2	
			Absent	Present	Absent	Present
			0	1	0	1
			Present	29	1	29
Any mood disorder	27	24	Rater 1		Rater 2	
			Absent	Present	Absent	Present
			4	3	4	3
			Present	24	0	24
Major depression	26	22	Rater 1		Rater 2	
			Absent	Present	Absent	Present
			5	4	5	4
			Present	22	0	22
Any anxiety disorder	20	19	Rater 1		Rater 2	
			Absent	Present	Absent	Present
			9	3	9	3
			Present	17	2	17
Posttraumatic stress-disorder	15	15	Rater 1		Rater 2	
			Absent	Present	Absent	Present
			14	2	14	2
			Present	13	2	13
Panic disorder	4	5	Rater 1		Rater 2	
			Absent	Present	Absent	Present
			26	0	26	0
			Present	4	1	4
Any substance disorder	3	4	Rater 1		Rater 2	
			Absent	Present	Absent	Present
			27	0	27	0
			Present	3	1	3
Eating disorder	11	9	Rater 1		Rater 2	
			Absent	Present	Absent	Present
			19	3	19	3
			Present	8	1	8

*Note.* = analyses were only performed for categories of disorders when there were at least three observations according to both raters.

*Appendix. B. Contingency table for SCIDs personality disorders (n = 30)*

Personality disorder	Total classifications Rater 1	Total classifications Rater 2	Matrix			
			Rater 1		Rater 2	
			Absent	Present	Absent	Present
Any PD	22	22				
					Rater 1	
					Absent	Present
			Rater 2		22	0
					Present	8
Cluster C	15	15			Rater 1	
					Absent	Present
			Rater 2		14	1
					Present	14
Avoidant PD	9	9			Rater 1	
					Absent	Present
			Rater 2		20	1
					Present	8
Obsessive-compulsive PD	8	8			Rater 1	
					Absent	Present
			Rater 2		22	0
					Present	8
Cluster B	9	8			Rater 1	
					Absent	Present
			Rater 2		21	1
					Present	8
					Present	0
Borderline PD	9	8			Rater 1	
					Absent	Present
			Rater 2		21	1
					Present	8
Cluster A	8	5			Rater 1	
					Absent	Present
			Rater 2		22	3
					Present	5
Paranoid PD	6	3			Rater 1	
					Absent	Present
			Rater 2		24	3
					Present	3

*Note.* = analyses were only performed for categories of disorders when there were at least three observations according to both raters. PD = Personality disorders.



# Chapter 5

## The perception and needs of psychologists towards blended care

Ies Dijkman,  
Geert Jan Dinant  
Mark Spigt

*Telemed J E Health 2017;23(12),1-13*

## Abstract

**Background.** Blended care, meaning that online (Internet) components are used in combination with face-to-face contact, in mental health is increasingly encouraged, but research about the needs of psychologists is scarce.

**Methods.** We assessed the perceptions, design, and barriers towards blended care among members of the Dutch Association of Psychologists through an e-mailed survey. Mean scores (SD) and answer percentages were calculated. Mann-Whitney tests were performed to investigate differences between users and nonusers and primary- and secondary care professionals.

**Results.** Generally, psychologists (63% response rate) had a positive perception towards blended care and they intended to use it in future ( $M = 3.71$ ,  $SD = 1.19$ ). Users of blended care and secondary care professionals were more positive toward blended care than non-users and primary care professionals. Online psychoeducation, diary forms and exercises for different therapeutic approaches and communication technology configurations were most welcomed. Still, quite some barriers were mentioned before professionals would use blended care.

**Conclusions.** Psychologists had a positive perception towards blended care, as long as attention is paid to the perceived barriers. Results of this survey could be used in the development of online components that correspond to the needs of professionals. Reviewing the needs of psychologists, e-health components of different therapeutic approaches, apart from cognitive behavioral therapy, are welcomed. Future research is necessary to gain insight in the (cost) effectiveness of blended care for different types of patients (e.g., transdiagnostic interventions) and of different therapeutic approaches.

## Introduction

There has been a steady increase in the use of online self-help programs for mental disorders in the past decade (1-4). Usually, it concerns standardized psychological treatments via the Internet, with minimal support from a mental health professional (1, 2). However, despite the fact that many online treatments have been shown to be effective and efficient (3-5), the transition from face-to-face contact to online interventions is progressing relatively slowly (6). In addition, it has been shown that, although the initial uptake and acceptability of the programs is usually good, dropout rates are high (1, 7). To overcome these issues, adding face-to-face therapist support to online interventions is proposed (1, 5, 8). Stand-alone online treatment therefore often develops into so-called 'blended care', meaning that online components are used in combination with face-to-face contact (9, 10).

Blended care can have advantages such as having more options for self-management, between-session communication between patient and therapist, a more independent patient-therapist relationship and better preparation of face-to-face sessions (9, 11). Dutch healthcare policy makers actively encourage this form of treatment and it is expected that blended care will play an increasingly important role in mental healthcare in the coming years (12).

Despite these high expectations research regarding the perceptions and needs of healthcare professionals is scarce. A study of van der Vaart *et al.* (2016) showed that primary care professionals would probably use online interventions in the future and that they were moderately enthusiastic about this form of treatment. Especially nonusers experienced several barriers to use blended care in their practice. The usefulness, the expected effort, and the facilitation in their practice predicted whether they would use such interventions in future (13).

A different study of van der Vaart *et al.* (2014) showed that secondary care professionals and their patients generally had a more positive perception towards blended care (9). Professionals opted for a flexible database instead of a fixed online protocol, to be able to adapt the treatment to the problems and needs of the patients (9). Especially practical online components (such as assignments, diaries and, psychoeducation) were considered useful, while face-to-face contact was deemed more appropriate for process-related components (introduction, evaluation, and discussing thoughts and feelings) (9).

The available studies on this topic suffered from a low response rate (13) or had limited external validity due to the selected population (e.g., secondary care only) (9). The present study therefore aimed for a large response rate among psychologists working in different settings, to study the perceptions and barriers towards blended care. In addition, the goal of the present study was to investigate specific wishes for the way blended care should be designed.

## Methods

### *Recruitment and study population*

Members from the Dutch Association of Psychologists (Nederlands Instituut van Psychologen, [NIP]) were approached for this survey. The Dutch Association of Psychologists is the largest association of psychologists in the Netherlands, where psychologists from all levels in the psychological health sector are registered. Only practicing psychologists in the Dutch province of Limburg who had registered their e-mail address and telephone number were invited for participation. Excluded were occupational psychologists, psychologists working with persons who have an intellectual disability, working in a nursing home and health professionals outside the Dutch province of Limburg or with missing contact information. Psychologists were first contacted through e-mail. After two weeks, nonresponders were contacted by telephone or again by e-mail in order to increase the response rate.

### *Questionnaire*

The questionnaire consisted of 27 open and closed questions, see the Appendix. To eliminate ambiguity regarding blended care, a definition was provided based on the study of Wentzel *et al.* (2016) (10). The professionals' demographics were assessed first (e.g., age, gender, echelon, and being a blended care user or not). Second, the psychologists' perceptions were assessed. To investigate the perceptions of the professionals, the Diffusion of Innovation Theory of Rogers (2003) was used as a theoretical framework (14). According to this theory there are defined categories of perceived characteristics that are essential predictors for the adoption of an innovation, including its relative advantage (perceived efficiency of the innovation in comparison to alternatives), compatibility (degree to which an innovation is compatible with the beliefs and preferences of the healthcare professional), complexity (perceived difficulty to use the innovation), trialability (perceived ability to use the innovation on a trial basis) and observability (the perceived frequency one uses an innovation or knows people who use the innovation).

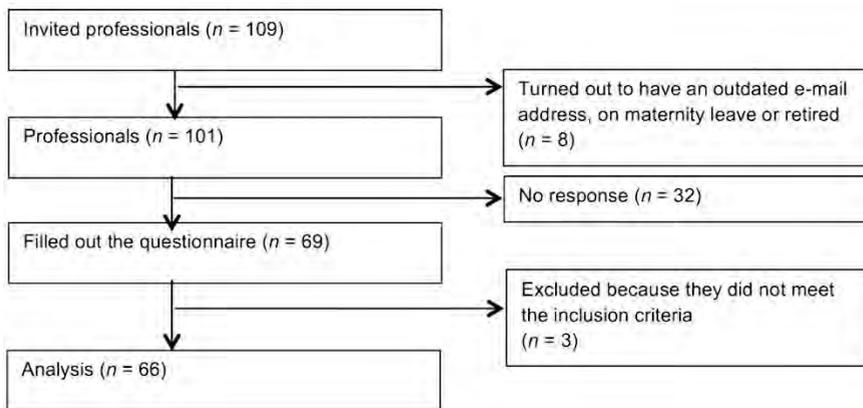
Carper *et al.* (2013) developed a questionnaire, based on this theory, to study clinician perceptions of computer-based psychological treatments and added a component future use intentions (likelihood of using the innovation in future) (15). Our questionnaire was based on the questionnaire of Carper *et al.* (2013) (15).

To reduce the burden to fill in the questionnaire, we selected questions of each category that were suitable for both users and nonusers. Third, based upon earlier research findings (16, 17), we examined the likelihood of use of different types of online components (e.g., online psycho-education, online diary forms), communication technology configurations (e.g., using e-mail, video calls or mobile technology) and possible exercises specifically applied for different therapeutic approaches (e.g., exercises based on Cognitive Behavioral Therapy, positive psychology or Acceptance and Commitment Therapy).

In addition, we investigated the barriers that prevent the use of blended care, based upon other findings of earlier research (6, 18, 19). Finally, professionals could answer to open questions (i.e., whether they had further needs or comments).

### Data analysis

Mean scores, standard deviations, and answer percentages were calculated to analyze the perceptions and barriers of mental health professionals toward blended care. Since the assumption of normality and homogeneity of variances was violated on most subscales, Mann-Whitney tests were performed to investigate differences between users versus nonusers and primary versus secondary care professionals. The answers to the open question were analyzed and arranged into themes, and a typical citation was selected per theme.



## Results

The inclusion process is shown in Figure 1. A total of 109 mental health professionals were invited to participate, 69 responded (response rate = 63%). Three professionals were excluded after filling in the questionnaire, since it turned out they did not meet the inclusion criteria after all (e.g., one was working in a nursing home, one worked with persons with an intellectual disability, and one was not a practicing psychologist). The characteristics of the included professionals ( $n = 66$ ) are shown in Table 1. All possible psychological degrees were represented in the survey (i.e., psychologists, healthcare psychologists, psychotherapists and clinical (neuro) psychologists). Professionals from different settings and working with different types of psychological disorders, such as depressive, anxiety, personality or developmental disorders responded. However, no psychologists worked in a general practice office. Our responders were mostly experienced

professionals, who did not use blended care in their treatment (i.e., 62% did not use blended care).

**Table 1.** Characteristics of mental health professionals ( $n = 66$ )

Characteristics		$n = 66$ (%)
Gender		
	Male	24 (36%)
	Female	42 (64%)
Age (M, SD)		
		49 (SD = 13)
		Range: 27 – 83 years
Professional background		
	General Psychologist, MSc	15 (22%)
	Healthcare psychologist and/or Psychotherapist	38 (58%)
	Clinical psychologist/Clinical Neuropsychologist	13 (20%)
Mean work experience (years SD)		
		20 (SD = 12)
		Range: 3-52 years
Working with		
	Child and adolescents	8 (12%)
	Adult care	57 (86%)
	Elderly care	1 (2%)
Echelon		
	Primary care psychologist (Generalistische Basis GGZ)	24 (37%)
	Secondary care (Specialistische GGZ)	30 (45%)
	Different (e.g., Medical rehabilitation)	12 (18%)
General severity of complaints of patient population		
	No mental disorders (subclinical symptoms)	1 (1.5%)
	Mild to moderate mental disorders	29 (44%)
	Severe to complex mental disorders	30 (45%)
	Guidance in psychiatric department of a general hospital (very complex care)	1 (1.5%)
	Different	5 (8%)
Blended care user		
	Nonuser	41 (62%)
	User	25 (38%)

### *Perception of mental health professionals towards blended care*

In general, the professionals had quite a positive perception towards blended care (Table 2). They were rather positive about intended use, compatibility, observability, and complexity. Of note is that 41% did not try or know where to find a blended care platform (trialability). In total, 30% did not believe that blended care had a relative advantage (in terms of its effectiveness and expected time savings), compared to treatment as usual.

**Table 2.** The perception of mental health professionals towards blended care (n = 66)

Subscale	Mean (SD)	Disagree <sup>a</sup>	Neutral <sup>b</sup>	Agree <sup>c</sup>
Relative advantage ( <i>Effectiveness and time-saving</i> )	3.17 (0.77)	30%	21%	49%
Compatibility ( <i>Positive perception and fits well with way of delivering therapy</i> )	3.52 (0.94)	24%	12%	64%
Complexity ( <i>Easy to use and knowing enough about computers to use blended care</i> )	3.44 (0.71)	9%	30%	61%
Observability ( <i>Knowing people/colleagues using blended care</i> )	3.46 (1.18)	26%	11%	64%
Trialability ( <i>Tried blended care or know where to find a program</i> )	2.96 (1.21)	41%	24%	35%
Future intentions use ( <i>Plan to use blended care in future</i> )	3.71 (1.19)	18%	20%	62%
Total	3.38 (0.76)	15%	38%	47%

Measured on a scale from 1 to 5 (1 = totally disagree to 5 = totally agree), <sup>a</sup> represents a score of  $\leq 2.5$ , <sup>b</sup> score  $> 2.5$  or  $< 3.5$ , <sup>c</sup> score of  $\geq 3.5$ .

Additional analyzes showed that users differed significantly from nonusers on all subscales, where users were significantly more positive than nonusers (Table 3). Nevertheless, 44% of the nonusers believed they would use blended care in the future. No significant differences were found between primary and secondary care professionals, except for observability (Table 3). Generally, secondary care professionals scored more positive on the subscales than primary care professionals.

### *Likelihood of using specific types of online components in the future*

When looking at the likelihood of using specific online components, it showed that most professionals scored between 'neutral and very likely' on all online components, except for the likelihood of using 'serious games' in the future (e.g., 39% did not intend to use serious games in the future), see Table 4. Generally professionals were less likely to use communication technologies such as chat or video, except for using or stay using (secured) e-mail to communicate with patients or colleagues and using mobile technology. Exercises based on mindfulness, cognitive behavioral therapy (CBT), Acceptance and Commitment Therapy (ACT) and positive psychology were most likely used in the future.

In total 14 persons (21%) used the option to report other online components beyond the ones mentioned in the questionnaire (results not shown in the Table). Most notable answers pertained to the need for psychoeducation in different languages than Dutch, modules, exercises, or apps in order to provide care after a short intensive treatment or when patients are on a waiting list, modules aimed for guiding the elderly, using (diagnostic) questionnaires, token economy systems (e.g., providing positive reinforcement to target behavior change), and using audiotaped relaxation exercises instead of relaxation exercises from screen.

**Table 3.** The perception towards blended care among users vs. nonusers and primary vs. secondary care professionals ( $n = 66$ )

Subscale	M (SD)	Disagree <sup>a</sup>	Neutral <sup>b</sup>	Agree <sup>c</sup>	P-value <sup>d</sup>	Mdn	r
<i>Relative advantage (Effectiveness and time-saving)</i>							
Users	3.52 (0.59)	12%	24%	64%	.003	3.5	-.36
Nonusers	2.96 (0.79)	41%	20%	39%		3.0	
Primary care	3.08 (0.83)	33%	25%	42%	.461	3.0	-.10
Secondary care	3.22 (0.77)	27%	20%	53%		3.5	
<i>Compatibility (Positive perception and fits well with way of delivering therapy)</i>							
Users	4.00 (0.69)	4%	16%	80%	.001	4.0	-.41
Nonusers	3.22 (0.95)	37%	10%	53%		3.5	
Primary care	3.25 (0.99)	38%	8%	54%	.259	3.5	-.16
Secondary care	3.60 (0.90)	17%	17%	66%		4.0	
<i>Complexity (Easy to use and knowing enough about computers to use blended care)</i>							
Users	3.80 (0.54)	0%	20%	80%	.001	4.0	-.41
Nonusers	3.22 (0.72)	15%	37%	48%		3.0	
Primary care	3.38 (0.85)	17%	13%	70%	.635	3.5	-.07
Secondary care	3.38 (0.65)	7%	47%	46%		3.0	
<i>Observability (Knowing people/colleagues using blended care)</i>							
Users	4.14 (0.78)	4%	8%	88%	.000	4.0	-.46
Nonusers	3.04 (1.19)	39%	12%	49%		3.0	
Primary care	3.21 (1.20)	29%	8%	63%	.046	3.5	-.27
Secondary care	3.77 (1.23)	20%	7%	73%		4.0	
<i>Trialability (Tried blended care or know where to find a program)</i>							
Users	4.02 (0.82)	4%	24%	72%	.000	4.0	-.69
Nonusers	2.32 (0.91)	63%	25%	12%		2.0	
Primary care	2.81 (1.03)	42%	29%	29%	.099	3.0	-.23
Secondary care	3.37 (1.29)	30%	20%	50%		3.25	
<i>Future intention use (Plan to use blended care in future)</i>							
Users	4.52 (0.77)	4%	4%	92%	.000	5.0	-.56
Nonusers	3.22 (1.13)	27%	29%	44%		3.0	
Primary care	3.54 (1.29)	25%	13%	62%	.404	4.0	-.11
Secondary care	3.80 (1.27)	20%	20%	60%		4.0	
<b>Total</b>							
Users	4.00 (0.43)	0%	16%	84%	.000	4.08	-.67
Nonusers	3.00 (0.66)	25%	51%	24%		2.92	
Primary care	3.21 (0.80)	21%	33%	46%	.161	3.38	-.19
Secondary care	3.52 (0.81)	17%	30%	53%		3.67	

Rated on a scale from 1 to 5 (1 = totally disagree to 5 totally agree). <sup>a</sup> Represents a score of  $\leq 2.5$ , <sup>b</sup> score  $>2.5$  or  $<3.5$ , <sup>c</sup> score of  $\geq 3.5$ . <sup>d</sup> Mann-Whitney test to investigate differences between users vs. nonusers and primary vs. secondary care professionals.

**Table 4:** Intention to use online components in future ( $n = 66$ )

	Mean (SD)	Unlikely <sup>a</sup>	Neutral <sup>b</sup>	Likely <sup>c</sup>
<b>Types of online components</b>				
Online psycho-education (via text or video)	4.09 (1.00)	11%	4%	85%
Online diary form (e.g., registration thoughts or tracking daily activities)	4.11 (0.98)	11%	6%	83%
Online exercises (e.g., exercises for cognitive behavioral therapy or mindfulness)	3.92 (0.97)	9%	18%	73%
Online modules (e.g., module depression or anxiety: modules often consist of psycho-education and several exercises)	3.41 (1.30)	30%	8%	62%
Serious games (computer games containing a treatment-element)	2.95 (1.22)	39%	26%	35%
Total <sup>d</sup>	3.70 (0.86)	9%	30%	61%
<b>Communication technology configurations</b>				
Communicate with patient or colleagues via (secured) e-mail	3.98 (0.98)	9%	8%	83%
Mobile technology via sms or app (e.g., by measuring mood/behavior in the natural environment and moment itself)	3.33 (1.22)	27%	20%	53%
Communicate with patient or colleagues via video calls	3.00 (1.10)	41%	18%	41%
Communicate with patient or colleagues via chat	2.86 (1.14)	44%	21%	35%
Total <sup>d</sup>	3.30 (0.83)	18%	38%	44%
<b>Possible exercises applied specifically for different therapeutic approaches</b>				
Mindfulness exercises.	3.89 (0.99)	14%	9%	77%
Exercises based on cognitive behavioral therapy (CBT)	3.92 (1.04)	9%	17%	74%
Exercises from the Acceptance and Commitment Therapy (ACT).	3.71 (1.13)	17%	18%	65%
Exercises based on positive psychology	3.59 (1.20)	23%	14%	63%
Exercises based on schema therapy	3.26 (1.03)	27%	26%	47%
Exercises based on interpersonal therapy.	3.05 (1.09)	30%	29%	41%
Exercises from psychodynamic psychotherapy.	2.65 (1.13)	47%	26%	27%
Total <sup>d</sup>	3.44 (0.70)	8%	42%	50%

Measured on a scale from 1 to 5 (1 = very unlikely to 5 = very likely). <sup>a</sup> represents a score of  $\leq 2$ , <sup>b</sup> score = 3, <sup>c</sup> score of  $\geq 4$ . <sup>d</sup> score of  $\leq 2.5$  is unlikely, a score  $>2.5$  or  $<3.5$  is neutral and a score of  $\geq 3.5$  is likely.

### *Barriers and necessary preparation activities before using a blended care platform*

As shown in Table 5, quite some preparation activities were found to be necessary before using a new blended care program; most scores were rated between ‘absolutely necessary to neutral’, except for the improvement of the computer skills (see Table 5). The 3 highest percentage scores were as follows: receiving information regarding the online components (83% thought this was necessary), followed by knowing the perceptions of the patients (necessary in 79% of the cases), and the clarification of legal issues/liability (78% thought this was necessary).

**Table 5.** Necessary preparation activities before using a new blended care platform ( $n = 65$ )

Barriers	Mean (SD)	Not necessary <sup>a</sup>	Neutral <sup>b</sup>	Necessary <sup>c</sup>
Obtaining information online components.	1.95 (0.89)	8%	9%	83%
Determine perceptions patients	1.95 (0.96)	6%	15%	79%
Clarify legal issues/liability.	2.11 (0.85)	8%	14%	78%
Obtaining information effectiveness platform.	2.12 (0.98)	13%	12%	75%
Determine impact time investment	2.17 (0.91)	8%	18%	74%
Determine associated incomes and costs	2.29 (1.04)	12%	17%	71%
Providing more secure network	2.18 (0.90)	8%	23%	69%
Gather information existing ethical / clinical guidelines	2.25 (0.83)	5%	26%	69%
Follow training use platform	2.43 (1.02)	15%	26%	59%
Ascertain whether care organization agrees to use platform.	2.77 (1.41)	28%	23%	49%
Arranging better IT support	2.88 (1.26)	35%	22%	43%
Improve computer skills	3.32 (1.16)	46%	31%	23%
Total <sup>d</sup>	2.37 (0.65)	5%	32%	63%

Measured on a scale from 1 to 5 (1 = absolutely necessary 5 = absolutely not necessary), the total number is lower than 66 due to the missing values of one person. <sup>a</sup> represents a score  $\geq 4$ , <sup>b</sup> score = 3, <sup>c</sup> score  $\leq 2$ ; <sup>d</sup> a score  $\geq 3.5$  is not necessary, a score  $>2.5$  or  $<3.5$  is neutral and a score  $\leq 2.5$  is necessary

### *Further comments regarding blended care*

A total of 13 respondents answered the general open question at the end of the questionnaire (20%). The answers could roughly be categorized into four themes as follows: (1) emphasis on using blended care in addition to rather than instead of face-to-face contact, (2) involvement of healthcare insurers and fear of a focus on financial efficiency, (3) limited system flexibility, and (4) suitability for patients. Table 6 presents one illustrative citation per theme.

**Table 6.** Most notable comments regarding blended care

Category	Number of professionals	Illustrative citation
Emphasis using blended care in addition to rather than instead of face-to-face contact	4	"...Very satisfied using e-health especially for support/in addition to face-to-face treatment. It cannot completely replace face-to-face treatment in my opinion. It can be useful to provide exercises, planning's, work out schedules, readings etc. at home."
Involvement healthcare insurers and fear of focus on financial efficiency	3	"..The healthcare insurers and government have a bad reputation nowadays when it comes to patient care and humanity versus management and (financial) efficiency".
Limited system flexibility	1	"...In my quest for blended care providers I encounter limited amount of flexible modules, where I have to connect to an (expensive) software system, where not all important modules are integrated (for me)."
Suitability patient	1	"...I also notice that not every patient (especially depending on the nature of psychological problems) is open minded to this form of treatment".

## Discussion

### *Summary of the most important results*

The aim of the present study was to assess the perceptions and needs of professionals towards blended care. Generally, professionals were positive about blended care and they intended to use it in the future. Blended care users and secondary care professionals were more positive toward blended care than nonusers and primary care professionals. Different online components, such as online psychoeducation, diary forms, and exercises of different therapeutic approaches were most likely used in the future. When looking at the preferred communication technology configurations, especially facilitation of (secured) e-mail between patients and colleagues and mobile technology, were welcomed. Still, quite some preparation activities were mentioned before professionals would use a new program, such as receiving information regarding the online components, knowing the perceptions of the patients and the clarification of legal issues/liability. Professionals emphasized on using online components in addition to rather than instead of face-to-face contact. They had concerns about the role of healthcare insurance companies and feared the focus on financial efficiency instead of humanity and patient care.

### *Results in relation to literature, implications for practice and future research*

In line with earlier findings within primary care (13), a relatively high amount of professionals planned to use blended care in the future. This could be the result of the fact that healthcare professionals nowadays are highly encouraged by the government and healthcare insurance companies to use this form of treatment. In addition, in correspondence with earlier findings, nonblended care users were less positive toward blended care than blended care users (13). The present study not only assessed the perceptions of primary care professionals, but examined and compared the perceptions of professionals in different settings. Generally, no significant differences between primary and secondary care professionals were found. This might be explained by the fact that the sample size was not large enough to reach statistical significance. Still, it was rather surprising that the secondary care professionals were more positive than primary care professionals, since earlier research showed that Internet-based treatment was more acceptable for patients with mild and moderate symptoms (as seen in primary care) than for patients with severe conditions (as seen in secondary care) (18). Secondary care psychologists were more positive for example about the compatibility and observability. It could be that the online components would better fit to a secondary care setting, where often a more in-depth approach is used instead of a holistic/generalized one.

Results showed that the psychologists were rather positive and had the intention to use different (practical) e-health components in addition to face-to-face contact in future. These results fit within the recommendations of van der Vaart et al., (9), who found that the complexity of the problems of patients often hinders following a strict protocol,

and therefore a flexible database with practical online components to tailor treatment to patients was suggested (9). In the past years most developers and researchers focused on online CBT, since CBT is easily structured and, therefore, it lends itself to an online format (e.g. 16, 17, 20). Reviewing the needs of the psychologists developing online components based on different therapeutic approaches apart from CBT, such as exercises based on mindfulness, ACT and positive psychology would be welcomed as well.

During the development of new online components a Center for e-health Research and Disease Management (CeHRes) roadmap could be used as a holistic framework to improve uptake and impact of the e-health intervention (21). The roadmap could be used as a practical approach to plan, coordinate and execute the development of the e-health intervention by using five steps, where stakeholder participation is essential (21). The first step (contextual inquiry) includes gathering information from intended users and the environment in which the technology is implemented (21). Results of the current study could be used during this step; however, we examined the needs of psychologists only and following this roadmap, information of all important stakeholders should be taken into account. In the design of the components developers could make use of persuasive technologies (22, 23). For example a study of Mohr et al. improved adherence rates of a multi-modal eMental Health intervention using, among others, a persuasive technologies framework in the design of the intervention.

For example, using simple and brief e-health components that involve monitoring and feedback, developing components that can be inserted into daily routine, by making tools that simplify tasks, providing tailored feedback to the patients, using media such as video to promote vicarious learning, and developing specific features that can affect human relationships to improve acceptability of the intervention (23, 24). Still, in the development and research of blended care there are a lot of interesting and important challenges in the future (16, 17). Obviously, the (cost) effectiveness of blended care needs to be studied and is the focus of a set of future studies (25-27). It would be interesting to investigate, which types or which parts of blended care are responsible for the effects (16, 17). In addition, it would be interesting to gain insight in the effectiveness of blended care treatment for different therapeutic approaches apart from CBT and for other disorders besides anxiety and mood disorders (e.g., treatment of a combination of disorders/trans-diagnostic interventions) (16, 17).

Finally, attention should be paid to the barriers of the psychologists that were observed during this study, with the highest scores found for receiving information regarding the online components, knowing the perceptions of the patients and the clarification of legal issues/liability. In the past years several studies focused on the ethical and legal issues regarding telepsychology (e.g. 20, 28, 29-31). To meet the psychologists in the need for more clarification about these issues, existing guidelines were developed (e.g., 30, 31). These guidelines provide psychologists information about informed consent procedures, standard operating procedures, and emergency issues, necessary to guide and facilitate clinicians in the online provision of mental health services (e.g. 30). To provide

guidance for therapists in when and for whom blended treatment could be suitable, therapists could use a Fit For Blended Care instrument that is developed by Wentzel et al. (2016) (10). This instrument assists in a guided dialogue between therapists and patients, discussing possible (practical) prerequisites, barriers (e.g., patients risk of crisis), facilitators, and it provides an advice overview in order to increase the treatment success (10). This information could be provided by the e-health providers, when professionals consider blended care or the information could be made available during trainings to the professionals.

### *Strengths and Limitations*

Studies among healthcare professionals often suffer from very low response rates. The response rate in this study was high (63%) increasing the chance of having a representative research population. However, practice nurses (PNs) specialized in mental healthcare and working in a general practice office were underrepresented. Van der Vaart et al. (13), investigated the perceptions of PNs and showed that they were more positive than primary care psychologists. Therefore, we expect that our observed general positive perception towards blended care would also apply if we had included more PNs. Second, up until now, there is limited amount of research available that focus on the perceptions of mental health professionals towards blended care (9, 13). To our knowledge, all available studies were conducted in the Netherlands.

Therefore, one may wonder whether the general positive perception of professionals towards blended care could be generalized across countries. Literature shows that many factors influence the professionals' acceptance of eHealth applications (32, 33). Technical conditions, legal contexts, and health policy issues differ per country and seem to play an important role in the use of e-health services (33). In addition, with an increased focus on self-management and a more independent therapist-patient relationship, blended care will have an influence on the fit within the patient-therapist interaction, which is considered as one of the essential factors for successful embedded e-health initiatives into routine care (34). In the Netherlands, the technical conditions are good; the Internet is frequently used and e-health technology is highly accessible within the practices (13). Blended care would also fit within the patient-professional interactions and it is highly stimulated by government and healthcare insurance companies.

It is questionable, however, whether the professionals in different countries (e.g., more traditional countries, or counties with less Internet access) would be open minded to this form of treatment. Third, despite our large response rate the number of participants was quite small. This was especially a problem for the comparison between different groups, such as echelon differences or differences between different types of patients, and therefore the latter was not assessed in this study. We expect that with more participants, significant differences between primary and secondary care users would be found. Finally, it should be noticed that we mostly examined intended behavior, which

could differ from actual behavior. It could be that professionals had difficulty in imagining how blended care and the proposed elements would look like or what the possibilities are (9). This could also explain the less positive perceptions towards blended care of nonblended care users in comparison to blended care users.

## **Conclusions**

This study showed that professionals from different healthcare settings had a positive perception towards blended care, as long as the online components were used in addition to rather than instead of face-to-face treatment and attention is paid to the barriers that prevent the use of blended care. The results of this survey could be particularly useful in the development of online programs that would correspond to the needs of psychologists. For example, while in the past years online CBT programs were mainly developed and evaluated, psychologists also opt for development of online components of different therapeutic approaches, apart from CBT. To help professionals in using blended care, more information about the online components, perceptions of the patients and legal issues/liability must be provided by e-health providers. Further research is necessary, to gain insight in the effectiveness of blended care for different types of patients and different therapeutic approaches.

## References

1. de Graaf LE, Huibers MJ, Riper H, Gerhards SA, Arntz A. Use and acceptability of unsupported online computerized cognitive behavioral therapy for depression and associations with clinical outcome. *J Affect Disord* 2009; 116: 227-231.
2. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. *Psychol Med* 2003; 33: 217-227.
3. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clin Psychol Rev* 2012; 32: 329-342.
4. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? *Clin Psychol Rev* 2011; 31: 89-103.
5. Spek V, Cuijpers P, Nyklicek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychol Med* 2007; 37: 319-328.
6. Whitfield G, Williams C. If the Evidence Is So Good - Why Doesn't Anyone Use Them? A National Survey of the Use of Computerized Cognitive Behaviour Therapy. *Behav Cogn Psychother* 2004; 32: 57-65.
7. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. *Psychol Med* 2009; 39: 705-712.
8. Gerhards SA, Abma TA, Arntz A *et al.* Improving adherence and effectiveness of computerised cognitive behavioural therapy without support for depression: a qualitative study on patient experiences. *J Affect Disord* 2011; 129: 117-125.
9. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJ. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. *BMC Psychiatry* 2014; 14: 355.
10. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing Online and Face-to-Face Therapy: How to Benefit From Blended Care in Mental Health Care. *JMIR Ment Health* 2016; 3: e9.
11. Månsson KN, Ruiz ES, Gervind E, Dahlin M, Andersson G. Development and initial evaluation of an internet-based support system for face-to-face cognitive behavior therapy: a proof of concept study. *J Med Internet Res* 2013; 15: e280.
12. Schalken F, Blankers M, Donker T *et al.* *Handboek online hulpverlening: met internet zorg en welzijn verbeteren*. Houten: Bohn Stafleu van Loghum, 2013.
13. van der Vaart R, Atema V, Evers AW. Guided online self-management interventions in primary care: a survey on use, facilitators, and barriers. *BMC Fam Pract* 2016; 17: 27.
14. Rogers EM. *Diffusion of innovations (5th ed.)*. New York: Free Press, 2003.
15. Carper MM, McHugh RK, Barlow DH. The dissemination of computer-based psychological treatment: a preliminary analysis of patient and clinician perceptions. *Admin Policy Ment Health*. 2013; 40: 87-95.
16. Thewissen V, Gunther N. E-mental health: state of the art. *Tijdschrift voor Psychotherapie* 2015; 41: 374-392.
17. Blankers M, Donker T, Riper H. E-mental health in Nederland: wetenschappelijke evidentie en uitdagingen voor de praktijk. *De Psycholoog* 2013; 48: 12-23.
18. Gun SY, Titov N, Andrews G. Acceptability of Internet treatment of anxiety and depression. *Australas Psychiatry* 2011; 19: 259-264.
19. Wells M, Mitchell KJ, Finkelhor D, Becker-Blease KA. Online mental health treatment: concerns and considerations. *Cyberpsychol Behav* 2007; 10: 453-459.
20. Proudfoot JG. Computer-based treatment for anxiety and depression: is it feasible? Is it effective? *Neurosci Biobehav Rev* 2004; 28: 353-363.
21. van Gemert-Pijnen JE, Nijland N, van Limburg M *et al.* A holistic framework to improve the uptake and impact of eHealth technologies. *J Med Internet Res* 2011; 13: e111.

22. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: Key issues, process model, and system features. *Commun Assoc Inf Syst* 2009; 24: 28.
23. Fogg BJ. *Persuasive technology: using computers to change what we think and do*. Boston: Morgan Kaufmann Publishers, 2003.
24. Mohr DC, Duffecy J, Jin L *et al*. Multimodal e-mental health treatment for depression: a feasibility trial. *J Med Internet Res* 2010; 12: e48.
25. Kooistra LC, Wiersma JE, Ruwaard J *et al*. Blended vs. face-to-face cognitive behavioural treatment for major depression in specialized mental health care: study protocol of a randomized controlled cost-effectiveness trial. *BMC Psychiatry* 2014; 14: 290.
26. Romijn G, Riper H, Kok R *et al*. Cost-effectiveness of blended vs. face-to-face cognitive behavioural therapy for severe anxiety disorders: study protocol of a randomized controlled trial. *BMC Psychiatry* 2015; 15: 311.
27. Kemmeren LL, van Schaik DJ, Riper H, Kleiboer AM, Bosmans JE, Smit JH. Effectiveness of blended depression treatment for adults in specialised mental healthcare: study protocol for a randomised controlled trial. *BMC Psychiatry* 2016; 16: 113.
28. Childress CA. Ethical issues in providing online psychotherapeutic interventions. *J Med Internet Res* 2000; 2:e5.
29. Baker DC, Bufka LF. Preparing for the telehealth world: Navigating legal, regulatory, reimbursement, and ethical issues in an electronic age. *Prof Psychol Res Pract* 2011; 42: 405.
30. Hsiung RC. Suggested principles of professional ethics for the online provision of mental health services. *Stud Health Technol Inform* 2001; 84: 1296-1300.
31. American Psychological Association. Guidelines for the practice of telepsychology. *Am Psychol* 2013; 68: 791-800.
32. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health care provider adoption of eHealth: systematic literature review. *Interact J Med Res* 2013; 2: e7.
33. Santana S, Lausen B, Bujnowska-Fedak M *et al*. Online communication between doctors and patients in Europe: status and perspectives. *J Med Internet Res* 2010; 12: e20.
34. Murray E, Burns J, May C *et al*. Why is it difficult to implement e-health initiatives? A qualitative study. *Implement Sci* 2011; 6: 6.
35. Central Committee on Research Involving Human Subjects (CCMO). <http://www.ccmo.nl/en/your-research-does-it-fall-under-the-wmo>. Available at: <http://www.webcitation.org/6nuA6jRNr> (last accessed May 15, 2017).

*Questionnaire needs of healthcare professionals regarding blended care*

This questionnaire was originally administered in Dutch and translated in English for the publication of the article. It is not formally translated according to standard procedures.

1. Are you currently guiding or treating patients with psychological problems?
  - Yes
  - No
2. What is your professional background?
  - Psychologist in training
  - General Psychologist, MSc
  - Psychotherapist
  - Healthcare psychologist
  - Clinical psychologist/Clinical neuropsychologist
  - Different, please provide further explanation
3. What is your sex?
  - Male
  - Female
4. What is your age?
5. In what kind of echelon are you *mainly* working (if you work in different echelons please select the echelon where you are mainly active).
  - General practice with practice nurse specialized in mental healthcare
  - Primary mental healthcare [in Dutch: Generalistische basis GGZ]
  - Specialized mental healthcare [in Dutch: Specialistische GGZ]
  - Different (please provide further explanation)
6. How much working experience do you have within mental healthcare?
7. With what kind of patients do you mainly work (if you are working with different patients please select the main patient population)
  - Child and adolescents
  - Adult care
  - Elderly care
  - Different, please provide further explanation
8. The patients within your practice/institution *mainly* have:
  - No psychological disorders (subclinical symptoms)
  - Mild to moderate psychological disorders
  - Severe to complex psychological disorders
  - Crisis/residential care (very complex care)
  - Different, please provide further explanation

9. What kind of psychological disorders do your patients *mainly* have?

In this study blended care refers to a combination of face-to-face consultations and online components, where the online components and face-to-face consultations both are not standalone treatment pathways but are interconnected to each other. The online components are carefully selected and adjusted to treatment. The online components can be used flexibly and in a dynamic way. The patient and therapist decide together how the online components could be used optimally.

Imagine for example a database where you can flexibly select psycho-education, diary forms, modules or online exercises (e.g., from cognitive behavioral therapy, Acceptance and Commitment Therapy or meditation exercises) in addition to treatment.

10. Do you use blended care for the treatment/guidance of your patients?

- Yes
- No (little or not at all)

11. Please answer the questions below based on your beliefs about blended care. If you have not heard of blended care before now, base your answers on what you expect to be true.

- Using blended care would save me more time than other treatments  
Totally disagree, disagree, neutral, agree, totally agree
- I think that blended care would be effective for a lot of my patients  
Totally disagree, disagree, neutral, agree, totally agree
- I have a positive perception about blended care  
Totally disagree, disagree, neutral, agree, totally agree
- I believe that blended care would fit with the way I would like to deliver treatment  
Totally disagree, disagree, neutral, agree, totally agree
- I think it is easy to use blended care for my patients  
Totally disagree, disagree, neutral, agree, totally agree
- I know enough about computers to use blended care  
Totally disagree, disagree, neutral, agree, totally agree

12. Please answer the questions below based on your beliefs about blended care. If you have not heard of blended care before now, base your answers on what you expect to be true.

- I know colleagues who have considered to use blended care  
Totally disagree, disagree, neutral, agree, totally agree
- I know people who use blended care  
Totally disagree, disagree, neutral, agree, totally agree
- I plan to use / stay using blended care in future for my patients  
Totally disagree, disagree, neutral, agree, totally agree

- I have tried blended care  
Totally disagree, disagree, neutral, agree, totally agree
  - I know where to find online components to try blended care  
Totally disagree, disagree, neutral, agree, totally agree
13. Imagine that you are going to work with a new blended care platform, where you can work with the following eHealth configurations. How likely would you use the following configurations in addition to face-to-face treatment?
- Online psycho-education (via text or video)  
Very unlikely, unlikely, neutral, likely, very likely
  - Online dairy forms (e.g. for tracking thoughts or daily activities)  
Very unlikely, unlikely, neutral, likely, very likely
  - Online exercises (e.g. exercises based on cognitive behavioral therapy or mindfulness)  
Very unlikely, unlikely, neutral, likely, very likely
  - Online module (e.g. module depression or anxiety: modules often consist of psycho-education and different assignments)  
Very unlikely, unlikely, neutral, likely, very likely
  - Serious games (computer games with a treatment-element)  
Very unlikely, unlikely, neutral, likely, very likely
14. Are there, apart from the above-mentioned components, other components that you as a healthcare professional would need?
- No
  - Yes, please provide further explanation
15. Imagine that you are going to work with a new blended care platform, where you can work with the following communication technology configurations. How likely would you use the following technologies in addition to regular face-to-face treatment?
- Communicate with patient or colleagues through chat  
Very unlikely, unlikely, neutral, likely, very likely
  - Communicate with patient or colleagues through (secured) e-mail.  
Very unlikely, unlikely, neutral, likely, very likely
  - Communicate with patient or colleagues through video calls  
Very unlikely, unlikely, neutral, likely, very likely
  - Mobile technology via sms or app (e.g. by measuring the mood/behavior in the natural environment or at the moment itself).  
Very unlikely, unlikely, neutral, likely, very likely
16. Imagine that you are going to work with a new blended care platform for which exercises from different therapeutic approaches are available online. How likely would you use the following online exercises?

## Supplementary material

- Exercises based on cognitive behavioral therapy  
Very unlikely, unlikely, neutral, likely, very likely
  - Exercises based on positive psychology  
Very unlikely, unlikely, neutral, likely, very likely
  - Exercises based on Acceptance and Commitment Therapy (ACT)  
Very unlikely, unlikely, neutral, likely, very likely
  - Mindfulness exercises  
Very unlikely, unlikely, neutral, likely, very likely
  - Exercises based on schema focused therapy  
Very unlikely, unlikely, neutral, likely, very likely
  - Exercises from psychodynamic psychotherapy  
Very unlikely, unlikely, neutral, likely, very likely
  - Exercises based on interpersonal therapy.  
Very unlikely, unlikely, neutral, likely, very likely
17. Imagine that you are going to work with a new blended care platform. Which preparatory activities would then be necessary?
- Obtain more information about the available online components  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary
  - Obtain more clarity regarding legal issues / liability  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary
  - Obtain more information about the effectiveness of the blended care program  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary
  - Follow a training to use the blended care program  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary
  - Providing a more secure network  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary
  - Gather information about existing ethical / clinical guidelines regarding blended care  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary
  - Improve my computer skills  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary
  - Arranging better IT support  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary
  - Ascertain whether my care organization agrees to use the blended care platform  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary
  - Determine the associated incomes and costs of the blended care platform  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary
  - Determine the impact on my time investment (preparation and implementation of the program).  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary

- Determine the perceptions of my patients towards a blended care program  
Totally unnecessary, unnecessary, neutral, necessary, totally necessary

This was the end of the online survey. Thank you very much for participation. Do you have any further comments?

- Yes
- No





## General discussion



## Major findings

The main goal of this dissertation was to evaluate experiences with and the concurrent validity of TeleScreen, an eDiagnostic system to improve routine assessments of clinical and personality disorders based on the DSM-IV-TR, in patients with a suspicion of a mental disorder, in primary care. Patients, GPs and PNs were (moderately) enthusiastic about the system. Patients showed no preference for face-to-face consultation with a psychologist above the online system. GPs and PNs felt strengthened in their control function. In most cases the system gave a different echelon advice when compared to the echelon suggested by GPs and PNs, prior to the evaluation by the psychologist. Nevertheless, GPs and PNs generally accepted the results of the examination and the advice given. TeleScreen appeared to be a valid source of information for professionals in the assessment and classification of DSM-IV-TR disorders in both a population of primary care patients and patients referred to a specialized mental healthcare setting. Still, for a number of disorders misclassifications were found, resulting in the advice that TeleScreen cannot fully replace clinical assessment. Psychologists from different healthcare settings were positive about blended care, as long as the online components are applied as an addition to, rather than instead of face-to-face treatment, and attention is paid to perceived barriers. In addition, we gained insight into preferred eHealth components to support treatment. This knowledge can be used for the development of online elements that fit to the needs of psychologists. For example, eHealth components for other therapeutic approaches, besides cognitive behavioral therapy, are welcomed by clinicians in their treatment of patients.

## Reflection on the main findings

### *What is the best reference standard?*

Our results show that TeleScreen classifies some mental disorders better than others. The decision what reference standard to use, is of major importance in the assessment of validity. At the time of our research, we could not compare TeleScreen with other online questionnaires, since TeleScreen was the first online instrument in the Netherlands. A non-online alternative would be the SCID-I and SCID-II, since these are widely regarded as gold standard in the assessment of clinical and personality disorders and are therefore often used in mental healthcare settings (1, 2). In the second phase of our studies it was possible to compare TeleScreen with SCID-I and SCID-II in a population of patients referred to a specialized treatment center for personality disorders. Comparing TeleScreen to SCID-I and SCID-II was however logistically too complex in general practice, since performing SCID interviews with one patient takes 120–180 minutes, which would have created a considerable burden, and a possible delay in referral for the patient, which

would probably be ethically unacceptable. Therefore, the best available method to determine the presence or absence of a mental disorder was using the judgment of a trained and supervised psychologist, with a Master of Science in Psychology degree and a basic certificate of psychodiagnostics. All patients were contacted by the psychologists. During the consultation with the psychologist, the complaints the patient experienced were evaluated. Psychologists could check the classifications generated by the system, ask additional questions and clarify ambiguous test results. The thus obtained DSM-IV-TR classification and echelon advice was used as a reference standard. In this dissertation DSM-IV-TR classification is used instead of a diagnosis, since a diagnosis would contain more elements (such as complicating factors and factors involved in success or failure of treatment) than a description of symptoms (3). Although the judgment of the psychologist contains more elements than a pure DSM-IV-TR classification (e.g. suicide risk is assessed, it takes into account earlier treatment history, comorbidity and the course of the disorder), it is still not a definite diagnosis.

During our study in general practice, we were struggling with the kind of gold standard or reference standard to use, which is often considered to be a challenge in the evaluation of diagnostic procedures (4). We were in doubt whether the use of a psychologist's judgment was a valid reference standard instead of the well-known SCID. In addition, the fact that the judgment of psychologists was partly based on results of TeleScreen, while an independent examination would be preferable (5), could have led to biased test results. In the end we used both the SCID and the psychologists judgement in our studies. This makes it possible to compare the two 'gold standards'. Our evaluations showed that the main conclusions between the studies were not very different. Therefore, we would argue that future studies on the validity of eDiagnostics in primary care, would not have to be complicated by the demand to use the SCID as a gold standard.

### *What is acceptable validity?*

Sensitivity and specificity are often used as measures of the criterion validity of a test (5). When TeleScreen would reach 100% sensitivity, this indicates that the clinician can be confident that a negative test result indicates the absence of a disease. When TeleScreen would reach 100% specificity, the clinician can be sure that with a positive test result the patient does have the disease. Tests with 100% sensitivity and 100% specificity are rare in practice, especially within the field of mental healthcare (6). Often a trade-off between sensitivity and specificity is necessary, obtaining higher sensitivity at the expense of lower specificity or vice versa, by adjusting the threshold for the test to be called positive.

There are no uniform criteria to decide whether an instrument is valid for its clinical purpose in a given situation. Acceptable validity depends on the aim of the test, its costs and consequences of false positive and false negative diagnostic conclusions (5, 7). TeleScreen is designed as an aid to facilitate healthcare providers with a (first) assessment and classification of mental disorders. The choice between high sensitivity or high

specificity differs per mental disorder and depends on, among others, the severity and prevalence of the specific disorder. As missed disorders can affect the patient's quality of life or can even be fatal, one would want to avoid false negative diagnostic test results, which especially applies to more severe, high risk disorders (e.g. bipolar disorder or psychotic disorder). This means that higher levels of sensitivity with relatively lower levels of specificity are preferable for more severe, high risk disorders. This would necessarily result in a higher number of false positive diagnostic test results. False positive diagnostic test results are in principle less problematic in this situation, as professionals could correct for the misclassifications, so that (unnecessary) classification or stigmatization can be minimized. Other questionnaires aimed at assisting the professional in the identification of mental disorders also consider this to be appropriate (e.g. 6, 8, 9, 10). High specificity may be preferred in case of more common disorders, such as the major depression or anxiety disorders. This would decrease the demand on the professional judgement of the clinician.

### *Diagnostic conclusions resulting from TeleScreen*

When reviewing the entire diagnostic process of mental health in primary care, including the use of TeleScreen, from the beginning until the end, one could say that different working hypotheses/diagnoses about a person's mental complaints are made at different points in time. The first diagnostic test result comes from the GP or PN who requests an assessment with TeleScreen, and often formulates his or her own working hypothesis based on his or her own experiences with the patient, prior to the assessment. It is up to the GP or PN to share this working hypothesis via the system with the TelePsy consultant. Subsequently, the online questionnaire is filled out by the patient and an indication for a most likely mental disorder is automatically generated by TeleScreen, irrespective of the working hypothesis of the GP or PN. This step, leads to the second diagnostic test result that has to be checked by a professional. Third, a psychologist interprets the results and makes his or her own provisional DSM-IV-TR classification based on TeleScreen outcomes and a consultation with the patient. This can be seen as a third diagnostic test result. Since TeleScreen is the main subject of this thesis, we mainly focused on the improvement of TeleScreen (second diagnostic test result).

### *Improving TeleScreen*

Classifications of disorders that showed low validity, may be improved by refining TeleScreens' questions. For example, questions could be broadened, so that more patients would recognize themselves in the questions. Improved specificity can be reached by sharpening questions, so that less patients would recognize themselves in the questions. Another possible method to optimize sensitivity and specificity, is by changing cut-off points. Up until now, cut-off points of TeleScreen were based on the formal DSM-IV-TR criteria (11). One could adjust the diagnostic threshold (if desired) for those disorders

that showed low levels of sensitivity, meaning that less symptoms are required to meet the criterion of a mental disorder. A method to determine the optimal sensitivity and specificity cut-off point, is the Receiver Operating Characteristics (ROC) analysis (5). After fitting the ROC curve, it is possible to calculate the Area Under the Curve (AUC), which has a maximum of 1.0 (5). One disadvantage of the AUC as indicator is that both false positive and false negative diagnostic test results are equally weighted, while the two may have different implications (5). As previously mentioned, there are situations in which a false negative diagnostic test result weights more heavily than false positive diagnostic test results (e.g. in case of more severe, high risk mental disorders). A solution might be to adjust this optimum, to see whether it is possible to only reduce the number of false positive diagnostic test results, without affecting the false negative diagnostic test results. However, adjusting cut-off points for one setting (such as general practice), could have implications for the usage of TeleScreen in other healthcare settings. For example, the validity of TeleScreen was different in the primary care setting for social phobia, when compared to the specialized care setting for patients with personality pathology. This makes it difficult to determine which kinds of adjustments are required. It is also questionable whether false positive or false negative diagnostic test results can really be avoided when it comes to self-report questionnaires, since they are influenced by many more factors, such as factors with regard to human processes (e.g. language comprehension, information processing, memory deficits or social desirability) (12, 13), factors with regard to the mental disorder (e.g. lack of insight) or factors as a result of the DSM-IV-TR or DSM-5 (e.g. diagnostic overlap of symptoms).

In our opinion, it seems best to stick to the current cut-off points based on the DSM-IV-TR or DSM-5 instead of adjusting the DSM cut-off points. If one would deviate from DSM, one loses its anchor points, which could be a problem when TeleScreen is used in different healthcare settings. A deviation from DSM-IV-TR or DSM-5 should only be done when TeleScreen could be used as an absolute diagnostic tool, meaning that definite diagnoses can be made after filling out TeleScreen, and the clinician is not needed to interpret results. However in practice this is not feasible, since diagnostic assessment and decision making is not so straightforward. When using DSM-IV-TR or DSM-5 cut-off points as anchor points, a clinician can at least be confident that there is no discrepancy between classifications set by TeleScreen and DSM-IV-TR or DSM-5. The clinician can be confident that the patient does or does not recognize the required symptoms of a mental disorder based on the DSM-IV-TR or DSM-5. Hereafter the diagnostic process starts. In future, the classification of mental disorders, where TeleScreen shows low levels of validity may be improved, by refining questions to enhance its clarity. Upcoming changes in the next version of TeleScreen, due to implied changes of the DSM-IV-TR to the DSM-5, served as an opportunity to refine and re-examine the questions.

### *Diagnostic process*

The diagnostic process starts after a patient fills out the TeleScreen. Further elaboration of complaints is deemed needed in a clinical interview to extend the working hypothesis and set the most suitable preliminary DSM-IV-TR or DSM-5 classification. The clinician's judgment is important to interpret results, to determine whether the patient understood questions correctly, whether there is a tendency to over- or underreport symptoms, and whether the patient has sufficient insight into his or her symptoms and mental health problems. An important aspect when dealing with uncertainty or complex diagnostic decisions as described in medical decision and medical problem solving models, is that clinicians use their gut-feelings as a specific form of intuition in diagnostic reasoning (14). Gut-feelings are part of non-analytical cognitive processes and it is opted to be used next to analytical processes, that is: evidence-based medicine to avoid cognitive biases that arise when relying solely on one's own knowledge (14). It is important that the complaints of the patient are considered within the context of the person, also taking the story behind the complaints into account instead of solely focusing on DSM symptoms (15, 16). Clinicians should be aware that the system can generate false positive and false negative diagnostic test results. Various factors, such as the introspective ability of the patient, social desirability (17), the level of education, diagnostic category, and level of suffering all have an influence on the validity of self-report questionnaires. Also, self-report questionnaires rely on retrospective assessments and are therefore vulnerable to recall biases (18). Nowadays, the Experienced Sampling Method, also known as Ecological Momentary Assessment, gains popularity (19). Experienced Sampling Method refers to assessment strategies in which patients are asked to complete questionnaires at semi-random moments in real-life and at real-time; so in the natural environment and at the moment itself (19). Currently, TeleScreen does not make use of the Experienced Sampling method. In future it would be interesting to develop an abridged version of TeleScreen that makes use of the Experienced Sampling Method, for instance via a mobile application.

### *Implementation of TeleScreen*

Successful implementation of innovations is often complex. The nature of the innovation, the characteristics of GPs and patient, economic, organisational and political factors are involved in successful implementation (20, 21). In the area of eMental health, professionals are usually sceptical about using new eHealth solutions for various reasons (22-24). One reason is that eHealth innovations are often developed for and not in collaboration with end-users. A top-down approach of developing eHealth technologies seems to lack important insight and feedback from stakeholders and may result in a tilt of investments relative to impact (25), usability problems and high attrition rates of end-users (26). During the development and implementation of TeleScreen, focus groups were organized to stay in close contact with both GPs and PNs. In addition, several mental healthcare

institutions were involved in the development of TeleScreen. In this way there was close contact with end-users and the system could be further developed where necessary.

Still, within the field of eDiagnostics, there is resistance from professionals, for instance with regard to screening questionnaires and routine outcome monitoring questionnaires (15, 16). A frequently expressed argument against the use of (online) questionnaires to assess mental disorders, is that it would stimulate a disease-oriented approach, implying a 'cure', which could cause high prescription levels of psychotropic medication, while a more person-centred approach could be needed in general practice (16). This means that the complaints of the patient should be considered within the context of the person; the story behind the complaints is leading and not the complaint itself. Also, it is suggested that research and treatment is too often focused on symptom and complaint reduction, instead of acceptance of complaints, especially when it is applied as screening (15). It is therefore suggested to focus on contextual factors and include positive mood states in addition to negative mood states and symptomatology, rather than being solely focused on negative symptomatic mood states (15, 19).

In our view eDiagnostics could be considered when a person-oriented approach is not working and a disease-oriented approach is necessary, to avoid missing psychological disorders or as a guidance for consultation within general practice (16). TeleScreen is not a screening questionnaire, since it does not include assessment of all persons in the population; a healthcare professional requests the use of eDiagnostics (27). In addition, filling out the online questionnaire (TeleScreen) should be seen as one aspect of the entire diagnostic process. The contact of patients with GP and/or PN prior to the assessment and contact with psychologists after assessment are important, so that the background and context of the person could be taken into account, offering the best of both worlds.

The reluctant attitude of some professionals in general practice is in contrast with our results where positive reactions were found among service users (28). Taking the literature into account the evaluations of GPs and PNs were surprisingly positive. In our view, there are two key features of TeleScreen that help in the positive perception of service users. First, TeleScreen is not time-consuming for GPs and PNs and it has a minimal effect on their workload (29). When TeleScreen is used in addition with a consultation of a professional of TelePsy, GPs or PNs do not need to interpret the questionnaire themselves. They receive additional information about their patients, just like they would receive additional information from laboratory test results, which increases their ability to reach a working hypothesis supplemented by a provisional DSM-IV-TR or DSM-5 classification of the patient and an advice on the echelon pathway to follow (28, 29). Second, the fact that the instrument was developed in close contact with end-users probably contributed to the user-friendliness of the system for GPs and PNs. Still, it would be interesting to investigate opinions of dedicated nonusers. A similar research approach could be used within general practice, as we did in chapter five, evaluating the perceptions and needs of psychologists towards blended care (30).

### *Developing eHealth content*

TelePsy was founded in the Netherlands in 2010 in collaboration with various GPs and mental healthcare professionals. Initially, TelePsy focused on supporting healthcare professionals by developing TeleScreen, which was the main tool investigated in this thesis. After this, TelePsy gradually evolved into an online platform for healthcare professionals. They offered a platform including a large number of questionnaires that could also be used for effect assessment. Thereafter, TelePsy started to develop online eHealth components, including psycho-education, exercises, journals and modules, to assist healthcare professionals in the treatment of patients with mental health problems.

The roadmap of the Center for eHealth Research and Disease Management (CeHRes) described in the publication by van Gemert-Pijnen et al. (2011) can be used as a theoretical model to develop eHealth content (23). This holistic framework is developed to improve the uptake and impact of eHealth technologies and includes the input of different stakeholders to develop feasible, sustainable and beneficial eHealth technologies for end-users and all stakeholders involved (23). The roadmap is used as a theoretical and practical approach using five steps to plan, coordinate and execute the development of eHealth components. The first step (contextual inquiry) includes gathering information from intended users and the environment in which the technology is implemented (23). Our results of the study on perceptions and needs towards blended care can be used as a starting point (30). The next advice would be to further develop eHealth components, in close contact with end-users. When designing content, developers could make use of persuasive technologies (31, 32), as for example used in a study of Mohr et al. to improve adherence rates of an eMental health intervention (33). After the development reaches a stable phase, one could determine whether goals are achieved (as described in the final step of the roadmap) and during this phase scientific evaluation can be performed (23), for example to gain insight into the feasibility and cost-effectiveness of blended care treatment for different patient groups.

### **Strengths of our studies**

Selecting the right study population is important for the external validity of a study. One could for example choose a population with a broad spectrum of diseases, also ranging from mild to severe diseases, or one could aim for patients in which the clinicians would normally apply the test (34). We conducted our studies in real-world care settings, with real patient populations. This contributes to the external validity of the research. We used actual treatment-seeking patients in both general practice and a specialized healthcare setting. Patients with a broad spectrum of diseases were included and the exclusion criteria were limited to a few (e.g. patients who were unable to perform the diagnostic tests such as patients with insufficient understanding of the Dutch language, a cognitive

impairment or severely limited reality testing capabilities). The fact that we used actual treatment seeking patients is in contrast to studies that, for example, used internet banners to recruit participants (9) or used visitors that entered an online website to fill out the online questionnaire (35). This means that our results would most probably also apply to patient populations outside our study population.

On the other hand, the setting where we did our research has some specific characteristics. In the Netherlands, most patients are guided or treated by their GPs and a small number of patients are referred to mental healthcare settings (36). The population of patients in our study probably had quite severe problems. It consisted of patients in both a specialized setting with personality pathology and persons in a primary care setting. When looking at the primary care setting, a large number of patients were advised to be treated in secondary mental healthcare, which implies treatment of the more complex disorders (28, 37). Also, GPs requested the help of TelePsy, presumably because they suspected psychopathology, which may also indicate a more severe patient population. We assume that our findings could be generalized to other primary care settings in the Netherlands, but in countries with a different primary care system the results may be different.

In addition, we performed surveys on users in a gatekeeper system. It is questionable whether results can be generalized to countries with a non-gatekeeper system, such as Belgium, France, Germany, Finland and Sweden (38, 39). It is known that the acceptance of eHealth interventions depends on technical conditions, legal contexts, and health policy issues (40). In the Netherlands the conditions are good. There are incentives from the government to use eHealth interventions. In addition, the Internet is frequently used and eHealth technology is highly accessible within practices (41). One may wonder whether GPs in non-gatekeeper systems, or GPs in countries with less incentives from the government to use eHealth, are open minded to use this instrument. The current positive evaluations may therefore over-estimate the satisfaction scores of GPs. A replication study in such a population would be interesting.

Second, in line with the above, the fact that the research was conducted in clinical practice contributed to the relevance and practical clinical value of the research. The insights into the validity of TeleScreen lead to further development of the instrument, by refining questions and decreasing the demand of the professional judgment of the clinician. In addition, while a lot of questionnaires were only applicable or studied in one healthcare setting, TeleScreen was studied and can be used in different healthcare settings. This increases its relevance for practice and there is a need for an instrument that can be used in different echelons (42).

There is great potential to use the same eHealth system among different disciplines who work in different echelons. One way to do so would be to transfer answers to TeleScreen questions filled out by patients in one setting to a healthcare professional in a different setting. This might enhance the efficiency of care of patients who receive treatment within different echelons, because patients do not have to answer similar

questions twice. To start with using questions in general practice, at the starting point of care, to quickly triage where the patient should go, would seem the most logical (29). This way patients receive more information about the nature and course of their problem and they know what to expect (29). It would be ideal if the same answers could be used in the next echelon, to go in depth with the patients, combining the results of TeleScreen with an intake or an interview with a healthcare professional. In this way communication between healthcare professionals and between patient and healthcare professional may be improved. This idea fits within the current trend in which patients get more control over and insight into their patient record, taking the information with them to a specialist. A specialist could then add his or her specialty to the treatment of the patient.

Third, our studies were performed in an efficient way. The majority of studies in this thesis are observational in nature. We mostly used routine care data or online questionnaires among service users, so that data could be gathered relatively easily. The burden of administration for participants, clinicians and researchers was minimized. By using routine care data, less time was needed and more patients could be included. This is in line with the current trend to use routinely collected or 'big' data to study trends and effects (e.g. 43), instead of performing randomized clinical trials (RCT). Especially within the field of information technology (IT), which is known to progress in a relatively rapid way, research methods based on big data gain popularity (4). One may wonder whether an RCT is still lacking in the current research line. Reviewing our results, we believe that an RCT would give us more insight in the clinical effects of TeleScreen. The disadvantage of an RCT is that it requires a tremendous amount of work and is therefore high in costs (44). An observational design (such as a before-and-after evaluation) (45) would be an alternative to study trends with regard to the cost-efficacy of TeleScreen, however this would be a less powerful design. For example it is impossible to produce hard evidence that TeleScreen is cost-effective, at most it could be an incentive for further research. So, if possible, an RCT would be preferable.

## Limitations

To perform adequate scientific research, the evaluation of the questionnaire should be done at a stable phase, meaning that questions of the questionnaire to be investigated should not change during the research period. This often conflicts with IT practices, where progress is often rapid and the evaluation often takes longer than the development of new applications (4). Apart from IT practices that are known to develop in a fast way, scientific research in general often includes gathering new knowledge and expertise during the research period. In our studies, not only the online application evolved quickly, there was also a transition from the DSM-IV-TR to the DSM-5, which had a considerable effect on TeleScreen. For example with regard to clinical disorders, several changes have been made (e.g. with regard to the posttraumatic stress-disorder, dysthymia and eating disorders) and

therefore continuation of the current research line by means of replication studies would be of interest. Although we aimed to choose the most efficient design, it still took time before the scientific paper was published and when published, data were outdated. The advantage, however, was that it was published in a peer-reviewed journal, which is known to improve the quality of the article (46) and research can be easily replicated. A solution to this research versus practice dilemma may be to use the original publication as a basis to explain the study methods and discussion of results. One could provide an update every six months with new test results, published as an appendix on the website of the journal or the website of TelePsy. This fits within the recommendation to establish a comprehensive scenario for such fields where progress is often rapid, with relatively stable overall frameworks, for which new data could be inserted subsequently (4).

Second, collaboration between companies and an academic institution could increase the chance of effective implementation of knowledge and improving the quality of healthcare (47), however the fact that the principal investigator was an embedded researcher also ensued some challenges. Vindrola-Padros et al. (2017) described several lessons to be learned from the literature with regard to embedded researchers (47). First, a connection with academia remains important to maintain a critical perspective. In addition, regular research meetings should be scheduled between the host institution and academic staff. Also the embedded researcher should consider experiences and points of view of subgroups within the organisation and should be reflective on his or her own position as an individual and part of organisational context. A way of dealing with such different roles is to set clear guidelines at the start to manage expectations. One could make use of the concordat of Brewster et al. (2017), for enabling partners in the evaluation to work together effectively (48). One of the main challenges is designing research that could potentially lead to negative results or highlight undesirable qualities of the organisation.

The advantage of the fact that the principal investigator was an embedded researcher was the possibility to understand the problems of the organisation and clinical practice. By using routinely collected data, an effective research design could be developed, which could also be easily embedded in the organisation promoting a sustainable research intervention. TeleScreen remains subject to further development and research. Once in a while validity research is therefore replicated. In addition, the development of components of the eHealth platform was based on the observed needs of psychologists, as described in chapter five. There are plans to organize focus groups with professionals and patients from different settings, to help in the development of the online components that would fit to the needs of end-users. Taking the literature into account, it would have been useful to use a concordat to manage expectations in advance, to define the role of the researcher, types of studies, study timeframes and feedback processes (48).

## Future research

Many patients with mental health disorders do not receive effective treatment on time, which could cause deterioration of the mental disorder and high healthcare costs (49-51). Unnecessary referrals between institutions, medication use, specialist treatments, and thus costs, may be avoided when disorders are recognized at an early phase and patients are directly referred to healthcare professionals best suited for the detected disorder. However, cost savings are not obvious and whether the system performs better than usual care has not been investigated. For example, TeleScreen may lead to overdiagnosis and assessment costs versus no costs of the GP making a preliminary diagnosis. Future research should therefore be focused on the cost-effectiveness of TeleScreen, when compared to care as usual, preferably through an RCT.

Second, in line with the above, we now assume that if the preliminary DSM classifications are correct, referrals will be correct as well. However, patients are not systematically monitored after their initial DSM classification. Therefore, we cannot evaluate whether the best possible result was achieved with the DSM classification, referral and treatment that followed. To assess the impact of TeleScreen, an RCT would be needed, to compare the intervention with care as usual. How many patients will still have the same or even worse complaints after one year? And if so, could we have foreseen that? Are patients better treated in a different echelon or with a different treatment? What kind of (blended) treatment is best for what kind of patient group? Currently, we are not monitoring mental healthcare. As a result, we do not know whether using eDiagnostics is really better and what kinds of (treatment) processes could be improved. Research may lead to significant improvements in healthcare, for both patient and healthcare professionals.

Third, this new way of eDiagnostics raises the issue whether there are ways to improve, change or adapt TeleScreen. For the development of online questionnaires, it remains a challenge to use clear language for patients and avoid healthcare-related terminology (52). Filling out TeleScreen could take 52 minutes (i.e. that was the mean time in a severe population of persons with personality disorders) and results are dependent on the comprehension of questions by patients (17). This may, however, be a challenge especially for the current group of patients with a suspicion of a mental disorder, who are known to suffer from concentration problems quite often. Also TeleScreen is not suitable for every patient, think of immigrants, persons who have difficulty with reading, or persons who find it difficult to understand information. Reducing administration time, improving accessibility and increased comprehension of questions might be achieved by using multimedia techniques, such as video, animations and images instead of, or next to text-based questions. Trimbos Institute, that is the Netherlands Institute of Mental Health and Addiction, developed a Visual Screener, for most anxiety, mood, alcohol-related disorders and suicidal thoughts (53). One study was published using this system, where agoraphobia was assessed though one question accompanied by an image (53). Using

multimedia techniques to assess mental disorders may improve accessibility among different patient groups and may improve comprehension of questions, which in turn may lead to better support for both professional and patient.

Another option to improve accessibility would be to develop different versions of TeleScreen, for different patient groups (e.g. different languages or different ages). Currently an informant version of TeleScreen Junior for children at ages 4 to 18 is available and a pilot study has been conducted to assess its validity (data not published). There are plans to validate both the informant version of TeleScreen Junior based on DSM-5 for the ages of 4 to 18, and the self-report version of TeleScreen for the ages of 12 to 18. An informant version of TeleScreen (adult version) would be of interest as well. This may especially be useful in the assessment of persons with limited introspective ability and limited reality testing capacities (for instance patients with autism spectrum disorders, psychotic disorders, personality disorders and bipolar disorder) (54, 55).

## References

1. Lobbetael J, Leurgans M, Arntz A. Inter-rater reliability of the structured clinical interview for DSM-IV Axis I disorders (SCID I) and Axis II disorders (SCID II). *Clin Psychol Psychother* 2011; 18: 75-9.
2. Widiger TA, Samuel DB. Evidence-based assessment of personality disorders. *Psychol Assess* 2005; 17: 278-87.
3. Nederlandse Vereniging voor Psychiatrie. *Vijftien veel gestelde vragen over DSM-5*. [https://www.boompsychologie.nl/actueel-item/80-1007\\_NVvP-beantwoordt-15-veelgestelde-vragen-over-de-DSM-5](https://www.boompsychologie.nl/actueel-item/80-1007_NVvP-beantwoordt-15-veelgestelde-vragen-over-de-DSM-5). Archived at: <http://www.webcitation.org/6wpYQK8Wg>, 2014
4. Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. *BMJ* 2002; 324: 477-480.
5. De Vet H, Terwee C, Mokkink L, Knol D. *Measurement in medicine: a practical guide*. New York: Cambridge University Press, 2011.
6. Ayearst LE, Bagby RM. Evaluating the psychometric properties of psychological measures. In M. M. Antony & D. H. Barlow (Eds.). *Handbook of assessment and treatment planning for psychological disorders*. New York: The Guilford Press, 2010.
7. Smits N, Smit F, Cuijpers P, De Graaf R. Using decision theory to derive optimal cut-off scores of screening instruments: an illustration explicating costs and benefits of mental health screening. *Int J Methods Psychiatr Res* 2007; 16: 219-29.
8. Zimmerman M, Mattia JI. A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. *Arch Gen Psychiatry* 2001; 58: 787-94.
9. Donker T, van Straten A, Marks I, Cuijpers P. A brief Web-based screening questionnaire for common mental disorders: development and validation. *J Med Internet Res* 2009; 11: e19.
10. Sheehan DV, Lecrubier Y, Sheehan KH *et al*. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;59: Suppl 20:22-33;quiz 34-57.
11. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders 4th edition, Text Revision (DSM-IV-TR)*: American Psychiatric Association, 2000.
12. Dozois DJ, Dobson KS. Depression. In M. M. Antony & D. H. Barlow (Eds.). *Handbook of assessment and treatment for psychological disorders (2nd ed.)*. New York: Press G, 2010.
13. Kraemer HC, Measelle JR, Ablow JC *et al*. A new approach to integrating data from multiple informants in psychiatric assessment and research: mixing and matching contexts and perspectives. *Am J Psychiatry* 2003; 160: 1566-77.
14. Stolper E, Van de Wiel M, Van Royen P *et al*. Gut feelings as a third track in general practitioners' diagnostic reasoning. *J Gen Internal Med* 2011; 26: 197-203.
15. Hafkenscheid A, van Os J. ROM van geïndividualiseerde behandeldoelen. *Psychopraktijk* 2014; 6: 29-32.
16. Lucassen P, Postma S, olde Hartman T. Ggz-vragenlijsten staan persoonsgerichte zorg in de weg. *Huisarts Wet* 2017; 60: 112-4.
17. Chan D. So why ask me? Are self-report data really that bad. In: CE Lance RJ Vandenberg. *Statistical and Methodological Myths and Urban Legends. Doctrine, Verity and Fable in the Organizational and Social Sciences*. New York, NY: Taylor & Francis Group, 2009
18. Hufford MR. Special methodological challenges and opportunities in ecological momentary assessment. In: Stone AA, Shiffman S, Atienza A, Nebeling L. *The science of real-time data capture: Self-reports in health research*. New York: Oxford University Press, 2007.
19. Verhagen SJW, Berben JA, Leue C *et al*. Demonstrating the reliability of transdiagnostic mHealth Routine Outcome Monitoring in mental health services using experience sampling technology. *PLoS One* 2017; 12: e0186294.
20. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. *Med J Aust* 2004; 15: 180: S57-60.
21. Vogt F, Armstrong D, Marteau TM. General practitioners' perceptions of the effectiveness of medical interventions: an exploration of underlying constructs. *Implement Sci* 2010; 5: 17.

22. Chaudhry B, Wang J, Wu S et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006; 16; 144: 742-52.
23. van Gemert-Pijnen JE, Nijland N, van Limburg M et al. A holistic framework to improve the uptake and impact of eHealth technologies. *Journal Med Internet Res* 2011; 13: e111.
24. Emmelkamp P. Technological innovations in clinical assessment and psychotherapy. *Psychother Psychosom* 2005; 74: 336-43.
25. Kelders SM, van Gemert-Pijnen JE, Werkman A, Seydel ER. Usage and effect of a web-based intervention for the prevention of overweight; a RCT. *Stud Health Technol Inform* 2010; 160: 28-32.
26. Eysenbach G. The law of attrition. *Journal Med Internet Res* 2005; 31; 7: e11.
27. Thombs BD, Coyne JC, Cuijpers P et al. Rethinking recommendations for screening for depression in primary care. *CMAJ* 2012; 184: 413-8.
28. Dijkman I, Dinant GJ, Spigt M. eDiagnostics: a promising step towards primary mental health care. *Fam Pract* 2013; 30: 695-704.
29. Gidding L, Spigt M, Brouwer E et al. Experiences of general practitioners, practice nurses, psychologists and patients with stepped collaborative care for depression: a focus group study in a large primary healthcare organization. *Eur J Pers Cent Healthc* 2014; 2; 170-178.
30. Dijkman I, Dinant GJ, Spigt M. The Perception and Needs of Psychologists Toward Blended Care. *Telemed J E Health* 23; 983-995.
31. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems* 2009; 24: 28; 485-500.
32. Fogg BJ. *Persuasive technology: Using computers to change what we think and do*. Boston: Morgan Kaufmann Publishers, 2003.
33. Mohr DC, Duffecy J, Jin L et al. Multimodal e-mental health treatment for depression: a feasibility trial. *J Med Internet Res* 2010; 19; 12: e48.
34. Guyatt G, Sackett D, Haynes B. Evaluating diagnostic tests. In Haynes RB, Sackett DL, Guyatt GH, Tugwell P. *Clinical epidemiology: How to do clinical practice research*, Philadelphia:Lippincott, Williams & Wilkins, 2006.
35. Nguyen DP, Klein B, Meyer D, Austin DW, Abbott JA. The Diagnostic Validity and Reliability of an Internet-Based Clinical Assessment Program for Mental Disorders. *J Med Internet Res* 2015; 21; 17:e218.
36. Verhaak PF, van Dijk CE, Nuijen J, Verheij RA, Schellevis FG. Mental health care as delivered by Dutch general practitioners between 2004 and 2008. *Scand J Prim Health Care* 2012; 30: 156-62.
37. Dijkman I, Dinant GJ, Spigt M. The concurrent validity of a new eDiagnostic system for mental disorders in primary care. *Fam Pract* 2016; 33: 607-16.
38. Verhaak PF, van den Brink-Muinen A, Bensing JM, Gask L. Demand and supply for psychological help in general practice in different European countries: access to primary mental health care in six European countries. *Eur J Public Health* 2004; 14: 134-40.
39. Kulu-Glasgow I, Delnoij D, de Bakker D. Self-referral in a gatekeeping system: patients' reasons for skipping the general-practitioner. *Health Policy* 1998; 45: 221-38.
40. Santana S, Lausen B, Bujnowska-Fedak M et al. Online communication between doctors and patients in Europe: status and perspectives. *J Med Internet Res* 2010; 15; 12: e20.
41. van der Vaart R, Atema V, Evers AW. Guided online self-management interventions in primary care: a survey on use, facilitators, and barriers. *BMC Fam Pract* 2016; 09; 17: 27
42. Tyrer P, Reed GM, Crawford MJ. Classification, assessment, prevalence, and effect of personality disorder. *Lancet* 2015; 21; 385: 717-26.
43. Markowitz A, Blazskiewicz K, Montag C, Switala C, Schlaepfer TE. Psycho-informatics: Big Data shaping modern psychometrics. *Med hypotheses* 2014; 82: 405-11.
44. Feinstein AR. An additional basic science for clinical medicine: II. The limitations of randomized trials. *Ann Intern Med* 1983; 99: 544-50.
45. Guyatt GH, Tugwell PX, Feeny DH, Drummond MF, Haynes RB. The role of before-after studies of therapeutic impact in the evaluation of diagnostic technologies. *J Chronic Dis* 1986; 39: 295-304.
46. Spigt M, Arts IC. How to review a manuscript. *J Clinical Epidemiol* 2010; 63: 1385-90.

47. Vindrola-Padros C, Pape T, Utley M, Fulop NJ. The role of embedded research in quality improvement: a narrative review. *BMJ Qual Saf* 2017; 26:70-80.
48. Brewster L, Aveling EL, Martin G, Tarrant C, Dixon-Woods M. What to expect when you're evaluating healthcare improvement: a concordat approach to managing collaboration and uncomfortable realities. *BMJ Qual Saf* 2015; 24: 318-24.
49. Rost K, Zhang M, Fortney J *et al.* Persistently poor outcomes of undetected major depression in primary care. *Gen Hosp Psychiatry* 1998; 20: 12-20.
50. Essau CA. Course and outcome of major depressive disorder in non-referred adolescents. *J Affect Disord* 2007; 99: 191-201.
51. Essau CA. Course and outcome of somatoform disorders in non-referred adolescents. *Psychosomatics* 2007; 48: 502-9.
52. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient Educ Couns* 2006; 61: 173-90.
53. van Ballegooijen W, Riper H, Donker T *et al.* Single-item screening for agoraphobic symptoms: validation of a web-based audiovisual screening instrument. *PLoS One* 2012;7: e38480.
54. Kessler D, Lloyd K, Lewis G, Gray DP. Cross sectional study of symptom attribution and recognition of depression and anxiety in primary care. *BMJ* 1999; 318: 436-9.
55. Jorm AF. Mental health literacy: public knowledge and beliefs about mental disorders. *The Br J Psychiatry* 2000; 177: 396-401.



## Summary



**Chapter 1** starts with an introduction of the setting of the studies presented in this thesis, that is mental disorders in primary care. We first discuss the importance of (early) recognition of mental disorders, in which the GP plays a crucial role. Next, a description of the Dutch Mental Healthcare setting is described. Most importantly, a referral from the GP to a reimbursed mental healthcare trajectory must nowadays always be accompanied by a suspected DSM-IV-TR (or DSM-5) disorder. Hereafter, we describe eMental health and the use of online screening questionnaires in general practice. The terms diagnosis and classification are explained. We elaborate on the most important evidence-based instruments that can be used in (Dutch) GP practices. Among others, an eDiagnostic system called TeleScreen was described, which forms the core instrument of the studies presented in this thesis. TeleScreen can assist (healthcare) professionals in the diagnostics and classification of mental disorders, which may be used to perform extensive triage at the very start of a patient's trajectory. First, experiences of patients, general practitioners (GPs) and practice nurses (PNs) with regard to TeleScreen were examined. Second, the validity of TeleScreen in both primary care and specialized mental health care settings were studied. Finally, perceptions, wishes and barriers of psychologists toward blended care were examined.

The following research questions are addressed in this thesis:

1. What are experiences of patients, GPs and PNs with regard to TeleScreen?
2. What is the concurrent validity of TeleScreen when compared with the psychologists' judgments, in patients presenting in primary care, with a possible mental disorder?
3. What is the concurrent validity of TeleScreen, when compared with the results of the Structural Clinical Interview for DSM-IV-TR Axis-I (SCID-I) and Axis-II disorders (SCID-II), for patients with personality disorders?
4. What are the perceptions of psychologists, and their wishes and barriers towards blended care?

**Chapter 2** describes an evaluation study among service users (patients, GPs and PNs) of TeleScreen. We also compared the echelon advice of GPs and PNs that was given with the referral to TeleScreen to the echelon advice generated by TeleScreen. In total 252 patients (69% response rate) and 49 professionals (GPs and PNs) responded (72% response rate per practice). Results showed that patients showed no clear preference for a face-to-face consultation with a psychologist over an eDiagnostic system. GPs and PNs felt strengthened in their control function. In most cases, the system gave a different echelon advice than the GPs and PNs. Nevertheless, GPs and PNs accept the results of the examination and the advice given. It can be concluded that using the Internet to classify mental health disorders in primary care seems promising. Patients were quite satisfied with TeleScreen and GPs and PNs were very satisfied with such a system for their patients. This system, where eDiagnostics is used before a referral to a mental healthcare

institution instead of afterwards, is new and could change the management of mental health care.

**Chapter 3a and chapter 3b** describe the validity of TeleScreen when compared with the psychologists' judgments, in patients presenting in primary care, with a possible mental disorder. In Chapter 3a data of 675 primary care patients in the Netherlands were included in the analysis. Validity was examined using sensitivity, specificity, positive predictive values, negative predictive values and Cohen's kappa values. Furthermore, 100 patient records were randomly selected to investigate the inter-rater reliability among psychologists.

Generally, results showed that the inter-rater agreement between psychologists was fair to good or excellent. TeleScreen could be used to correctly estimate the echelon advice and correctly identify and exclude most clinical disorders, such as mood, anxiety and substance-related disorders. TeleScreen was less accurate in identifying patients with Asperger's disorder, sexual disorder and adjustment disorder. For several classifications the sensitivities and specificities were good, but the PPVs were rather low. The system was able to determine the absence of a personality disorder and the sensitivities and specificities for most specific personality disorders were good. Still, the PPVs for several specific Axis II classifications were rather low. Finally, the system was inaccurate in identifying the global assessment of functioning (Axis V of the DSM-IV-TR) of the patient.

In **Chapter 3b** data were updated. We examined the validity of TeleScreen by comparing routinely collected results of 496 patients from Limburg GP practices with the judgment of a psychologist. Results show that TeleScreen can correctly recognize, exclude and predict various clinical disorders. It was able to adequately determine the absence of an Axis II disorder and it gave a correct advice for the echelon. However, for a number of mental disorders, the system generated false positives and false negatives and it was inaccurate in assessing the general assessment of functioning (Axis V of the DSM-IV-TR).

In conclusion, TeleScreen can be seen as a valid screening instrument in general practice for most DSM-IV-TR classifications. However, for several classifications the system generates too many false positives or false negatives and therefore TeleScreen cannot, at this point of time, replace clinical assessment. A healthcare professional (in our studies a psychologist) is needed to interpret results and reduce the amount of false positives or false negatives.

In **Chapter 4**, we describe the concurrent validity of TeleScreen when compared with well-validated structured interviews for personality disorders and their co-occurring clinical disorders, specifically the SCID-I and SCID-II. A total of 89 patients referred for specialized treatment of personality pathology filled out TeleScreen, and mental health problems were also independently assessed through a SCID-I and SCID-II interview. The concurrent validity was examined using sensitivity, specificity, PPVs, NPVs and Cohen's Kappa. TeleScreen showed moderate to good validity for any personality disorder, borderline

personality disorder, avoidant personality disorder and obsessive-compulsive personality disorder with the exception of the dependent and paranoid personality disorder. Clinical disorders also showed moderate to good values, except for social phobia, dysthymia and eating disorders. Overall, it can be concluded that TeleScreen appears to be a valid source of information for professionals in the assessment and classification of DSM-IV-TR disorders for individuals with personality disorders.

Blended care, meaning that online (internet) components are used in combination with face-to-face contact, in mental health is increasingly encouraged, but research about the needs of psychologists is scarce. **Chapter 5** aimed to study the perceptions of psychologists and barriers with regard to blended care. In addition, specific wishes for the way blended care should be designed were investigated. Opinions of psychologists of the Dutch Association of Psychologists were collected through an e-mailed survey. Results showed that psychologists from different healthcare settings (63% response rate) generally had a positive perception towards blended care and that they intended to use it in the future. Users of blended care and secondary care professionals were more positive towards blended care than nonusers and primary care professionals. Online psychoeducation, diary forms, exercises of different therapeutic approaches and communication technology configurations were most welcomed. Still, quite some barriers were mentioned before professionals would use blended care. For instance, in order to help professionals in using blended care, more information about the online components, perceptions of the patients and legal issues or liability must be provided by eHealth providers. It could be concluded that psychologists from different healthcare settings had a positive perception towards blended care, as long as attention is paid to perceived barriers. Reviewing the needs of psychologists, eHealth components of different therapeutic approaches apart from cognitive behavioral therapy were welcomed. The results of this survey can be particularly useful in the development of online programs that correspond to the needs of psychologists. Future research is necessary to gain insight in the (cost)effectiveness of blended care for different types of patients (e.g. transdiagnostic interventions) and of different therapeutic approaches.

**Chapter 6** is the general discussion of this thesis. This final chapter starts with the major findings of the studies presented in this thesis. Next, we discuss the methodological considerations, in particular the type of reference standard to use to determine the validity of TeleScreen. Hereafter, acceptable validity was discussed. There are no uniform criteria to decide whether an instrument is valid for its clinical purpose in a given situation. The choice between high sensitivity or high specificity is described per situation. Next, different steps of the diagnostic process are discussed. This made it possible to discuss at which points improvements can be made. Hereafter, improvements of TeleScreen are described and recommendations are provided for professionals on how to use TeleScreen. Professionals that use the instrument should for instance be aware that it is a self-report questionnaire which has its limitations. Next, successful implementation of new

innovations in (mental) healthcare are discussed and later specifically applied to TeleScreen. Within the field there is also resistance from professionals with regard to online questionnaires. The main reasons are described, along with observed features of TeleScreen, which may have helped in the positive perception of service users. More information about the company TelePsy is given, that developed TeleScreen and hereafter gradually evolved into an eHealth platform. Next, the chapter describes a developed framework of van Gemert-Pijnen et al. (2011) that can be used during the development of eHealth components. Regarding strengths and limitations, one strength was that we conducted our studies in real-world care settings with real patient populations, which contributes to the external validity of the studies and the relevance and practical clinical value of the research. In addition, our studies were performed in an efficient way; the majority of studies in this thesis were observational in nature. Limitations with regard to observational designs and the question whether a randomized controlled trial should have been conducted are discussed. In addition, the transition from DSM-IV-TR to DSM-V and the impact on our research are mentioned. Also, both advantages and disadvantages from the fact that the principal investigator was an embedded researcher of the company TelePsy were discussed. The chapter ends with recommendations with regard to future research.







## Samenvatting



**Hoofdstuk 1** start met een beschrijving van de setting waar de studies zoals beschreven in deze dissertatie zich hebben afgespeeld, namelijk psychische stoornissen binnen de eerstelijns huisartsenzorg. Eerst wordt dieper ingegaan op het belang van (vroegtijdige) herkenning van psychische stoornissen, waarbij de huisarts een cruciale rol vervult. Daarna wordt de setting, Nederlandse Geestelijke Gezondheidszorg (GGZ) beschreven. Een belangrijk punt bijvoorbeeld is dat een verwijzing van de huisarts tegenwoordig noodzakelijk is voor een vergoed GGZ traject en vergezeld moet gaan met een vermoeden van een DSM-benoemde stoornis. Hierna wordt eMental health beschreven en het gebruik van screeningsvragenlijsten binnen de eerstelijns huisartsenzorg. De termen diagnose en classificatie worden beschreven. Vervolgens worden de meest belangrijke wetenschappelijk onderbouwde instrumenten beschreven die gebruikt worden binnen de (Nederlandse) eerstelijns huisartsenzorg. Onder andere is er een eDiagnostiek systeem ontwikkeld, genaamd TeleScreen, hetgeen het voornaamste instrument is dat onderzocht is in deze dissertatie. TeleScreen kan (gezondheidszorg) professionals ondersteunen bij de diagnostiek en classificatie van psychische stoornissen, hetgeen gebruikt kan worden om uitvoerige triage toe te passen aan het begin van een mogelijk GGZ traject. Eerst zijn ervaringen van patiënten, huisartsen en praktijkondersteuners GGZ (POH-GGZ) ten opzichte van TeleScreen onderzocht. Ten tweede werd de validiteit van TeleScreen onderzocht in zowel een eerstelijns huisartsensetting als een gespecialiseerde GGZ setting. Tenslotte is dieper ingegaan op de houding, wensen en barrières van psychologen ten opzichte van blended care.

De volgende onderzoeksvragen zijn aan bod gekomen in dit proefschrift:

1. Wat zijn de ervaringen van patiënten, huisartsen en POH-GGZ met TeleScreen?
2. Wat is de concurrente validiteit van TeleScreen indien vergeleken met het oordeel van een psycholoog bij patiënten in de eerstelijns huisartsenzorg met een vermoeden van een psychische stoornis?
3. Wat is de concurrente validiteit van TeleScreen, indien vergeleken met resultaten van een gestructureerd klinisch interview voor DSM-IV-TR As I stoornissen (SCID-I) en As II stoornissen (SCID-II) bij patiënten met persoonlijkheidsstoornissen?
4. Wat zijn de houdingen, wensen en barrières van psychologen ten opzichte van blended care?

**Hoofdstuk 2** beschrijft een evaluatiestudie onder gebruikers (patiënten, huisartsen en POH-GGZ) van TeleScreen. Naast de evaluatie werd het echelonadvies van de huisarts en POH-GGZ, hetgeen gegeven is voorafgaand aan de uitkomsten van TeleScreen, vergeleken met die van TeleScreen. In totaal hebben 252 patiënten (antwoordpercentage 69%) en 49 professionals (huisartsen en POH-GGZ) gereageerd (antwoordpercentage per praktijk 72%). Resultaten tonen aan dat patiënten geen duidelijke voorkeur hebben voor een face-to-face consult met een psycholoog boven het eDiagnostiek systeem. Huisartsen en POH-GGZ voelen zich versterkt in hun regiefunctie. In de meeste gevallen geven huisartsen en POH-GGZ een ander echelonadvies dan het eDiagnostiek systeem. Desondanks

kunnen huisartsen en POH-GGZ zich wel vinden in de resultaten en het gegeven advies. Het kan geconcludeerd worden dat het gebruik van Internet om psychische stoornissen te classificeren veelbelovend is. Dit systeem, waarbij eDiagnostiek gebruikt wordt, voorafgaand in plaats van na een verwijzing naar een GGZ-instelling is nieuw en kan het management van zorg veranderen.

In **hoofdstuk 3a en 3b** zijn we dieper ingegaan op de validiteit van TeleScreen indien vergeleken met het oordeel van een psycholoog, bij patiënten in de eerstelijns huisartsenzorg met een vermoeden van een psychische stoornis. In hoofdstuk 3a zijn gegevens van 675 patiënten binnen de Nederlandse eerstelijns huisartsenzorg geïncludeerd in de analyse. De validiteit werd onderzocht door middel van sensitiviteit, specificiteit, positief voorspellende waarden (PVW), negatief voorspellende waarden (NVW) en Cohen's kappa waarden. Daarnaast werden 100 patiëntendossiers at random geselecteerd om de interbeoordelaarsbetrouwbaarheid onder de psychologen te onderzoeken.

De interbeoordelaarsbetrouwbaarheid tussen de psychologen is in het algemeen redelijk tot goed of excellent. TeleScreen kan het echelonadvies correct inschatten en de meeste klinische stoornissen correct herkennen en uitsluiten, zoals verschillende stemming, angst en aan middelen gebonden stoornissen. TeleScreen kan de stoornis van Asperger, seksuele stoornis en aanpassingsstoornis minder goed identificeren. Voor verschillende classificaties is de sensitiviteit en specificiteit goed, maar de PVW laag. Het systeem kan de afwezigheid van een persoonlijkheidsstoornis bepalen en de sensitiviteit en specificiteit van de meeste specifieke persoonlijkheidsstoornissen is goed. Echter, de PVW van verschillende specifieke As II classificaties is laag. Tot slot is het systeem minder goed in het correct bepalen van As V van de DSM-IV (global assessment of functioning; GAF-score) van de patiënt indien vergeleken met het oordeel van de psychologen.

In **hoofdstuk 3b** heeft er een replicatieonderzoek plaatsgevonden. Wederom is de validiteit van TeleScreen onderzocht, waarbij routinematig verzamelde gegevens van 496 patiënten uit huisartsenpraktijken in Limburg, vergeleken zijn met het oordeel van de psycholoog. TeleScreen kan diverse klinische stoornissen correct herkennen, uitsluiten en voorspellen. Het is mogelijk om de afwezigheid van een As II stoornis te bepalen en het geeft een juiste verwijzing voor het echelon. Echter, voor een aantal psychische stoornissen genereert het systeem teveel fout positieve en fout negatieve classificaties en daarom kan TeleScreen op dit moment de klinische beoordeling niet vervangen. Een zorgprofessional (in deze studies een psycholoog), is nodig om de resultaten te interpreteren en het aantal fout positieven en fout negatieven te reduceren.

In **hoofdstuk 4** wordt de concurrente validiteit in kaart gebracht door TeleScreen te vergelijken met een gestructureerd interview om persoonlijkheidsstoornissen en klinische stoornissen in kaart te brengen, namelijk de SCID-I en SCID-II. In totaal vulden 89 patiënten die werden verwezen voor specialistische behandeling van persoonlijkheidsproblematiek TeleScreen in en werden klachten onafhankelijk middels een SCID-I en SCID-II

interview in kaart gebracht. De concurrente validiteit werd onderzocht door middel van sensitiviteit, specificiteit, PVW, NVW en Cohens Kappa. TeleScreen toont matig tot goede validiteit voor enigerlei persoonlijkheidsstoornis, een borderline persoonlijkheidsstoornis, ontwijkende en obsessief-compulsieve persoonlijkheidsstoornis, met uitzondering van de afhankelijke en paranoïde persoonlijkheidsstoornis. Klinische stoornissen tonen eveneens matig tot goede validiteit, behalve voor de sociale fobie, dysthymie en eetstoornissen. In het algemeen kan hieruit geconcludeerd worden dat TeleScreen een valide bron van informatie is voor professionals bij het in kaart brengen en classificeren van DSM-IV-TR stoornissen bij personen met persoonlijkheidsstoornissen.

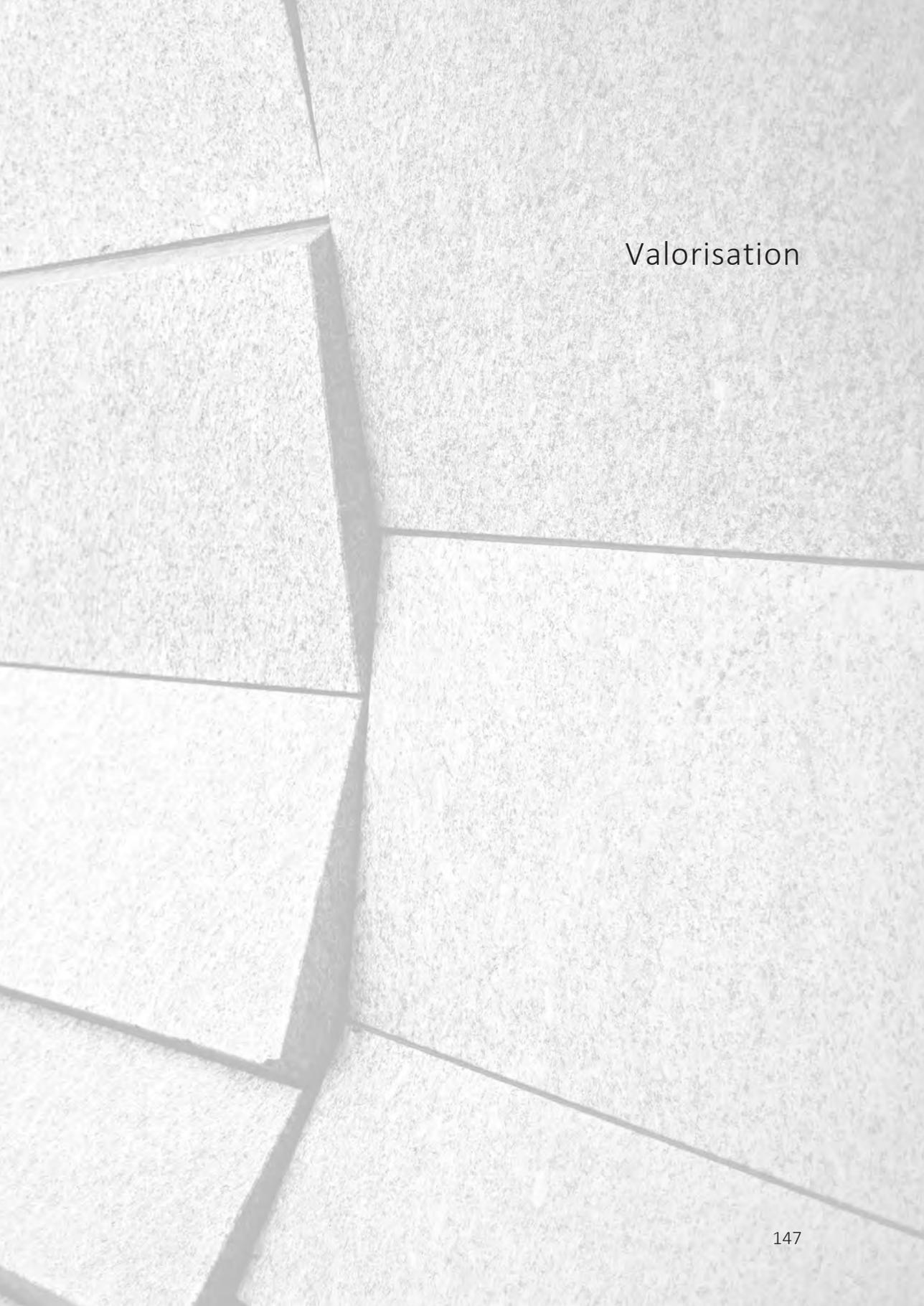
Blended care, dat wil zeggen dat online (internet) componenten gebruikt worden in combinatie met face-to-face contact, in de GGZ wordt in toenemende mate aangemoedigd, maar onderzoek naar de behoeften van psychologen ontbreekt. **Hoofdstuk 5** heeft als doel om de houdingen van psychologen, barrières en specifieke wensen ten aanzien van blended care in kaart te brengen. Meningingen van psychologen die aangesloten zijn bij het Nederlandse Instituut van Psychologen (NIP) werden verzameld middels een vragenlijst die per e-mail werd verstuurd. Psychologen uit verschillende GGZ takken (63% antwoordpercentage) hebben in het algemeen een positieve houding ten opzichte van blended care en zij hebben de intentie om blended care in de toekomst te gebruiken. Gebruikers van blended care en mensen uit de specialistische GGZ zijn positiever ten opzichte van blended care dan niet-gebruikers en psychologen uit de generalistische basis GGZ. Online psycho-educatie, dagboekformulieren, oefeningen van verschillende therapeutische stromingen en communicatie methoden, zoals e-mail en mobile technologie, zijn het meest welkom. Verschillende barrières worden genoemd en voorbereidende activiteiten worden echter noodzakelijk alvorens professionals een nieuw blended care platform zouden gaan gebruiken. Bijvoorbeeld, om professionals te helpen bij het gebruik van blended care is het nodig dat er meer duidelijkheid komt omtrent de inhoud van de online componenten, de attitudes van de patiënten en juridische aspecten en aansprakelijkheid, bij voorkeur aangeboden door eHealth providers. Het kan geconcludeerd worden dat psychologen van verschillende GGZ takken een positieve houding hebben ten opzichte van blended care, zolang aandacht wordt besteed aan de waargenomen barrières. Als we kijken naar de behoeften van psychologen, dan zijn eHealth componenten van verschillende therapeutische stromingen naast cognitieve gedragstherapie welkom. De resultaten van dit onderzoek kunnen bruikbaar zijn voor de ontwikkeling van online programma's die overeenkomen met de behoeften van psychologen. Toekomstig onderzoek is nodig om meer inzicht te krijgen in de (kosten)effectiviteit van blended care voor verschillende typen patiënten (bijvoorbeeld transdiagnostische interventies) en voor verschillende therapeutische stromingen.

**Hoofdstuk 6** is de algemene discussie van dit proefschrift. Het laatste hoofdstuk start met de belangrijkste bevindingen van de studies die gepresenteerd worden in deze dissertatie. Hierna wordt dieper ingegaan op de methodologische afwegingen, in het bijzonder

welk type referentiestandaard te gebruiken om de validiteit van TeleScreen te bepalen. Voorts wordt acceptabele validiteit bediscuteerd. Er zijn geen uniforme criteria om te beslissen of een instrument valide is voor klinische doeleinden in een specifieke situatie. De keuze tussen hoge sensitiviteit of hoge specificiteit wordt beschreven per situatie. Dan worden de verschillende stappen binnen het diagnostische proces beschreven. Dit maakt het mogelijk om te kijken op welke punten verbeteringen gemaakt kunnen worden. De verbeteringen omtrent TeleScreen worden beschreven en er worden aanbevelingen gedaan voor professionals voor het gebruik van TeleScreen. Professionals die het instrument gebruiken moeten er bijvoorbeeld van bewust zijn dat het gaat om een zelfrapportagevragenlijst, hetgeen beperkingen met zich meebrengt. Hierna wordt dieper ingegaan op succesvolle implementatie van innovaties binnen de (geestelijke) gezondheidszorg. Binnen de zorg is er ook weerstand van professionals ten opzichte van online vragenlijsten. De belangrijkste redenen worden beschreven, samen met eigenschappen van TeleScreen die er vermoedelijk voor hebben gezorgd dat gebruikers positief zijn. Hierna wordt meer informatie over het bedrijf TelePsy gegeven, het bedrijf dat TeleScreen heeft ontwikkeld en dat daarna gradueel is ontwikkeld in een eHealth platform. Vervolgens zijn we in het hoofdstuk ingegaan op het theoretische denkkader van Gemert-Pijnen et al. (2011), hetgeen gebruikt kan worden tijdens de ontwikkeling van eHealth componenten. De sterke kanten en beperkingen van de studies in dit proefschrift zijn vervolgens beschreven. Een sterk punt is bijvoorbeeld dat de studies zijn uitgevoerd binnen een gezondheidszorg setting met een bestaande patiëntenpopulatie, hetgeen bijdraagt aan de externe validiteit van de studies, de relevantie en praktische toepasbaarheid. Daarnaast zijn de studies uitgevoerd op een efficiënte manier; de meeste studies waren observatief van aard. De beperkingen die dit met zich meebrengt en de vraag of er geen randomized controlled trial zou moeten hebben plaatsgevonden zijn eveneens besproken. Daarnaast zijn de beperkingen van de studies omschreven. Bijvoorbeeld de transitie van de DSM-IV-TR naar de DSM-V en de impact die dat heeft gehad op het onderzoek. Daarnaast zijn zowel voor als nadelen beschreven van het feit dat de onderzoekster een medewerker was van het bedrijf TelePsy. Het hoofdstuk sluit af met aanbevelingen voor toekomstig onderzoek.







## Valorisation



This valorisation chapter elaborates on the societal value of the research, where the acquired scientific knowledge is translated into knowledge that could be used or is suitable for a broader public than the scientific community only. The chapter describes the relevance for clinical practice and elaborates on the innovative character of the findings. Next, different stakeholders for whom the results of the studies can be considered are described. Finally, we look at new activities and further dissemination of the acquired knowledge.

### *Relevance for clinical practice*

Mental disorders are very common. In the Netherlands approximately 1.896.700 people suffer from a mental disorder per year (18%) [1]. Mental disorders are associated with high burden of disease [2], reduced quality of life, work disability and high societal costs [3-5]. Research showed that mental disorders can be missed or misinterpreted, for instance due to unfamiliarity of GPs with psychological issues or patients' feeling of shame to express complaints [6-11]. Such pitfalls can enhance the under-recognition of mental issues, which in turn could result in delay of receiving adequate care, incorrect referrals or ineffective treatment [7, 12-14]. Early assessment seems necessary to identify mental disorders at an early stage, to avoid aggravation of complaints and to make sure that patients receive the appropriate treatment. With TeleScreen extensive triage at the very start of a patient's trajectory is possible.

In 2014 the Dutch government stimulated a shift from specialized mental healthcare to primary care and general practice with the practice nurse, specialized in mental healthcare (PN). Since then, only patients with a suspicion of a mental disorder based on the DSM-IV-TR (or DSM-5) are allowed to be referred to (reimbursed) mental healthcare; that is a primary care psychologist or specialized mental healthcare worker [15]. Since then, a referral from the GP must always be accompanied with a suspected diagnosis from DSM-IV-TR (or DSM-5). The aim of the transition was to avoid expensive specialized mental healthcare; treating patients quickly and effectively and avoiding unnecessary specialized treatments. GPs play a central role in this system. GPs act as gatekeepers to (specialized) mental healthcare. However, they often feel uncertain or are unfamiliar with psychological interventions, which may result in incorrect treatment, incorrect referrals and long waiting lists [7]. TeleScreen, which was at the center of the studies presented in this thesis, is developed to support GPs in this difficult task.

### *Private public partnership*

The collaboration between TelePsy and Maastricht University has been very fruitful. In developing and selling their applications TelePsy needs scientific research. Healthcare organisations that consider using TelePsy eDiagnostics, want proof of the validity of the system. To be credible, the validity must be investigated by an independent scientific

institute. The careful scientific analysis of the validity of the system has led to much more insight into the strengths and weaknesses of TeleScreen, which made it possible to specifically adjust the system where needed.

Maastricht University also greatly benefited from the collaboration. The main advantage, which could be felt on a daily basis, was that data collection was guaranteed. Many academic PhD trajectories are challenged by problems with data collection. The data were retrieved from ongoing tests that were routinely done. If we needed more patients in the analysis, all we had to do was wait a few more weeks. We mostly used routine care data or online questionnaires among service users, so that data could be gathered relatively easily. The burden of administration for participants, clinicians and researchers was therefore minimized; less time was needed and more patients could be included.

Another benefit from the collaboration between TelePsy and Maastricht University was the guaranteed utilisation. The system that was investigated was instantly used in routine care. GPs and (specialized) mental healthcare institutions in the Netherlands, but also in Germany, Belgium, and the United Kingdom (UK) increasingly use tools that TelePsy develops. Research served to develop the tools. This is in large contrast with healthcare innovations that are primarily and solely initiated by academia. All too often the use of the innovation stops when the scientific evaluation is completed.

TeleScreen was studied in a real patient population with treatment seeking patients, which increased the innovative character of the studies. Often, research is conducted for example after patients are recruited via internet banners or visitors that entered an online website to fill out the questionnaire [16, 17], allowing certain types of patients to be selected (e.g. the ones who are motivated or interested in eHealth). TeleScreen was examined both within general practice and within specialized mental healthcare. This is unique and increases the added value of the current research.

### *Stakeholders*

There are different stakeholders for whom the results of the studies can be considered relevant, these are: patients, healthcare professionals, insurance companies and scientists.

*Patients.* Based on the findings of the evaluation study, the opinion of patients with regard to the e-tool was investigated [18]. We conclude that patients were quite satisfied with the system (e.g. mean score = 7.5). They felt understood, agreed with the results of the diagnostic assessment and there was no clear preference for a face-to-face consultation with a psychologist over the eDiagnostic system [18]. One of the advantages for patients from a societal perspective is that patients get more insight in their complaints and their treatment options at the start of a (possible) healthcare trajectory. Potentially, patients are directly sent to healthcare professionals best suited for their problem. Eventually, this may lead to more effective care and a reduction in waiting time.

Diagnostic assessment often takes several consultations, and could therefore only be done at (specialized) healthcare institutions. By introducing this online system, assessment is more easily accessible. eDiagnostics can be performed within general practice and can be completed by patients at home. There are no geographical restrictions and patients do not have to travel for the completion of the test [19]. In an early phase, patients can now know what to expect. Early assessment can avoid a harmful or ineffective treatment, for instance focusing on clinical disorders, while a treatment for a personality disorder may be necessary [20].

However, TeleScreen is not suitable for every patient. Think of patients with insufficient understanding of the Dutch language, a cognitive impairment, or severely limited reality testing capabilities [18, 21]. Also, not every patient may be comfortable with expressing their problems in writing [22]. The eDiagnostic tool is mainly developed to advise and support the healthcare professionals about their patients. The results of TeleScreen are directly sent to the healthcare professional, instead of the patients. Despite the benefits for the professional, it means that the control remains within the healthcare professional. There is no direct feedback moment for patients, so patients cannot directly use the outcomes to prepare themselves for a consultation. This may be a point of discussion or improvement of the system in future.

*Healthcare professionals.* TeleScreen was initially developed by TelePsy as a system for GPs and PNs, to perform diagnostics before referral to a mental healthcare institution. The idea behind the tool was to make sure that patients were directly sent to the institution best equipped for the patient's problem. After a while, also (specialized) mental healthcare institutions started to use the system, for instance for internal referral or to know what kind of patients were on their waiting lists. When patients are directly sent to the institution best equipped for the problem, waiting lists and thus aggravation of complaints or costs may be reduced or prevented. The innovation also has a practical benefit. The eDiagnostic tool is easily accessible for professionals. Within a short timeframe, the report can be accessible for professionals and patients. By using the consultation function that TelePsy offers, GPs do not need to interpret the questionnaires themselves or change their behaviour. They receive additional information about their patients, which strengthens their gatekeepers function.

*Insurance companies.* Insurance companies may also benefit from the conducted research and tools that TelePsy develops. When patients are directly sent to healthcare institutions best suited for the problem, waiting lists and costs may eventually be reduced. Also, due to its scalability, eMental health has the potential to reach a large audience [19]. Therapeutic face-to-face interventions are often still necessary [23], but may be more time-efficient as the therapist can build on results that are filled out in advance. Possibly the number of face-to-face sessions will be reduced, while the clinical effects are

maintained [24-26]. This may improve the cost-effectiveness of treatment and may reduce the healthcare costs.

*Scientists.* TeleScreen is one of the first comprehensive eDiagnostic systems, covering most axes of the DSM-IV-TR (or DSM-5), in the Netherlands or even worldwide. The careful scientific analysis of the validity of the system has led to much more insight into the strengths and weaknesses of such a system, which made it possible to specifically adjust the system where needed [21, 27]. Also, we gained knowledge when it comes to the perception and needs of psychologists with regard to blended care [28]. Dissemination within international literature has taken place because the articles have been published in peer-reviewed international journals and there was also worldwide attention for TeleScreen (see: Reuters Health [29]). This makes implementation of the innovation easier in other markets, such as the UK.

### *New activities and further dissemination*

*General Practice.* TeleScreen is used by approximately 70% of GPs in the South of the Netherlands (Limburg). From 2014 onwards, the consultation has been reimbursed by healthcare insurance companies. Since the communication between patient and healthcare professional is largely offered through the online platform and via a telephone call, there are no geographical restrictions. Therefore, also GPs outside Limburg could ask consultation of TelePsy for help. For further implementation, PR and communication is important. It is important that care providers are aware of the purpose of the instrument and how it can be used in practice. This can be achieved through training. Recently, TelePsy has offered accredited training courses in the area of the DSM-5 and on TeleScreen, which can increase implementation throughout the Netherlands.

*Mental healthcare.* When looking at mental healthcare, several institutions in the Netherlands already use the instrument. For some institutions, special triage teams were developed, meaning that healthcare professionals interpret the instrument themselves on a daily basis. Other institutions use the instrument during an intake with the patient. The communication within an institution could potentially be improved: using the instrument at the beginning of a trajectory (triage) to decide where the patient should go to or decide who is on their waiting list seems most logical. Hereafter the same answers can be used during a follow-up intake or as part of the diagnostic process where in-depth assessment can take place. For further implementation, communication and PR is also important. A training can be followed at TelePsy, before professionals start using the instrument. In addition (free) webinars are organized to promote communication and disseminate the product.

Now, a substantial body of research was conducted based on TeleScreen among patients aged 18 years and older. However, different versions of TeleScreen can be of added value as well. Think of translation to different languages (e.g. German or English) or different

target groups, such as the TeleScreen Junior (both self-report or an informant version). Currently an informant version of TeleScreen Junior for children at ages 4 to 18 is available and a pilot study has been conducted to assess its validity (data not published). There are plans to validate both the informant version of TeleScreen Junior based on DSM-5 for the ages of 4 to 18, and the self-report version of TeleScreen for the ages of 12 to 18. An informant version of TeleScreen (adult version) would be of interest as well. This may especially be useful in the assessment of persons with limited introspective ability and limited reality testing capacities (for instance patients with autism spectrum disorders, psychotic disorders, personality disorders and bipolar disorders).

## References

1. de Graaf R, ten Have M, van Gool C, van Dorsselaer S. Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc Psychiatry Psychiatr Epidemiol* 2012; 47: 203-13.
2. Whiteford H, Ferrari A, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the global burden of disease study 2010. *PLoS one* 2015; 10: e0116820.
3. Alonso J, Angermeyer MC, Bernert S *et al.* Disability and quality of life impact of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta psychiatr Scand* 2004; 420: 38-46.
4. Wittchen HU, Jacobi F. Size and burden of mental disorders in Europe—a critical review and appraisal of 27 studies. *Eur Neuropsychopharmacol* 2005; 15: 357-76.
5. Smit F, Cuijpers P, Oostenbrink J *et al.* Costs of nine common mental disorders: implications for curative and preventive psychiatry. *J Ment Health Policy and Econ* 2006; 9: 193-200.
6. Postel MG, de Haan HA, De Jong CA. E-therapy for mental health problems: a systematic review. *Telemed J E Health* 2008; 14: 707-14.
7. Gidding L, Spigt M, Brouwer E *et al.* Experiences of general practitioners, practice nurses, psychologists and patients with stepped collaborative care for depression: a focus group study in a large primary healthcare organization. *Eur J Pers Cent Healthc* 2014; 2: 170-8.
8. Verhaak PF, Schellevis FG, Nuijen J, Volkers AC. Patients with a psychiatric disorder in general practice: determinants of general practitioners' psychological diagnosis. *Gen Hosp Psychiatry* 2006; 28: 125-32.
9. Kessler D, Lloyd K, Lewis G, Gray DP. Cross sectional study of symptom attribution and recognition of depression and anxiety in primary care. *BMJ* 1999; 318:436-9.
10. Joling KJ, van Marwijk HW, Piek E *et al.* Do GPs' medical records demonstrate a good recognition of depression? A new perspective on case extraction. *J Affect Disord* 2011; 133: 522-7
11. Janssen EH, van de Ven PM, Terluin B *et al.* Recognition of anxiety disorders by family physicians after rigorous medical record case extraction: results of the Netherlands Study of Depression and Anxiety. *Gen Hosp Psychiatry* 2012; 34: 460-7.
12. Lecrubier Y. Widespread underrecognition and undertreatment of anxiety and mood disorders: results from 3 European studies. *J Clin Psychiatry* 2007; 68: 36-41.
13. Alonso J, Codony M, Kovess V *et al.* Population level of unmet need for mental healthcare in Europe. *Br J Psychiatry* 2007; 190: 299-306.
14. Fernández A, Haro JM, Martínez-Alonso M *et al.* Treatment adequacy for anxiety and depressive disorders in six European countries. *Br J Psychiatry* 2007; 190: 172-3.
15. Bakker P, Jansen P. *Generalistische Basis GGZ. Verwijsmodel en productbeschrijvingen*. Enschede: Bureau HHM, 2013.
16. Donker T, van Straten A, Marks I, Cuijpers P. A brief Web-based screening questionnaire for common mental disorders: development and validation. *J Med Internet Res* 2009; 11: e19.
17. Nguyen DP, Klein B, Meyer D, Austin DW, Abbott JA. The Diagnostic Validity and Reliability of an Internet-Based Clinical Assessment Program for Mental Disorders. *J Med Internet Res* 2015;21; 17: e218.
18. Dijkman I, Dinant GJ, Spigt M. eDiagnostics: a promising step towards primary mental health care. *Fam Pract* 2013; 30: 695-704.
19. Schalken F, Blankers M, Donker T, Duijnstra U, Van Gemert M, van Hoogenhuyze C. *Handboek online hulpverlening: met internet zorg en welzijn verbeteren*. Bohn Stafleu van Loghum: Houten, 2013.
20. Tyrer P, Reed GM, Crawford MJ. Classification, assessment, prevalence, and effect of personality disorder. *Lancet* 2015; 385: 717-726.
21. Dijkman I, Dinant GJ, Spigt M. The concurrent validity of a new eDiagnostic system for mental disorders in primary care. *Fam Pract* 2016; 33: 607-616.

22. Rochlen AB, Zack JS, Speyer C. Online therapy: review of relevant definitions, debates, and current empirical support. *J Clin Psychol* 2004; 60: 269–83
23. Spek V, Cuijpers P, Nyklicek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychol Med* 2007; 37: 319-328.
24. Kooistra LC, Wiersma JE, Ruwaard J *et al*. Blended vs. face-to-face cognitive behavioural treatment for major depression in specialized mental health care: study protocol of a randomized controlled cost- effectiveness trial. *BMC Psychiatry* 2014; 14: 290.
25. Kemmeren LL, van Schaik DJ, Riper H, Kleiboer AM, Bosmans JE, Smit JH. Effectiveness of blended depression treatment for adults in specialised mental healthcare: study protocol for a randomised controlled trial. *BMC Psychiatry* 2016; 16: 113
26. Romijn G, Riper H, Kok R *et al*. Cost-effectiveness of blended vs. face-to-face cognitive behavioural therapy for severe anxiety disorders: study protocol of a randomized controlled trial. *BMC Psychiatry* 2015; 15: 311.
27. Dijkman I, Dinant, G. J, Spigt M. (2017). E- diagnostiek voor psychische stoornissen. *Huisarts Wet* 60: 432-435.
28. Dijkman I, Dinant G. J, Spigt M. (2017). The Perception and Needs of Psychologists Toward Blended Care. *Telemed J E Health* 23: 1-13.
29. Huggins C.E. Online system may ID mental health disorders. Reuters health news, 2013.



The image shows a close-up of a wall or ceiling composed of large, rectangular panels with a fine, pebbled texture. The panels are light-colored, possibly off-white or light grey, and are separated by dark, recessed grout lines. The lighting is soft, creating subtle shadows that emphasize the three-dimensional quality of the textured surface. The overall appearance is clean and modern.

Dankwoord



Graag bedank ik alle mensen die me hebben gesteund bij het tot stand komen van dit proefschrift.

Ik wil beginnen met alle deelnemers van de studies, oftewel de huisartsen, praktijkondersteuners GGZ, (klinisch) psychologen, psychotherapeuten en patiënten, die hebben meegewerkt aan dit onderzoek. In het bijzonder wil ik de focusgroepen van de huisartsen en POH-GGZ bedanken, die vanaf het begin mee hebben gedacht met het onderzoek en hebben geholpen bij de ontwikkeling van TeleScreen. De kennis die is vergaard vanuit de dagelijkse praktijk heeft ertoe geleid dat dit onderzoek kon worden uitgevoerd. Deze kennis heeft ons veel gebracht als het gaat om het in kaart brengen van sterke en zwakke punten van het systeem, waardoor het systeem kon worden aangepast voor hulpverleners en patiënten. Mijn drijfveer voor het onderzoek is o.a. vroegtijdige signalering van klachten en voorkomen dat patiënten 'van het kastje naar de muur worden gestuurd'. Goede diagnostiek is mijns inziens van groot belang om een juiste behandeling en zorgtraject te indiceren. Zonder jullie hulp waren we nergens.

Mijn promotieteam wil ik in het bijzonder bedanken. Mark, ik heb veel van je geleerd. Zodra ik het overzicht niet meer had, hakte je geduldig de taken in stukjes, waardoor het behapbaar werd en overzichtelijk bleef. Naast je manier van redeneren, waar ik veel van heb geleerd, heeft dit er mede toe geleid dat ik het proefschrift niet als last ging zien. De koffie was daarbij heerlijk. Geert-Jan, bedankt voor de inspirerende bijeenkomsten, je vertrouwen en prikkelende discussies omtrent onderzoek en zorg in de geestelijke gezondheidszorg en de huisartsenzorg. Ik heb de manier van onderzoek binnen huisartsgeneeskunde gewaardeerd, waarbij geprobeerd wordt op een praktische manier (aansluitend aan de praktijk) toch veel informatie te vergaren. Dit zorgt ervoor dat de kennis haar weg naar de praktijk vindt en vice versa. Daarnaast wil ik mijn andere collega's bij huisartsgeneeskunde bedanken voor de gezellige tijd. Er heerst een ontspannen en prettige sfeer binnen het team.

In het bijzonder bedank ik Marco. Als gepassioneerde ondernemer maak jij je hard voor zaken. Zonder jouw idee om vroegtijdig klachten in kaart te brengen ter ondersteuning van een verwijzing, geen product en dus geen onderzoek. Jij hebt mij de kans gegeven om überhaupt te kunnen promoveren. Je hebt er daarnaast voor gezorgd dat mijn ambitie om tevens werkzaam te blijven in de praktijk, verwezenlijkt kon worden. Door deze kansen, bleef ik inspiratie houden voor het onderzoek. Hartelijk dank voor je vertrouwen.

Ook wil ik stilstaan bij Hilde. Hilde, als projectmanager stond jij aan de wieg van TeleScreen en samen hebben we veel plezier gehad bij de procesgroepen: bedankt voor de leuke tijd! Bij de start van mijn promotie heb ik de luxe gehad dat er reeds een subsidie was gekregen van Stichting Robuust in samenwerking met CZ en VGZ. TelePsy heeft de rest van mijn promotie gefinancierd door mij onderzoektijd te geven en toegang te geven tot de database: waarvoor hartelijk dank!

Daarbij wil ik ook Scelta noemen. Farid: jij wist me te vinden, samen met je collega's van Scelta, en zorgde voor het opzetten van een mooi en leuk onderzoek: bedankt! Het was een fijne samenwerking. Ik hoop dat we in de toekomst nog vaker samen zullen optreden.

Naast mijn PhD-traject heb ik altijd parttime gewerkt bij voornamelijk TelePsy, maar tegelijkertijd ook huisartspraktijk Centrum Heerlen, de Wijkpraktijk en Psychotherapie Praktijk Lionarons. Ik wil al mijn collega's bedanken voor de interesse in het onderzoek, de ambitie en flexibiliteit om er voor alle patiënten iets moois van te maken.

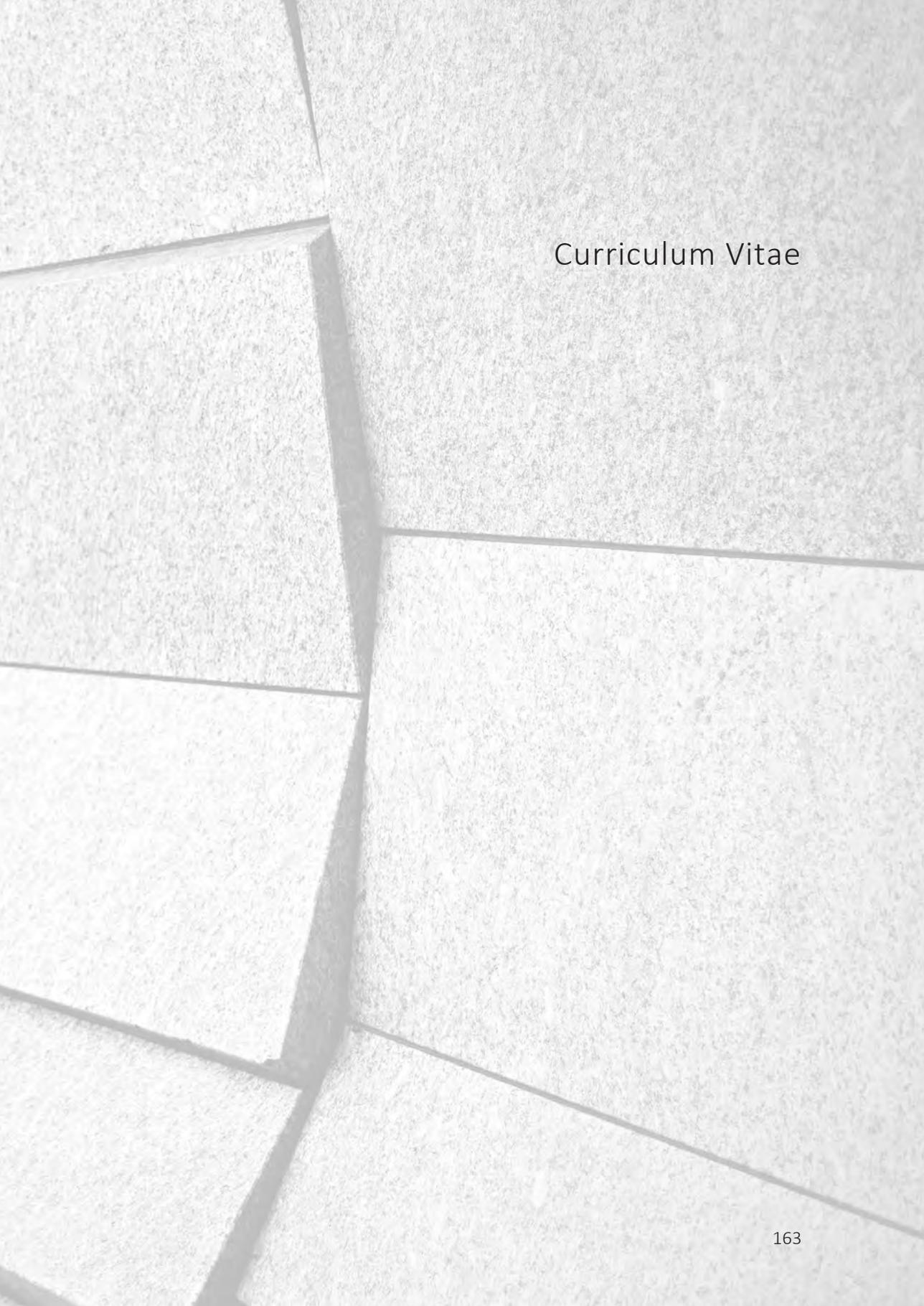
Ik bedank daarnaast graag de leden van de leescommissie, prof. dr. Bart Rutten, prof. dr. Jean Muris, prof. dr. Henrica de Vet en dr. Vivianne Thewissen, en daarbij de andere leden van de corona voor het lezen en beoordelen van het proefschrift.

In het bijzonder bedank ik uiteraard mijn familie, mijn moeder en mijn zus en mijn vader die er helaas niet bij kan zijn, maar het proefschrift binnen no-time zou hebben uitgelezen zodra hij het in zijn handen had gehad. Mijn schoonfamilie wil ik bedanken voor de onvoorwaardelijke steun en trots.

Tot slot wil ik mijn man bedanken. Len: je bent een rots in de branding. Samen hebben we twee prachten van kinderen, Taeke en Moos, en dat doen we met beide een drukke baan en deze promotie toch maar goed. Je bent er altijd voor me, weet goed te relativeren: samen zijn we een team! Ik hou van je.







## Curriculum Vitae





Ies Dijkman was born on 21 November, 1985 in Nijmegen, the Netherlands. She graduated from secondary school in 2005. Ies studied her bachelor Psychology (both Biological and Cognitive direction) and followed additional courses at the Faculty of Health, Medicine and Life-Sciences (Maastricht University) and at the University of Oslo (Norway). She received the degree of Master of Science in Psychology in 2011. During her studies she followed a 9 month clinical internship at RIAGG and wrote her Master's thesis about social anxiety disorders: about likeability and the personality trait openness to experience.

Hereafter, Ies had a part-time position as a psychologist at TelePsy and a part-time position as an external PhD-student at Maastricht University. She started her PhD at Maastricht University in 2012, under the supervision of prof. dr. Geert-Jan Dinant and dr. Mark Spigt. During her PhD trajectory she evaluated experiences of general practitioners (GPs), practice nurses (PNs) and patients with regard to TeleScreen. She was involved in focus groups of GPs and PNs. Subsequently, she studied the concurrent validity of TeleScreen both in primary care and specialized mental healthcare. Finally she studied perceptions of psychologists and barriers with regard to blended care. Ies presented her work at national and international conferences (e.g. NVVP voorjaarscongres 2015, NHG wetenschapsdag 2017, and DGPPN Kongress 2016).

As a triage-psychologist working for TelePsy she performed psychodiagnostics and advised general practitioners about their patients (2012-2017). In addition, she was project manager of TeleScreen Junior, an instrument developed for adolescents and youth (2014) at TelePsy. Meanwhile she performed additional clinical work. She worked as a psychological test assistant (end of 2014) at a clinical practice. In 2015 she worked as a PN where she guided patients with psychological, psychosocial and/or psychosomatic problems. From 2017 onward Ies works as a Research and Development Manager of TelePsy where she performs both coordinating and executive activities. She works on the content of the eHealth platform. Additionally, she manages a team of psychologists and gives supervision. Finally, she works for one day a week as a psychologist in a clinical practice (from 2018 onward).

