Social Media as a tool to contribute to evaluation practices by the Dutch justice system concerning recidivism

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Management of Technology**

Faculty of Technology, Policy and Management

by

Gökhan Anlagan

Student number: 1513672

To be defended in public on September 23rd 2019

Graduation committee

Chairperson	: Prof.dr. W.A.G.A. (Harry) Bouwman, Section ICT
First Supervisor	: Dr. L. (Laurens) Rook, Section ETI

This page was intentionally left blank

PREFACE

The very moment I started writing this sentence, it really got to me; I am finishing my thesis. Honestly, it was not unlike the past couple of weeks, or even days, were not already alarming and preparing me for this moment, it still feels a bit weird when thinking about what the end now actually means to me. I feel I have been privileged to close the 'study' chapter of my life knowing I did something I loved. For that, I will forever be thankful for two individuals.

You may already have this voice in your head almost screaming at you that these two are obviously either the members of my graduation committee, or my mom and dad. Well, undeniably without my parents I would not be here, but this is not their moment.

Harry Bouwman deserves to be acknowledged first and foremost simply for the fact that he accepted me to graduate - period. With retirement lurking just around the corner, I am grateful really you still took the time to guide me with my thesis as my Chair. Your insights during my research have shaped this thesis into the scientifically sound product it is. All the energy that you put in my thesis taught me more about research than I could learn in a lifetime.

As a bridge between the two key figures that guided me with my thesis, I would like to thank Nadir and Fehmi for letting me use their laptops for a software, Maggie for doing what he does best, Igor and Appie for their noninterest, and Nuran for disturbing me throughout it all!

Laurens Rook however is the person that deserves the most praise I could possible give to anybody for this thesis. When I was working on my research proposal, I knew that undoubtedly you would be the perfect First Supervisor - I must say I was right. From the ridiculously long sentences of mine that u corrected, to the unique insights you gave me that eventually formed my Discussion; I could not have asked for a better mentor. It was only natural for you to be a part of the end of my journey here at TU Delft, with all you did for me simply because you are a good person - I truly appreciate it all, thank you again.

This page was intentionally left blank

ABSTRACT

The Dutch justice department could benefit greatly from the personal digital data from social media since psychological attributes are known to be strong predictors for recidivism, and assessing the risk of recidivism (with the RISc) in the Netherlands is a very cost- and labour intensive process with low prediction power. The research question "How can personal digital data extracted from social media, as an alternative for existing ways to measure personality attributes, efficiently and accurately contribute to determination of the recidivism probability of an individual?" is raised to give the Dutch justice department recommendations on how the assessment of the risk of recidivism can be improved based on the predictability of psychological characteristics from social media data. We performed a meta-analysis ($n = 11$) to explore (1) the strength of the predictability of social media data of the Big Five personality traits, and (2) how potential moderators influence the accuracy of the prediction. Main findings were the point estimates of the random effects model (Agreeableness 0.26; Extraversion 0.36; Conscientiousness 0.27; Openness 0.30; Neuroticism 0.31 all with $p < 0.001$) and the highest significant R^2 values ($p < 0.05$) from the moderator analysis for Agreeableness ($R^2 = 0.75$), Extraversion ($R^2 = 0.72$), Openness ($R^2 = 0.92$), and Neuroticism ($R^2 = 0.25$) for the moderator 'Activity', and for Conscientiousness ($R^2 = 0.64$) for the moderator 'Social Media Platform'. This study gives new insights which will help the Dutch justice department make the assessment of recidivism (1) relatively effortless, (2) cheaper, (3) more accurate, and (4) without cognitive bias.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Context

Rapid technological developments have influenced the way the World Wide Web is being used. There is an immense amount of data available for organizations to explore and exploit. This is called ‘Big Data', and can be viewed as one of the biggest IT trends of the last couple of years (Gandomi and Haidar 2015). Organizations use Big Data as an asset; capture, store and analyse data, to increase the organizational performance by making decisions based on these data. Big Data can be defined as data sets that are too complex for traditional data processing applications to deal with (Madden 2012), and is already being used in a vast area of different type of organizations in all kind of fields (McAfee et al. 2012). The data users generate are useful for a varying range of industries (Hayes and Joseph 2003) because it benefits organizations worldwide greatly by providing them an insight in what the users want (Labrinidis and Jagadish 2012).

Not only the emergence of large user data sets, but more importantly the access to them has had great influence on social science researchers; studying psychological attributes got a whole new dimension. Furthermore, the upcoming of large user data sets brought social and computer sciences together. These data sets created an opportunity for researchers to study human behaviours on social media. Researchers are now able to determine psychological attributes and behaviours based on the analysis of digital data created by users with their social media use (Schwartz and Ungar 2015). The digital data researchers use to predict psychological attributes are in the form of demographic data (e.g. age, gender), activity data (e.g. number of friends, likes), language (e.g. tweets), and pictures (Schwartz and Ungar 2015). For a variety of aspects in life, like job performance, social status, health, relationships, subjective well-being and online behaviour; personality can be a predictive factor according to studies on this area (Komarraju et al. 2009; Judge et al. 1999; Anderson et al. 2001). In psychological studies, personality is a crucial subject (Ozer and Benet-Martinez 2006).

The traditional way of mapping out one's personality is usually done on the Big Five scale, which can be assessed with different kind of inventories that require subjects to answer a certain amount of questions that apply to themselves (Schmitt et al. 2007). The result of a personality test mentioned earlier gives the score of an individual on five personality traits; Agreeableness, Conscientiousness, Extraversion, Openness to Experience, and Neuroticism (McCrae and John 1992). However, in psychology an addition of three other personalities exist, called the 'dark personalities'. The dark personalities are the traits Machiavellianism, Narcissism, and Psychopathy; called the Dark Triad (Paulhus and Williams 2002). Unlike the Big Five personality traits, where a single inventory (test) is used to assess all of the traits, with the Dark Triad each personality trait is assessed with a different, specific test. Machiavellianism can be assessed with the MACH-IV test (Christie and Geis 1970), Narcissism with the Narcissistic Personality Inventory (Raskin and Hall 1979), and Psychopathy with the Psychopathy Checklist-Revised (Hare 1991). While the inventories that assess the Big Five personality traits, or the ones that assess any of the Dark Triad traits, have become the standard tests over the course of time, and are being used worldwide for a wide variety of reasons from everyday life to science and business, they don't come without downsides. Personality tests require effort from both the subjects being tested, and the examiner. Preparing questionnaires, handing them out (or sending by mail), filling the questionnaire in, checking results of questionnaires, and doing it all over again to determine the reliability. Beside the amount of effort required, especially for the subject being tested, there is always the risk of the subject not being honest; the subject can be lying deliberately, or just more guessing answers, rather than answering it truthfully (Mischel 1963). Instead of making use of surveys, using personal digital data from social media to predict personality traits presents to be a rapid, cost-effective alternative. Using digital data instead of surveys also enables to reach a larger population, since the effort required to gather data is way less than having people filling in questionnaires.

An area where the use of personal digital data from social media for the assessment of one's personality has not yet found ground, is the Dutch justice department. Whether a suspect in a legal-case still ongoing, or an already convicted criminal, one's personality traits can contribute, or in some cases even be a necessity in decisions or judgements to be made. Decisions made regarding the detainees (e.g. type of regime, treatment, parole) are now largely based on the assessment of the inmate based on the reports made by custodians, by supervising warden and the prison director (Van Wingerden et al. 2011; Stevens 2010). For their reintegration process, to return to society, an inmate must go through several meetings with an assigned probation officer, after which the officer must decide, if permission is granted, and under which conditions a detainee can start leaving the institution (Van der Knaap and Alberda 2009). For suspects in certain type of cases (violence/murder, sexual offense), a psychologist contributes to the assessment of that person. While these people, forming a judgement about another person, are expected to be objective and to be experts in their field, the fact remains that human beings are prone to be biased (unintentionally) in many occasions. The halo effect is a good example, which is a form of cognitive bias where the brain allows specific traits of a person to influence the overall evaluation of him/her (Nisbett and Wilson 1977). This effect can be seen as a behaviour, which is usually performed unconscious. It affects the way people interpret information about someone/something of whom they have a positive impression. This is the case where an attractive person is judged more successful and popular, than an unattractive one. There also is also the horn effect, which is another form of a cognitive bias, where when the first (and most important) impression of someone is negative, all the other positive characteristics get ignored and that person is seen only in negative light (Sigall and Ostrove 1975). There are even more cognitive biases that cause contentious issues in the legal system. Eyewitnesses are known to have given false information by making up details that are untrue, because they have never encoded the initial information, but select some new detail (inspired by news broadcasting for example) and believe it to be true (McCloskey and Zaragoza 1985). Not only does it occur that eyewitnesses give misinformation, the relationship between the confidence and accuracy is uncertain. According to Loftus (2019) one can't say anything about the relationship between confidence and accuracy of eyewitnesses since at times the relationship is strong (the more confident a person is in his/her answers the higher the chances are it's true), is non-existent, and even weak (people being confident about their wrong assessments). Studies show that judges and jury members are good at assessing whether a subject is accurate or not (Beaudry et al. 2015). Cognitive biases can result in people being charged and tried in court more harshly for a felony than normally. On the other hand, a suspect may play the part of a "mad-man" in order to get a less harsh punishment, maybe even in the form of a treatment, rather than a prison sentence. Also, the unwilling nature of suspects or detainees to cooperate with obligated meetings with the psychologist, may result in a false assessment in the end. Wrongful assessing an inmate for having a high chance for recidivism can have major consequences; such as ankle-band as electronic supervision, extension of supervision duration, or even denying participation in the reintegration process. The other way around, when an inmate is expected to not resort to criminal habits, while the chance for recidivism is actually high (but unknown to the Dutch justice department), wrongful assessment can result in the inmate causing harm to society again. In an attempt to eliminate human errors in judgement, one can turn to digitally generated data that can be interpreted accordingly.

Recidivism is difficult to predict, and psychometric tests of adult personality and psychopathy like the MMPI (Minnesota Multiphasic Personality Inventory) thus far were only able to report a weak relationship between MMPI-based typologies and recidivism (Megargee and Bohn 1979). Even though the recidivism number in the Netherlands is not the highest found globally at 48%, it is still significantly higher than in better performing countries like Norway with 20% (Fazel and Wolf 2010). Also, the current method to assess the risk of recidivism correlates poorly with the actual recidivism number; $r = 0.30$ (Wartna et al. 2008). Since the fate of both suspects and inmates are to a large extent based on the assessment of their probability of recidivism, additional research in determining this factor can prove to be valuable in (1) reducing the recidivism number in the Netherlands, (2) create a more objective approach without the interference of human biases/errors which will bring more fairness in verdicts, (3) reduce the amount of effort (whole workday) that is required to assess recidivism, (4) find a stronger predictor for recidivism than the current one ($r = 0.30$), and (5) make it possible to assess the risk of recidivism also for suspects in ongoing cases (not only convicted individuals) since they are judged based on this risk whether they can await their trial in freedom, or not.

1.2 Research Framework

1.2.1 Main Objective

The **main objective** of this research project is to provide theoretically grounded recommendations on how the Dutch justice department can improve the assessment of the recidivism probability of suspects and inmates. The recommendations will be based on the predictability of personal digital data from social media that can be retrieved about the person in matter, on his/her psychological characteristics. This research contributes, on how personal social media data can be collected in an alternative way taking into consideration the current privacy regulations.

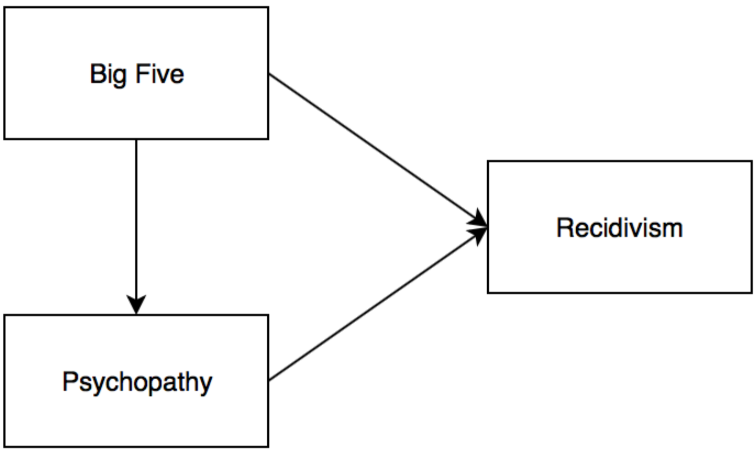

Fig. 1.1 Conceptual Model

1.2.2 Research Question and Sub-Questions

The objective leads to the following **research question**:

How can personal digital data extracted from social media, as an alternative for existing ways to measure personality attributes, efficiently and accurately contribute to determination of the recidivism probability of an individual?

From the research question, **sub-questions** are derived that outline how our research will be conducted. We will first tackle four sub questions to gain practical background information for our research project. To start with, we will research our construct 'Recidivism'; what it means exactly:

1. What is Recidivism?

After we have obtained a clear definition of recidivism, we need a deeper insight in the current assessment of recidivism and personality by the Dutch Justice department (with a focus on the Big Five personality traits). Since we want to compare whether the new method proposed and investigated in our research project is a more accurate, and less effortful alternative than the current one, we need to answer the following questions beforehand:

2. Which methods for recidivism assessment of Dutch prisoners and suspects are being used at the moment?
    a. If currently attention is being paid to the Big Five personality traits, how are the traits implemented for the assessment of recidivism?
    b. If currently no attention is being paid to Big Five personality traits, what is the current assessment method for predicting recidivism? How could personality be assessed with the Big Five personality traits, both online and offline, for assessing recidivism?

As practical background information, we also need to check whether it is allowed to scrape personal digital data from social media in the case of Dutch prisoners; hence, we have to explore the regulatory conditions for retrieving personal digital data from social media.

3. What are the regulatory conditions for acquiring personal digital data from social media?
    a. How do privacy regulations affect the collection of personal digital data from social media?
    b. What are the legal boundaries for the use of personal digital data scraped from social media?

Since our research focus is on people who are incarcerated, or awaiting trial as a suspect, we will explore the regulations around seized electronic devices (e.g. phone or laptop) since they are the source of information from where the Justice department can get the data they need.

4. What are the regulations around seized electronic devices from inmates/suspects?
    a. If social media data retrieved from seized electronic devices can be used in a legal-case, can it also be used for personality assessment? If not, why not, and how can it be regulated to be used in the future?

   b. If social media data retrieved from seized electronic devices can't be used in a legal-case, why not, and how can it be regulated to be used in the future for personality assessment?

After an exploration of the current recidivism assessment methods and the regulatory conditions concerning the collection of personal digital data from social media, we next will focus on our two other constructs; Psychopathy and Big Five. First, we will examine the construct Psychopathy; what it is, how it is assessed, whether it can be used to determine recidivism, and whether personal digital data from social media can be used to assess Psychopathy.

5. What is Psychopathy?
   a. How is one's Psychopathy assessed?
   b. What is the correlation between Psychopathy and recidivism?
   c. How can personal digital data be extracted from social media to assess one's Psychopathy?
   d. How does personality assessment through personal digital data from social media relate to the personality assessment through the Psychopathy Checklist?

Next, we will focus on our other construct, the Big Five; what it is, how it is assessed, whether it can be used to determine recidivism, and whether personal digital data from social media can be used to assess one's Big Five personality traits.

6. What are the Big Five personality traits?
   a. How are the Big Five personality traits assessed?
   b. What are the correlations between the Big Five personality traits and recidivism?
   c. How can personal digital data be extracted from social media to assess one's Big Five personality traits?
   d. How does personality assessment through personal digital data from social media relate to the personality assessment through the Big Five personality traits questionnaire?

The final remaining correlation, which is relevant for our research, is that between our earlier mentioned constructs; Psychopathy and Big Five.

7. How can Psychopathy be assessed from the Big Five personality traits making use of personal digital data from social media?

Our research above will bring us to our final sub-question that we need to address, to be able to say how personal digital data extracted from social media can, as an alternative for existing ways to measure personality attributes, efficiently and accurately contribute to determination of the recidivism probability of an individual.

8. What is the predictive power of personal social media data over psychological characteristics?

We will perform a meta-analysis to study the predictive power of personal social media data over psychological characteristics, and how different variables influence the accuracy of the prediction. The method for collecting the data, and the analysis, is explained next.

Both **data collection** and **data analysis** for our research project will be done in the following manner:

Data collection

- The data that we will collect for our meta-analysis will be in the form of so-called 'effect sizes'; gathered from existing published studies, found through the literature search, that eventually will be selected for our meta-analysis

Data analysis

- Reliability in a meta-analysis is established by having two or more individuals performing the literature search, selecting and coding the studies, and then checking the inter-rater agreement between them (McHugh 2012). However, due to the nature of this research project (begin a thesis project), the meta-analysis will be performed by one individual.
- Validity (construct validity) will be tested by way of testing the heterogeneity of the collected effect sizes from the studies that are included in the meta-analysis, which is the conventional way to test validity in a meta-analysis (Smith and Robertson 1993).

1.2.3 Outline of Research

So, with this research we wish to provide theoretically grounded recommendations on how the Dutch justice department can improve the assessment of the recidivism probability of suspects and inmates, based on the predictability of personal digital data from social media that can be retrieved about the person in matter, on his/her psychological characteristics. To achieve this main goal of ours, we built our research around our main research question. From our main question, we designed sub-questions that guide us in the process of better understanding our three main constructs (Recidivism, Psychopathy, and the Big Five), how they are related to one another, and how social media data can be used to assess all three of them.

After this chapter, we lay-out some practical background information in the next one (Ch. 2) regarding how recidivism is currently assessed in the Netherlands (2.1), and what laws and regulations apply to extracting and processing online social media data both in general, and with the focus on situations concerning suspects and inmates (2.2). Thus, we answer our first four sub-questions in Chapter 2.

In Chapter 3, we present our literature review, where we, after the introduction section (3.1) where we show our search strategy, we focus on the two constructs Psychopathy and Big Five. For our fifth sub-question, we explore what Psychopathy is, how it is measured, how it is related to recidivism, and what the current studies say about the possibilities to assess Psychopathy from personal social media data (3.2). In the next section (3.3), we treat our sixth sub-question by addressing our other construct, the Big Five, and show the fundaments, how it is measured, how it is related to recidivism, and what the current studies say about the possibilities to assess the Big Five from personal social media data. Before we conclude Chapter 3, we explore how Psychopathy and the Big Five are related following our seventh sub-question (3.4). At the end (3.5), our conclusions set the next chapter up, and explain in great detail our motivation why we chose to do a meta-analysis.

With Chapter 4 we provide our method-chapter where first give a brief historic introduction of the method itself (meta-analysis) and explain what it is (4.1), explore the different variants (4.2), and then present step by step how we performed it, which answers our eighth and final sub-question (4.3 - 4.8).

In Chapter 5 we discuss the scientific and practical relevance of our work, the limitations we encountered in this project, our recommendations for future studies, and our final conclusions of our research.

This research contributes, not to theory per se, but more on how personal social media data can be collected in an alternative way taking into consideration the current privacy regulations.

CHAPTER 2: PRACTICAL BACKGROUND

While the core focus of this study is on the relationship between personality traits and personal digital data from social media, our eventual goal is to investigate what the practical use can be of using personal digital data from social media to assess the risk of recidivism, and how this is different from the current way of assessing it.

Initially we will start this chapter by defining what recidivism is, and then explore how recidivism is currently assessed in the Netherlands (2.1) in the amount of effort required to assess recidivism, and how accurately it is. When we know the details about the current process of assessing recidivism in the Netherlands, we can in the end conclude whether our proposed method (assessing it using personal digital data from social media) is more accurate, and requires less effort.

After having covered the traditional methods part in Section 2.1, we shift our focus to the method we would like to investigate and propose in the end with our research study. We need some more practical background information first, since we will be dealing with personal digital data from social media. All different kind of laws and regulations that apply to the extraction and processing of personal digital data from social media in the Netherlands from suspects and inmates whether it be online from their account, or from their electronic devices (e.g. phone or laptop) that were seized, will be examined (2.2).

At the end of this chapter we conclude what point(s) of improvement(s) can be made that relate to the current way of assessing recidivism in the Netherlands, and whether there are restrictions both in the extraction and processing of personal digital data from social media of suspects and inmates both online and offline (2.3).

2.1 Recidivism

This section we will start by giving a clear definition of our concept 'Recidivism', and then continue with how the assessment of recidivism is currently performed in the Netherlands.

2.1.1 Definition Recidivism

It is important to explore the definitions of recidivism, and the different ways it can be measured. There seems to be some variations throughout studies. According to Babinski et al. (2001) recidivism can be assessed both through self-reported delinquent behaviour and official records; where limitation of self-report delinquent behaviour is the possibility of socially desirable answers, and of official records is that they do not report undetected crimes. Recidivism can be distinguished in (1) criminal-legal recidivism where a person has already been lawfully sentenced for a previously committed crime whenever that person commits a crime, (2) penal recidivism where a prison sentence is pronounced to an individual who had already been sentenced by the same sanction before, and (3) criminological recidivism where a person doesn't necessarily need to have been convicted before, but has committed a criminal act previous to the new act (Međedović et al. 2012). In our research, when we talk about recidivism, we talk about criminal-legal recidivism. Note that this matches the definition being used in the Dutch justice system correctly (Wartna et al. 2014).

2.1.2 Recidivism in the Netherlands

Estimating the risk for recidivism in the Netherlands is currently only being done on already convicted inmates during their stay in prison with a tool called 'Recidive Inschattingsschalen'

(RISc) (Van der Knaap and Alberda 2009). Not only does the RISc estimate the risk for recidivism, it also maps out which criminogenic factors (characteristic and situations of people associated with criminal-behaviour) are the cause for this risk. The RISc was developed in 2002 as a part of the then introduced new policy-program called 'Terugdringen Recidive' (TR) whose main goal was to reduce recidivism under adult convicts through a combination of activities both in the prison system and during probation. The RISc is one of the methods being used to realize the goal of TR to reduce recidivism by assessing the risk for recidivism for convicts.
The protocol for working with the RISc is as follows: a certified probation officer fills in the RISc according to specific information gained from the conversation with the subject to assess the extent of the problems based on 12 criminogenic factors. Other than the conversation, the probation officer also obtains information from the case file of the subject. The total workload for a certified probation officer to assess, and process the RISc for an individual is at least a whole workday (8 hours) - which is very cost- and labour-intensive.

The RISc consists of three parts: basic-diagnostic, in-depth diagnostic and indication statement (in Dutch: basisdiagnostiek, verdiepingsdiagnostiek en indicatiestelling). Basic-diagnostic is the first step where the risk of recidivism is based on and where criminogenic factors are mapped out. The probation officer can choose to perform an in-depth diagnostic when the basic-diagnostic fails to provide sufficient insight. In the final step of the RISc, the indication statement, an assessment is made what the possibilities are for the subject to participate in certain interventions (e.g. motivation, behaviour or personality). The probation office hopes to come to a fitting probation trajectory for the subject by using the RISc. In the end, the goal of the RISc is not only to predict the risk for recidivism, but also to reduce this risk with the help of interventions.

The basic-diagnostic consists of quantitative and qualitative items. Quantitative items are being scored and are based on characteristics of the subject or its situation and cover areas that contribute to the risk of recidivism (Adviesbureau Van Montfoort & Reclassering Nederland 2004). The total of 61 items are divided into 12 scales that represent a static or dynamic criminogenic factor:

1. Criminal History
2. Current offense and offense pattern
3. Housing and living
4. Education, work and learning
5. Income and dealing with money
6. Relations with partner and family
7. Relations with friends and acquaintances
8. Drug use
9. Alcohol use
10. Emotional well-being
11. Thinking pattern, behaviour and skills
12. Attitude

The quantitative items of the 12 scales are scored on a categorical scale ranging from 0, 1, to 2. Where 0 indicates the absence of problems, a higher score corresponds to the presence of serious problems. A rough scale score is calculated on each scale by adding the items scores together. Since the scales 1 and 2 are combined, the actual RISc profile consists of 11 scale scores. The rough scale scores are converted into a weighted score based on the English instrument 'Offender Assessment System', also known as OASys (Howard et al. 2003). The sum of all the weighted scale-scores is the total score which is the indication for the risk of recidivism. Since the second scale (current offense and offense pattern) can't be assessed for someone who has not been convicted yet, the RISc can only be applied to already convicted people. The weighted scale-scores of the RISc are divided into 3 categories that assess the extent to which criminogenic factor is present; i.e. criminogenic problem is not present, it is present, and it is extensively present. The total scores are also divided into three categories, which reflect the severity of the risk of recidivism; i.e. low risk, average risk, and high risk.

The Big Five personality traits are not used whatsoever currently for the assessment of recidivism in the Netherlands. To determine the predictive validity of the RISc, data from the WODC-Recividemonitor, a research project that measures recidivism in the Netherlands, was used (Wartna et al. 2008). The data from the WODC-Recividemonitor originally comes from Onderzoek- en Beleidsdatabase Justitiële Documentatie; which is an encrypted and anonymised copy of the original Justitiële Documentatie- systeem (JDS). The JDS has a record of everyone who has ever been involved in a criminal case. To examine the relationship between RISc total score and recidivism, the correlations between the weighted RISc score and all forms of recidivism were calculated. According to Cohen (1988) correlation of 0.10 is weak, 0.30 is normal, and 0.50 strong. The RISc total score has correlations around 0.30 with recidivism (Wartna et al. 2008).

Since we would like to see whether the Dutch justice department could improve the current practice of assessment of the recidivism probability of suspects and inmates, based on the predictability of the personality traits of individuals derived from personal digital data shared on social media, we need to examine laws and regulations that apply to the extraction and exploitation of such sensitive info.

2.2 Personal Digital Data

This section is divided in sub-sections each dealing with different laws or regulations; European Convention on Human Rights, The Dutch Constitution, General Data Protection Regulation, and Dutch Police Act 2012 (2.2.1 - 2.2.4). Due to possible privacy restrictions, the different types of laws and regulations are explored in how they (will) affect the use of personal digital data from social media of suspects and inmates in the Netherlands. The digital data can be acquired both online (algorithms) and offline (from seized electronic devices from suspects). Therefore, we will also further investigate whether data from seized electronic devices (e.g. phones, laptops) can be used to assess one's personality; what the regulations and restrictions are that come in to play (2.2.5). In the end (2.2.6), we will give a short summary of this section where we present an overview of the details of laws and regulations that apply to our research.

2.2.1 European Convention on Human Rights

Both the confiscation of objects and the investigation into these electronic devices with the sole purpose of prosecution of criminal offences is an attack on the privacy of people. Whether it be on the privacy of the user or owner of these devices, or of third parties whose data is stored within, the European Convention on Human Rights (ECHR) has set human rights that also protects human rights in Europe in the form of Article 8 (European Court of Human Rights 2018):

"Article 8 of the Convention- Right to respect for private and family life

> 1. Everyone has the right to respect for his private and family life, his home and his correspondence.
> 2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others."

Since the term 'private life' is a complex one to explain and it is not possible to give a clear definition of it, the ECHR has already stated in 1992 that it is impossible and probably also undesirable to define the term 'privacy' and whether private life is affected by certain government actions must be looked at on a case-by-case basis (ECHR December 16, 1992). Article 8(2) states the two criteria in which interference by a public authority is justified. The ECHR assesses whether these two conditions (in accordance with law and necessity in a democratic society) are met.

According to the European Court, the condition 'in accordance with law' has four further requirements. During a case in Prezhdarovi (Bulgaria) in 2014 the Court stated that in accordance with law meant that the impugned measure should have some basis in domestic law, the domestic law must be accessible to the person concerned, the person affected must be able (if needed with appropriate legal advice) to foresee the consequences of the domestic law for him, and the domestic law must be compatible with the rule of law (ECHR September 30, 2014).

With the 'necessity in a democratic society' the ECHR lets the member states decide whether this criteria is met with their own interpretation. The Court merely assesses if there is an urgent social need, but leaves the member states a certain margin of appreciation.
What is vital for ECHR is that there is no arbitrary violation in the private life of the citizens or any kind of misuse of power. Even in a case where Article 8 ECHR is violated it does not necessarily mean that the evidence gathered through it is considered invalid. Violation of Article 8 ECHR is not per se a violation of Article 6 ECHR; the protection of the right for a fair trial (ECHR June 1, 2010).

2.2.2 The Constitution of the Kingdom of the Netherlands

The protection of privacy as a fundamental right in the Netherlands has been included in Article 10 in The Dutch Constitution (Grondwet 2018):

> "1. Everyone shall have the right to respect for his privacy, without prejudice to restrictions laid down by or pursuant to Act of Parliament.
> 2. Rules to protect privacy shall be laid down by Act of Parliament in connection with the recording and dissemination of personal data.
> 3. Rules concerning the rights of persons to be informed of data recorded concerning them and of the use that is made thereof, and to have such data corrected shall be laid down by Act of Parliament."

Privacy is elaborated as any information concerning an identified or identifiable natural person according to Article 1(a) of Wet Bescherming Persoonsgegevens (Wet bescherming persoonsgegevens 2000). The protection afforded by the Dutch Constitution is just as much absolute as the protection afforded by Article 8 ECHR.

2.2.3 General Data Protection Regulation

General Data Protection Regulation (GDPR) is a new European Union (EU) regulation made by the European Parliament and Council of the European Union on data protection and privacy for citizens within the EU and the European Economic Area (EEA) introduced 14 April 2016 and implemented 25 May 2018 (Carey 2018; EU Directive 2016/680). The GDPR is superseding the Data Protection Directive 95/46/EC on which the Dutch Data Protection Act (Wet bescherming persoonsgegevens) was largely based which was effective until 25 May 2018.

For working with personal data in the police and criminal justice sector there is a separate Data Protection Directive included in the GDPR that provides rules on personal data exchanges at national, European, and international levels (EU Directive 2016/680).

2.2.4 Police Act 2012

Before the introduction of the GDPR, along with the separate Directive that gives guidelines for working in the police and criminal justice sector, the Police Data Act (Wet politiegegevens) provided rules on how to deal with personal data in the context of the justice system. Where normally the Dutch Data Protection act provided guidelines in data protection, the exemption is made in Article 2(2c) of the Dutch Data Protection act where it states that the act doesn't apply to the processing of personal data for the purpose of performing the police responsibilities referred to in Sections 3 and 4 (1) of the Police Act 2012 (Wet bescherming persoonsgegevens 2000; Politiewet 2012). Processing personal data by the law enforcement is permitted when done for certain purposes as explained in Sections 8, 9 and 10 of the Police Act. Section 8 states the purpose can be to carry out daily police tasks. According to Section 9 the purpose can also be to maintain law and order by contributing to an investigation. Finally, by Section 10, processing personal data is allowed to gain insight in the involvement of certain people that committed, or planned crimes which are due to their size or gravity, or their coherence with other crimes, a serious threat to the law and order, or with actions which are due to their nature or frequency or the organized way in which they are committed, seriously violate the public order.

2.2.5 Electronic devices

Electronic devices (e.g. laptop, smartphone) are confiscated to help law enforcements with further investigation. These devices can be greatly helpful in providing more insight in the network of a suspect, locations the suspect has been, and the communication that has been done with it. The legal basis to confiscate any object in the Netherlands is laid down in the Criminal Law (Wetboek van Strafrecht) by several Articles (art. 96 Sv., art. 96b Sv., art 97 Sv. and art 110 Sv) and the confiscation can only be done by law enforcers (opsporingsambtenaren) with a warrant obtained by a judge (rechter commisaris) in case of a property search (e.g. house, office, car). In any case, for a confiscation, the person in matter must be a suspect for a crime according to art. 67 lid 1 Sv. and requires provisional detention for this act.

2.2.6 Summary

Both in the European Convention on Human Rights and The Constitution of the Kingdom of the Netherlands have set human rights that protects the privacy of its citizens. Since the protection afforded by the Dutch Constitution is just as much absolute as the protection afforded by Article 8 ECHR, we can conclude from Article 8 ECHR that interference by a public authority is allowed when it is in the interests of national security, or even for the prevention of disorder or crime, or for the protection of the rights and freedoms of others. GDPR supersedes previous regulations on data protection and privacy of citizens in the EU, and has a special directive included for working with personal data in the police and criminal justice sector. Just like in the Dutch Police Act, processing personal data by law enforcement is permitted when it is done to carry out police tasks, to maintain law and order by contributing to an investigation, or to gain insight in the involvement of certain people that committed, or planned crimes with a certain level of gravity and threat to law and order and public order. Confiscation of any electronic device in the Netherlands can only be carried out when the person in matter of whose electronic devices are confiscated is a suspect for a crime that requires provisional detention.

2.3 Conclusion

Recidivism in the Netherlands is the case where a person has already been lawfully sentenced for a previously committed crime whenever that person commits a crime. The risk of recidivism is in the Netherlands currently assessed with the RISc instrument, which also maps out the criminogenic factors that are the cause for the risk. There are three steps in the RISc: (1) basic-diagnostic; a certified probation officers fills in the RISc based on information gained from conversation(s) with the subject and its case-file (2) in-depth diagnostic; only applied when basic-diagnostic fails to provide sufficient insight (3) indication statement; assessment of what the possibilities are for the subject to participate in certain interventions. There is no attention being paid to the Big Five personality traits in the RISc, but can easily be implemented in the basic-diagnostic step by having inmates fill out questionnaires. An important notice is the fact that the RISc can only be applied to already convicted people, since the second scale (current offense and offense pattern) can't be assessed for someone who has not been convicted yet. The RISc total scores given to the subject in the basic-diagnostic part have correlations of 0.30 with recidivism.

This chapter provided us supplementary information to answer especially the last part of our research question: *whether recidivism can be assessed more efficiently and accurately*. Our main objective is to provide recommendations on how to assess recidivism through personality assessment with personal digital data from social media, but for now, in this chapter, we focused on the current way of recidivism assessment. We found that, while there is room for improvement in assessing recidivism in the Netherlands in the correlation area, the bigger improvement can be in the amount of effort required to assess the risk of recidivism, and without human error/bias. The current protocol is very cost- and labour-intensive, while technological developments are such that quicker and potentially more accurate solutions may be available. Also, recidivism can only be assessed for already convicted individuals for now (due to the RISc scale). By using personal digital data from social media, recidivism can also be assessed for suspects now that we know that there are no legal boundaries in assessing digital information from a suspect or inmate. In the next chapter (Ch. 3) we will focus on the theoretical part of this research to explore in more depth what Psychopathy is, what the Big Five personality traits are, how both of them can be assessed from digital data from social media platforms, and how they are related to recidivism.

CHAPTER 3: LITERATURE REVIEW

In the previous chapter (Ch. 2) we studied how risk of recidivism is currently assessed in the Netherlands, and looked for areas for improvement for assessing the risk more efficiently and accurately. We found that the current method was not very accurate in predicting recidivism, that it takes a whole day of work (8 hours) for a certified probation officer, and can only be executed for already convicted individuals (no suspects awaiting trial). Since there is room for improvement in assessing the risk of recidivism in the Netherlands, and there are no legal boundaries that restrict acquiring and processing any kind of digital data from suspects or inmates in the Netherlands both online and offline, we will continue our research in this chapter by exploring the possibilities to extract personal social media data to eventually assess recidivism. We will not investigate how personal digital data can be used to directly determine recidivism, but to measure personality attributes (Psychopathy and Big Five) which will in turn be used to determine recidivism (Fig. 3.1).

We will start our theoretical chapter by providing the literature search strategy used in our research (3.1). Following this section, there are two sections (3.2 and 3.3) explaining the fundaments of Psychopathy, and the Big Five personality traits, and how they can be measured. From these sections, it becomes clear that we need to make the relationship between Psychopathy and the Big Five explicit in order to answer our research question. Therefore, after these two sections, we investigate the relationship between Psychopathy and the Big Five (3.4). In the end (3.5) we give an overall conclusion to this chapter, and introduce the next chapter (Ch. 4); a method chapter where we examine what a meta-analysis is first, and then present how we conducted our own meta-analysis.

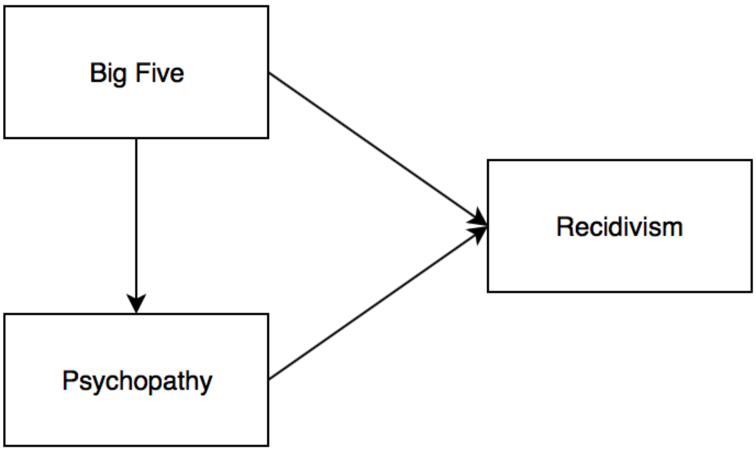

Fig. 3.1 Conceptual Model

3.1 Literature Search

Studies on the relationship between our core concepts Psychopathy, Big Five and Recidivism were explored by conducting an extensive literature search in databases like Google Scholar, Web of Science and Scopus. We constructed five groups of keywords for our literature search; three groups for our constructs (Psychopathy, Big Five, and Recidivism), and two more for social media and analytic approaches with respect to online data extraction. A combination of the keywords from each group was used to search for relevant studies (Table 3.1). Next to the databases, we also used the Google search engine to find additional papers. We looked for the terms we used for our search in the title, abstract, and keyword section of the scholarly literature.

Table 3.1 Keywords for literature search

Psychopathy	Big Five	Recidivism	Social Media	Analytics
Psychopathy	Big Five	Recidivism	Social media	Data extraction
Psychopathy Checklist	Personality traits		Social network	Data mining
PCL	Personality		Facebook	Digital data
PCL-R	Agreeableness		Instagram	Content analysis
Dark Triad	Extraversion		YouTube	Machine learning
	Conscientiousness		Twitter	Big data analytics
	Openness to Experience		Snapchat	Text analysis
	Neuroticism		LinkedIn	Text mining
			Sina Weibo	Image analysis
			WeChat	

To explore Psychopathy (3.2) in more depth (3.2.1 - 3.2.2), we used the keywords under Psychopathy first. After that, to see how psychopathy is related to recidivism (3.2.3), we used the keywords under Psychopathy and Recidivism. Finally, to see whether Psychopathy can be assessed using personal digital data from social media (3.2.4), we combined the keywords under Psychopathy, along with the ones under Social Media and Analytics.

For the Big Five personality traits (3.3) we followed a similar approach, in that we used the keywords under Big Five to study more about the Big Five personality traits (3.3.1 - 3.3.2), and used the same keywords combined with the one under Recidivism, to see how the Big Five personality traits are related to recidivism (3.3.3). In the end, we used the keywords under Big Five again, but this time along with the keywords under Social Media and Analytics, to see whether its possible to assess the Big Five from personal digital media data (3.3.4).

For the final section (3.4), we combined the keywords under Psychopathy and Big Five to investigate the relationship between Psychopathy and the Big Five personality traits.

3.2 Psychopathy

Psychopathy as a socially aversive personality is one of the three interrelated higher-order personality constructs, along with Machiavellianism and Narcissism, that make up the so-called 'Dark Triad of Personality' (Paulhus and Williams 2002). An individual scoring high on psychopathy shows positive relations to unwanted behaviors such as aggression (Kerig and Stellwagen 2010), substance abuse (Benning et al. 2003), and, most importantly for our research, criminal-legal recidivism (Asscher et al. 2011). That is why in this section we will first explore what Psychopathy is and how it is assessed, then examine the relationship between Psychopathy and Recidivism, and finally see whether personal digital data from social media can be used to assess one's Psychopathy (Fig. 3.2)

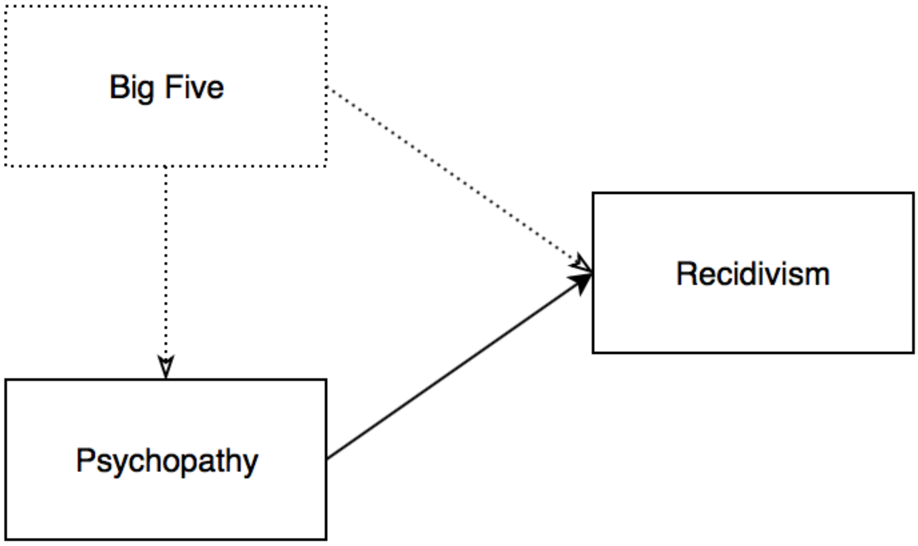

Fig. 3.2 Psychopathy and Recidivism: the relationship between the constructs we will examine in Section 3.2

3.2.1 Psychopathy

Psychopathy is described by Patrick et al. (2009) as a personality disorder characterized by antisocial behavior, impaired empathy and remorse, along with egotistical traits. In the penal forensic system, Psychopathy is the most prominent clinical finding according to Hare (1998) and there is a significant difference between the behavior of offenders diagnosed as psychopaths from that of other criminals. Psychopaths account for most of the violent crimes around the world and are characterized by starting their criminal careers at an early age, being skilled in multiple forms of crime, to be the most undisciplined members of the prison system and don't benefit in general from rehabilitation programs (Hare 1998). Furthermore, these psychopaths present the highest criminal recidivism rates; around three times higher than other offenders (Hemphill et al. 1998). While the prevalence of psychopathy in the overall population is around 1% (Hare 1998), among inmate populations it is estimated at 15-20% (Konrad 2002). There is evidence that psychopaths (someone suffering from chronic mental disorder, which causes violent or abnormal social behaviour) commit much higher numbers of criminal and violent offences than non-psychopaths (Hare and Jutai 1983).

The construct Psychopathy is constructed with observations by Cleckley (1941), operationalised in Hare's Psychopathy Checklist (Lilienfeld and Andrews 1996; Hare 1991). Describing psychopathy by a list of criteria was first done by Cleckley (1941) where anyone fitting enough of these criteria would count as a psychopath or sociopath.

3.2.2 Measuring Psychopathy

To measure Psychopathy, Hare (1970) developed the Psychopathy Checklist (PCL), which is regarded as one of the best developed instruments (the most validated measure) to determine Psychopathy among people (Schroeder et al. 1983; Hare 1983; Cooke et al. 2001; Vitacco et al. 2005). The PCL got revised by Hare (1991) and became the Psychopathy Checklist-Revised (PCL-R). To assess the presence of Psychopathy in an individual, the most commonly used psychological assessment tool is the PCL-R (Venables et al 2014). This PCL-R consists of a 20-item symptom rating scale and is used by qualified examiners (e.g. psychologists) to score subjects on a three-point scale according to specific criteria through both background information of subject, and an interview with the subject, to compare a subject's degree of Psychopathy with that of a prototypical psychopath. Each of the 20 items is given a score of 0, 1, or 2 based on how well it applies to the subject being tested. The total score indicates how closely the test subject matches the score of a prototypical psychopath; a maximum score of 40. Where people with no criminal backgrounds score around 5, non-psychopathic criminal offenders tend to score around 22. With a score of 30 or above the subject is diagnosed as a psychopath. The 20 items included in the PCL-R are seperated into Factor 1 and Factor 2. Factor 1 items relate to personal relationship and emotional states (e.g. selfishness, lack of remorse and insensitivity toward others) and correlate with low empathy and low nervousness (Huchzermeier et al. 2007; Harpur et al. 1989; Zágon and Jackson 1994; Verona et al. 2001). Factor 2 items refer to criminal versatility, recidivism, antisocial personality disorder, and thrill thriving (Harpur et al. 1989).

Self-report measures

With the gradual expansion of Psychopathy research (including non-forensic samples as well) several self-report measures were developed to assess the construct. These self-report measures assess affective-interpersonal and antisocial aspects. However, only three known measures exist that to some extent relate to the PCL-R factors (Hare et al. 1989; Levenson et al. 1995; Williams et al. 2007; Lilienfeld and Andrews 1996; Lilienfeld et al. 2005).

The first one is the Levinson's Self-Report Psychopathy Scale (LSRP) which also measures the same two facets the PCL-R does (Factor 1 and 2) and does this with responses given on a 4-point Likert scale; 16 items for primary psychopathy and 10 items for secondary giving a total of 26 items (Levenson et al. 1995).The LSRP, which is a valid and reliable scale, tells more about actions concerning community life, rather than examining the criminal activity of an individual scoring high on Factor 2 (Brinkley et al. 2001).

The second measure, the Self-Report Psychopathy scale (SRP), was developed after initial PCL and was constructed to determine the same constructs as the PCL (Hare et al. 1989; Hare 1985). However, the SRP was not successful in measuring the factors of the PCL (Hare 1991). Therefore, the SRP-II, a 60-item revised version of Hare's SRP was developed, from which also an abridged version of 31 items exists; 9 items as a scale to determine the PCL-R factors, 13 items as a scale for assessing the behavioural factor, and nine items to tap both factors (Hare 1991). But again, studies on the SRP-II were not successful representing the PCL-R factors (Benning et al. 2005; Williams and Paulhus 2004). The final version of the Self-Report Psychopathy scale, the SRP-III, is a 40-items scale that has 31 of the items from the SRP-II with an addition of 9 new items. The SRP-III is still being researched and currently there exist the 31-, 62- and 64-tems versions of it.

Besides the two earlier mentioned self-reports, there is also the Psychopathic Personality Inventory (PPI), and the revised version the PPI-R, which are used more often than the other two measures. The original version, the PPI consists of 187 items, whereas the PPI-R has 154 items divided over 8 subscales that don't include antisocial or criminal items (Lilienfeld et al. 2005). The PPI correlates moderately (at best) with scores on the PCL-R (Bloningen et al. 2010).

Table 3.2 Tools to measure Psychopathy

	Total Items	Scale	Assessment
PCL-R	20	3-point Likert	Qualified examiner
LSRP	26	4-point Likert	Self
SRP-II	60	5-point Likert	Self
SRP-III	31/40/62/64	5-point Likert	Self
PPI	187	4-point Likert	Self
PPI-R	154	4-point Likert	Self

To summarize, the PCL-R, assessed by a qualified examiner, is the most validated measure to determine Psychopathy, whereas the LSRP, SRP-II/III, and PPI(-R) are self-report measures developed with non-criminal, non-psychiatric samples that thus work better for non-forensic-research.

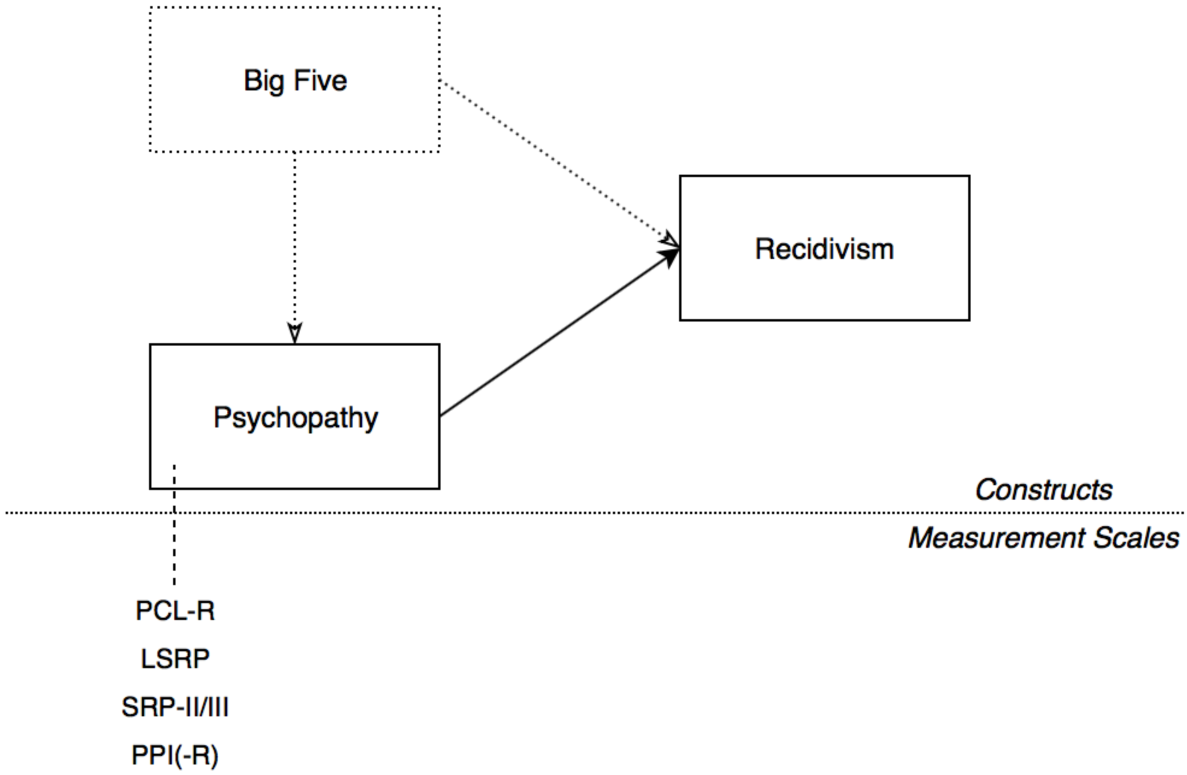

Fig. 3.3 Psychopathy can be assessed with different inventories

3.2.3 Psychopathy and Recidivism

Hart et al. (1988) showed that released psychopaths had higher rates for recidivism than other inmates. In another study, Harris et al. (1991) showed that psychopaths had a high recidivism (77%) compared to the much lower rate for non-psychopaths (21%). PCL-R scores correlate on an average of 0.27 with recidivism (Hemphill et al. 1998). Studies have shown that correlations between PCL-R Factor 2 and recidivism are higher than the correlations between PCL-R Factor 1 and recidivism (Hare et al. 1991). The ability of the PCL-R to predict recidivism has considerable cross-cultural generalizability (Hare et al. 2000). The relationship between the PCL-R Factor 1 and 2 and recidivism was also explored by Međedović et al. (2012) in two studies. The first study showed high correlations for criminal-legal recidivism (0.32 and 0.30) and for penal recidivism (0.31 and 0.32). High correlations were also found in the second study; 0.23 and 0.42 for criminal-legal recidivism and 0.24 and 0.19 for penal recidivism.

Our extensive literature search resulted in a handful ($n = 13$) of papers that studied the correlation between Psychopathy and recidivism. We focused on research that determined Psychopathy with the PCL-R, included a sample of the age 16 or higher, and that assessed criminal-legal recidivism. The results are laid out in the table below (Table 3.3).

Table 3.3 Correlation between Recidivism and PCL-R Factor 1 and 2

Study	N	Age	Factor 1	Factor 2
1. Harris et al. (1993)	618	Adult	0.22	0.34
2. Serin (1996)	81	18-59	0.14	0.36
3. Firestone et al. (1998)	78	Adult	0.09	0.23
4. Salekin et al. (1998)	78	Adult	0.26	0.14
5. Firestone et al (1999)	251	Adult	0.22	0.46
6. Grann et al. (1999)	318	16-72	0.23	0.45
7. Firestone et al. (2000)	192	Adult	0.27	0.47
8. Loucks and Zamble (2000)	81	Adult	0.25	0.47
9. Wilson (2000)	199	Adult	0.39	0.48
10. Kroner and Loza (2001)	78	18-64	0.20	0.36
11. Kroner and Mills (2001)	87	18-55	0.19	0.22
12. Loza and Loza-Fanous (2001)	68	18-64	0.24	0.37
13. Serin and Brown (2001)	263	18-59	0.17	0.31

So, the most trusted, accurate method to assess Psychopathy of an individual in a forensic environment is the PCL-R. Factor 2 of the PCL-R in particular correlates strongly with recidivism, and thus seems a reliable way to assess recidivism among suspects and prison inmates (Fig. 3.4).

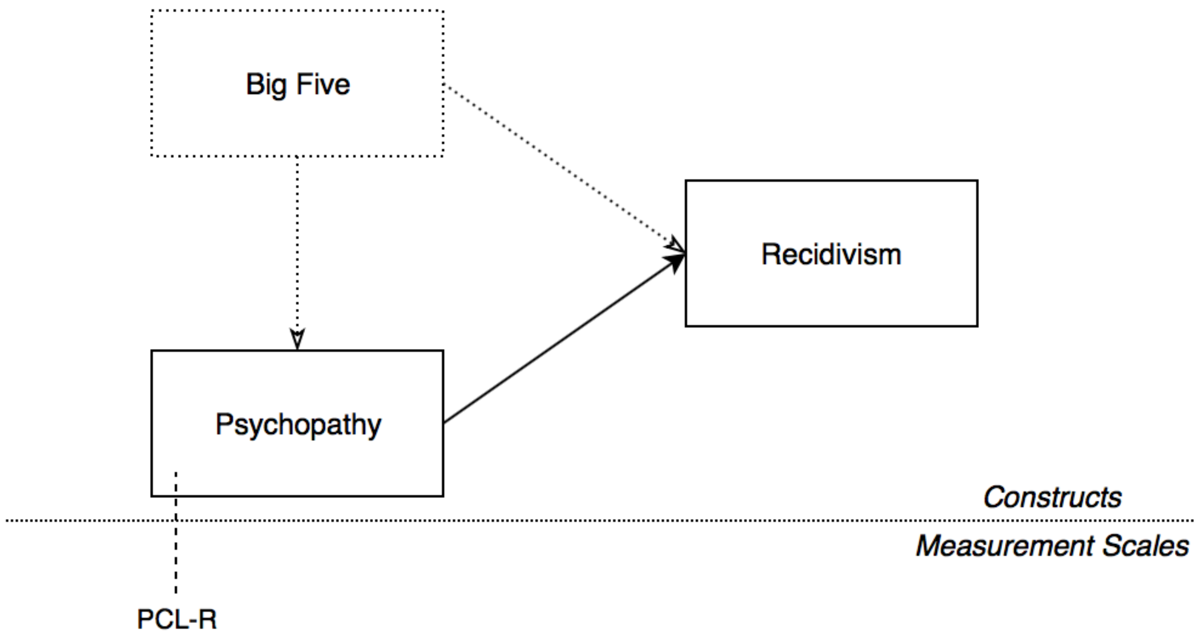

Fig. 3.4 Psychopathy assessed with the PCL-R is a good predictor for recidivism

3.2.4 Psychopathy and personal digital data from social media

Our literature search left us without any research that studied the relationship between Psychopathy measured with PCL-R and personal digital data from social media platforms. Only a few papers investigated how Psychopathy was linked to personal digital data from social media, but in none of the studies, Psychopathy was assessed with the PCL-R. However, many studies, instead, investigated the relationship between the Big Five personality traits and personal digital data from social media. Where the traits of the Dark Triad focus on undesirable personality traits that are affiliated with (for example) manipulation and misuse of others, the Big Five personality traits explain characteristics that apply to most people and interpersonal situations (Furnham et al. 2013). Interestingly, according to several studies, the Big Five personality traits are able to fully predict the qualities described by the Dark Triad (Brunell et al. 2008; Miller et al. 2001). It thus is worth investigating whether the Big Five traits can predict the construct Psychopathy as well.

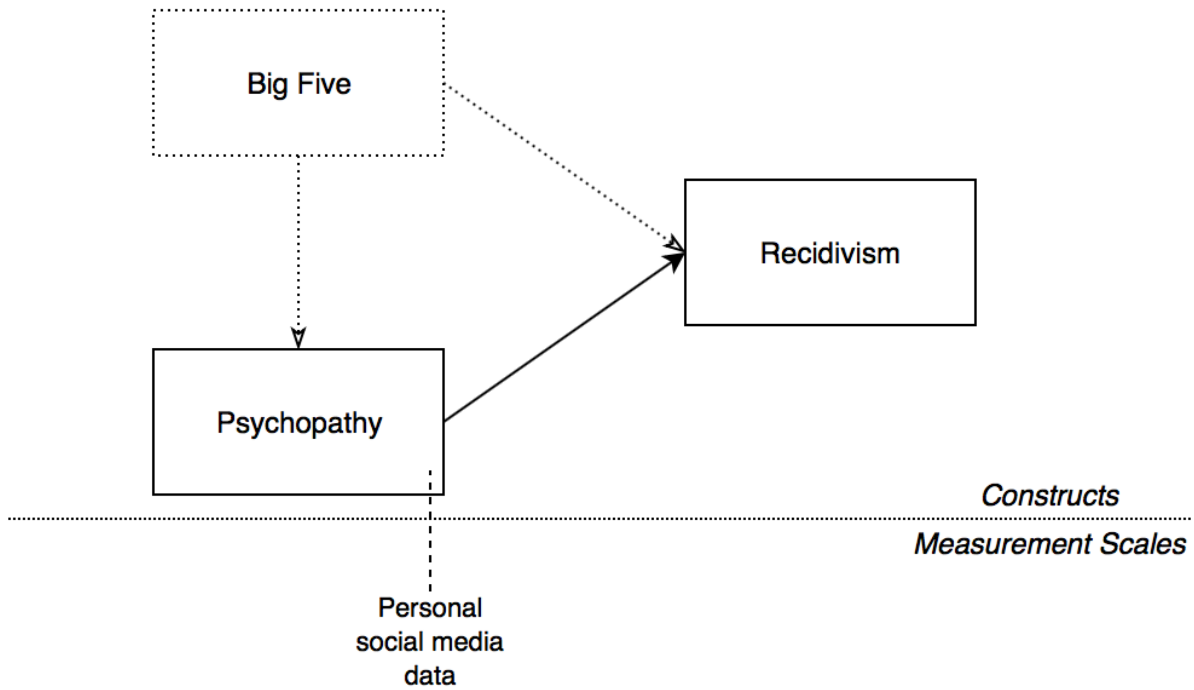

Fig. 3.5 Psychopathy can't be assessed using personal social media data

3.2.5 Conclusion

The construct of Psychopathy is one of the three traits (Machiavellianism and Narcissism being other two) of the Dark Triad, and is usually determined with the PCL-R. The PCL-R is carried out by qualified examiners (e.g. psychologists) to score subjects based on file information, and semi-structured interview to determine whether an individual can be specified as a psychopath. Unlike the PCL-R, also self-report scales (e.g. LSRP, SRP, PPI) exist to assess Psychopathy, but they are developed for non-criminal, non-psychiatric samples, and thus work better for non-forensic research. Studies have shown that Psychopathy, assessed with the PCL-R, seems to be a good predictor for recidivism. Factor 2 of the PCL-R in particular correlates strongly with recidivism, and thus seems a reliable way to assess recidivism among suspects and prison inmates. Unfortunately, no studies investigated whether personal digital data from social media can be used to assess Psychopathy.

However, plenty of studies examine the possibilities of using personal digital data from social media with the help of the Big Five. If we can find out that the Big Five can be assessed through personal digital social media data, and if we can find the relationship between the Big Five and Psychopathy, it follows that we can answer our research question. In the following section (3.3), we will first dive deeper in the Big Five personality traits, see how they can be measured, how they are related to recidivism, and how they can be assessed from personal digital data from social media. After that we will explore the relationship between the Big Five and Psychopathy in a new section (3.4).

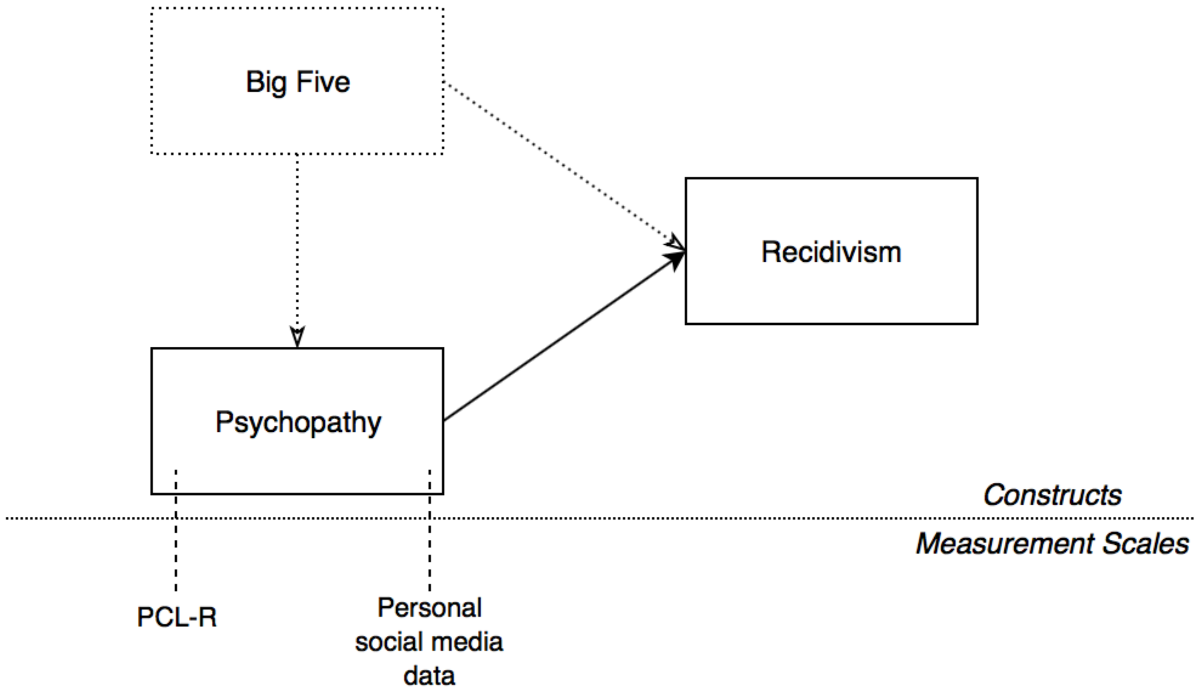

Fig. 3.6 Not possible to assess recidivism using personal social media data to determine Psychopathy

3.3 The Big Five

Although many models exist to describe one's personality, the most common used and accepted one is the Big Five model (McCrae and Costa 1987; John et al. 2008). In this section, we will first explain the Big Five personality traits in more detail (3.3.1), see how they are measured (3.3.2), then see how the Big Five is related to recidivism (3.3.3), and finally explore how the Big Five can be assessed from personal digital data from social media (Fig. 3.7).

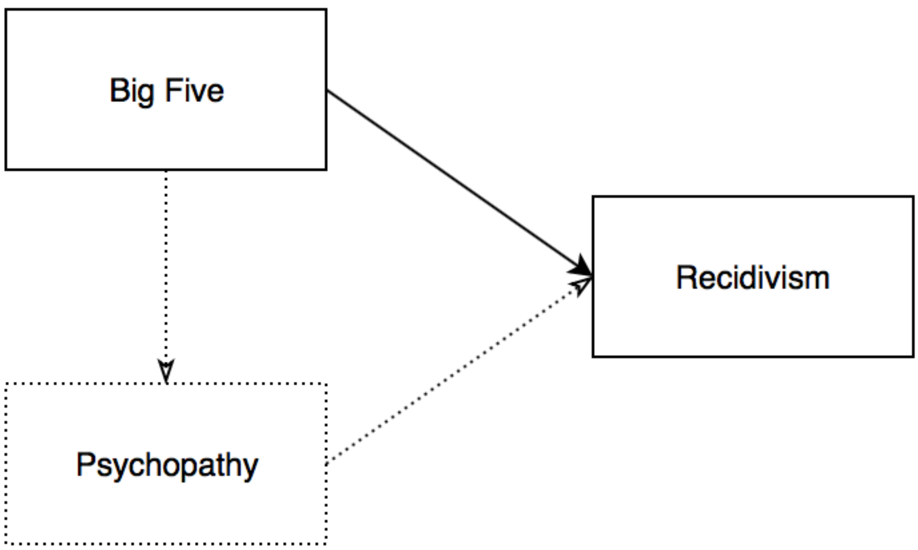

Fig. 3.7 Big Five and Recidivism: the relationship between the constructs we will examine in Section 3.3

3.3.1 The Big Five personality traits

The Big Five personality traits consist of 5 dimensions in total that form the personality structure which is supported with both longitudinal and cross-cultural evidence (McCrae and John 1992; McCrae and Costa 2003). The first dimension is Agreeableness, with traits including being courteous, flexible, helpful, nurturing, trusting, good-natured, forgiving, tolerant, cooperative and soft-hearted. Individuals scoring high for Agreeableness are trusting, empathic, and peace-keepers who are generally optimistic and trusting of others. While people low in Agreeableness are arrogant, manipulative, and not concerned about others.

Extraversion is the second dimension, with traits such as being sociable, assertive, outgoing, amicable and active. Extroverts tend to be friendly, outgoing, social, and energetic.

Conscientiousness being the third dimension, has traits such as being responsible, organized, hardworking, and achievement-oriented. It relates to the control of impulses, ability to plan, organize, and complete behavioural tasks. Conscientious people are reliable, and tend to be high achievers who work hard.

Common traits associated with the fourth dimension, Openness to Experience, includes traits like being imaginative, cultured, curious, broad-minded, artistically sensitive, curious, intelligent and original. Openness to Experience refers to an individual’s interest in culture, and desire for new activities and emotions. People with high scores tend to be artistic, sophisticated in taste and appreciate different views, ideas and experiences.

The final dimension, Neuroticism, which assesses emotional stability and adjustment, are being anxious, depressed, angry, sensitive, embarrassed, emotional, insecure, and worried. People scoring high on Neuroticism are moody, tense, and undergo negative emotions rather quickly.

The conventional way to assess the earlier mentioned five personality traits is by having the subject being tested fill in a questionnaire. Since there are a few different inventories that can be used to determine the Big Five of an individual, we will go through them in the next sub-section.

3.3.2 Measuring the Big Five

Below are several inventories to measure the Big Five personality traits elaborated: NEO Personality Inventory, Big Five Inventory and International Personality Item Pool.

The groundwork for measuring all of the Big Five personality traits with a questionnaire was laid by McCrae and Costa (1985) with their research and development of the NEO Personality Inventory (NEO PI). The initial personality assessment with this NEO PI was for the original three factors; Neuroticism, Extraversion and Openness to Experience (Costa and McCrae 1985). Each of these three factors included six facet sub-scales; a more detailed and specified aspect of a broader personality trait (Costa and McCrae 1985). The other two factors, Agreeableness and Conscientiousness, along with their six facet sub-scales were introduced in the Revised NEO-PI (NEO PI-R) by Costa and McCrae (1991). An overview of all the Big Five personality traits, along with their features as described in the NEO PI-R is provided in Table 3.4.

Table 3.4 Big Five personality traits including their facets as included in the NEO PI-R

Agreeableness	Extraversion	Conscientiousness	Openness to Experience	Neuroticism
Trust	Warmth	Competence	Fantasy	Anxiety
Straightforwardness	Gregariousness	Order	Aesthetics	Hostility
Altruism	Assertiveness	Dutifulness	Feelings	Depression
Compliance	Activity	Achievement Striving	Actions	Self-consciousness
Modesty	Excitement Seeking	Self-Discipline	Ideas	Impulsiveness
Tendermindedness	Positive Emotion	Deliberation	Values	Vulnerability to Stress

The final version of the NEO inventories is the NEO Personality Inventory-3 (NEO PI-3) and intended to make the inventory applicable to a wider portion of the population (Costa and McCrae 2010). Both the NEO PI-R and the NEO PI-3 forms consist of 240 items (five-point Likert scale). Since the NEO PI has been lengthy for many research applications, by requiring around 30-40 minutes to fill out the form according to Costa and McCrae (2010), a shorter measure was developed in the form of the 60-item NEO Five-Factor Inventory (NEO FFI) by Costa and McCrae (1992). The NEO FFI is a shortened form of the earlier mentioned NEO PI-R and takes about 10-15 minutes to fill out. Just like how the NEO PI-R got revised in 2005, a revised version of the NEO FFI, called the NEO FFI-3, was also published where 15 of the 60 items were replaced (Costa and McCrae 2010).

The Big Five Inventory (BFI) was constructed by John et al. (1991) to address the need at the time for a short instrument measuring the Big Five traits. Since the goal was to create a short inventory that could assess the five dimensions efficiently without the need for more differentiated measurement of individual facets, the 44-item BFI (BFI-44) was developed. Instead of using single adjectives as items, the BFI uses short phrases based on trait adjectives which are typical markers of the Big Five (John et al. 1991). This very same BFI-44 was later reduced to a 10-item version (BFI-10) by Rammstedt and John (2007) to present an inventory where subject time is extremely limited. The BFI-10 is particularly designed for large scale-scale assessments with limited time resources.

The International Personality Item Pool (IPIP) is a public domain personality inventory originally developed by Hendriks (1997) along with his colleagues and students at the University of Groningen in the Netherlands. In a paper delivered by Goldberg (1999) the IPIP project became international after the translation of the then 1311 items to English. Currently the IPIP has over 3000 items, with 400 scales to measure constructs like those in existing inventories which can be found on the IPIP website (https://ipip.ori.org). To assess one's personality on the Big Five model there is the IPIP-NEO (International Personality Item Pool - Neuroticism, Extraversion & Openness); a personality questionnaire. According to the website op the IPIP project (https://ipip.ori.org), the IPIP-NEO inventory contains 300 items and takes about 30-40 minutes for most people to complete. While the shortened version, which measures the same traits as the original but more efficiently, contains 120 items of the original and takes about 10-20 minutes to complete. There is a shorter version called the Mini-IPIP which contains 20 items and is developed and validated across 5 studies (Donnellan et al. 2006). Lastly, there is an even more brief measure in the form of the Ten-Item Personality Inventory (TIPI) for when extremely short measures are needed, or personality is not the most important domain (Gosling et al. 2003).

Table 3.5 Summary of the tools to measure the Big Five personality traits

	Total items	Minutes to complete
NEO PI-3	240	45-60
NEO PI-R	240	45-60
NEO-FFI	60	10-15
NEO FFI-3	60	10-15
BFI-44	44	7-10
BFI-10	10	<1
IPIP-NEO	300	30-40
IPIP-NEO (short)	120	10-20
Mini IPIP	20	3-5
TIPI	10	<1

There are thus a few inventories, along with their revised and/or shortened versions, to assess the Big Five personality traits of an individual (Fig. 3.8).

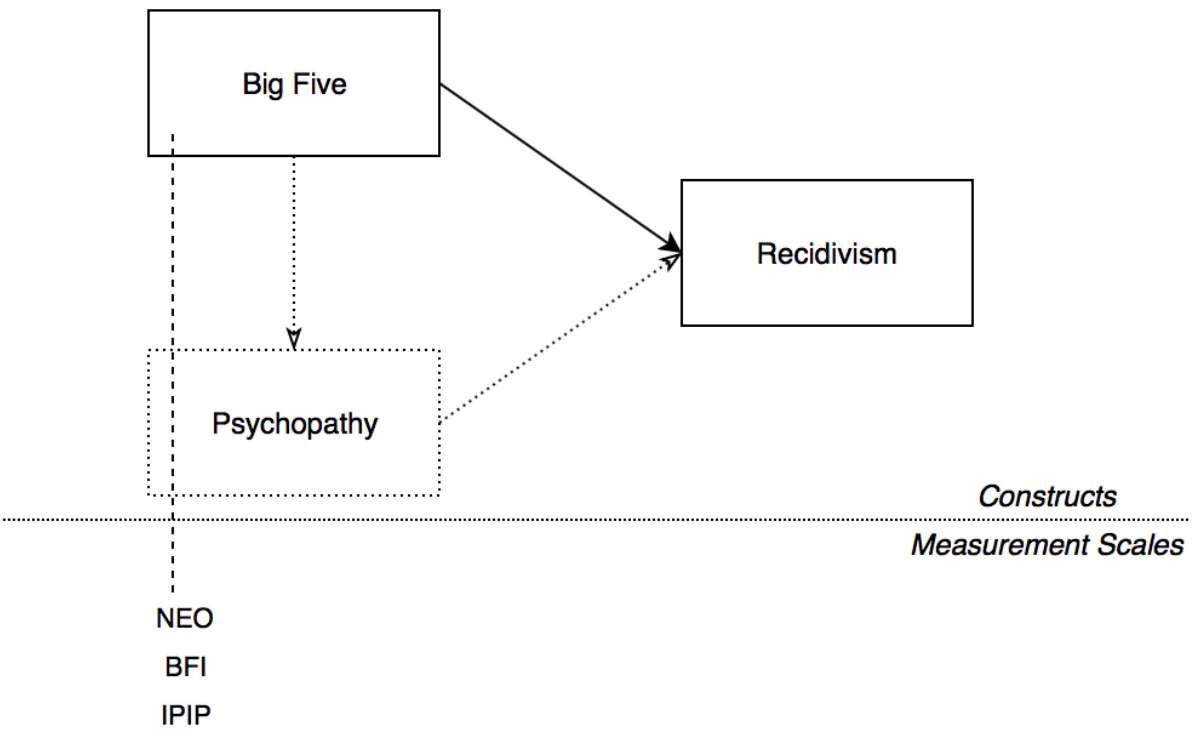

Fig. 3.8 The Big Five personality traits can be assessed with different inventories

3.3.3 Big Five and Recidivism

We could only identify three studies with our literature search that explored the relationship between the Big Five and recidivism. At the end of this sub-section the results from these studies are summarized in Table 3.6.

Clower and Bothwell (2001) examined the predictive validity of the Big Five with respect to recidivism. Their study included 51 inmates from the Lafayette Parish Correctional Center where the NEO-FFI was administered to the participants whose so called 'rap sheets' were also obtained from the correctional facility stating the number and types of crimes each had been arrested for. Multiple regression analysis showed that the combination of low conscientiousness and low openness to experience were associated with a significant increase in the number of arrests.

Van Dam et al. (2005) examined which of the two personality models PEN (Eysenck 1977) or Big Five, differentiated best between Dutch juvenile offenders and college students, and between recidivists and non-recidivists. The three basic PEN dimensions of personality are Psychopathy, Extraversion and Neuroticism (Eysenck 1977). The samples consisted of 96 male adolescents for the offenders with ages between 13 and 25 years old, and 204 male adolescents for the college students with ages between 15 and 24 years old. Since in our own study we focus on the Big Five personality traits, we will only elaborate both on the assessment of the Big Five traits, and their results, not the ones from PEN. The dimensions of the Big Five were measured by the Short Big Five Questionnaire (SBF) which consists of 30 adjectives that represent Extraversion, Conscientiousness, Agreeableness, Emotional Stability and Resourcefulness; where Emotional Stability corresponds with reversed Neuroticism, and Resourcefulness with Openness to Experience (Gerris et al. 1988). There were two types of recidivism investigated: self-report and official record. Self-report was measured by the Self-report list for delinquent behaviour (SRDB) which consists of 20 items each representing a criminal act, where having committed one or more of the 20 acts was considered a recidivist (Boendermaker 1998). Official records were obtained from the Criminal Justice Department of the Ministry of Justice. To examine the predictive validity of the two personality models, MANOVA's were carried out where univariate analyses for official record recidivism revealed that recidivists scored significantly higher than non-recidivists on PEN's Extraversion. For self-report recidivism, univariate analyses showed that recidivists scored significantly higher on Neuroticism, and lower on lower on Agreeableness of the Big Five model than non-recidivists.

Međedović et al. (2012) conducted two studies to explore personality-related determinants of recidivism. One study was conducted in one correctional institution (113 male participants), and the other in another one (112 male participants). In both studies personality was measured with the NEO-FFI and two types of recidivism was investigated: criminal-legal recidivism and penal recidivism. Data were analysed with hierarchical linear regression, where age and education level were introduced at the first level, and the five personality traits on the second. The first study showed that Agreeableness had significant negative β coefficient (-0.18) in the prediction of criminal-legal recidivism. Same as the first one, in the second study Agreeableness was the only trait to have significant predictive contribution (-0.22) in the prediction of criminal-legal recidivism.

Table 3.6 Correlations between the Big Five and Recidivism

	Agreeableness	**Extraversion**	**Conscientiousness**	**Openness**	**Neuroticism**
Clower and Bothwell (2001)	-	-	Negative	Negative	-
Van Dam et al. (2005)	Negative	-	-	-	Positive
Međedović et al. (2012)	Negative	-	-	-	-

Using the information from the three earlier mentioned studies we can conclude that the Big Five personality traits can't be used as indicators for recidivism.

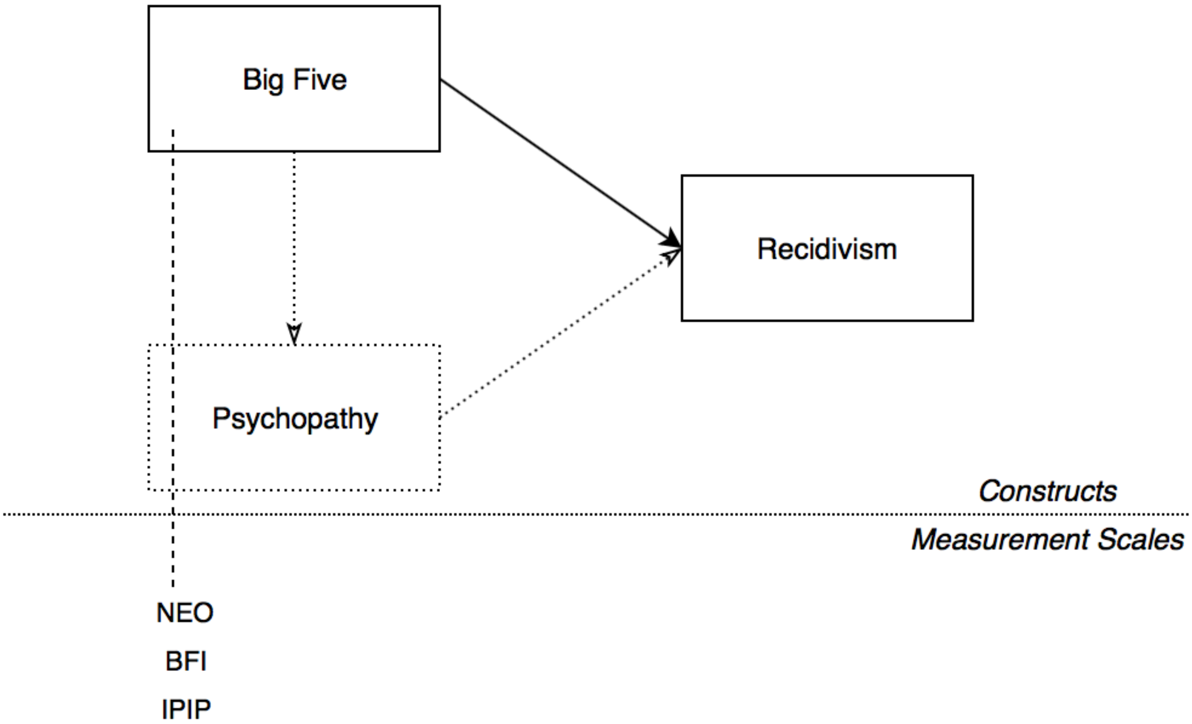

Fig. 3.9 Recidivism can't be determined with the Big Five personality traits

3.3.4 Big Five and personal digital data from social media

Usually there are two known approaches used to acquire and exploit personal digital data from social media to assess one's Big Five. The first approach is the (semi-)automated data mining approach, where algorithms are utilized to acquire information from public profiles (Kosinski et al. 2016). The participants are asked to complete a personality questionnaire before the data mining process starts. These questionnaires are used to evaluate the models that are developed, after data mining algorithms and machine learning analyse and correlate the activity of the user to personality traits (Bachrach et al. 2012). The other approach is in the form of an application (connected to the social media platform) which is installed and approved by the participants on their electronic device to share personal activity data. Like the previous method, participants are asked to complete a personality questionnaire beforehand (this time through the application). While this approach promises to deliver more data than the automated data mining approach, it does require participants to install the application and approve that their data will be collected through it. Social media activity and demographic features are extracted with the application, which in return allows the application to correlate patterns of behaviour to personality traits (Ortigosa et al. 2014).

People with a large amount of 'friends' on social media platforms have a high score for Extraversion, and they also seem to be engaged much more in social media (Kuss and Griffiths 2011; Kosinski et al. 2014). Whereas Seidman (2013) has found out that neurotic individuals reveal more information about themselves online and use social media to gain information about others, Schwartz et al. (2013) showed that the posts of these very same people contain more negative words than usual. The Conscientiousness with respect to personal digital data has been explored by Kosinski et al. (2014) and concluded that individuals who take a great deal in being cautious with their online profile (e.g. post and 'like' less), were the ones scoring high on Conscientiousness. People with larger social networks online, and who express many 'likes', appear to score high for Openness for

Experience (Quercia et al. 2012). People scoring high for Agreeableness seem to show less negative feeling with the content they share (Schwartz et al. 2013).

The research design of researchers studying the use of digital data from people using social media to retrieve the Big Five personality traits has been consistent thus far. The researchers assess the Big Five of the subjects with self-reports questionnaires. After that, they collect the digital data from the social media platforms. This data is then processed to acquire variables for the models for prediction. At the end, strength of the prediction for the Big Five is determined from the features. One point where the studies do vary, is in the type of social media data (e.g. activity, language), used. Another point is the social media platforms which are used (e.g. Facebook, Instagram, YouTube etc.). Their findings, the accuracy of their predictability, vary across traits, and are all written down in detail.

In their study, Liu et al (2014) used micro-blogging behaviours (language/text) of users of Chinese Sina Weibo to predict the personality of the users. They used different combinations of digital data (e.g. activity, language, pictures). By using bloggers found on Google's Blog Search engine (blogsearch.google.com) Yarkoni (2010) analysed their word use to assess personality. Facebook was used as the social media platform by Schwartz et al. (2013) to predict personality using textual features from status updates, and by Kosinski et al. (2013) also to predict personality but with using Facebook Likes. The prediction for each of the personality traits is not equally accurate in all the studies. Both Yarkoni (2010) and Kosinski et al. (2013) had considerable higher correlations for Openness to Experience than any other trait. Besides the type of social media platform, and the type of social media data, studies also vary considerably in sample size (from under one hundred to tens of thousands).

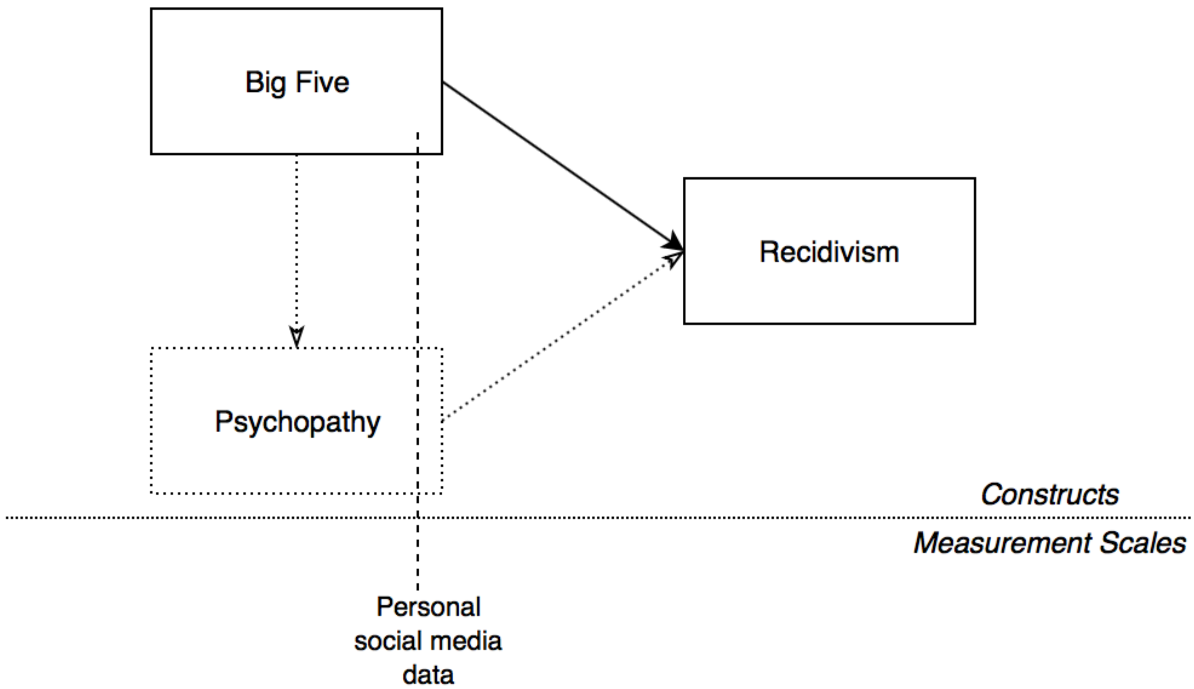

Fig. 3.10 Studies have shown a strong link between personal social media data and the Big Five

3.3.5 Conclusion

The most well regarded and accepted theoretical framework to describe one's personality is the Big Five model, which consists of 5 dimensions; Agreeableness, Extraversion, Conscientiousness, Openness to Experience, and Neuroticism. To determine the Big Five personality traits of an individual, there are many instruments, along with their abbreviations and revisions, which are all valid and reliable (e.g. NEO-PI, BFI, IPIP). Depending on the type of research conducting, the composition of the sample, the maximum amount of time and effort the assessment should take, one can choose any of the earlier mentioned instruments to measure the Big Five. Since in our research we are interested whether we can use personal social media data to determine the risk of recidivism, we took the time to see how the Big Five traits are related to recidivism, but were left with three papers that had contradicting results. Only Agreeableness seems to correlate negatively with recidivism. However, we did find out that the Big Five traits could be determined from personal digital data from social media.

We found many studies that in their core, had the same objective; to predict the Big Five personality traits using personal social media data. However, they varied in the type of social media platform used (e.g. Facebook, Instagram), the type of digital data extracted (e.g. demographic information, activity, pictures), and the sample size used (from under one hundred to tens of thousands). Extraction of the digital data can be done in one way or another; either by data mining (with algorithms) or with the use of an application to be installed by the participant. So now we have the promise we can use personal social media data, to determine the Big Five, and we know from the previous section (3.2) that recidivism is correlated with Psychopathy. We will explore the last remaining relationship, the one between the Psychopathy and the Big Five, in the next section.

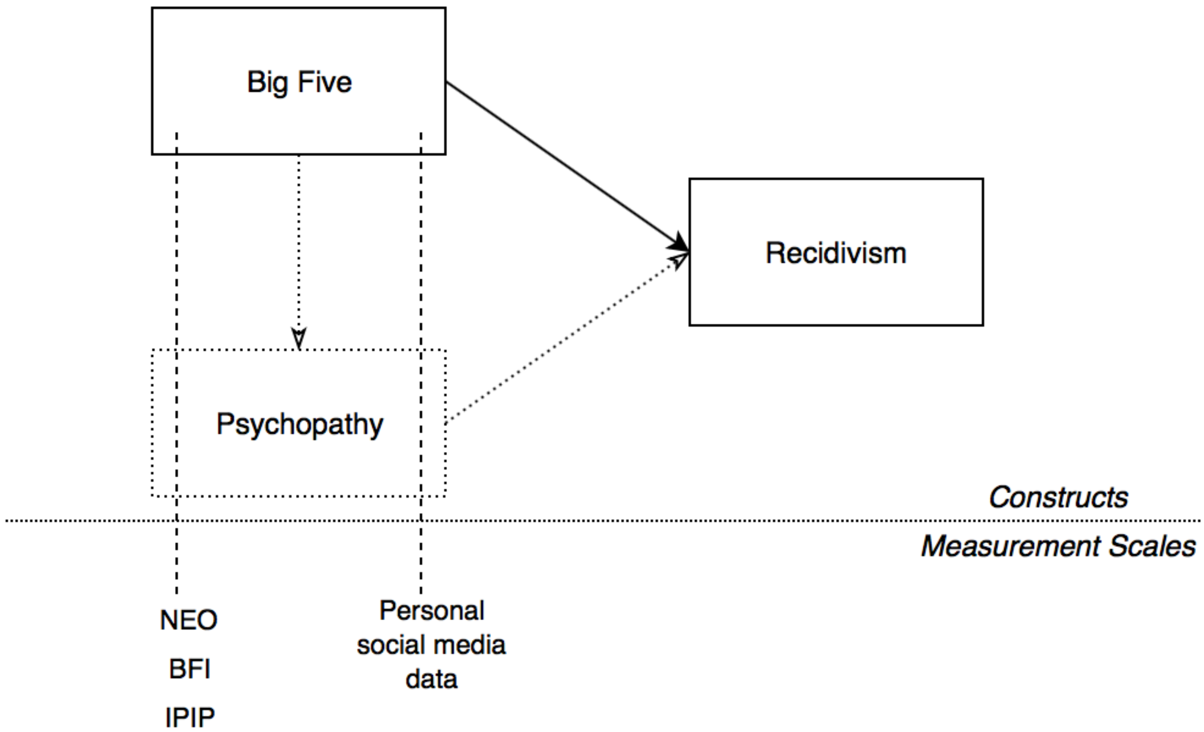

Fig. 3.11 Not possible to assess recidivism using personal social media data to determine the Big Five

3.4 Psychopathy and the Big Five

We have examined three papers where the correlation between the Dark Triad (Psychopathy, Machiavellianism, and Narcissism) and the Big Five is studied. In this section, we will briefly go through the three studies and present their results focusing only on the correlations between Psychopathy and the Big Five (Fig. 3.12)

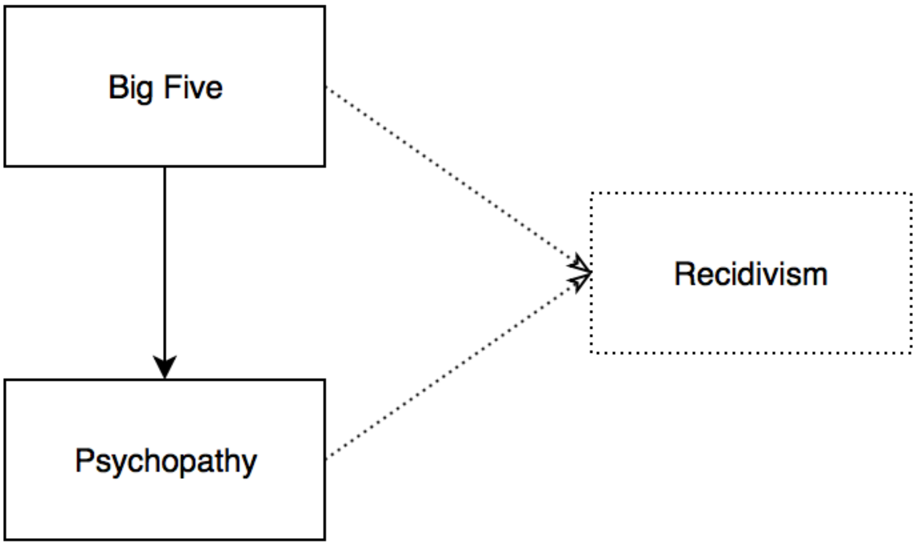

Fig 3.12 Big Five and Psychopathy: the relationship between the constructs we will examine in Section 3.4

3.4.1 Psychopathy and the Big Five

The relation between the Dark Triad and the Big Five was investigated by Paulhus and Williams (2002) by using a sample of 245 undergraduate psychology students whose traits were measured with the BFI (for the Big Five), NPI (for Narcissism), MACH-IV (for Machiavellianism), and SRP-III (for Psychopathy). Since our interest only lays in the relationship between Psychopathy and the Big Five, we will only discuss the results for those two constructs. Psychopathy seemed to correlate negatively with Agreeableness, Conscientiousness and Neuroticism, and positively with Extraversion and Openness to Experience. The results of the correlations between Psychopathy and the Big Five can be found in Table 3.7.

Table 3.7 Correlations between Psychopathy and Big Five from Paulhus and Williams (2002)

	Psychopathy
Agreeableness	-0.25
Extraversion	0.34
Conscientiousness	-0.24
Openness to Exp.	0.24
Neuroticism	-0.34

Jakobwitz and Egan (2006) recruited 82 persons (from the general population) and had them fill out the NEO-FFI-R (for the Big Five), the MACH-IV (for Machiavellianism), the LSRP (for psychopathy), and the NPI (for narcissism) in order to examine to what extent the Dark Triad traits reflect the same underlying construct, and to what extent the Big Five personality traits could capture the constructs of the Dark Triad. Correlations between the Dark Triad measures and the NEO-FF-I were calculated and principal components analysis with Varimax rotation of the derived factors was calculated to simplify measures of the Dark Triad and personality, and to examine whether all three scales of the Dark Triad reflected the same underlying construct. We narrowed our focus to the results between Psychopathy, this time expressed as Primary and Secondary Psychopathy (corresponding to Factor 1 and 2 respectively of the PCL-R) in the study, and the Big Five, and presented them in Table 3.8. Significant negative correlations between Agreeableness and both Primary and Secondary Psychopathy, and significant positive correlations between Neuroticism and again both Primary and Secondary Psychopathy were found. From the factor analysis, using Varimax rotation, it could be concluded from the first factor that low score on Agreeableness is associated with high scores on both Primary and Secondary Psychopathy. The second factor contrasted a high negative loading for Conscientiousness, and a high positive loading for Secondary Psychopathy and Neuroticism. Both Openness to Experience and Extraversion loaded on separate factors thus are entirely unrelated to Psychopathy. So, in contrast to the work described previously by Paulhus and Williams (2002), no correlation was found for both Openness to Experience, and Extraversion, and Neuroticism was positively correlated, instead of negatively.

Table 3.8 Correlations between Psychopathy and Big Five from Jakobwitz and Egan (2006)

		Psychopathy
	Primary (Factor 1)	**Secondary (Factor 2)**
Agreeableness	-0.43	-0.23
Extraversion	0.08	0.04
Conscientiousness	-0.21	-0.19
Openness to Exp.	-0.21	-0.21
Neuroticism	0.30	0.47

O'Boyle et al (2015) examined the relationship between the three traits of the Dark Triad and the Big Five by way of a meta-analysis which included 310 independent samples in total drawn from 215 sources. Table 3.9 summarizes the results for the correlations between Psychopathy and the Big Five traits, and shows that Psychopathy was negatively related to both Agreeableness and Conscientiousness.

Table 3.9 Correlations between Psychopathy and Big Five from O'Boyle et al. (2015)

	Psychopathy
Agreeableness	-0.42
Extraversion	0.04
Conscientiousness	-0.31
Openness to Exp.	0.04
Neuroticism	0.05

3.4.2 Summary

The results of the three different studies, with both overlapping and contradictory results, are summarized in Table 3.10. Both Agreeableness and Conscientiousness seem to correlate significantly negatively with Psychopathy throughout the studies. Since Psychopathy can be described by a set of interpersonally aversive qualities (e.g. social manipulatives and disrespect for other people's feelings), it was to be expected to be negatively associated with Agreeableness and Conscientiousness (traits based on respect, harmony, abidance to societal order). On one hand, one can think that, since individuals scoring high on Extraversion thrive on excitement, and are action oriented people, and so psychopaths would score high on Extraversion as well. However, other psychopathic traits like emotionality and the inability to emphasize with others, will reduce the amount of reward gained from interacting with others, thus lower the score for Extraversion. Neuroticism also has two sides for it; psychopaths can be seen as impulsive in nature, lack control of impulse and thus burst out into extreme actions, but on the other side the lack of anxiety can be the very reason how they perform the most vicious acts that are contributed to psychopaths. Openness to Experience has facets that can be associated with psychopathy (e.g. active imagination and being open for the unusual), but it also has facets that would correlate negatively (e.g. openness to feelings and one's personal values).

Table 3.10 Correlations between Psychopathy and the Big Five from three different studies

	Paulhus and Williams (2002)	**Jakobwitz and Egan (2006)**		**O'Boyle et al. (2015)**
		Primary (Factor 1)	**Secondary (Factor 2)**	
Agreeableness	-0.25	-0.43	-0.23	-0.42
Extraversion	0.34	0.08	0.04	0.04
Conscientiousness	-0.24	-0.21	-0.19	-0.31
Openness	0.24	-0.21	-0.21	0.04
Neuroticism	-0.34	0.30	0.47	0.05

So, in the end, whereas Agreeableness and Conscientiousness correlate negatively with Psychopathy, the association of Extraversion, Neuroticism and Openness to Experience is far less certain.

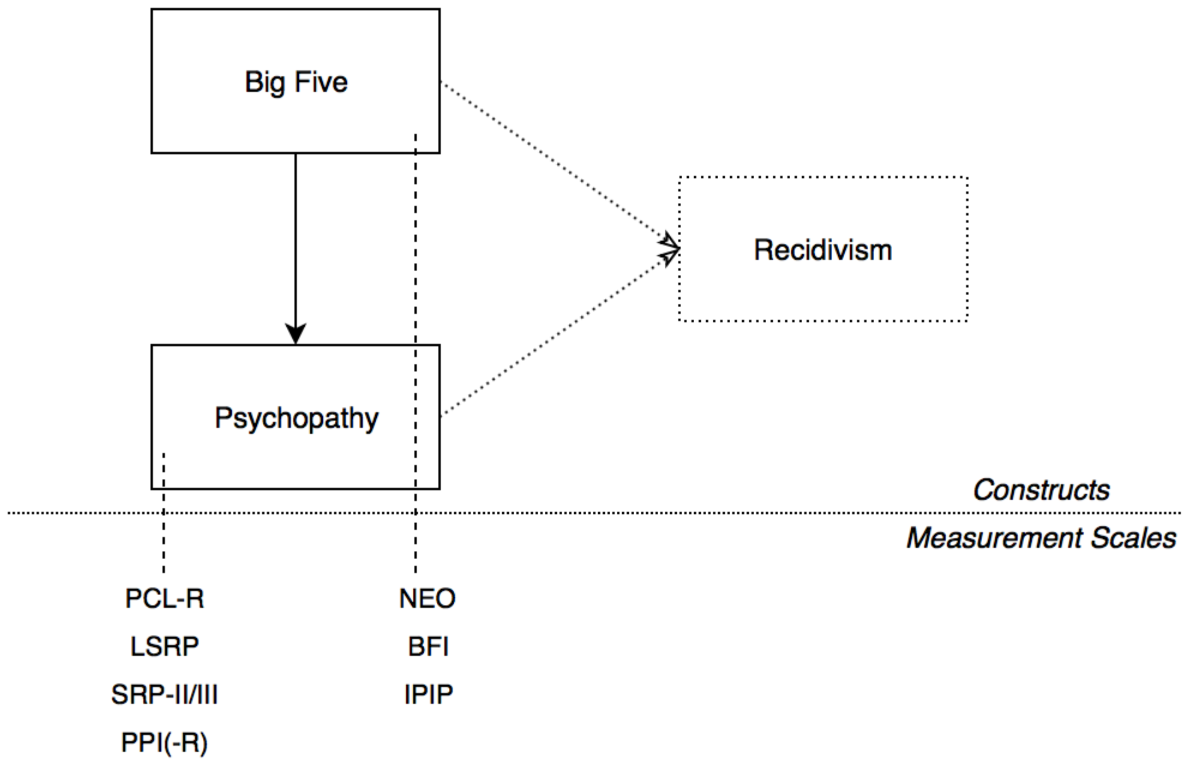

Fig. 3.13 There is a strong link between the Big Five and Psychopathy

3.5 Conclusion

In the previous chapter (Ch. 2) we studied how currently risk of recidivism is assessed in the Netherlands, and looked for areas for improvement for assessing the risk more efficiently and accurately. We found that the current method was not very accurate in predicting recidivism, takes a whole day of work (8 hours) for a certified probation officer, and can only be executed for already convicted individuals (no suspects awaiting trial). In this chapter, we studied the possibilities to extract personal social media data to eventually assess recidivism. We investigated how personal digital data from social media can be used to determine personality attributes (Psychopathy and Big Five), and whether these personality attributes can be used to predict recidivism.

Psychopathy as a construct is most accurately determined with the PCL-R which is divided in Factor 1 and Factor 2 items. The Factor 2 items have shown to be good indications to predict recidivism with correlations around 0.35-0.40; more than the 0.30 from the RISc (currently used to assess recidivism is the Netherlands). However, we could not find any papers where the link between Psychopathy and personal social media data was studied. Since it was impossible to extract and exploit personal social media data to measure Psychopathy, to predict recidivism, we investigated the Big Five personality traits.

For the Big Five, we only found the trait Agreeableness to correlate negatively with recidivism. Unlike Psychopathy, the Big Five is not a strong predictor for recidivism. We did however found both Agreeableness and Conscientiousness to correlate negatively with Psychopathy. Since our main goal in our research is to explore the possibilities to assess personality attributes using personal social media data to predict recidivism, we looked for studies that used personal social media data to determine the Big Five personality traits. We found many studies that investigated the possibilities of using personal social media data to determine the Big Five personality traits. While these studies have shown that the Big Five traits could indeed be assessed from personal social media data, it must be noted that they found different results because they used different social media platforms, data types, models, and sample sizes. We want to explore what the reasons may be for the different results from the studies that wished to determine the Big Five from personal social media data. By performing a meta-analysis, we will be able to find out why studies had varying results; to which variables it can be accounted to. So, with the meta-analysis, not only we will see how accurately social media data can predict the Big Five, we also aim to find out what variables influence the accuracy of the prediction, thus are a significant predictor for the Big Five assessed from personal social media data. In the next chapter (Ch. 4) we have our method chapter where we elaborate in detail what a meta-analysis is, and how we conducted our own for our research.

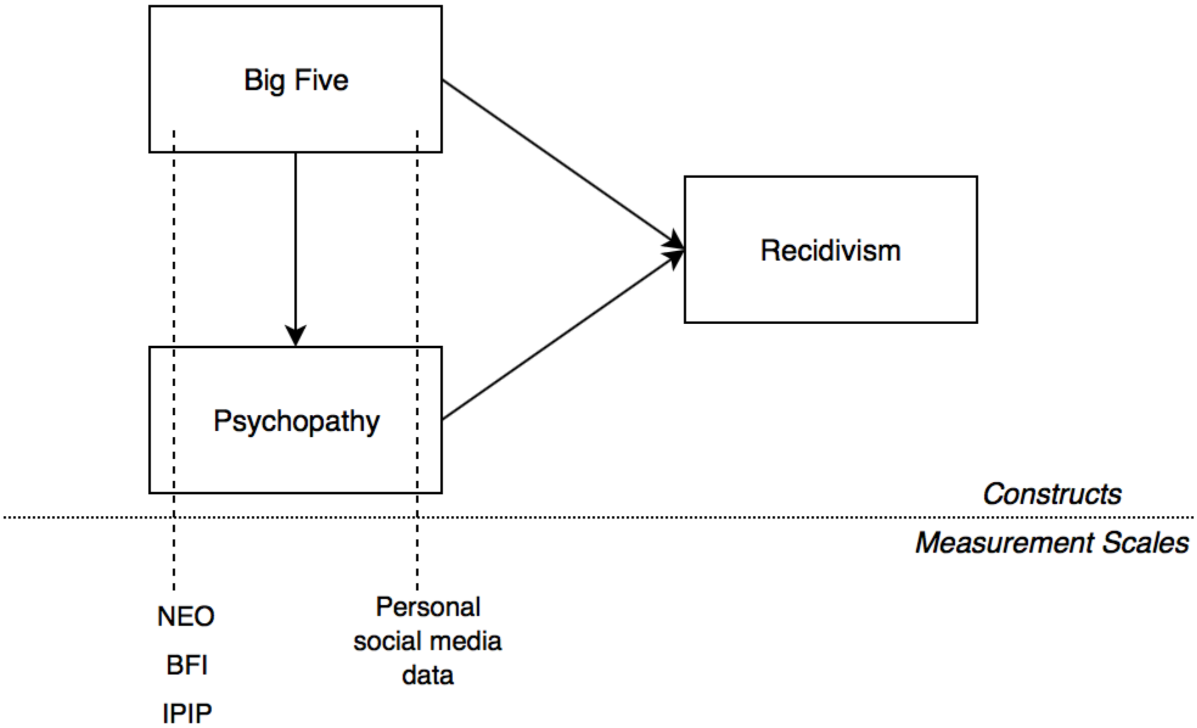

Fig. 3.14 The focus in our meta-analysis will be the relationship between personal social media data and the Big Five

CHAPTER 4: METHOD META-ANALYSIS

In the previous chapter (Ch. 3) we discovered that the Big Five can be assessed from personal social media data. However, studies have reported different results since they varied in many aspects. Due to the great variety of methods used, we will perform a meta analysis to explore how accurately the Big Five can be predicted with social media data, and what variables influence the accuracy of prediction.

Before we start with our meta-analysis, we will first start with a brief historical introduction of the origin of meta-analysis, to see how it is formed, and how its strengths and weaknesses have been tackled (4.1). After the introduction, we will have a section where we will elaborate on the different types of meta-analyses, and classify our own accordingly so we can conduct it properly (4.2). The following six sections (4.3 - 4.8) are each dedicated to a particular step in our own meta-analysis (totalling six steps) following the guidelines by Lipsey and Wilson (2001). At the end of this chapter, in section 4.8 (the sixth step), we will give our conclusions and interpretations of our meta-analysis. After having conducted our meta-analysis in this chapter, we have our final chapter (Ch. 5); Discussion.

4.1 Introduction

Eysenck (1952) argued that psychotherapy was not beneficial for patients, and that the majority of patients with mental health problems get better anyway, whether or not treated with psychotherapy. Some two decades later, several hundreds of studies on psychotherapy not only produced results that supported Eysenck's claim, but also proved him wrong. It was not until 1977 when Glass, along with his colleague Smith, countered the findings of Eysenck in a convincing manner by statistically standardizing and averaging the effect of a few hundred studies, totaling a sample size of 50.000 patients, thus giving birth to a new method they called "meta-analysis".

Meta-analysis can be understood as a review strategy, since it reviews research literature to summarize the results in a particular research domain (Cooper 1986). What sets meta-analysis apart, is the statistical analysis of the distribution of findings across studies; it is a statistical procedure for combining data from multiple studies. According to Lipsey and Wilson (2001), meta-analysis is a method to encode and analyse statistical data that summarizes study findings, and is only applicable to empirical research studies that produce quantitative findings. Since the aim of a meta-analysis is to combine and compare results of different research studies, findings need to be conceptually comparable and configured in similar statistical form for both practical and conceptual reasons. The findings of studies in the meta-analysis are so-called effect-size: a statistical concept that measures the strength of the relationship between variables. The effect size statistics make it possible to compare studies that use different operationalisations (measurement procedures), thus yielding different numerical values, because it is a form of standardization (Durlak and Lipsey 1991). The statistical standardization, produced by the effect size statistic, makes it possible to interpret, combine, and compare the different numerical values of different study findings (Lipsey and Wilson 2001). Since there are many possibilities to define an effect size statistic, it is important for the researcher to select one that contributes to appropriate standardization for the numerical comparison and analysis across studies.

The theoretical and statistical foundation of meta-analysis seemed sound when it was first introduced in 1977 by Smith and Glass: it introduces a practical way to summarize research findings, it presents crucial research findings in a more comprehensive and refined way than with a traditional review procedure, it is capable of finding effects which are unable to retrieve through other approaches to summarize studies, and it is an organized method to handle information from a large number of study findings (Lipsey and Wilson 2001).

However, meta-analysis was not initially received without criticism (especially from Eysenck), and even up till now, has its critics. Hunt (1997) argued that if the quality of the studies that were included in the meta-analysis was not good, the meta-analysis itself could never be better than the included studies itself -- the 'garbage in, garbage out' principle. While there are different ways to set extremely strict methodological criteria for the inclusion of study findings, the fact remains that there still exist difficult trade-offs, because there is very little consensus among researches on how methodological quality is measured. This issue is either tackled by keeping strict methodological criteria, thus summarizing merely a narrow research domain having little generality, or by coding methodological characteristics to see their influence on study findings later on in the statistical analysis.

Another point of criticism with is the so-called the 'apples and oranges' issue - i.e., when studies included in the meta-analysis don't actually deal with the same constructs and relationships (Sharpe 1997). There exists a grey area where results from studies are not exact replicas, but they are related more in a general meaning. Over the course of time however, with more research in meta-analysis along with technical advances, it became possible to assess homogeneity of gathered effect sizes (Lipsey and Wilson 2001). The meta-analysts can thus test empirically the degree to which studies show different results that it should not be assumed they are comparable.

Since its introduction, meta-analysis itself has been researched more thoroughly, mainly also due to the reasons of criticism mentioned earlier. The result of research into the domain of meta-analysis has produced several practical guides, for researchers wishing to perform a meta-analysis themselves. Lipsey and Wilson (2001) have outlined six steps that one should follow to perform a meta-analysis. Before we start with our own meta-analysis, we first need to know with what type of meta-analysis we will deal with. Therefire we will explore the different types of meta-analyses first in the next section (4.2)

4.2 Types of Meta-Analyses

Before laying out the fundamental six steps to perform a meta-analysis, we will shortly focus on the different types of meta-analyses. Since the different types of meta-analyses vary in terms of purpose, unit of analysis, treatment of study variation, and products (Bangert-Drowns, 1986), it is important to determine beforehand the type a meta-analyst deals with. Two major categories can be distinguished, along with two subcategories:

1. *Group Contrast Meta-Analysis* (revolve around a group contrast)
    a) Treatment effective meta-analysis: surveys a research on a defined treatment domain. The difference between the treatment and control group mean is represented by a standardized effect size.
    b) Group differences meta-analysis: surveys research on differences between (naturally occurring) groups (e.g. males and females). The difference between the means of the groups on a variable of interest is, like treatment effective meta-analysis, represented by effect size.

2. *Correlation Association Meta-Analysis* (revolve around correlational relationships)
    a) Test validity meta-analysis: research on test validity by examining the correlation between a test/measure and a criterion variable. Unlike with the group contrast meta-analysis, the primary statistical indicator is not the effect size, but the product-moment correlation.

b) Covariation meta-analysis: research on the covariation of two or more variables of interest (e.g. relationship between alcohol use and domestic violence).

However, according to Lipsey and Wilson (2001), there is an addition of two more forms in which research findings may be categorized:

3. *Central Tendency Description*: research describing a variable measured on a single sample of respondents. (e.g. a survey study reporting the percentage of men having migraine).

4. *Pre-Post Contrast*: like the Central Tendency Description, is also from of a single sample comparing the central tendency (e.g. mean) on a variable. But, this time the same variable is measured at one time, and again at another moment in time, to investigate the change (e.g. how much faster athletes are at the end of a training program than at the beginning).

Since we are interested in the correlation between an established method (Big Five assessed with questionnaires) and a new method (Big Five assessed with personal social media data), our meta-analysis falls perfectly in the category of 'test validity meta-analysis' of the *Correlation Association Meta-Analysis*. In order to statistically analyse the study findings in a research paper, one must go through the steps of a meta-analysis one by one with great care; the steps involved in a meta-analysis are like links in a chain, and the finished product is only as good as the weakest link (Durlak and Lipsey 1991). In the following section, we will start with our meta-analysis, conducting it following the guidelines provided by Durlak and Lipsey (1991) and Lipsey and Wilson (2001).

4.3 Step 1: Research Question and Inclusion Criteria

Specific research questions, formal hypotheses, and the major variables of importance need to be made explicit. By inspecting the relevant literature before the meta-analysis, the researcher can sharpen the research questions and better anticipate issues that may come up later on in the research (Nurius and Yeaton 1987). So, meta-analysis needs to start with an accurate statement of the field that will be explored; this statement will determine the choice of studies that will be taken in the meta-analysis. It is important that the problem statement is straightforward and complete, but does not need to be highly detailed before specific inclusion and exclusion criteria are developed. We had already constructed our research question (Ch. 1) following our main objective:

How can personal digital data extracted from social media, as an alternative for existing ways to measure personality attributes, efficiently and accurately contribute to determination of the recidivism probability of an individual?

Specific inclusionary criteria need to be determined in order to provide a definition of the population of studies to be assessed. Some recurring issues in defining the relevant population of studies are: whether to include unpublished studies as well, whether to include all studies or only those that meet certain methodological criteria, or what the time-period needs to be covered in the literature search (Glass et al. 1981; Smith et al. 1980). Following our main objective and our research question, we developed the following inclusion criteria which studies had to meet to be included in our meta-analysis:

1. All the Big Five personality traits are assessed by way of a standardized self-report measure (e.g. BFI, BFI-10, IPIP, Mini-IPIP, TIPI).
2. The relationship between the Big Five and personal digital data from social media is measured on an individual level.
3. Digital data is collected automatically from the social media platform.
4. Statistical data reporting how accurately the Big Five is predicted is given.
5. The reported data are statistically independent (i.e. studies using overlapping samples won't be included).
    a) Studies in the same research paper did not use the same social media platform.
    b) Studies in the same research paper did not use the same type of social media data to assess the Big Five.

After determining the inclusion criteria, it was time to start the literature search (4.4), bearing the inclusion criteria in mind, to find relevant studies that we could include in our meta-analysis.

4.4 Step 2: Literature Search

To identify and obtain relevant a unbiased sample of studies a search strategy needs to be realized. Literature can be found using multiple groups of keywords in online databases (e.g. Scopus, Web of Science), with searches on the world wide web (https://www.google.com), and by checking the citations of the studies we found from the databases. In the end, there will be studies that meet the inclusionary criteria, but won't be taken in the final meta-analysis. They will be overlooked or will lack the statistical information. To test whether the included studies are a representative sample of the available evidence, the presence of publication bias will be checked on a later stage in the meta-analysis (4.7). In this section, we will explain the search strategy that we used to find relevant studies to include in our meta-analysis. The search strategy resembles the one we used in Chapter 3 for our literature review, but the purpose and scope is of course different this time. We are now interested in finding studies that we can include in our meta-analysis - i.e., studies that have explored the correlation between the Big Five assessed with a self-report inventory, and assessed using personal social media data.

Studies on the relationship between the Big Five personality traits and personal social media data were explored by conducting a literature search in three online databases - Google Scholar, Web of Science and Scopus. We used multiple combinations of keywords that we firstly divided into three groups: one referring to the Big Five personality traits, the other to social media platforms, and the last one to analytic approaches with respect to online data extraction. A combination of the keywords from each category was used to search for relevant studies (Table 4.1).

Table 4.1 Keywords for literature search

Big Five	Social Media	Analytics
Big Five	Social media	Data extraction
Big 5	Social network	Data mining
Five Factor	Facebook	Content analysis
5 Factor model	Instagram	Machine Learning
Personality	Twitter	Big data analytics
Agreeableness	Snapchat	Text analysis
Conscientiousness	LinkedIn	Text mining
Extraversion		Image analysis
Openness		
Neuroticism		

First, we looked for the terms for our search in the title, abstract, and keyword section of the scholarly literature, and comprised our initial list of potential studies for the meta-analysis. Then we examined the reference list of the studies selected from our initial broad search. Finally, we inspected the citations of the included publications to find additional studies. Our study search yielded a total of 1021 articles initially, which got reduced to 825 potential articles for our meta-analysis after the removal of duplicates.

The next step was to read the abstracts of all the 825 studies to assess whether they met our specific inclusion criteria (4.3). After reading the abstracts of all the 825 papers, we were left with 50 articles which showed relevance to our study. Evidently, we have excluded a major part of our initial selection of studies, which can be accounted for having used (1) not the correct keywords for our search, and/or (2) looked in the wrong databases. We do have to note here that we excluded many studies because they used social media platforms not relevant for our research (e.g. Sina Weibo because it is not used in the Netherlands), or they did not extract social media data online, but had participants fill out questionnaires about their social media use.

The total of 50 papers, which we were left with in the end, were fully read so we could be certain they met our specific inclusion criteria. Besides determining whether the 50 papers met our inclusion criteria by reading them in full, we also extracted certain information from each paper, and coded them according to our coding scheme, which is described in the next section (4.5).

4.5 Step 3: Coding

Studies need to be coded for all the characteristics that potentially influence study findings. Because the importance of particular variables varies across research areas, the meta-analyst needs to specify the variables vital for a certain research area and then code for each of them. However, since it is impossible to specify all the variables to be coded in the meta-analysis, Durlak and Lipsey (1991) suggested that coding for the following has proven to be useful: study context, methodological characteristics, subject/client/sample characteristics, characteristics of tasks or interventions, and effect size. The meta-analyst needs to determine and report the level of intercoder agreement attained in coding the studies according to Stock et al. (1982).

In this section, we will code the studies ($n = 50$) from the articles whose abstract showed relevant significance for our research study; promising to fit for all the inclusion criteria described earlier (4.3). All the 50 papers, collected with our literature search (4.4), were fully read to assess whether they met our inclusion criteria, and to code for the five characteristics which we have distinguished that could potentially influence study findings (i.e. higher/lower correlation). Based on the coding procedure for the characteristics in the next sub-section, we will code the studies for potential moderators in the statistical analysis (4.7).

4.5.1 Coding Process

We will elaborate on the five characteristics we chose to code our studies for, what they entail, and why we chose them:

1) Study Quality
The sources of the studies analysed were either peer-reviewed journals, or conference proceedings. We assessed the study quality of both of the sources differently; peer-reviewed journals were ranked based on Scimago Journal & Country Rank (https://www.scimagojr.com/index.php), conference proceedings were ranked using CORE Conference Ranking (http://portal.core.edu.au/conf-ranks/).

For ranking the journals, we used the SCImago Journal Rank indicator (SJR), developed by Scimago which shows the visibility of the journals that are in the Scopus database (SCImago n.d). Journal quality was ranked as high when it fit in the highest quartile (1), as medium when it fit in the second quartile (2), and as low when it was in the third and fourth quartile (3 and 4), or not ranked at all.

Assessing the study quality for papers studies published in conferences, we ranked conferences as high quality corresponding to a A* or A score on CORE, medium quality corresponding to a B, and low quality corresponding to a C or nothing at all.

2) Big Five Scale
The Big Five traits need to be assessed by way of a standardized self-report measure (e.g. NEO-FFI or BFI) according to our first inclusion criterion (4.3).

3) Social Media Platform
By using different social media platforms, one encounters different privacy settings which allow certain type of personal digital data to be assessed freely (publicly), only by a select group of people (e.g. friends or followers), or not at all. We distinguish between social media platforms where digital data is public domain (e.g. Twitter) and where it is private to either everyone, or to a select group of people (e.g. Facebook).

4) Type of Social Media Data used
To predict the Big Five, studies have used different king of social media data. Studies vary greatly in the type of digital data they use to predict the Big Five traits. Before separating studies in categories of digital data used, we first distinguished whether studies used just one type of personal digital data, or multiple. After that, we could distinguish four types of digital data that was extracted by studies to predict the Big Five traits:

1. Demographics (e.g. gender)
2. Activity (e.g. likes)
3. Language (e.g. messages)
4. Pictures (e.g. images)

5) Sample Size
Durlak and Lipsey (1991) advised to always code the studies in the meta-analysis for the sample size. Since we were also dealing with studies reporting data using sample sizes varying from not even 100 hundred participants to tens-of-thousands, sample size could, therefore, potentially be a determining factor for study findings, and it needed to be accounted for via coding as well.

4.5.2 Excluded Studies

An overview of the articles (along with their studies) that we coded for the characteristics described earlier, are presented in Appendix A. Since some articles produced multiple studies, our final list contained 50 papers, providing data over 61 studies (numbered 1-50 in Appendix A).

While coding the studies, we also assessed whether they met out earlier mentioned inclusion criteria (4.3). Based on these criteria we had to exclude many studies, and were eventually left with 11 studies to be included in our meta-analysis. Before we continue to the next section (4.6), where the effect sizes gathered from these 11 studies will be discussed, we will first explain why each one of the 50 studies are excluded from our meta-analysis based on our five inclusion criteria. Whenever a study did not meet either our first, second, or third criterion, it was excluded from our meta-analysis right away:

Criterion 1 ($n = 2$)

- 30(#2). Tandera et al. (2017) [study #2]: Big Five is predicted with 'Apply Magic Sauce' (a prediction API)
- 44(#2). Guntuku et al. (2017) [study #2]: Big Five is predicted using picture analysis

Criterion 2 ($n = 4$)

- 7(#2). Farnadi et al. (2016) [study #2]: YouTube is used as a platform to gather personal digital data from
- 10. Majumder et al. (2017): No social media is used
- 17. Gou et al. (2014): Digital data is not linked to personality
- 19. Chapsky (2011): Not only Facebook, but also Netflix and Last FM is used to predict personality

Criterion 3 ($n = 5$)

- 1. Okumura and Okumura (2015): Examiners are used to collect digital data
- 2(#1). Wall et al. (2016) [study #1]: Coders are used to collect digital data
- 2(#2). Wall et al. (2016) [study #2]: Observers are used to collect digital data
- 21(#1). Kosinski et al. (2014) [study #1]: Questionnaire is used to collect digital data information of participants
- 50. Qiu et al. (2012): Authors collected digital data manually

Papers that did not meet our fourth criterion (because it did not report effect-sizes, or reported insufficient information to compute correlations) were not excluded immediately. We contacted the authors by e-mail to obtain missing information. In case of non-response, or unwillingness to share the requested information, the study was excluded from the meta-analysis:

Criterion 4 ($n = 15$)

- 6. Solinger et al. (2014): Accuracy is given - authors did not respond
- 11. Xue et al. (2018): Misses statistical information - authors did not respond
- 15. Celli and Lepri (2018): Regression is given - authors did not respond
- 16. Ting and Varathan (2018): Accuracy, mean and Cohens kappa is given - authors did not respond
- 20. Wald et al. (2012): Regression and accuracy given - authors did not respond
- 21(#2). Kosinski et al. (2014) [study #2]: Spearman's rank correlation is given - authors responded, but could not provide with the necessary data
- 24. Bachrach et al. (2012): Pearson's r is given, but for every trait, a different combination of digital data is used - authors did not respond
- 26. Souri et al. (2018): F-measure and accuracy is given - authors did not respond
- 29. Noë et al. (2016): Misses statistical information - authors responded but could not provide the necessary data
- 33. Da Silva and Paraboni (2018): F1-scores are given - authors did not respond
- 36. Schwartz et al. (2013): Square root R is given - authors did not respond
- 38. Howlader et al. (2018): Regression is given - authors responded saying they did not have the dataset anymore
- 39. Rumagit and Girsand (2018): Accuracy is given - authors did not respond
- 45(#2). Carducci et al. (2018) [study #2]: Regression is given - authors did not respond
- 49. Ferwerda and Tkalcic (2018): Spearman's rho is given - authors responded saying they could not give the dataset since it was not anonymous

In a situation where just one effect size results from a subject sample for a given distribution; effect sizes are almost always statistically independent. However, according to Wolf (1990) not only effect sizes for subsamples from the same study share dependencies, but also effect sizes from different studies performed by the same authors. Even though dependencies are usually small, in meta-analysis independencies need to be defined at the sample or study level. When we encountered multiple studies with non-independent data (our fifth criterion), following the guidelines by Lipsey and Wilson (2001), we included the study which had a bigger sample in our meta-analysis. Whenever studies had the same sample size, we, again following the guidelines by Lipsey and Wilson (2001), analysed the better performing study (the one with higher correlations for at least 3 personality traits):

Criterion 5 ($n = 24$)

- 25(#1). Bhardwaj et al. (2016) [study #1]: Used the same sample as Bhardwaj et al. (2016) study #2, but had lower correlations for three of the five personality traits.
- 3. Hagger-Johnson et al. (2011), 4. Collins et al. (2015), 5(#2). Park et al. (2015) [study #2], 5(#3). Park et al. (2015) [study #3], 7(#1). Farnadi et al. (2016) [study #1], 8. Youyou et al. (2015), 9. Yu and Markov (2017), 13. Alsadhan and Skillicorn (2017), 14. Farnadi et al. (2014), 22(#1). Nave et al. (2018) [study #1], 22(#2). Nave et al. (2018) [study #2], 23. Quercia et al. (2011), 27. Pratama and Sarno (2015), 30(#1). Tandera et al. (2017) [study #1], 31. Tadesse et al. (2018), 32. Yuan et al. (2018), 34. Vaidhya et al. (2017), 35. Alam et al. (2013), 43. Farnadi et al. (2013), 45(#1). Carducci et al. (2018) [study #1], 46. Farnadi et al. (2018), and 48. Segalin et al. (2017): All used the same data pool (myPersonality app) to get their sample as 42. Kosinski et al. (2013), but had a lower sample size.

4.5.3 Included Studies

In short, after reading 50 papers (including 61 studies) in full, we could exclude 39 of them based on our inclusion criteria. In the end, we were left with a total of 11 articles that met our inclusion criteria - providing data from 11 different studies:

- 7(#3). Farnadi et al. (2016) [study #3]
- 12. Kleanthous et al. (2016)
- 18. Kulkarni et al. (2018)
- 25(#2). Bhardwaj et al. (2016) [study #2]
- 28. Tsai et al. (2017)
- 37. Sumner et al. (2012)
- 40. Ferwerda et al. (2015)
- 41. Golbeck et al. (2011)
- 42. Kosinski et al. (2013)
- 44(#1). Guntuku et al. (2017) [study #1]
- 47. Kim and Kim (2018)

In order to ensure a transparent and complete reporting of our meta-analysis - from our initial broad search, all the way to the narrowed down selection of studies included - we expressed our article selection in the form of a flowchart (Fig. 4.1). This flowchart is based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Moher et al. 2009). After having comprised our selection of studies for our meta-analysis, the next step was to gather the effect sizes (4.6) needed for the statistical analyses (4.7).

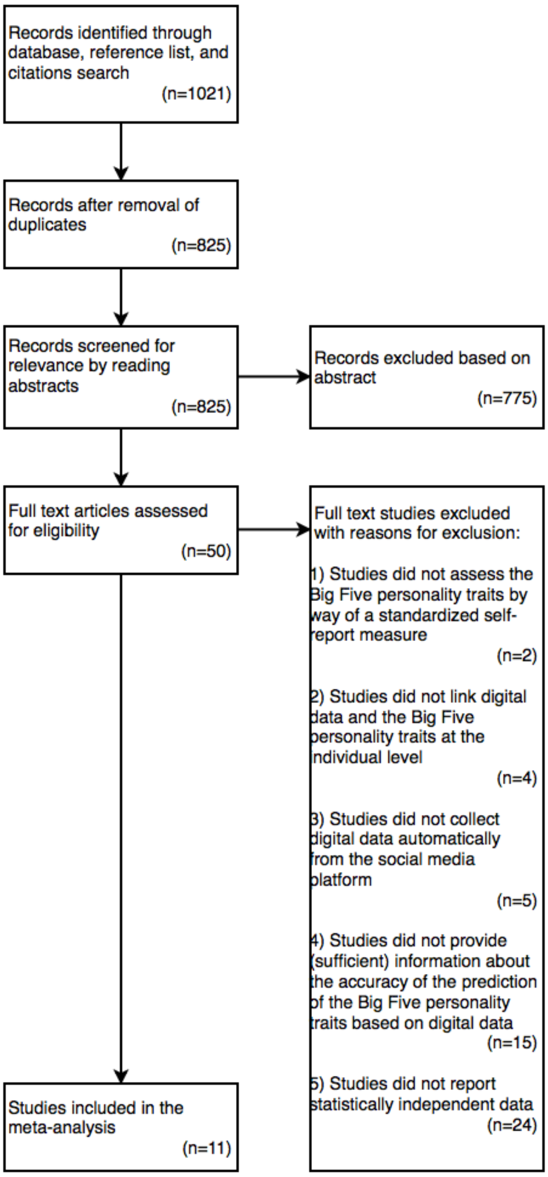

Figure 4.1 Representation of how we selected our studies

4.6 Step 4: Index of Effect Size

After having coded studies for the characteristics relevant for our meta-analysis, we reached the critical stage where the research findings needed to be coded on a numerical scale in order to compare and analyse resulting values. Usually, these research findings are relations of some sort between two constructs. The reported research findings in studies selected for the meta-analysis were coded into values of an effect size, and used later on for analysis (4.7). Both the nature of the study findings and the way they are presented statistically determine which effect size statistics is appropriate for the meta-analysis. We will start this section with necessary background information about the nature of effect size statistics: what it is, how it is computed, and how to use it (4.6.1). After that, we will present the effect size statistics gathered in our own research (4.6.2) and divide them into three different categories based on the different methods used to acquire the effect size statistics (4.6.3 - 4.6.5).

4.6.1 Effect Size

In a meta-analysis, each study finding needs to be encoded as a value on the same effect size statistic, so that the type of effect size statistics is the same across all studies. Only this way a meaningful analysis can be conducted. Study findings focusing on the covariation between multiple variables can be presented as either of the four cases: correlation between two continuous variables, correlation between a dichotomous and continuous variable, correlation between two dichotomous variables, and correlation of mixed pairings of dichotomous and continuous variables (Lipsey and Wilson 2001). In our research, since we deal with the correlation between two continuous variables, the product-moment correlation coefficient is the correct effect size statistic to use. The correlation coefficient for the relationship between variable *x* and *y* can be expressed as:

$$r = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y} \tag{4.1}$$

where σ_{xy}^2 is the covariance between *x* and *y*, and σ_x is the standard deviation of *x* and σ_y the standard deviation of *y*. Since the correlation coefficient is already a standardized index, it can immediately be used as an effect size statistic in the meta-analysis (Lipsey and Wilson 2001). The range of the correlation is between -1.00 and +1.00. Since each study finding is based on a given subject sample, and the amount of individuals in a sample will almost always vary across studies, various effect size values are based on distinct sample sizes. The problem with comparing results from varying samples sizes is that sampling error is greater for effect sizes gathered from small samples, than for large samples (Durlak and Lipsey 1991). By weighing each effect size value by its standard error (standard deviation of the sampling distribution), the problem of having studies with varying sample sizes in meta-analysis is tackled (Lipsey and Wilson 2001). Since a larger standard error results in a less precise effect size value, the weights are computed as the inverse of the squared standard error value. The outcome of this method provides the so called *inverse variance weight*, which besides the effect size, is also computed with each research finding in a meta-analysis (Lipsey and Wilson 2001). However, it's quite difficult to assess the standard error to calculate the right inverse variance weight. Therefore, meta-analysis is generally conducted using one of the few effect size statistics where the standard error formulation is known. An example of an effect size statistics that is already worked out is the product-moment correlation (Pearson's r), which is also the effect size that we have used in our meta-analysis. Since the standard error is used to determine the inverse variance weight, and formulating this standard error is problematic (Rosenthal 1994) with the product-moment correlation coefficient; correlations are transformed (usually) using Fisher's Zr-transform (Hedges and Olkin, 1985):

$$ES_{z_r} = 0.5log\left(\frac{1+r}{1-r}\right) \tag{4.2}$$

where *r* is the correlation coefficient. For the calculation of the converted values, one can use the above equation, or look the values up in a table of Zr-transformed values (e.g. Cooper and Hedges 1993). To convert the Zr-transformed correlation or mean correlation back into a standard correlation form, the so-called inverse of the Zr-transformation can be used (Hedges and Olkin 1985):

$$r = \frac{e^{2ES_{Zr}-1}}{e^{2ES_{Zr}+1}} \tag{4.3}$$

where r is the individual or mean correlation, ES_{Z_r} is the individual or mean Zr-transformed correlation. The correlation coefficient as an effect size then becomes:

$$ES_r = r \tag{4.4}$$
$$ES_{Z_r} = 0.5log\left(\frac{1+ES_r}{1-ES_r}\right) \tag{4.5}$$
$$SE_{Z_r} = \frac{1}{\sqrt{n-3}} \tag{4.6}$$
$$\omega_{Z_r} = \frac{1}{SE_{Z_r}^2} = n-3 \tag{4.7}$$

where ES_r is the effect size, SE_{Z_r} is the standard error of the effect size, ω_{Z_r} is the inverse variance weight of the effect size, r is the correlation, and n the total sample size. Usually the results of studies investigating the correlation among variables are reported as Pearson's r, thus allowing the coding of the effect size by just recording the correlation. However, it may very well be that the correlation information is reported in a different form, which forces the meta-analyst to estimate the desired correlation coefficient from p-values. Whenever the results of a meta-analysis are not reported as a correlation coefficient, it is usually done in the form of means and standard deviations. Since in our study we will be dealing with two continuous variables, the correct computational formula for Pearson's r (ES_r) becomes:

$$ES_r = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(N\sum x_i^2 - (\sum x_i)^2)(N\sum y_i^2 - (\sum y_i)^2)}} \tag{4.8}$$

where individual level data for each variable (x and y) and total sample size (N) is needed.

4.6.2 Effect Sizes in our Meta-Analysis

After having comprised our list of studies for our meta-analysis, we had to determine the effect size for each personality trait. From the 11 studies, most ($n = 8$) used multiple features (e.g. categories of words used, or different structural features) to determine the Big Five traits, some ($n = 2$) studies reported just a single correlation for every trait, and the rest ($n = 1$) reported findings in the form of AUC (Area Under the Curve). Since, according to Lipsey and Wilson (2001), when studies study the correlation between variables present their findings directly as Pearson's r, effect size coding comes down to recording that correlation (Equation 4.4). We did not have to convert our effect sizes to Fisher's Z, since it is not necessary for meta-analytic random effects model (4.7) - see Hunter and Schmidt (2004). Appraising the magnitude of the effect size was done by Cohen (1988) where effect sizes smaller than 0.20 are classified as small, between 0.20 and 0.80 as medium, and larger than 0.80 as large. The analogous values for the correlation effect size (in our case) are classified as small for correlations smaller than 0.10, medium between 0.10 and 0.40, and large for greater than 0.40. We will show in the following sub-sections how we gathered, chose, or found the effect sizes for our studies that we included in our meta-analysis.

Table 4.2 Studies in meta-analysis

Multiple Features ($n = 8$)
○ Farnadi et al. (2016) [#3]
○ Kulkarni et al. (2018)
○ Bhardwaj et al. (2016) [#2]
○ Tsai et al. (2017)
○ Ferwerda et al. (2015)
○ Golbeck et al. (2011)
○ Guntuku et al. (2017) [#1]
○ Kim and Kim (2018)
Single Correlation ($n = 2$)
○ Kleanthouse et al. (2016)
○ Kosinski et al. (2013)
AUC ($n = 1$)
○ Sumner et al. (2012)

We divided the studies that we will analyse into the three categories (Multiple Features, Single Correlation, and AUC) mentioned in the beginning of this sub-section, as also visualized by Table 4.2. We will explain our process of determining the effect sizes for each personality trait per category, and per study.

4.6.3 Multiple Features

By using different features to find the correlation between personal digital data from social media and personality, Kulkarni et al. (2018), Bhardwaj et al. (2016), Tsai et al. (2017), Ferwerde et al. (2015), Guntuku et al. (2017), and Kim and Kim (2018) provided different correlations between each feature and the Big Five personality traits. According to Lipsey and Wilson (2001), a meta-analyst should select one of the effect sizes for inclusion in the analysis (e.g. the best effect size) and omit the other effect sizes.

Both Farnadi et al. (2016) and Golbeck et al. (2011) not only used different features, but also used more than one model to determine the correlations. Following the instructions from Lipsey and Wilson (2001), again, we chose the model that had the higher correlations between the Big Five determined with a questionnaire, and with personal social media data, for our meta-analysis.

Table 4.3 The Effect Sizes (Pearson's r) of studies that used multiple features to determine the Big Five traits with social media data

	Agreeableness	Extraversion	Conscientiousness	Openness	Neuroticism
Farnadi et al. (2016) [#3]	0.39	0.41	0.48	0.42	0.43
Kulkarni et al. (2018)	0.16	0.14	0.29	0.18	0.09
Bhardwaj et al. (2016) [#2]	0.75	0.91	0.75	0.39	0.90
Tsai et al. (2017)	0.23	0.22	0.20	0.22	0.19
Ferwerda et al. (2015)	0.15	0.23	0.16	0.28	0.22
Golbeck et al. (2011)	0.48	0.55	0.60	0.65	0.53
Guntuku et al. (2017) [#1]	0.23	0.22	0.15	0.17	0.40
Kim and Kim (2018)	0.20	0.27	0.24	0.15	0.13

4.6.4 Single Correlation

Klaenthous et al. (2016), like Bachrach et al. (2012) and Kosinski et al. (2013), used previous research to identify social media activity that can predict the Big Five. In the end, just a single correlation (in the form of Pearson's r) was produced between each of the Big Five personality traits assessed with the questionnaire, and the computational model used.

The other study that provided just one single correlation per personality trait, Kosinski et al. (2013), used a linear regression model, and applied 10-fold cross-validation to express the prediction accuracy of regression by the Pearson correlation coefficient between predicted and actual attribute values.

Table 4.4 The Effect Sizes (Pearson's r) of studies that that provided just a single correlation (per trait) between the Big Five traits assessed with social media data and with a questionnaire

	Agreeableness	Extraversion	Conscientiousness	Openness	Neuroticism
Kleanthous et al. (2016)	0.03	0.26	0.28	0.16	0.01
Kosinski et al. (2013)	0.30	0.40	0.29	0.43	0.30

4.6.5 AUC

In their study, Sumner et al. (2012) used different models for the prediction of the personality traits. For the best performing model (Kaggle) the results are given in the Area Under the Curve (AUC) values. We converted these values to Pearson's r with the conversion table given by Rice and Harris (2005).

Table 4.5 The Effect Sizes (Pearson's r) of the study that initially gave the results of the correlations between the Big Five traits assessed with social media data and with a questionnaire in AUC

	Agreeableness	Extraversion	Conscientiousness	Openness	Neuroticism
Sumner et al. (2012)	0.18	0.25	0.19	0.16	0.20

4.7 Step 5: Statistical Analyses

In this stage of the meta-analysis, after having assembled a set of statistically independent effect sizes, summed up in Table 4.6, we proceed with the analysis of the effect size data which consists two separate parts: the initial analyses of the effect sizes, and moderator analyses.

The initial analyses will be performed to test the effect sizes for outliers using Grubb's test (4.7.1), to express the distribution of effect sizes and estimate the population mean with a corresponding confidence interval (4.7.2), and, finally, to determine the adequacy of the mean effect size for representing the entire distribution of effects for publication bias (4.7.3) and for homogeneity of the effect sizes (4.7.4). We used visualization techniques (i.e. forest plot and funnel plot), since they are a practical way of effectively communicating the final results (Wand and Bushman 1998).

After the initial analyses, moderator analyses will be performed using meta-regression to examine the impact of moderator variables on study effect size (4.7.5). The results from the moderator analyses will give us the answer what the predictive power is of personal social media data over the Big Five personality traits, and how the different variables (moderators) influence the accuracy of the prediction.

We conducted a total of five different meta-analyses; one for each of the Big Five traits. The random-effects model is used for our meta-analysis, since the true effect size varied in the studies. As stated earlier (4.6.2), we did not convert the effect sizes into Fisher's Z, since we used the random-effects model (Hunter and Schmidt 2004). We made use of the software 'Comprehensive Meta Analysis' (Version 3.3.070) for all the analyses in this section, except for Grubb's test for outliers (6.1.1), where we used Excel.

Table 4.6 Effect sizes of the studies included in the meta-analysis

	Agreeableness	Extraversion	Conscientiousness	Openness	Neuroticism
Farnadi et al. (2016) [#3]	0.39	0.41	0.48	0.42	0.43
Kulkarni et al. (2018)	0.16	0.14	0.29	0.18	0.09
Bhardwaj et al. (2016) [#2]	0.75	0.91	0.75	0.39	0.90
Tsai et al. (2017)	0.23	0.22	0.20	0.22	0.19
Ferwerda et al. (2015)	0.15	0.23	0.16	0.28	0.22
Golbeck et al. (2011)	0.48	0.55	0.60	0.65	0.53
Guntuku et al. (2017) [#1]	0.23	0.22	0.15	0.17	0.40
Kim and Kim (2018)	0.20	0.27	0.24	0.15	0.13
Kleanthous et al. (2016)	0.03	0.26	0.28	0.16	0.01
Kosinski et al. (2013)	0.30	0.40	0.29	0.43	0.30
Sumner et al. (2012)	0.18	0.25	0.19	0.16	0.20

4.7.1 Outliers

According to Hedges and Olkin (1985) one should always examine the distribution of effect sizes to detect outliers in a meta-analysis. This is, because extreme effect size values have disproportionate influence on the values of the means, variances, and other statistics that may distort them in misleading ways. To identify possible outliers in our set of effect sizes (Table 4.6), we executed Grubb's test (Grubbs 1969). Outlier were found for three of the personality traits (Agreeableness, Extraversion, and Neuroticism) all for the effect size coming from the same study; Bhardwaj et al. (2016) [#2]. To identify the outliers, we followed the method provided by Grubbs (1969), and used the critical T value provided by Grubbs and Beck (1972). A short overview of the Grubb's test we executed can be seen below (Table 4.7); the Grubb's tests worked out in detail can be found in Appendix B (Fig. B.1 - B.5).

Table 4.7 Grubb's test for the effect sizes in our meta-analyses

		Agreeableness	**Extraversion**	**Conscientiousness**	**Openness**	**Neuroticism**
Run #1	max G	2.375	2.561	2.144	2.217	2.365
	critical T	2.234	2.234	2.234	2.234	2.234
Run #2	max G	1.914	2.100	-	-	1.468
	critical T	2.176	2.176	-	-	2.176

We did not exclude outliers for several reasons: (1) they are all from the same study which is due to variability in the measurement, (2) the sample size is relatively small ($n = 31$) compared to other studies, and (3) since we will use meta-regression (which is remarkably resilient to such outliers) for our effect sizes, meaning that we can just leave the outliers in (Doucouliagos and Stanley 2009).

4.7.2 Confidence Interval

The confidence interval is useful since it not only indicates the degree of precision of the estimate of the mean effect size. Also, whenever the confidence interval does not include zero, it means that the mean effect size is statistically significant at the level specified by the confidence interval. In order to assess the confidence interval for a correlation coefficient, the standard error of the correlation coefficient needs to be computed first (Urdan 2011):

$$SE_{ES} = \sqrt{\frac{1-ES^2}{n-2}} \tag{4.10}$$

where ES is the effect size, and n the sample size. The confidence interval (for a two tailed test) then becomes:

$$CI = ES \pm (SE_{ES})(t) \tag{4.11}$$

where t can be found in Table C.1 in Appendix C with given p-value and degrees of freedom (N-2). If there is no zero included in the confidence interval, this indicates that the mean effect size is statistically significant at $p \leq a$. By computing a z-test, the meta-analyst can obtain a direct test of the significance of the mean effect size:

$$z = \frac{|ES|}{SE_{ES}} \tag{4.12}$$

This formula results in a standard normal variate which, if it exceeds 1.96 it is statistically significant with $p \leq 0.05$, two-tailed and if it exceeds 2.58 it is significant with $p \leq 0.01$ two-tailed.

We determined the 95% confidence interval for each effect size using Equation 4.11. For an effective showcase of the precisions associated with each individual effect size, and their corresponding confidence intervals, we visualized the effect sizes and their confidence intervals using the forest plot (Fig. B.6 - B.10 Appendix B). Forest plots effectively display the precision associated with each individual effect size, their associated confidence intervals, and the general pattern of results. The software that we used (Comprehensive Meta Analysis) also provided us with the point estimate, along with its own 95% confidence intervals, of the random effects model, which we also processed in the earlier mentioned forest plots. We have summarized all the confidence intervals, for all the effect sizes of each personality traits of each study in a forest plot (Fig. 4.2).

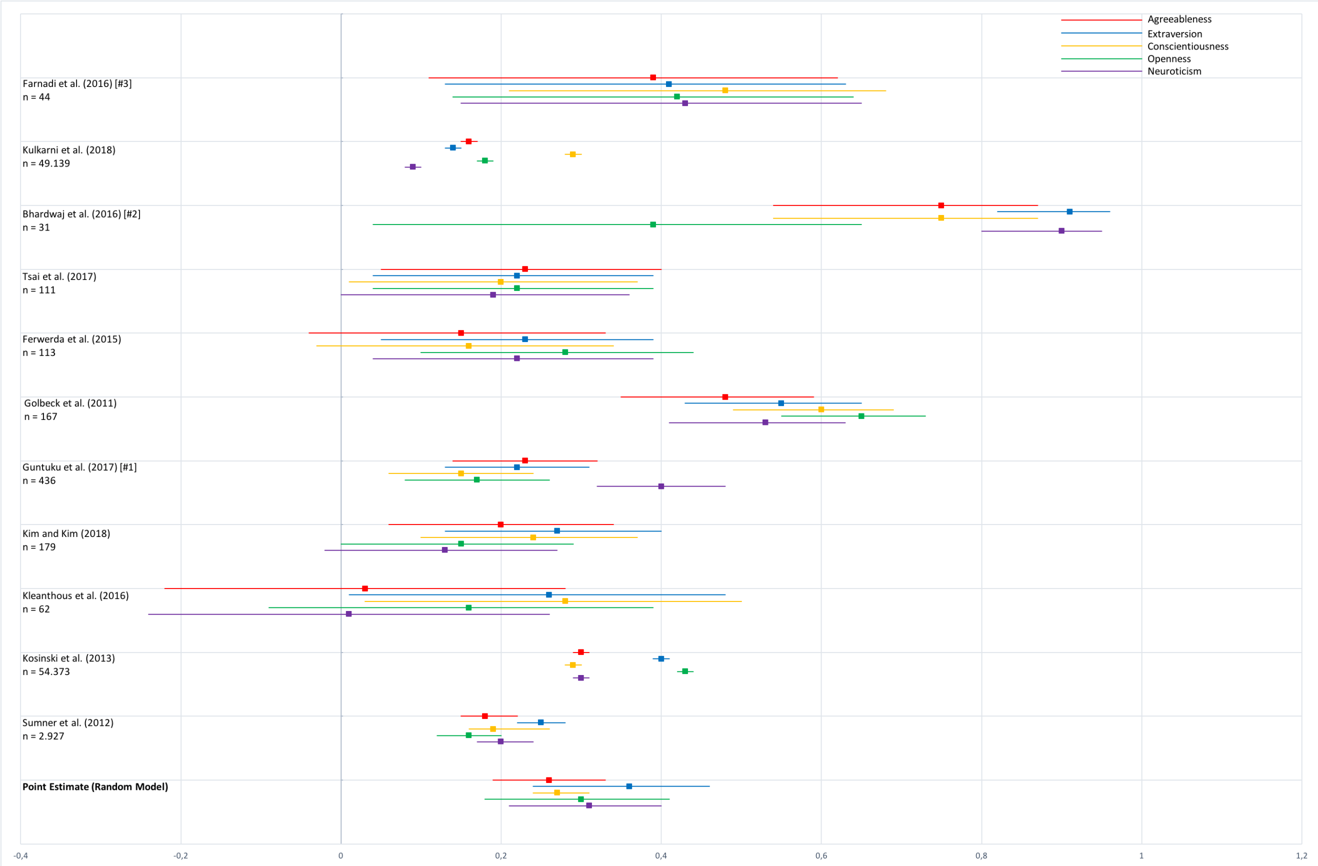

Fig. 4.2 Forrest plot of our effect sizes

4.7.3 Publication Bias

The presence of publication bias comes down to published studies being no longer a representative sample of the available evidence -- thus, distorting the results of the meta-analysis. We used several methods to search for the existence of a publication bias in our study. All of the five tests that we will discuss in this sub-section are again performed using our earlier mentioned software (Comprehensive Meta Analysis).

First, we plotted for each effect size the funnel plot; a visual method to detect publication bias (Fig. B.11 - B.15 Appendix B). The funnel plot is a great method for detecting potential publication bias; studies with high precision are plotted near the average, and studies with low precision will be spread evenly on both sides of the average. If the plot does not resemble a roughly funnel-shaped distribution, it may indicate the existence of a publication bias. While not exactly symmetrical, we can see the distributions of studies resembles a roughly funnel-shaped distribution for Agreeableness, Extraversion, and Openness. However, both Conscientiousness and Neuroticism may include publication bias in the effect sizes, which can be accounted due to the fact we let the outlier in in our meta-analysis earlier (4.7.1).

Next up, we looked at the fail-safe N for each trait; which had to be at least five times the number of reviewed studies plus ten. The results from our meta-analysis showed that this was the case for each trait (Agreeableness: $N = 5015$; Extraversion: $N = 7487$; Conscientiousness: $N = 8051$; Openness: $N = 8749$; Neuroticism: $N = 4296$) since they all are more than 65; five times 11 (observed studies) plus ten (Rosenthal 1979). The fail-safe N also indicates the number of studies that is required to refute significant meta-analytic means;

reduce the effect to a level not statistically different from zero (Rosenthal 1979). The complete outcome of the fail-safe N analysis can be found in Table B.1 in Appendix B.

Our third test was the Begg and Mazumdar's rank correlation test which reports the rank correlation (Kendall's tau) between the standardized effect size and the variance (or standard errors) of these effects (Begg and Mazumdar 1994). The results from Table B.2 (Appendix B) indicate that there is no concern for publication bias for each personality trait (Agreeableness $p = 0.18$; Extraversion $p = 0.09$; Conscientiousness $p = 0.14$; Openness $p = 0.11$; Neuroticism $p = 0.38$). However, as Begg and Mazumdar (1994) have stated themselves, the power of their test is extremely low when a small number of studies is assessed; test is very powerful with 75 studies, and only moderate with 25 studies.

Egger's regression test, like the Begg and Mazumdar's test, intends to quantify the bias captured by the funnel plot (Egger et al. 1997). Unlike the Begg and Mazumdar's test, Egger's test uses the actual values of the effect sizes and their precisions (the inverse of the standard error), rather than ranks, to predict the standardized effect (effect size divided by the standard error). The results of Egger's test (Agreeableness $p = 0.46$; Extraversion $p = 0.45$; Conscientiousness $p = 0.49$; Openness $p = 0.42$; Neuroticism $p = 0.35$) indicate no concern for publication bias (Table B.3 Appendix B).

Duval and Tweedie's Trim and Fill method was our last test to check our data for publication bias. The trim and fill method identifies and corrects funnel plot asymmetry caused by publication bias by first removing the smaller studies causing the funnel plot asymmetry, then using the trimmed funnel plot to estimate the true centre of the funnel, and finally replacing the omitted studies and their missing counterparts around the centre (Duval and Tweedie 2000). In the case of more small studies on the right than on the left of the funnel, there may be studies missing from the left. The results from our trim and fill tests for the effect sizes for each personality trait did not indicate a need for studies to be trimmed (Table B.4 - B.8 Appendix B). Hence, there is no concern for publication bias according to the trim and fill method for any of the set of data in our meta-analysis.

4.7.4 Homogeneity

In meta-analysis, in the most perfect scenario, the combined results of the studies included in the meta-analysis should all be undertaken in the exact same way, and with the very same experimental protocols. This most perfect scenario however never occurs, and does not need to be. But we do need to check whether the results found in the individual studies are similar enough to be able to state that the combined results will be a meaningful description of the set of studies. Homogeneity testing is based on a comparison of the observed variability in effect size values with an estimate of the variance that would be expected from subject-level sampling error alone according to Lipsey and Wilson (2001). When the effect sizes are not homogeneous, various descriptive variables can be examined to see whether they moderate (4.7.5) the effect size. According to Rosenthal and Rubin (1982) the question of the homogeneity of the effect size distribution (i.e., whether all of the effect sizes that are averaged into a mean value estimate the same population effect size) is an important one in meta-analysis. Any given effect size in a homogeneous distribution only differs by sampling error from the population mean. The Q statistic, on which the homogeneity test is based, is distributed as a chi-square with $k - 1$ degrees of freedom (k is number of effect sizes):

$$Q = \sum \omega_i (ES_i - \overline{ES})^2 \tag{4.13}$$

where ES_i is the individual effect size for $i = 1$ to k, $\overline{ES}$ weighted mean effect size over the k effect sizes, and ω_i the individual weight for ES_i. When Q exceeds the critical value for a chi-square with $k - 1$ degrees (Table C.2 in Appendix C), the null hypothesis of homogeneity is rejected; a statistically significant Q indicates a heterogeneous distribution. There is also an algebraically equivalent formula for Q that is computationally simpler to implement:

$$Q = (\sum \omega_i ES_i^2) - \frac{(\sum \omega_i ES_i)^2}{\sum \omega_i} \tag{4.14}$$

According to Mosteller and Colditz (1996), when the fixed effects assumptions are rejected (there is a heterogeneous distribution of effect size), whether for conceptual or statistical reasons, the 'Random Effects Model' needs to be adopted because of its generality. With a random effects model, the assumption is that each observed effect size differs from the population mean by subject-level sampling error plus a value that represents other sources of variability. The total variance associated with the distribution of effect size (v_i^*) is the sum of the variance associated with subject-level sampling error (v_i) and the one associated with the random effects variance (v_θ):

$$v_i^* = v_i + v_\theta \tag{4.15}$$

This new variance (v_i^*) is now used to compute the random effects mean size, confidence interval, significance test, and Q. We also get a new value for the inverse variance weight: v_i^*. A formula to obtain the estimate of the random effects variance component (v_θ) is:

$$v_\theta = \frac{Q-(k-1)}{\sum \omega_i - \left(\sum \omega_i^2 / \sum \omega_i\right)} \tag{4.16}$$

Whenever the effect size distribution is homogeneous, the formula above returns a negative value. It is important to note that whenever homogeneity of the distribution is rejected, one might need to start over from the second step (Weighted Mean) using a different statistical model, and repeat the entire procedure again.

We used different tests to determine the heterogeneity of the effect sizes of the studies which we all executed using the software 'Comprehensive Meta Analysis' mentioned earlier. First, we checked Cochran's Q (Equation 4.13), which is distributed as a chi-square statistic with k number of studies minus one degree of freedom. The chi-square test, where the null hypothesis assumes that all studies are homogeneous, gives a p-value to test this hypothesis. When the p-value of the test is low; the hypothesis is rejected and heterogeneity is present. Secondly, we used Tau (T), which is computed from T^2, as a first indicator of the extent of dispersion. Tau is an estimate of the standard deviation of the distribution of true effect size. Lastly, we will examine the I^2 statistic; indicator of the percentage of variance in a meta-analysis that is attributable to study heterogeneity (Higgins et al. 2003). We have summed up the results of the heterogeneity tests for all personality traits in the table below (Table 4.8). The results of the heterogeneity tests worked out in detail can be found in Appendix B (Table B.9 - B.13).

Table 4.8 Heterogeneity statistics for each trait with $p < 0.001$ for Q-values

	Point Estimate [95% CI]	**Q [df]**	I^2	T^2	T
Agreeableness	0.26 [0.19, 0.33]	608.98 [10]	98.36	0.01	0.10
Extraversion	0.36 [0.24, 0.46]	2130.72 [10]	99.53	0.04	0.19
Conscientiousness	0.27 [0.24, 0.31]	85.38 [10]	88.29	0.00	0.04
Openness	0.30 [0.18, 0.41]	2124.68 [10]	99.53	0.04	0.19
Neuroticism	0.31 [0.21, 0.40]	1336.52 [10]	99.25	0.02	0.15

We found the Q-values to be significant for each personality trait; they all exceed the critical value (29.5883) from Table C.2 (Appendix C). However, it is proven that the chi-squared test has low power as a comprehensive test for heterogeneity when the number of studies is small like in our meta-analyses (Gavaghan et al. 2000). Tau values for the personality traits indicate low true heterogeneity between studies. Given the I^2 values, the overall distribution of effect sizes can be accounted to true heterogeneity.

Due to methodological diversity always being present in meta-analyses, like in our case, statistical heterogeneity is something one can't escape from (Higgins et al. 2003). So, heterogeneity will always exist whether or not we can detect it using a statistical test. One may then argue: Why even bother testing heterogeneity? Interestingly, Borenstein et al. (2011) suggested to use I^2 as a criterion to decide whether it is worthwhile to conduct a moderator analysis; if I^2 is low, this indicates the absence of heterogeneity, thus nothing to be explored, if I^2 is large (like in our case), then a moderator analysis will most certainly be an interesting issue to explore.

4.7.5 Moderator Analyses

In the final stage of the statistical analysis, we can compare the mean effect sizes for studies grouped according to variables of interest to determine the factors that account for the differences in effect sizes (Hedges 1984). Variables describing various study features are used as independent variables to predict effect size. A regression model can be developed in which predictor variables representing study characteristics account for significant portions of the effect size heterogeneity found between studies. In meta-analysis, a tool most commonly used is called meta-regression (Van Houwelingen et. al 2002). There are three types of models concerning regression that can be distinguished with meta-analysis: simple regression, fixed effect meta-regression, and random effects meta-regression. Since both simple and fixed effect meta-regression don't allow for within-study variation, but random effects meta-regression does, the latter model was performed using the software 'Comprehensive Meta Analysis'. The random effects meta-regression model can be specified as:

$$y_j = \beta_0 + \beta_1 x_{1_j} + \beta_2 x_{2_j} + \ldots + \eta + \varepsilon_j \tag{4.17}$$

where y_j is the effect size in study j, β_0 the estimated overall effect size, the variables x_i $(i = 1 \ldots k)$ specify different characteristics of the study, and ε specifies the between study variation.

Given the interested in our research in the predictive power of personal social media data over the Big Five personality traits, and whether the use of different type of social media platforms and different type of digital data influences the accuracy of personality prediction, we had already performed our coding process earlier on (4.5) by extracting information from the studies like what social media platform they used, what kind of digital data, etc. From our coding process, we took for this sub-section (moderator analysis) specifically the information of studies concerning the social media platform, and digital data, which can be found in Table A.1 in Appendix A. To have a clear overview, we summarized the information we gathered for each study that we included in our meta-analysis in the table below (Table 4.9).

Table 4.9 Summary of the necessary information for the moderator analyses

Study	**Social Media Platform**	**Digital Data**
Farnardi et al. (2016)	Twitter	Activity
Kulkarni et al. (2018)	Facebook	Language
Bhardwaj et al. (2016) [#2]	LinkedIn	Activity Language
Tsai et al. (2017)	Facebook	Activity
Ferwerda et al. (2015)	Instagram	Pictures
Golbeck et al. (2011)	Facebook	Demographic Activity Language
Guntuku et al. (2017) [#1]	Twitter	Pictures
Kim and Kim (2018)	Instagram	Pictures
Kleanthouse et al. (2016)	Facebook	Activity
Kosinski et al. (2013)	Facebook	Activity
Sumner et al. (2012)	Twitter	Language

From the table above, we determined six possible moderating effects for our analysis, where we explored whether the use of a certain social media platform, the use of single or more type of digital data, or the use of a certain type of digital data (demographic, activity, language, and pictures), were a statistically significant predictor for the effect sizes we found. Specifically:

1. Social Media Platform (which social media platform was used)
2. Single vs. Multiple (the use of a single type of data, or a combination of multiple)
3. Demographic (whether demographic data was used or not)
4. Activity (whether activity data was used or not)
5. Language (whether text analysis was used or not)
6. Pictures (whether visual data in the form of pictures was used or not)

Using the software 'Comprehensive Meta Analysis' we gathered many statistical results from each meta-regression. However, we were only interested in the following three values:

- p: determines the statistical significance of the results
- $Q\ (df)$: chi-square statistic and degrees of freedom
- T^2: total amount of variance unexplained (called Tau-squared)
- R^2: variance that can be explained by moderator

To determine whether a moderator is a variable that can be used to effectively (to a certain extent) predict a personality trait, we investigated the R^2 value. This R^2 value represents the portion of variance, from the total amount of variance which was unexplained initially, that can be explained by the moderator:

$$R^2 = \frac{(T^2_{initial} - T^2_{moderator})}{T^2_{initial}} \tag{4.18}$$

where $T^2_{initial}$ is the initial Tau-squared value (gathered from previous sub-section (4.7.4), and $T^2_{moderator}$ is the new Tau-squared value from the meta-regression. When $T^2_{initial} < T^2_{moderator}$, R^2 is assumed 0.

We will now present the results for each of the six moderators in the form of a table, and briefly discuss the results one by one. To start with, we checked whether the use of a certain type of social media platform, which we divided in Facebook, Twitter, LinkedIn, and Instagram, as can be seen from Table 4.9, could predict personality traits much more accurate. Aside from Openness, the results were statistically significant for all the traits (Table 4.10). However, the choice for a social media platform was only a statistically significant predictor for Conscientiousness given the R^2 value.

Table 4.10 Statistics of meta-regression for Social Media Platform

	$Q\ (df)$	p	$T^2_{initial}$	$T^2_{moderator}$	R^2
Agreeableness	11.72 (3)	< 0.05	0.0104	0.0109	0.00
Extraversion	19.04 (3)	< 0.05	0.0369	0.0391	0.00
Conscientiousness	31.13 (3)	< 0.05	0.0013	0.0005	0.64
Openness	1.28 (3)	0.73	0.0368	0.0382	0.00
Neuroticism	24.46 (3)	< 0.05	0.0231	0.0241	0.00

The second moderator tested was the use of single versus multiple types of digital data to assess the Big Five. Only two studies, Bhardwaj et al. (2016) and Golbeck et al. (2011), used multiple different types of digital data to predict the Big Five, all other nine studies merely used a single, specific type (Table 4.9). While the results were statistically significant for all the Big Five traits, the variable (single vs. multiple) seemed to be a statistically significant predictor for Conscientiousness (Table 4.11).

Table 4.11 Statistics of meta-regression for Single vs. Multiple

	$Q\ (df)$	p	$T^2_{initial}$	$T^2_{moderator}$	R^2
Agreeableness	13.51 (1)	< 0.05	0.0104	0.0101	0.03
Extraversion	14.93 (1)	< 0.05	0.0369	0.0363	0.01
Conscientiousness	38.31 (1)	< 0.05	0.0013	0.0007	0.47
Openness	4.97 (1)	< 0.05	0.0368	0.0364	0.01
Neuroticism	19.35 (1)	< 0.05	0.0231	0.0223	0.03

Conscientiousness again, seemed to be the only trait that could be predicted with the next moderator we explored: Demographic (whether studies used demographic data such as gender). Other statistically significant results were found for Agreeableness and Openness, but they appeared not to be related to the effect sizes due to the low R^2 values (Table 4.12).

Table 4.12 Statistics of meta-regression for Demographic

	$Q\ (df)$	p	$T^2_{initial}$	$T^2_{moderator}$	R^2
Agreeableness	4.51 (1)	< 0.05	0.0104	0.0102	0.02
Extraversion	1.54 (1)	0.21	0.0369	0.0367	0.00
Conscientiousness	25.61 (1)	< 0.05	0.0013	0.0009	0.33
Openness	5.77 (1)	< 0.05	0.0368	0.0363	0.01
Neuroticism	2.85 (1)	0.09	0.0231	0.0228	0.01

In general, Activity (whether studies used activity data such as likes) scored best as a statistically significant predictor for the Big Five traits. The results were statistically significant for all Big Five dimensions, and aside from Conscientiousness, all traits were related to the effect size (Table 4.13).

Table 4.13 Statistics of meta-regression for Activity

	$Q\ (df)$	p	$T^2_{initial}$	$T^2_{moderator}$	R^2
Agreeableness	11.14 (1)	< 0.05	0.0104	0.0026	0.75
Extraversion	12.26 (1)	< 0.05	0.0369	0.0104	0.72
Conscientiousness	8.18 (1)	< 0.05	0.0013	0.0097	0.00
Openness	32.21 (1)	< 0.05	0.0368	0.0028	0.92
Neuroticism	5.15 (1)	< 0.05	0.0231	0.0172	0.25

There was only one statistically significant result for Language (whether studies used data such as status updated or tweets), which was for Conscientiousness. However, Language explained no variance at all, thus making it a moderator that could not predict any personality trait (Table 4.14).

Table 4.14 Statistics of meta-regression for Language

	$Q\ (df)$	p	$T^2_{initial}$	$T^2_{moderator}$	R^2
Agreeableness	0.51 (1)	0.48	0.0104	0.0058	0.45
Extraversion	2.38 (1)	0.12	0.0369	0.0191	0.48
Conscientiousness	4.03 (1)	< 0.05	0.0013	0.0099	0.00
Openness	0.53 (1)	0.47	0.0368	0.0152	0.59
Neuroticism	2.13 (1)	0.14	0.0231	0.0179	0.22

The final variable we explored, Pictures (whether studies used pictures like profile pictures or pictures shared), was only a statistically significant predictor for Conscientiousness but to a very low degree (Table 4.15)

Table 4.15 Statistics of meta-regression for Pictures

	$Q\ (df)$	p	$T^2_{initial}$	$T^2_{moderator}$	R^2
Agreeableness	1.05 (1)	0.31	0.0104	0.0105	0.00
Extraversion	1.66 (1)	0.20	0.0369	0.0373	0.00
Conscientiousness	6.42 (1)	< 0.05	0.0013	0.0012	0.11
Openness	1.08 (1)	0.30	0.0368	0.0370	0.00
Neuroticism	0.37 (1)	0.54	0.0231	0.0232	0.00

We summed up, for all the moderators, for which personality trait they appeared to be a statistically significant predictor according to our meta-regressions in the table below (Table 4.16). We noted the R^2 values where the moderators served as statistically significant predictors.

Table 4.16 R^2 values of statistically significant predictors for the Big Five traits for $p < 0.05$

	Agreeableness	Extraversion	Conscientiousness	Openness	Neuroticism
Social Media Platform	0	0	0.64	-	0
Single vs. Multiple	0.03	0.01	0.47	0.01	0.03
Demographic	0.02	-	0.33	0.01	-
Activity	0.75	0.72	0	0.92	0.25
Language	-	-	0	-	-
Pictures	-	-	0.11	-	-

Other than the statistics laid out in the tables above, we gathered a handful of other results from our meta-regressions, which we all put in Appendix B (B.9). Among others, along with the meta-regression, the software 'Comprehensive Meta Analysis' also provided scatterplots for each personality traits, for all the six moderators; see Appendix B (B.9).

To summarize, our moderator analyses shows that Agreeableness, Extraversion, Openness, and Neuroticism are best predicted with Activity; whereas the strength of the prediction is relatively low for Neuroticism. Conscientiousness on the other hand is best predicted with Social Media Platform, and to some extent with the moderators Single vs. Multiple and Demographic. This concludes the fifth step (Statistical Analyses) of our meta-analysis, and we can proceed with the final step, where we will discuss our interpretations of our research, and draw our conclusions.

4.8 Step 6: Conclusions and Interpretations

After a thorough analysis of the statistical data, we have reached the final step of our meta-analysis: interpreting results and drawing conclusions. To restrain ourselves from over-interpreting our findings or over-generalizing our conclusions, we will follow the three guidelines set by Durlak and Lipsey (1991). We will (1) restrict our conclusions specifically to the literature reviewed, (2) examine results from the meta-analysis testing important hypotheses that fail to reach significance closely, and (3) qualify study findings in relation to the limitations of the available studies, with respect to restricted or diminished data on specific variables of interest.

To put our results into the context in which we are interested - i.e., whether the Big Five personality traits determined from personal social media data can be used to assess the risk of recidivism -- we first need to recall the characteristics sought for in the studies with our literature search. Our literature search resulted in inclusion of only 11 studies in our meta-analysis. The low number of studies included can be accounted to our strict inclusion criteria, and the fact that many studies ($n = 15$) that lacked statistical info, which the authors could not share with us due to privacy. Coming back to our inclusion criteria, we specifically looked for studies where digital social media data was retrieved automatically, which is far less common than studies in which participants are asked in a self-report questionnaire about their social media usage (e.g. Wall et al. 2016). Furthermore, we looked for studies where social media platforms were used which are relevant in the Netherlands; this excluded social media platforms such as Sina Weibo which are mainly used in the Chinese market. So, the results from our meta-analysis directly apply to studies where the Big Five personality traits are predicted using digital data (demographic, activity, language, and pictures) from social media (Facebook, Twitter, Instagram, LinkedIn) collected automatically.

In our research, we found the mean predictive values of using personal social media data to predict the Big Five, and the 95% confidence interval (in brackets), to be: 0.26 [0.19; 0.33] for Agreeableness, 0.36 [0.24; 0.46] for Extraversion, 0.27 [0.24; 0.31] for Conscientiousness, 0.30 [0.18; 0.41] for Openness, and 0.31 [0.21; 0.40] for Neuroticism. Regarding the context of our results, we can therefore conclude that the correlations are moderate at best. When we analysed the additional moderators for our research, we found Activity (the use of activity data) to be a statistically significant predictor for Agreeableness ($R^2 = 0.75$), Extraversion ($R^2 = 0.72$), Openness ($R^2 = 0.92$), and Neuroticism ($R^2 = 0.25$) for $p < 0.05$. For Conscientiousness, we found Social Media Platform (choice for a particular social media platform), Single vs. Multiple (use of one or more type of data), Demographic (the use of demographic data), and Pictures (use of pictures) to be statistically significant predictors with R^2 values of 0.64, 0.47, 0.33, and 0.11, respectively, for $p < 0.05$. While we found significant results with all the moderators for Conscientiousness, we could not find statistically significant results for the other four traits both with Language and Pictures. The scientific and practical implications of these findings will be outlined in greater detail in the next chapter of this thesis.

Our interpretations and conclusions from our meta-analysis bring us to the end of this chapter, and bring us to our next, and final chapter (Ch. 5) where we will explain both the scientific and practical relevance of our study, discuss the limitations, present future work, and draw our conclusions from our research.

CHAPTER 5: DISCUSSION

This research was conducted to give recommendations on how the Dutch justice department could improve the assessment of the risk of recidivism of suspects and inmate based on the predictability of personal digital data from social media that can be retrieved about the person in matter, on its psychological characteristics. It was predicted that using different type of personal social media data would have an effect of the strength of the prediction of the Big Five, and from our main findings we could indeed conclude that this was true to a certain extent. The relevance of our research both in the scientific and practical field, along with its limitations are discussed in the following:

The results from the meta-analysis in our study not only show that social media data can predict the Big Five personality traits, but it also presents for each personality trait variable(s) that increase the prediction of the accuracy (e.g. Activity for Openness). Our study contributes to the claim that electronical devices we use nowadays (e.g. smartphones) not only are adding new elements to studies in cognitive science (Miller 2012), but have the potential to take a central role in psychological studies (Wilmer et al. 2017). Being able to use digital data to determine psychological attributes brings many possibilities than we could summarize here, but one that is relevant for our research is to use it to predict recidivism. Since two of the Big Five personality traits (i.e. Agreeableness and Conscientiousness) are known to correlate negatively with Psychopathy (O'Boyle et al. 2015), and studies show that Psychopathy is closely related to recidivism (e.g. Firestone et al. 2000; Loucks and Zamble 2000), our study gives new insights for research in how recidivism can be predicted in the Netherlands by the Dutch justice department with a rather unique approach.

With our study, we wished to improve the current way of assessing the risk of recidivism in the Netherlands on four areas where the current assessment method has major implications: efficiency, accuracy, scope, and fairness. Assessing the risk of recidivism (with the RISc) in the Netherlands is very cost- and labour intensive (Van der Knaap and Alberda 2009), with little success since the correlations with the actual recidivism rate are low at 0.30 (Wartna et al. 2008). Our study gives new insights suggesting that determining the risk of recidivism can be done more efficiently and accurately based on personality attributes derived from personal social media data. This insight has practical relevance for the prison system in the Netherlands, and could be further explored for practical applicability by practitioners. Another, practical contribution of our study could be that it provides a so-called "debiasing strategy" with respect to judgement and decision making in the Dutch justice system. Where cognitive biases can be understood as the systematic ways in which the choices and judgements of an individual are influenced by context and framing of information, debiasing is the effort to eliminate these cognitive biases (Fischoff 1981). Determining the risk of recidivism is prone to have many cognitive biases since people are judged by other individuals. Besides the more general known biases like the halo effect (Kahneman 2011), and the horn effect (Sigall and Ostrove 1975), many studies have shown that specifically legal systems are susceptible to biased judgements and erroneous decision making resulting in hundreds of innocent people having been wrongly convicted (Loftus 2019). Whether it be eyewitnesses (undeliberately) giving false information (McCloskey and Zaragoza 1985), or judges and jury members being bad at assessing whether someone is honest or not (Beaudry et al. 2015), it all strengthens the point that people are often not able to judge one another correctly. A practical insight from our study is that is makes sense to base decisions in the legal system on personal social media data, rather than on the interpretations of individual decision makers, of other individuals (witnesses, advisors) - because this would allow the Dutch justice department to eliminate (some of) the influence of cognitive biases in the judgement and decision making processes in the legal system.

There were some limitations to our study. First, there was no inter-rater reliability; degree of agreement among data collectors. In a meta-analysis, two or more researchers should normally first have trainings for data collection, then collect the data, and finally assess the effectiveness of their training and report the inter-rater reliability (McHugh 2012). Our meta-analysis was completely single-coded, which may have influenced the selection of our studies. Secondly, to find studies for our meta-analysis, we only looked in three major databases, with none being a Psychology database. The reason for us not searching in databases that are specialized in Psychology studies is that we did not have access to them with our TU Delft account. We may have missed out on prominent studies that could be included in our meta-analysis, which in turn could have affected the results we had, thus our conclusions. For instance, Harris and Bardey (2019) recently published their work (in 'Frontiers in Psychology') in which they found significant correlations between the pictures users share on Instagram, and the Big Five personality traits. On the contrary, we found the use of pictures to be a (weak) significant predictor only for Conscientiousness. Third, due to our strict inclusion criteria, we were obliged to exclude studies for our meta-analysis that did not investigate the relationship between personal digital data from social media, and each and every single one of the Big Five personality traits. Some studies only explored the relationship of one, or more (but not all) traits with personal social media data, and these studies were excluded due to our inclusion criteria. Celli and Rossi (2015) for example used Twitter as a social media tool to find the relationship between social media data in the form of activity (number of followers and retweets) and language (tweets) and only the Big Five personality trait Neuroticism. They found that neurotic people had the highest posting rate and retweeting score. These insights are valuable since they confirm our findings, and thus add strength to our claim that activity data is a good predictor for Neuroticism. Due to our first inclusion criterion, where we require studies to assess all the Big Five personality traits by way of a standardizes self-report measure (e.g. BFI, BFI-10, IPIP, Mini-IPIP, TIPI), we were obliged to exclude studies like Celli and Rossi (2015). Even though, throughout our report, we referred to our method as 'meta-analysis' (singular), we actually performed five separate meta-analyses (one for each of the Big Five traits), as mentioned in Section 4.7. In hindsight our first inclusion criterion could have been formulated in such a way that studies that studied one or more, but not all, of the Big Five traits could have been included.

For future work, our recommendations come in three ways. First off, our research did not cover the impact of cultural differences for the prediction of the Big Five traits with personal social media data. While the Big Five model maps out the personality of English speaking subjects well, since it is developed through lexical studies in English, the five personality traits are not sufficient to assess the personality of subjects from other languages (Ashton 2013). More recent cross-language studies have shown the existence of a sixth trait (i.e. Honesty-Humility) not only in other languages, but also in the English language (Ashton et al. 2004). In an attempt to find a structure across even more languages and cultures, De Raad et al (2014) found that three factors (i.e. dynamism, affiliation, and order) to have a higher chance to establish a structure that reappears across more languages and cultures. For example, using the five or six factors to study non-western (psycho-)lexically based trait structures (e.g. Chinese and South African) showed these structures to differ both from Western trait structures, as well as from each other (De Raad et al. 2014). Secondly, since our main objective was to explore to use personal digital data from social media to eventually assess the risk of recidivism in the Netherlands, we would advise future studies to investigate the possibilities to predict Psychopathy (instead of the Big Five) with social media data, since Psychopathy is strongly related to recidivism. As a sample, using ex-detainees, or even current detainees in the Netherlands who are in the final stage of their sentence, thus have the freedom to spend time outside the penitentiary facility, is an option one should take into consideration. Not only the correlation between Psychopathy assessed with the PCL-R, and predicted with personal social media data can be explored, but also the direct correlation between recidivism and personal social media data since one uses a sample from which the recidivism is known. Third, one should also explore the possibilities of using visual data (e.g. video) to assess the Big Five, since social media platforms focused on these type of data, like Snapchat, are rising in popularity. While Facebook is still the most used social media platform (±2.4 billion users), other platforms based on visual data such as YouTube (±1.0 billion), Instagram (±1.0 billion), and Snapchat (±0.2 billion) are catching up with Facebook -- since their growth is greater over the past couple of years (data retrieved from: https://www.statista.com). We included just three studies in our meta-analysis (Ferwerda et al. 2015, Guntuku et al. 2017, and Kim and Kim 2018) that used visual data (i.e. pictures) to assess the Big Five, while the trajectory of social media usage indicates that the portion of visual data (of the total social media data) will keep on growing the coming years (see also Harris and Bardey 2019).

Our study offers two contributions. First, we show different type of social media data to be significant predictors for the Big Five personality traits. In order to determine the Big Five from social media data, one can make use of just one, or more specific types of social media data to increase the accuracy of the prediction. Secondly, we present a novel method to assess the risk of recidivism in the Netherlands. While our study alone is not sufficient to actually determine the risk of recidivism with social media data, we provide clear recommendations where the Dutch justice department can improve the current way of assessing recidivism by making use of personal social media data to assess psychological attributes. Our insights will help the Dutch justice department make the assessment of the risk of recidivism (1) relatively effortless, (2) a lot cheaper, (3) more accurate, and (4) without cognitive bias. With further research in using psychological attributes assessed from social media data to determine recidivism, a reliable and valid assessment tool can be developed that will benefit the Dutch justice department greatly.

REFERENCES

Adviesbureau Van Montfoort, & Reclassering Nederland. (2004). *RISc versie 1.0. Recidive Inschattings Schalen. Handleiding.* Harderwijk, Nl: Flevodruk.

Anderson, C., John, O. P., Keltner, D., & Kring, A. M. (2001). Who attains social status? Effects of personality and physical attractiveness in social groups. *Journal of personality and social psychology*, *81*(1), 116.

Ashton, M. C. (2013). *Individual differences and personality*. Academic Press.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., ... & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *Journal of personality and social psychology*, *86*(2), 356.

Asscher, J. J., van Vugt, E. S., Stams, G. J. J., Deković, M., Eichelsheim, V. I., & Yousfi, S. (2011). The relationship between juvenile psychopathic traits, delinquency and (violent) recidivism: A meta-analysis. *Journal of child psychology and psychiatry*, *52*(11), 1134-1143.

Babinski, L. M., Hartsough, C. S., & Lambert, N. M. (2001). A comparison of self-report of criminal involvement and official arrest records. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *27*(1), 44-54.

Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012, June). Personality and patterns of Facebook usage. In *Proceedings of the 4th Annual ACM Web Science Conference* (pp. 24-32). ACM.

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin*, *99*(3), 388.

Beaudry, J. L., Lindsay, R. C., Leach, A. M., Mansour, J. K., Bertrand, M. I., & Kalmet, N. (2015). The effect of evidence type, identification accuracy, line-up presentation, and line-up administration on observers' perceptions of eyewitnesses. *Legal and Criminological Psychology*, *20*(2), 343-364.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088-1101.

Benning, S. D., Patrick, C. J., Hicks, B. M., Blonigen, D. M., & Krueger, R. F. (2003). Factor structure of the psychopathic personality inventory: validity and implications for clinical assessment. *Psychological assessment*, *15*(3), 340.

Benning, S. D., Patrick, C. J., Salekin, R. T., & Leistico, A. M. R. (2005). Convergent and discriminant validity of psychopathy factors assessed via self-report: A comparison of three instruments. *Assessment*, *12*(3), 270-289.

Blonigen, D. M., Patrick, C. J., Douglas, K. S., Poythress, N. G., Skeem, J. L., Lilienfeld, S. O., ... & Krueger, R. F. (2010). Multimethod assessment of psychopathy in relation to factors of internalizing and externalizing from the Personality Assessment Inventory: The impact of method variance and suppressor effects. *Psychological assessment*, *22*(1), 96.

Boendermaker, L. (1998). Eind goed, al goed?: De leefsituatie van jongeren een jaar na vertrek uit een justitiële behandelinrichting. *Onderzoek en beleid*.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Brinkley, C. A., Schmitt, W. A., Smith, S. S., & Newman, J. P. (2001). Construct validation of a self-report psychopathy scale: does Levenson's self-report psychopathy scale measure the same constructs as Hare's psychopathy checklist-revised?. *Personality and Individual Differences*, *31*(7), 1021-1038.

Brunell, A. B., Gentry, W. A., Campbell, W. K., Hoffman, B. J., Kuhnert, K. W., & DeMarree, K. G. (2008). Leader emergence: The case of the narcissistic leader. *Personality and Social Psychology Bulletin*, *34*(12), 1663-1676.
Carey, P. (2018). *Data protection: a practical guide to UK and EU law*. Oxford University Press, Inc..
Celli, F., & Rossi, L. (2015). Long chains or stable communities? The role of emotional stability in Twitter conversations. *Computational Intelligence*, *31*(1), 184-200.
Christie, R., & Geis, F. L. (1970). Studies in Machiavellianism. New York: Academic Press.
Cleckley, H. (1941). The mask of sanity; an attempt to reinterpret the so-called psychopathic personality.
Clower, C. E., & Bothwell, R. K. (2001). An exploratory study of the relationship between the Big Five and inmate recidivism. *Journal of Research in Personality*, *35*(2), 231-237.
Cohen, J. (1988). Statistical power analysis for the behavioral sciences 2nd edn.
Cooke, D. J., Kosson, D. S., & Michie, C. (2001). Psychopathy and ethnicity: Structural, item, and test generalizability of the Psychopathy Checklist—Revised (PCL-R) in Caucasian and African American participants. *Psychological assessment*, *13*(4), 531.
Cooper, H. (1986). The Integrative Research Review: A Systematic Approach Sage Publications: Beverly Hills, 1984, 143 pp. *Educational Researcher*, *15*(8), 17-18.
Cooper, H., & Hedges, L. V. (Eds.). (1993). *The handbook of research synthesis*. Russell Sage Foundation.
Costa, P. T., & McCrae, R. R. (1985). The NEO personality inventory.
Costa, P. T., & McCrae, R. R. (2010). The NEO Personality Inventory: 3. *Odessa, FL: Psychological assessment resources*.
Costa Jr, P. T., & McCrae, R. R. (1992). Neo personality inventory–revised (neo-pi-r) and neo five-factor inventory (neo-ffi) professional manual. *Odessa, FL: Psychological Assessment Resources*.
Costa Jr, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and individual Differences*, *12*(9), 887-898.
Da Silva, J. A. T., & Memon, A. R. (2017). CiteScore: A cite for sore eyes, or a valuable, transparent metric?. *Scientometrics*, *111*(1), 553-556.
De Raad, B., Barelds, D. P., Timmerman, M. E., De Roover, K., Mlačić, B., & Church, A. T. (2014). Towards a pan-cultural personality structure: Input from 11 psycholexical studies. *European Journal of Personality*, *28*(5), 497-510.
Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment*, *18*(2), 192.
Doucouliagos, H., & Stanley, T. D. (2009). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, *47*(2), 406-428.
Durlak, J. A., & Lipsey, M. W. (1991). A practitioner's guide to meta-analysis. *American Journal of Community Psychology*, *19*(3), 291-332.
Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463.
ECHR December 16, 1992, Publ. ECHR, Series A, vol. 251-B (Niemietz/Germany), par. 29.
ECHR June 1, 2010, 22978/05 (Gäfgen/Germany), par. 165
ECHR September 30, 2014, 8429/05 (Prezhdarovi/Bulgaria), par. 43
Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629-634.

EU Directive 2016/680. Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1527490238163&uri=CELEX:32016L0680
European Court of Human Right (2018). Guide on Article 8 of the European Convention on Human Rights: Retrieved from: https://www.echr.coe.int/Documents/Guide_Art_8_ENG.pdf
Eysenck, H. J. (1977). Crime and personality. London: Routledge and Kegan Paul.
Fazel, S., & Wolf, A. (2015). A systematic review of criminal recidivism rates worldwide: current difficulties and recommendations for best practice. *PloS one*, *10*(6), e0130390.
Firestone, P., Bradford, J. M., McCoy, M., Greenberg, D. M., Curry, S., & Larose, M. R. (1998). Recidivism in convicted rapists. *Journal of the American Academy of Psychiatry and the Law Online*, *26*(2), 185-200.
Firestone, P., Bradford, J. M., McCoy, M., Greenberg, D. M., Larose, M. R., & Curry, S. (1999). Prediction of recidivism in incest offenders. *Journal of Interpersonal Violence*, *14*(5), 511-531.
Firestone, P., Bradford, J. M., McCoy, M., Greenberg, D. M., Curry, S., & Larose, M. R. (2000). Prediction of recidivism in extrafamilial child molesters based on court-related assessments. *Sexual Abuse: A Journal of Research and Treatment*, *12*(3), 203-221.
Fischoff, B. (1981). *Debiasing* (No. PTR-1092-81-3). DECISION RESEARCH EUGENE OR.
Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of personality: A 10 year review. *Social and Personality Psychology Compass*, *7*(3), 199-216.
Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144.
Gavaghan, D. J., Moore, R. A., & McQuay, H. J. (2000). An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. *Pain*, *85*(3), 415-424.
Gerris, J. R., Houtmans, M. J. M., Kwaaitaal-Roosen, E. M., Schipper, J. C., Vermulst, A. A., & Janssens, J. M. A. M. (1998). Parents, adolescents and young adults in Dutch families: A longitudinal study. *Nijmegen: Institute of Family Studies, University of Nijmegen*.
Glass, G. V., Smith, M. L., & McGaw, B. (1981). *Meta-analysis in social research*. Sage Publications, Incorporated.
Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, *7*(1), 7-28.
Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, *37*(6), 504-528.
Grondwet 2018. Retrieved from: https://wetten.overheid.nl/BWBR0001840/2018-12-21
Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, *11*(1), 1-21.
Grubbs, F. E., & Beck, G. (1972). Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, *14*(4), 847-854.
Hare, R. D. (1970). Psychopathy: Theory and research.
Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and individual differences*, *1*(2), 111-119.
Hare, R. D. (1985). Comparison of procedures for the assessment of psychopathy. *Journal of Consulting and Clinical psychology*, *53*(1), 7.
Hare, R. D. (1991). *The Hare psychopathy checklist-revised: Manual*. Multi-Health Systems, Incorporated.

Hare, R. D. (1998). The Hare PCL-R: Some issues concerning its use and misuse. *Legal and criminological psychology*, *3*(1), 99-119.
Hare, R. D., Clark, D., Grann, M., & Thornton, D. (2000). Psychopathy and the predictive validity of the PCL-R: An international perspective. *Behavioral sciences & the law*, *18*(5), 623-645.
Hare, R. D., Harpur, T. J., & Hemphill, J. F. (1989). Scoring pamphlet for the Self-Report Psychopathy scale: SRP-II. *Unpublished document, Simon Fraser University, Vancouver, Canada*.
Hare, R. D., Hart, S. D., & Harpur, T. J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *Journal of abnormal psychology*, *100*(3), 391.
Hare, R. D., & Jutai, J. W. (1983). 12 CRIMINAL HISTORY OF THE MALE PSYCHOPATH: SOME PRELIMINARY DATA. *Prospective studies of crime and delinquency*, *2*, 225.
Harpur, T. J., Hakstian, A. R., & Hare, R. D. (1988). Factor structure of the Psychopathy Checklist. *Journal of consulting and clinical psychology*, *56*(5), 741.
Harpur, T. J., Hare, R. D., & Hakstian, A. R. (1989). Two-factor conceptualization of psychopathy: Construct validity and assessment implications. *Psychological Assessment: A Journal of consulting and clinical Psychology*, *1*(1), 6.
Harris, E., & Bardey, A. C. (2019). Do Instagram Profiles Accurately Portray Personality? An Investigation Into Idealized Online Self-Presentation. *Frontiers in psychology*, *10*, 871.
Harris, G. T., Rice, M. E., & Cormier, C. A. (1991). Psychopathy and violent recidivism. *Law and human behavior*, *15*(6), 625.
Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal justice and behavior*, *20*(4), 315-335.
Hart, S. D., Kropp, P. R., & Hare, R. D. (1988). Performance of male psychopaths following conditional release from prison. *Journal of consulting and clinical psychology*, *56*(2), 227.
Hayes, N., & Joseph, S. (2003). Big 5 correlates of three measures of subjective well-being. *Personality and Individual differences*, *34*(4), 723-727.
Hedges, L. V. (1984). Advances in statistical methods for meta-analysis. *New directions for program evaluation*, *1984*(24), 25-42.
Hedges, L.V., & Olkin, I. (1985). Statistical methods for meta-analysis.
Hemphill, J. F., Templeman, R., Wong, S., & Hare, R. D. (1998). Psychopathy and crime: Recidivism and criminal careers. In *Psychopathy: Theory, research and implications for society* (pp. 375-399). Springer, Dordrecht.
Hendriks, A. A. J. (1997). *The construction of the five-factor personality inventory (FFPI)* (Doctoral dissertation, Rijksuniversiteit Groningen).
Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557-560.
Howard, P., Clark, D., & Garnham, N. (2003). Evaluation and validation of the Offender Assessment System (OASys). *London: OASys Central Research Unit*.
Huchzermeier, C., Geiger, F., Bruß, E., Godt, N., Köhler, D., Hinrichs, G., & Aldenhoff, J. B. (2007). The relationship between DSM-IV cluster B personality disorders and psychopathy according to Hare's criteria: clarification and resolution of previous contradictions. *Behavioral sciences & the law*, *25*(6), 901-911.
Hunt, M. (1997). How science takes stock: The story of meta-analysis. Russell Sage Foundation.

Hunter, J. E., & Schmidt, F. L. (2004). Methods of meta-analysis: Correcting error and bias sin research findings. Sage.
Jakobwitz, S., & Egan, V. (2006). The dark triad and normal personality traits. *Personality and individual Differences*, *40*(2), 331-339.
John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The big five inventory—versions 4a and 54. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.
John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, *3*(2), 114-158.
Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, *52*(3), 621-652.
Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
Kerig, P. K., & Stellwagen, K. K. (2010). Roles of callous-unemotional traits, narcissism, and Machiavellianism in childhood aggression. *Journal of Psychopathology and Behavioral Assessment*, *32*(3), 343-352.
Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and individual differences*, *19*(1), 47-52.
Konrad, N. (2002). Prisons as new asylums. *Current Opinion in Psychiatry*, *15*(6), 583-587.
Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Machine learning*, *95*(3), 357-380.
Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802-5805.
Kroner, D. G., & Loza, W. (2001). Evidence for the efficacy of self-report in predicting nonviolent and violent criminal recidivism. *Journal of Interpersonal Violence*, *16*(2), 168-177.
Kuss, D. J., & Griffiths, M. D. (2011). Online social networking and addiction—a review of the psychological literature. *International journal of environmental research and public health*, *8*(9), 3528-3552.
Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, *5*(12), 2032-2033.
Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of personality and social psychology*, *68*(1), 151.
Li, L., Li, A., Hao, B., Guan, Z., & Zhu, T. (2014). Predicting active users' personality based on micro-blogging behaviors. *PloS one*, *9*(1), e84997.
Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal population. *Journal of personality assessment*, *66*(3), 488-524.
Lilienfeld, S. O., Widows, M. R., & Staff, P. A. R. (2005). Psychopathic personality inventoryTM-revised. *Social Influence (SOI)*, *61*(65), 97.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE publications, Inc.
Loftus, E. F. (2019). Eyewitness testimony. *Applied Cognitive Psychology*, *33*(4), 498-503.
Loucks, A. D., & Zamble, E. (2001). Predictors of criminal behavior and prison misconduct in serious female offenders. Correctional Service Canada.

Loza, W., & Loza-Fanous, A. (2001). The effectiveness of the Self-Appraisal Questionnaire in predicting offenders' postrelease outcome: A comparison study. *Criminal Justice and Behavior*, *28*(1), 105-121.

Madden, S. (2012). From databases to big data. *IEEE Internet Computing*, *16*(3), 4-6.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60-68.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*(1), 1.

McCrae, R. R., & Costa, P. T. (1985). Updating Norman's" adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of personality and social psychology*, *49*(3), 710.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, *52*(1), 81.

McCrae, R. R., & Costa, P. T. (2003). Personality in adulthood: A five-factor theory perspective. Guilford Press.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, *60*(2), 175-215.

McCrae, R. R., Martin, T. A., & Costa Jr, P. T. (2005). Age trends and age norms for the NEO Personality Inventory-3 in adolescents and adults. *Assessment*, *12*(4), 363-373.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, *22*(3), 276-282.

Međedović, J., Kujačić, D., & Knežević, G. (2012). Personality-related determinants of criminal recidivism. *Psihologija*, *45*(3), 277-294.

Megargee, E. I., & Bohn, M. J. (1979). *Classifying Criminal Offenders: New System Based on MMPI* (Vol. 82). SAGE Publications, Incorporated.

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on psychological science*, *7*(3), 221-237.

Miller, J. D., Lyman, D. R., Widiger, T. A., & Leukefeld, C. (2001). Personality disorders as extreme variants of common personality dimensions: Can the five factor model adequately represent psychopathy?. *Journal of personality*, *69*(2), 253-276.

Mischel, T. (1963). Psychology and explanations of human behavior. *Philosophy and Phenomenological Research*, *23*(4), 578-594.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, *151*(4), 264-269.

Mosteller, F., & Colditz, G. A. (1996). Understanding research synthesis (meta-analysis). *Annual review of public health*, *17*(1), 1-23.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of personality and social psychology*, *35*(4), 250.

Nurius, P. S., & Yeaton, W. H. (1987). Research synthesis reviews: An illustrated critique of" hidden" judgments, choices, and compromises.

O'Boyle, E. H., Forsyth, D. R., Banks, G. C., Story, P. A., & White, C. D. (2015). A meta-analytic test of redundancy and relative importance of the dark triad and five-factor model of personality. *Journal of personality*, *83*(6), 644-664.

Ortigosa, A., Carro, R. M., & Quiroga, J. I. (2014). Predicting user personality by mining social interactions in Facebook. *Journal of computer and System Sciences*, *80*(1), 57-71.

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.*, *57*, 401-421.
Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. *Development and psychopathology*, *21*(3), 913-938.
Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of research in personality*, *36*(6), 556-563.
Pescott, O. L., & Stewart, G. B. (2014). Assessing the impact of human trampling on vegetation: a systematic review and meta-analysis of experimental evidence. *PeerJ*, *2*, e360.
Politiewet 2012. Retrieved from: https://wetten.overheid.nl/BWBR0031788/2018-09-19
Poythress, N. G., Lilienfeld, S. O., Skeem, J. L., Douglas, K. S., Edens, J. F., Epstein, M., & Patrick, C. J. (2010). Using the PCL-R to help estimate the validity of two self-report measures of psychopathy with offenders. *Assessment*, *17*(2), 206-219.
Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., & Crowcroft, J. (2012, February). The personality of popular facebook users. In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (pp. 955-964). ACM.
Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality*, *41*(1), 203-212.
Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological reports*.
Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and human behavior*, *29*(5), 615-620.
Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, *86*(3), 638.
Rosenthal, R. (1994). Statistically describing and combining studies. *The handbook of research synthesis*, 231-244.
Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of cross-cultural psychology*, *38*(2), 173-212.
Schroeder, M. L., Schroeder, K. G., & Hare, R. D. (1983). Generalizability of a checklist for assessment of psychopathy. *Journal of Consulting and Clinical Psychology*, *51*(4), 511.
Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, *8*(9), e73791.
Schwartz, H. A., & Ungar, L. H. (2015). Data-driven content analysis of social media: a systematic overview of automated methods. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 78-94.
SCImago, (n.d). SJR - SCImago Journal & Country Rank [Portal]. Retrieved *28th of June 2019*, from http://scimagojr.com
Seidman, G. (2013). Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, *54*(3), 402-407.
Serin, R. C. (1996). Violent recidivism in criminal psychopaths. *Law and Human Behavior*, *20*(2), 207-217.
Serin, R. C., & Brown, S. L. (2001). Empirical guidelines for clinicians: Enhancing the practice of offender risk

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical psychology review*, *17*(8), 881-901.
Sigall, H., Ostrove, N. (1975). Beautiful but dangerous: effects of offender attractiveness and nature of the crime on juridic judgement. *Journal of Personality and Social Psychology, 31(3), 410*
Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American psychologist*, *32*(9), 752.
Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
Smith, M., & Robertson, I. T. (1993). Requirements of Measures: Validity and Meta-Analysis. In *The Theory and Practice of Systematic Personnel Selection* (pp. 117-145). Palgrave Macmillan, London.
Stevens, L. (2010). Voorlopige hechtenis en vrijheidsstraf. De strafrechter voor voldongen feiten?.
Urdan, T. C. (2011). *Statistics in plain English*. Routledge.
Van Dam, C., Janssens, J. M., & De Bruyn, E. E. (2005). PEN, Big Five, juvenile delinquency and criminal recidivism. *Personality and individual differences*, *39*(1), 7-19.
Van der Knaap, L.M., & Alberda, D.L. (2009). De predictieve valideit van de Recidive Inschattingsschalen (RISc). *Cashier 2009-12* Retrieved from: https://www.wodc.nl/onderzoeksdatabase/de-predictieve-validiteit-van-de-recidive-inschattingsschalen-risc.aspx
Van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in medicine*, *21*(4), 589-624.
Van Wingerden, S. G. C., Moerings, L. M., & Van Wilsem, J. A. (2011). Recidiverisico en straftoemeting. *Research Memoranda-Raad voor de Rechtspraak*.
Venables, N. C., Hall, J. R., & Patrick, C. J. (2014). Differentiating psychopathy from antisocial personality disorder: A triarchic model perspective. *Psychological medicine*, *44*(5), 1005-1013.
Verona, E., Patrick, C. J., & Joiner, T. E. (2001). Psychopathy, antisocial personality, and suicide risk. *Journal of abnormal psychology*, *110*(3), 462.
Vitacco, M. J., Neumann, C. S., & Jackson, R. L. (2005). Testing a four-factor model of psychopathy and its association with ethnicity, gender, intelligence, and violence. *Journal of consulting and clinical psychology*, *73*(3), 466.
Wang, M. C., & Bushman, B. J. (1999). Integrating results through meta-analytic review using SAS software. SAS institute.
Wartna, B. S. J., Blom, M., & Tollenaar, N. (2008). De WODC-recidivemonitor Den Haag.
Wet bescherming persoonsgegevens 2000. Retrieved from: https://wetten.overheid.nl/BWBR0011468/2018-05-01
Wartna, B. S. J., Tollenaar, N., Verweij, S., Timmermans, M., Witvliet, M., & Homburg, G. H. J. (2014). Terugval in recidive.
Wet politiegegevens. Retrieved from: https://wetten.overheid.nl/BWBR0022463/2018-05-01
Williams, K. M., Nathanson, C., & Paulhus, D. L. (2003, August). Structure and validity of the self-report psychopathy scale-III in normal populations. In *111th annual convention of the American Psychological Association*.
Williams, K. M., & Paulhus, D. L. (2004). Factor structure of the Self-Report Psychopathy scale (SRP-II) in non-forensic samples. *Personality and Individual Differences*, *37*(4), 765-778.

Williams, K. M., Paulhus, D. L., & Hare, R. D. (2007). Capturing the four-factor structure of psychopathy in college students via self-report. *Journal of personality assessment*, *88*(2), 205-219.
Wilmer, H. H., Sherman, L. E., & Chein, J. M. (2017). Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. *Frontiers in psychology*, *8*, 605.
Wilson, D. B., Gallagher, C. A., Coggeshall, M. B., & MacKenzie, D. L. (1999). A quantitative review and description of corrections-based education, vocation, and work programs. *Corrections Management Quarterly*, *3*, 8-18.
Wilson, N. (2000, November). Does the PCL-SV predict serious violent reoffending for offenders appearing before the National Parole Board? (Tech. Rep.). Hamilton, New Zealand: Corrections Psychological Service.
Wolf, F. M. (1990). *Methodological observations on bias*. Russell Sage Foundation.
Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, *44*(3), 363-373.
Zágon, I. K., & Jackson, H. J. (1994). Construct validity of a psychopathy measure. *Personality and Individual Differences*, *17*(1), 125-135.

Articles in Meta-Analysis

Reference marked with an asterisk indicate studies that were read in full for the meta-analysis - marked with a double asterisk indicate studies that are included in the meta-analysis. The list is not in alphabetic order, but the order used as in Appendix A.

1. *Okumura, N., & Okumura, M. (2015, November). A Construction of Knowledge Base for Personality Estimation based on Submitted Text Data in Twitter or Blogs. In *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (pp. 418-423). SCITEPRESS-Science and Technology Publications, Lda.

2. *Wall, H. J., Kaye, L. K., & Malone, S. A. (2016). An exploration of psychological factors on emoticon usage and implications for judgement accuracy. *Computers in Human Behavior*, *62*, 70-78.

3. *Hagger-Johnson, G., Egan, V., & Stillwell, D. (2011). Are social networking profiles reliable indicators of sensational interests?. *Journal of Research in Personality*, *45*(1), 71-76.

4. *Collins, S., Sun, Y., Kosinski, M., Stillwell, D., & Markuzon, N. (2015, March). Are you satisfied with life?: Predicting satisfaction with life from facebook. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 24-33). Springer, Cham.

5. *Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology*, *108*(6), 934.

6. *Solinger, C., Hirshfield, L., Hirshfield, S., Friendman, R., & Leper, C. (2014, June). Beyond Facebook Personality Prediction. In *International Conference on Social Computing and Social Media*(pp. 486-493). Springer, Cham.

7. **Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., ... & De Cock, M. (2016). Computational personality recognition in social media. *User modeling and user-adapted interaction*, *26*(2-3), 109-142.

8. *Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036-1040.

9. *Yu, J., & Markov, K. (2017, November). Deep learning based personality recognition from facebook status updates. In *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)* (pp. 383-387). IEEE.

10. *Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, *32*(2), 74-79.

11. *Xue, D., Wu, L., Hong, Z., Guo, S., Gao, L., Wu, Z., ... & Sun, J. (2018). Deep learning-based personality recognition from text posts of online social networks. *Applied Intelligence*, *48*(11), 4232-4246.

12. **Kleanthous, S., Herodotou, C., Samaras, G., & Germanakos, P. (2016, July). Detecting personality traces in users' social activity. In *International conference on social computing and social media*(pp. 287-297). Springer, Cham.

13. *Alsadhan, N., & Skillicorn, D. (2017, November). Estimating Personality from Social Media Posts. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 350-356). IEEE.

14. *Farnadi, G., Sitaraman, G., Rohani, M., Kosinski, M., Stillwell, D., Moens, M. F., ... & De Cock, M. (2014). How are you doing?: Emotions and personality in Facebook. In *2nd Workshop on Emotions and Personality in Personalized Services (EMPIRE 2014); workshop at the 22nd Conference on User Modelling, Adaptation and Personalization (UMAP 2014)* (pp. 45-56).

15. *Celli, F., & Lepri, B. (2018). Is Big Five better than MBTI? A personality computing challenge using Twitter data.

16. *Ting, T. L., & Varathan, K. D. (2018). Job recommendation using facebook personality scores. Malaysian Journal of Computer Science, 31(4), 311-331.

17. *Gou, L., Zhou, M. X., & Yang, H. (2014, April). KnowMe and ShareMe: understanding automatically discovered personality traits from social media and user sharing preferences. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 955-964). ACM.

18. **Kulkarni, V., Kern, M. L., Stillwell, D., Kosinski, M., Matz, S., Ungar, L., ... & Schwartz, H. A. (2018). Latent human traits in the language of social media: An open-vocabulary approach. PloS one, 13(11), e0201703.

19. *Chapsky, D. (2011, July). Leveraging online social networks and external data sources to predict personality. In 2011 International Conference on Advances in Social Networks Analysis and Mining (pp. 428-433). IEEE.

20. *Wald, R., Khoshgoftaar, T., & Sumner, C. (2012, August). Machine prediction of personality from Facebook profiles. In 2012 IEEE 13th International Conference on Information Reuse & Integration (Iri) (pp. 109-115). IEEE.

21. *Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. Machine learning, 95(3), 357-380.

22. *Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical preferences predict personality: evidence from active listening and facebook likes. Psychological science, 29(7), 1145-1158.

23. *Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011, October). Our twitter profiles, our selves: Predicting personality with twitter. In 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing (pp. 180-185). IEEE.

24. *Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012, June). Personality and patterns of Facebook usage. In Proceedings of the 4th annual ACM web science conference (pp. 24-32). ACM.

25. **Bhardwaj, S., Atrey, P. K., Saini, M. K., & El Saddik, A. (2016). Personality assessment using multiple online social networks. Multimedia Tools and Applications, 75(21), 13237-13269.

26. *Souri, A., Hosseinpour, S., & Rahmani, A. M. (2018). Personality classification based on profiles of social networks' users and the five-factor model of personality. Human-centric Computing and Information Sciences, 8(1), 24.

27. *Pratama, B. Y., & Sarno, R. (2015, November). Personality classification based on Twitter text using Naive Bayes, KNN and SVM. In 2015 International Conference on Data and Software Engineering (ICoDSE) (pp. 170-174). IEEE.

28. **Tsai, T. H., Chang, H. T., Chang, Y. C., & Chang, Y. S. (2017). Personality disclosure on social network sites: An empirical examination of differences in Facebook usage behavior, profile contents and privacy settings. Computers in Human Behavior, 76, 469-482.

29. *Noë, N., Whitaker, R. M., & Allen, S. M. (2016, August). Personality homophily and the local network characteristics of facebook. In 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM) (pp. 386-393). IEEE.

30. *Tandera, T., Suhartono, D., Wongso, R., & Prasetio, Y. L. (2017). Personality prediction system from facebook users. Procedia computer science, 116, 604-611.

31. *Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2018). Personality Predictions Based on User Behavior on the Facebook Social Media Platform. IEEE Access, 6, 61959-61969.

32. *Yuan, C., Wu, J., Li, H., & Wang, L. (2018, July). Personality Recognition Based on User Generated Content. In 2018 15th International Conference on Service Systems and Service Management (ICSSSM) (pp. 1-6). IEEE.

33. *Da Silva, B. B. C., & Paraboni, I. (2018, September). Personality Recognition from Facebook Text. In International Conference on Computational Processing of the Portuguese Language (pp. 107-114). Springer, Cham.

34. *Vaidhya, M., Shrestha, B., Sainju, B., Khaniya, K., & Shakya, A. (2017, November). Personality Traits Analysis from Facebook Data. In 2017 21st International Computer Science and Engineering Conference (ICSEC) (pp. 1-5). IEEE.

35. *Alam, F., Stepanov, E. A., & Riccardi, G. (2013, June). Personality traits recognition on social network-facebook. In Seventh International AAAI Conference on Weblogs and Social Media.

36. *Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. PloS one, 8(9), e73791.

37. **Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012, December). Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In 2012 11th International Conference on Machine Learning and Applications (Vol. 2, pp. 386-393). IEEE.

38. *Howlader, P., Pal, K. K., Cuzzocrea, A., & Kumar, S. D. (2018, April). Predicting facebook-users' personality based on status and linguistic features via flexible regression analysis techniques. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (pp. 339-345). ACM.

39. *Rumagit, R.Y., Girsang, A.S. (2018). Predicting Personality Traits Of Facebook Users Using Text Mining. In Journal of Theoretical and Applied Information Technology 96 (20)

40. **Ferwerda, B., Schedl, M., & Tkalcic, M. (2015, September). Predicting personality traits with instagram pictures. In *Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015* (pp. 7-10). ACM.

41. **Golbeck, J., Robles, C., & Turner, K. (2011, May). Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 253-262). ACM.

42. **Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802-5805.

43. *Farnadi, G., Zoghbi, S., Moens, M. F., & De Cock, M. (2013, June). Recognising personality traits using Facebook status updates. In *Seventh International AAAI Conference on Weblogs and Social Media*.

44. **Guntuku, S. C., Lin, W., Carpenter, J., Ng, W. K., Ungar, L. H., & Preoţiuc-Pietro, D. (2017, June). Studying personality through the content of posted and liked images on Twitter. In *Proceedings of the 2017 ACM on web science conference* (pp. 223-227). ACM.

45. *Carducci, G., Rizzo, G., Monti, D., Palumbo, E., & Morisio, M. (2018). TwitPersonality: Computing Personality Traits from Tweets Using Word Embeddings and Supervised Learning. *Information*, *9*(5), 127.

46. *Farnadi, G., Tang, J., De Cock, M., & Moens, M. F. (2018, February). User profiling through deep multimodal fusion. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 171-179). ACM.

47. **Kim, Y., & Kim, J. H. (2018). Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study. *Information Processing & Management*, *54*(6), 1101-1114.

48. *Segalin, C., Celli, F., Polonio, L., Kosinski, M., Stillwell, D., Sebe, N., ... & Lepri, B. (2017, October). What your Facebook profile picture reveals about your personality. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 460-468). ACM.

49. *Ferwerda, B., & Tkalcic, M. (2018). You Are What You Post: What the Content of Instagram Pictures Tells About Users' Personality. In *The 23rd International on Intelligent User Interfaces, March 7-11, Tokyo, Japan*.

50. *Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*, *46*(6), 710-718.

APPENDIX A

This appendix contains the list of studies that were read in full for our meta-analysis. The list includes 50 papers (including 61 studies in total) that we coded according to our coding procedure (4.5).

Table A.1 The 50 papers (61 studies) that we coded for our meta-analysis

Article/Study	**Quality**	**Big Five Scale**	**Social Media**	**Digital Data**	**Sample**
1. Okumura and Okumura (2015)	Low	NEO-FFI	Twitter	Language	23
2(#1). Wall et al. (2016)	High	IPIP	Facebook	Emoticon	92
2(#2). Wall et al. (2016)	High	IPIP	Facebook	Emoticon	54
3. Hagger-Johnson et al. (2011)	High	IPIP	Facebook	Demographic	694
4. Collins et al. (2015)	Low	IPIP	Facebook	Demographic Activity Language	9.242
5(#1). Park et al. (2015)	High	IPIP	Facebook	Language	4.824
5(#2). Park et al. (2015)	High	IPIP	Facebook	Language	2.324
5(#3). Park et al. (2015)	High	IPIP	Facebook	Language	1.943
6. Solinger et al. (2014)	Low	BFI	Facebook	Demographic Activity Pictures	20
7(#1). Farnadi et al. (2016)	High	IPIP	Facebook	Demographic Activity Language	3.371
7(#2). Farnadi et al. (2016)	High	TIPI	YouTube	-	404
7(#3). Farnadi et al. (2016)	High	BFI	Twitter	Activity	44
8. Youyou et al. (2015)	High	IPIP	Facebook	Activity	1.919
9. Yu and Markov (2017)	Low	IPIP	Facebook	Activity Language	250
10. Majumder et al. (2017)	High	not given	none	Language	2.467
11. Xue et al. (2018)	Medium	IPIP	Facebook	Language	115.864
12. Kleanthous et al. (2016)	Medium	IPIP	Facebook	Activity	62
13. Alsadhan and Skillicorn (2017)	High	IPIP	Facebook	Language	250
14. Farnadi et al. (2014)	Medium	IPIP	Facebook	Demographic Language	5.865
15. Celli and Lepri (2018)	Low	BFI	Twitter	Activity Language	900
16. Ting and Varathan (2018)	Low	BFI	Facebook	Activity	50
17. Gou et al. (2014)	High	IPIP	Twitter	Language	224
18. Kulkarni et al. (2018)	High	IPIP	Facebook	Language	49.139
19. Chapsky (2011)	Low	IPIP	Facebook	Demographic	615
20. Wald et al. (2012)	Low	BFI	Facebook	Demographic Activity Language	537
21(#1). Kosinski et al. (2014)	High	IPIP	Facebook	Activity	153.000
21(#2). Kosinski et al. (2014)	High	IPIP	Facebook	Demographic Activity	+354.000
22(#1). Nave et al. (2018)	High	IPIP	Facebook	Demographic Activity	22.252

22(#2). Nave et al. (2018)	High	IPIP	Facebook	Demographic Activity	21.929
23. Quercia et al. (2011)	High	IPIP	Twitter	Activity	335
24. Bachrach et al. (2012)	Low	IPIP	Facebook	Activity	180.000
25(#1). Bhardwaj et al. (2016)	High	BFI	Facebook	Activity	31
25(#2). Bhardwaj et al. (2016)	High	BFI	LinkedIn	Activity Language	31
26. Souri et al. (2018)	High	NEO-FF-R	Facebook	Demographic Activity	100
27. Pratama and Sarno (2015)	Low	IPIP	Twitter	Language	250
28. Tsai et al. (2017)	High	BFI	Facebook	Activity	111
29. Noë et al. (2016)	High	IPIP	Facebook	Demographic Activity	313.699
30(#1). Tandera et al. (2017)	Low	IPIP	Facebook	Activity Language	250
30(#2). Tandera et al. (2017)	Low	Apply Magic Sauce	Facebook	Activity Language	150
31. Tadesse et al. (2018)	High	IPIP	Facebook	Activity Language	250
32. Yuan et al. (2018)	Low	IPIP	Facebook	Language	250
33. Da Silva and Paraboni (2018)	High	BFI	Facebook	Language	1.039
34. Vaidhya et al. (2017)	Low	IPIP	Facebook	Language	250
35. Alam et al. (2013)	High	IPIP	Facebook	Language	250
36. Schwartz et al. (2013)	High	IPIP	Facebook	Language	74.941
37. Sumner et al. (2012)	Low	TIPI	Twitter	Language	2.927
38. Howlader et al. (2018)	Medium	IPIP	Facebook	Activity Language	115.872
39. Rumagit and Girsand (2018)	Low	BFI	Facebook	Language	345
40. Ferwerda et al. (2015)	Medium	BFI	Instagram	Pictures	113
41. Golbeck et al. (2011)	High	BFI	Facebook	Demographic Activity Language	167
42. Kosinski et al. (2013)	High	IPIP	Facebook	Activity	54.373
43. Farnadi et al. (2013)	High	IPIP	Facebook	Activity Language	250
44(#1). Guntuku et al. (2017)	Medium	BFI	Twitter	Pictures	436
44(#2). Guntuku et al. (2017)	Medium	none	Twitter	Pictures	4.132
45(#1). Carducci et al. (2018)	Low	IPIP	Facebook	Activity Language	250
45(#2). Carducci et al. (2018)	Low	BFI	Twitter	Activity	24
46. Farnadi et al. (2018)	High	IPIP	Facebook	Activity Language Pictures	5.670
47. Kim and Kim (2018)	High	BFI	Instagram	Pictures	179
48. Segalin et al. (2017)	High	IPIP	Facebook	Pictures	11.736
49. Ferwerda and Tkalcic (2018)	Low	BFI	Instagram	Pictures	193
50. Qiu et al. (2012)	High	BFI	Twitter	Language	142

APPENDIX B: META-ANALYSIS RESULTS

This appendix contains all the computations, raw data, and plots, that we gathered in our meta-analysis (Ch. 6) using Excel (B.1) and Comprehensive Meta Analysis (B.2 - B.7).

B.1 Grubb's Test

From top to bottom, we used the same order for our effect sizes as in Table 6.1 (6.1.1). The Grubb's test statistic is calculated by taking the absolute value of the effect size minus the effect size mean, divided by the standard deviation (Grubbs 1969):

$$G_i = \frac{|E_i - \bar{E}|}{S}$$

where G_i is the Grubb's statistic for study i, E_i the effect size for study i, $\bar{E}$ the mean effect size, and S the standard deviation. The critical value (for T) we located in Table I from Grubbs and Beck (1972).

The third effect size from the top, corresponding to Bhardwaj et al. (2016) [#2], was the outlier for Agreeableness, Extraversion, and Neuroticism. When we excluded this effect size, we can see that no more outliers were present with the remaining ten values. For both Conscientiousness and Openness there were no outliers at all.

Effet Sizes	Grubb's test statistics		Effet Sizes	Grubb's test statistics
0,39	0,548811495		0,39	1,210962137
0,16	0,617989414		0,16	0,585949421
0,75	2,375108569			
0,23	0,262876094		0,23	0,039063295
0,15	0,668719888		0,15	0,664076011
0,48	1,005385763		0,48	1,914101443
0,23	0,262876094		0,23	0,039063295
0,2	0,415067517		0,2	0,273443063
0,03	1,27748558		0,03	1,601595085
0,3	0,092237226		0,3	0,507822832
0,18	0,516528465		0,18	0,429696242
Mean	0,281818182		**Mean**	0,235
St. Dev.	0,197120176		**St. Dev.**	0,127997396
# values	11		**# values**	10
Critical Value	2.234		**Critical Value**	2.176

Fig. B.1 Grubb's test for Agreeableness

Effet Sizes	Grubb's test statistics
0,41	0,270703183
0,14	0,966202129
0,91	2,561268575
0,22	0,599711666
0,23	0,553900358
0,55	0,912061493
0,22	0,599711666
0,27	0,370655127
0,26	0,416466435
0,4	0,224891875
0,25	0,462277743
Mean	0,350909091
St. Dev.	0,218286717
# values	11
Critical Value	2.234

Effet Sizes	Grubb's test statistics
0,41	0,947251863
0,14	1,276730771
0,22	0,617772954
0,23	0,535403227
0,55	2,100428043
0,22	0,617772954
0,27	0,205924318
0,26	0,288294045
0,4	0,864882135
0,25	0,370663772
Mean	0,295
St. Dev.	0,121403826
# values	10
Critical Value	2.176

Fig. B.2 Grubb's test for Extraversion

Effet Sizes	Grubb's test statistics
0,48	0,766064266
0,29	0,204283804
0,75	2,144979946
0,2	0,663922364
0,16	0,868206168
0,6	1,378915679
0,15	0,91927712
0,24	0,45963856
0,28	0,255354755
0,29	0,204283804
0,19	0,714993315
Mean	0,33
St. Dev.	0,195806026
# values	11
Critical Value	2.234

Fig. B.3 Grubb's test for Conscientiousness

Effet Sizes	Grubb's test statistics
0,42	0,793480813
0,18	0,69218539
0,39	0,607772538
0,22	0,444574356
0,28	0,073157805
0,65	2,217244259
0,17	0,754088149
0,15	0,877893666
0,16	0,815990907
0,43	0,855383572
0,16	0,815990907
Mean	0,291818182
St. Dev.	0,16154369
# values	11
Critical Value	2.234

Fig. B.4 Grubb's test for Openness

Effet Sizes	Grubb's test statistics		Effet Sizes	Grubb's test statistics
0,43	0,483833456		0,43	1,101352972
0,09	0,876720774		0,09	0,97898042
0,9	2,364599597			
0,19	0,476557765		0,19	0,367117657
0,22	0,356508862		0,22	0,183558829
0,53	0,883996465		0,53	1,713215735
0,4	0,363784553		0,4	0,917794144
0,13	0,71665557		0,13	0,734235315
0,01	1,196851181		0,01	1,46847063
0,3	0,036378455		0,3	0,305931381
0,2	0,436541464		0,2	0,305931381
Mean	0,309090909		**Mean**	0,25
St. Dev.	0,249898161		**St. Dev.**	0,163435342
# values	11		**# values**	10
Critical Value	2.234		**Critical Value**	2.176

Fig. B.5 Grubb's test for Neuroticism

B.2 Forest Plots

We made use of the free online tool 'Forest Plot Generator' provided by Evidence Partners (https://www.evidencepartners.com/resources/forest-plot-generator/) to generate the forest plots for each personality trait.

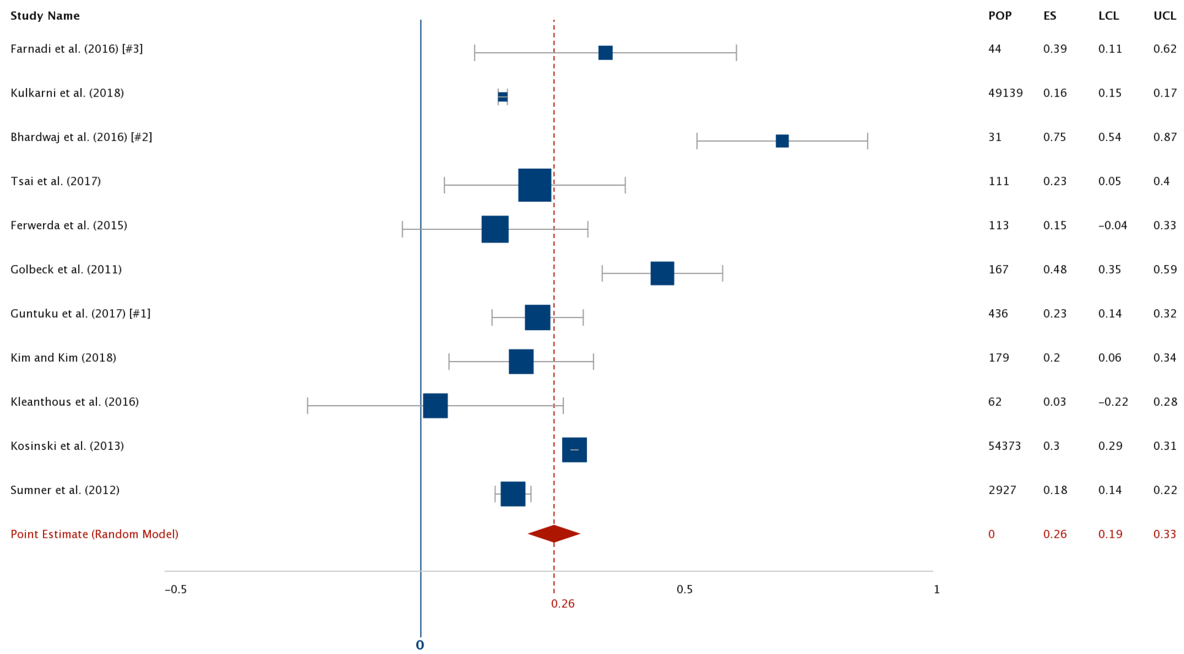

Fig. B.6 Forest plot for Agreeableness

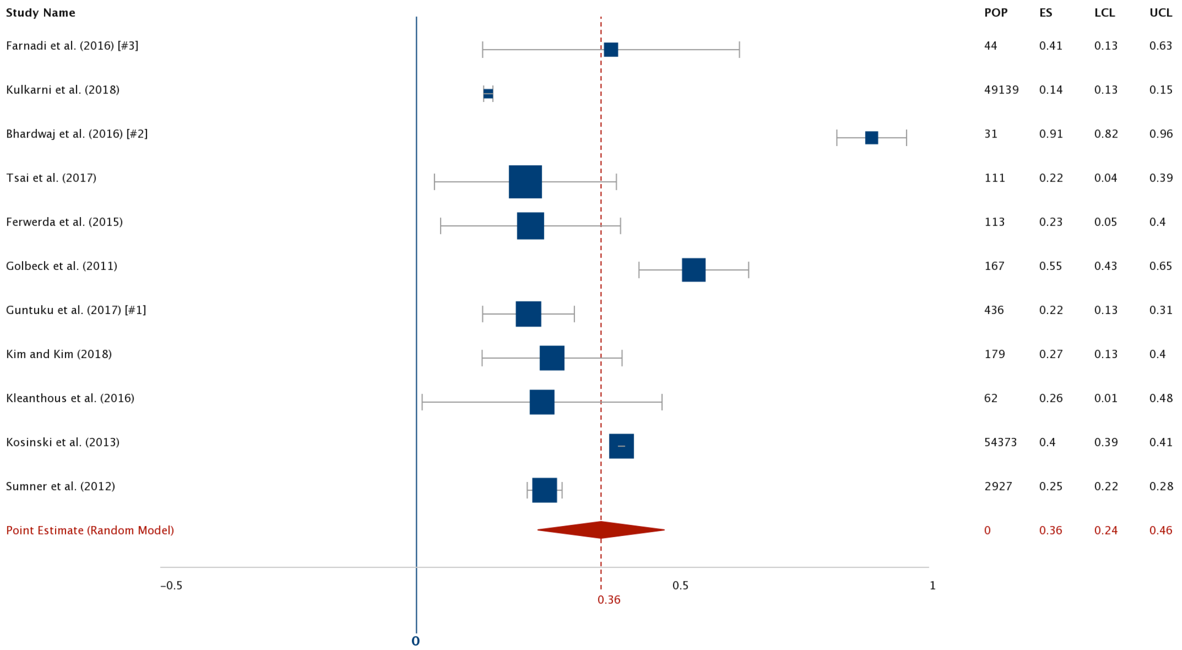

Fig. B.7 Forest plot for Extraversion

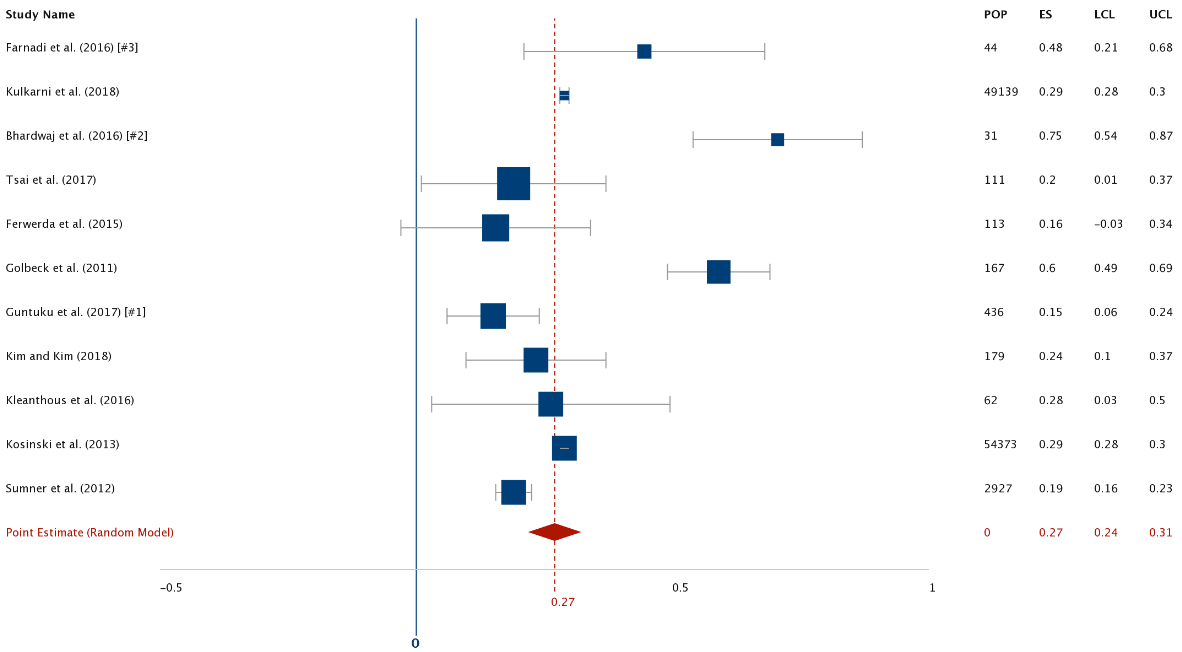

Fig. B.8 Forest plot for Conscientiousness

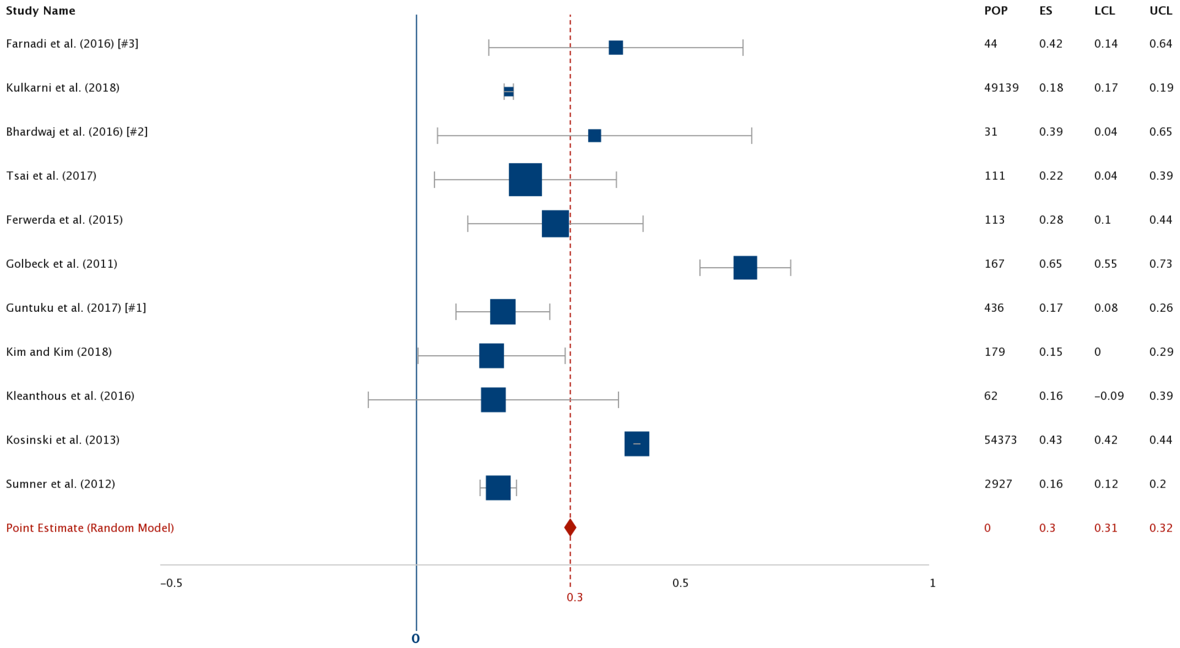

Fig. B.9 Forest plot for Openness

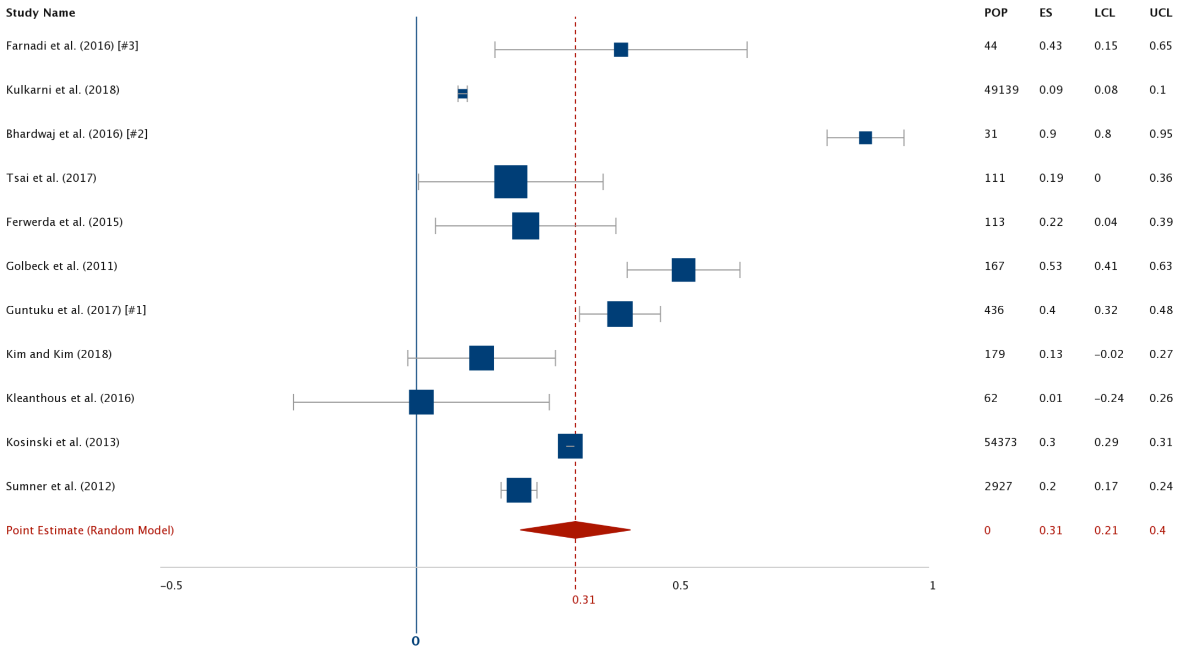

Fig. B.10 Forest plot for Neuroticism

B.3 Funnel plots

The funnel plot is a plot of a measure of study size (the standard error in our case) on the vertical axis as a function of effect size on the horizontal axis.

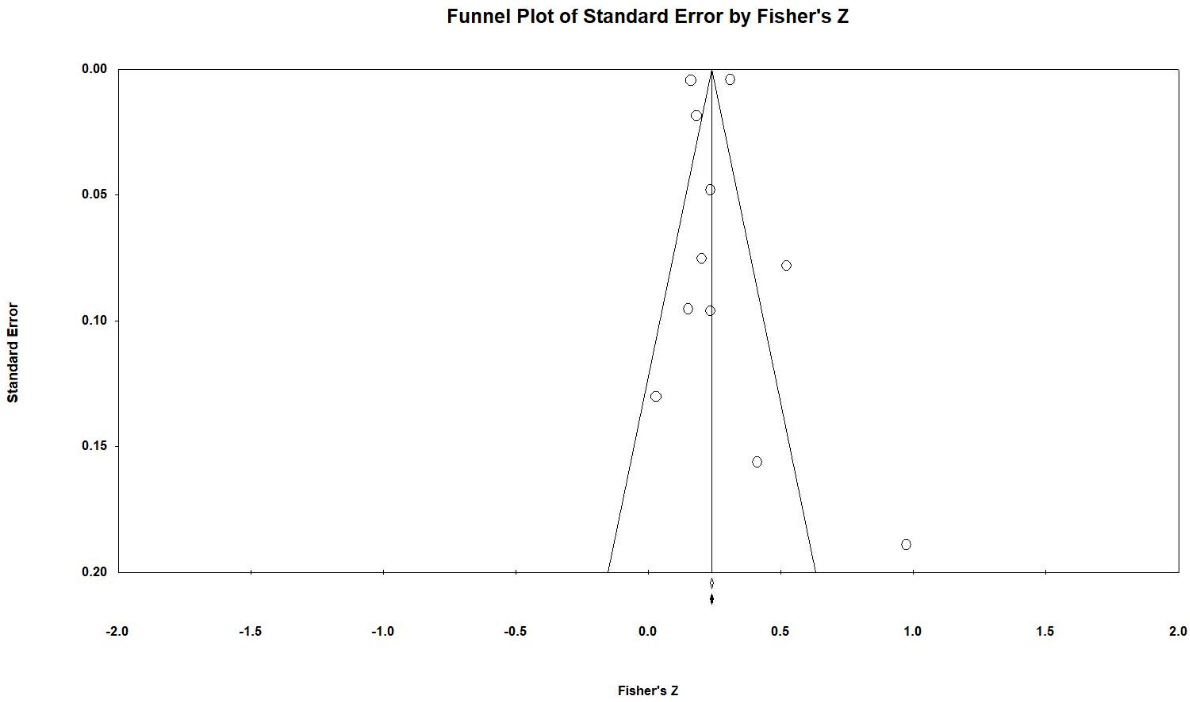

Fig. B.11 Funnel plot for Agreeableness

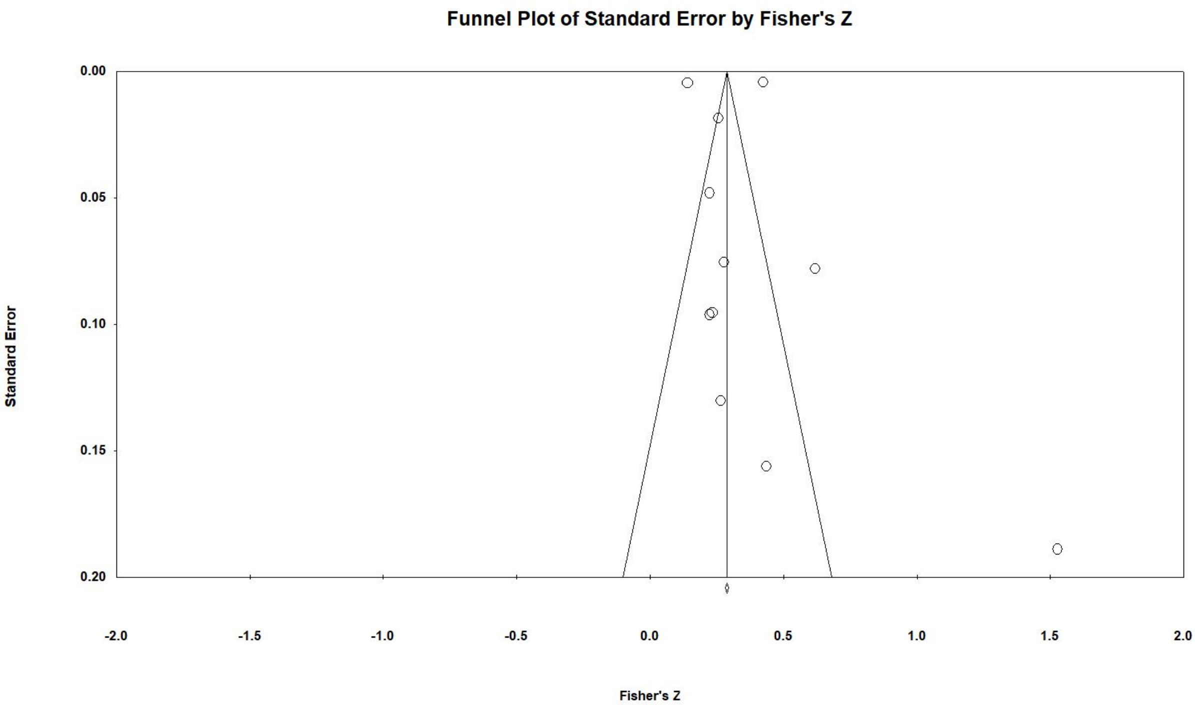

Fig. B.12 Funnel plot for Extraversion

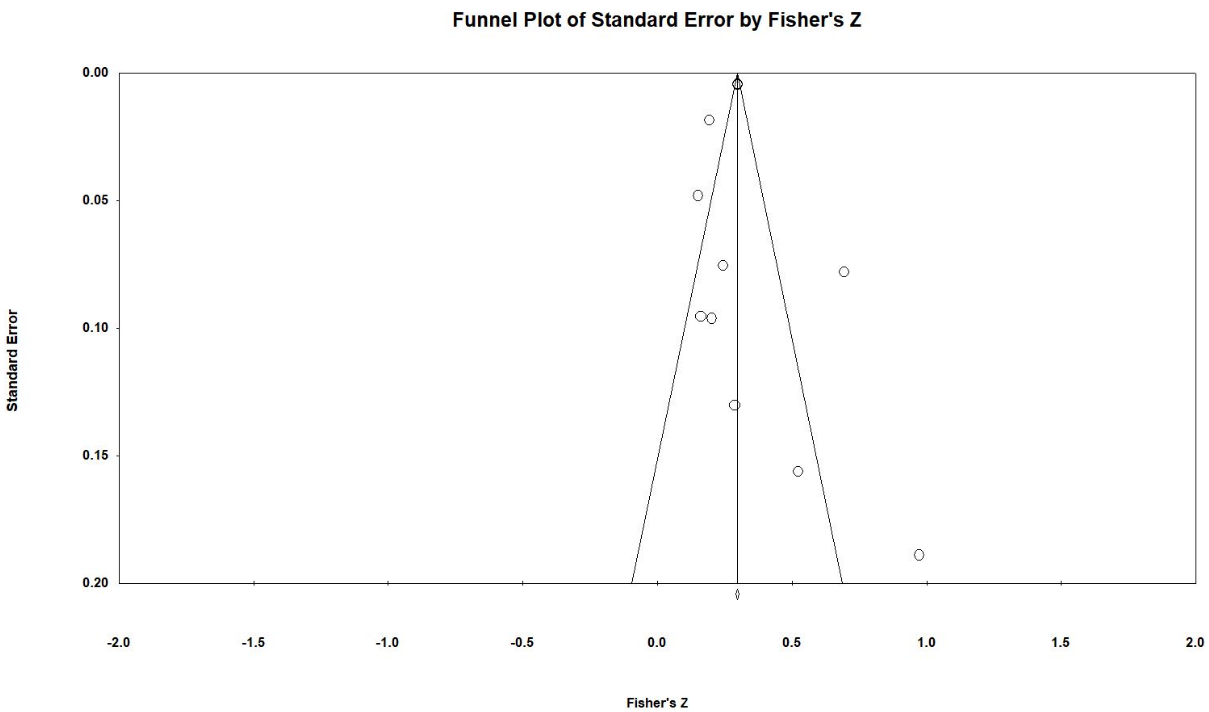

Fig. B.13 Funnel plot for Conscientiousness

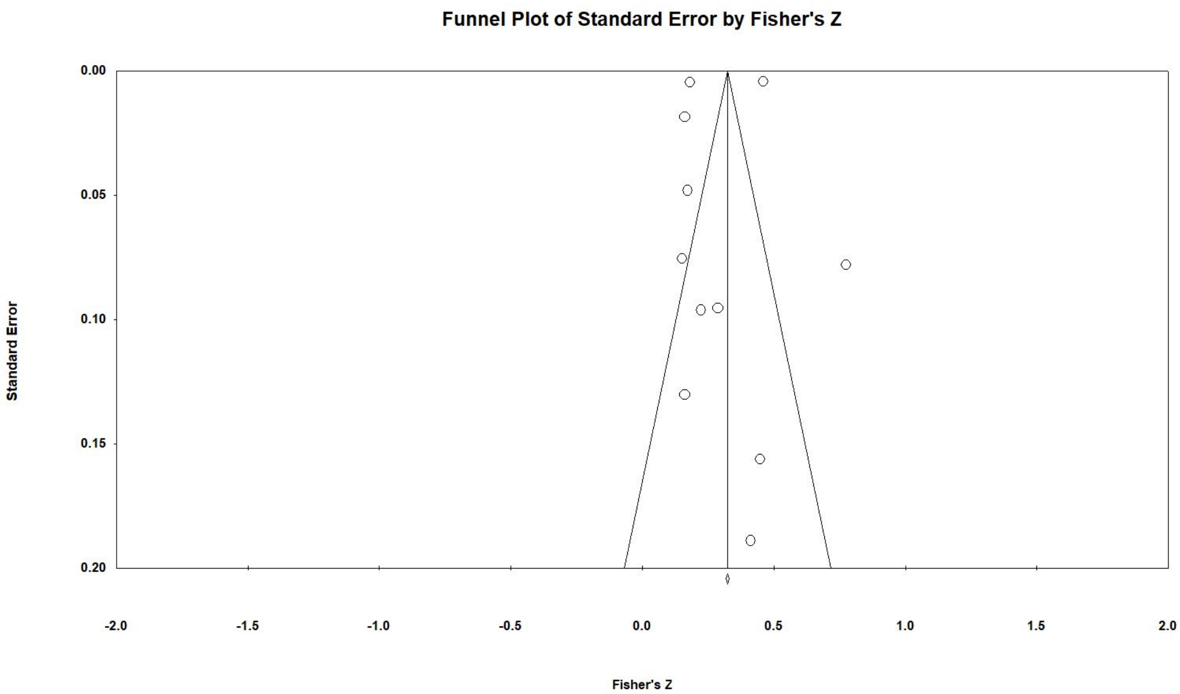

Fig. B.14 Funnel plot for Openness

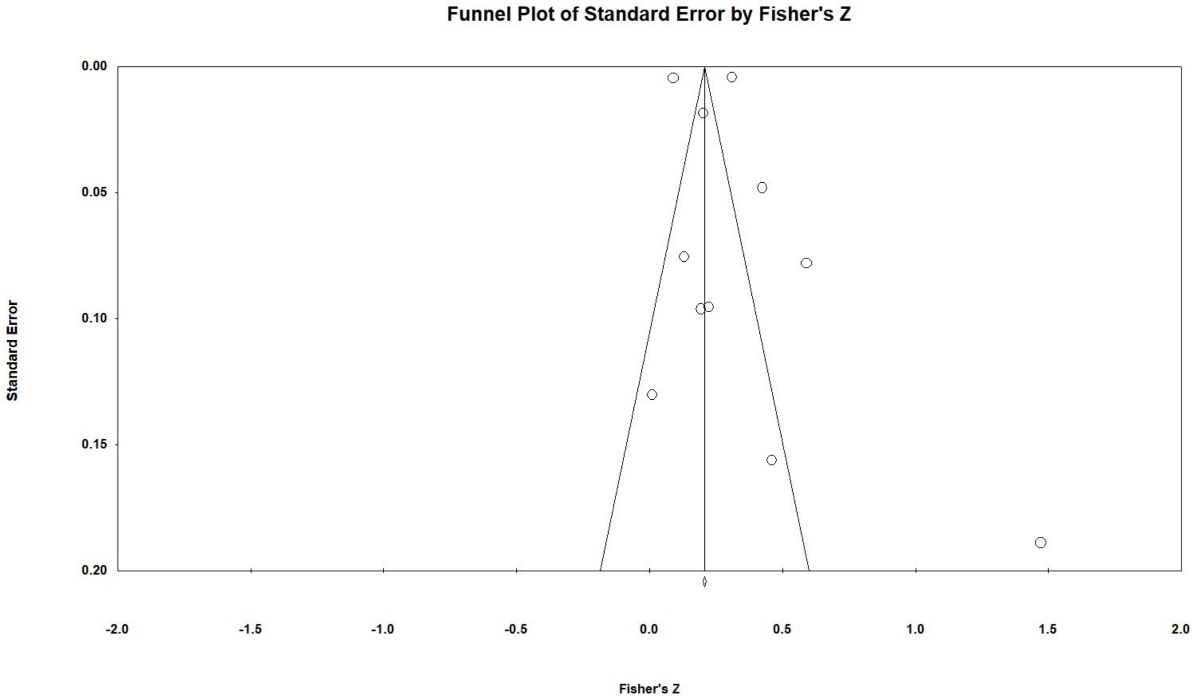

Fig. B.15 Funnel plot for Neuroticism

B.4 Fail Safe N

Table B.1 Results of the fail-safe N analysis

	Agreeableness	Extraversion	Conscientiousness	Openness	Neuroticism
Z-value	43.44212	53.60157	53.05866	55.30770	41.12690
P-value	0.00000	0.00000	0.00000	0.00000	0.00000
Alpha	0.05000	0.05000	0.05000	0.05000	0.05000
Tails	2.00000	2.00000	2.00000	2.00000	2.00000
Z for alpha	1.95996	1.95996	1.95996	1.95996	1.95996
# observed studies	11.00000	11.00000	11.00000	11.00000	11.00000
Fail-safe N	5394.00000	8217.00000	8051.00000	8749.00000	4833.00000

B.5 Begg and Mazumdar Rank Correlation

Table B.2 Results of the Begg and Mazumdar rank correlation test

	Agreeableness	Extraversion	Conscientiousness	Openness	Neuroticism
Kendall's S statistic (P-Q)	13.00000	19.00000	15.00000	17.00000	5.00000
Tau	0.21818	0.32727	0.25455	0.29091	0.07273
z-value for tau	0.93420	1.40130	1.08990	1.24560	0.31140
P-value (1-tailed)	0.17510	0.08056	0.13788	0.10646	0.37775
P-value (2-tailed)	0.35020	0.16112	0.27576	0.21291	0.75550

B.6 Egger's Regression of the Intercept

Table B.3 Results of the Egger's regression of the intercept test

	Agreeableness	Extraversion	Conscientiousness	Openness	Neuroticism
Intercept	0.28524	0.71923	-0.02953	-1.18343	1.69001
Standard error	2.96352	5.54098	1.11019	5.52423	4.35629
95% lower limit (2-tailed)	-6.41870	-11.81535	-2.54096	-13.68011	-8.16460
95% upper limit (2-tailed)	6.98919	13.25380	2.48190	11.31326	11.54462
t-value	0.09625	0.12980	0.02660	0.21422	0.38795
df	9.00000	9.00000	9.00000	9.00000	9.00000
P-value (1-tailed)	0.46272	0.44979	0.48968	0.41757	0.35354
P-value (2-tailed)	0.92543	0.89958	0.97936	0.83515	0.70707

B.7 Duval and Tweedie's Trim and Fill

Table B.4 Results of Duval and Tweedie's Trim and Fill for Agreeableness

	Observed values	Adjusted values
Studies trimmed	0	0
Fixed Effects		
Point Estimate	0.23365	0.23365
Lower Limit	0.22799	0.22799
Upper Limit	0.23929	0.23929
Random Effects		
Point Estimate	0.25832	0.25832
Lower Limit	0.18665	0.18665
Upper Limit	0.32724	0.32724
Q Value	608.98056	608.98056

Table B.5 Results of Duval and Tweedie's Trim and Fill for Extraversion

	Observed values	Adjusted values
Studies trimmed	0	0
Fixed Effects		
Point Estimate	0.28119	0.28119
Lower Limit	0.27567	0.27567
Upper Limit	0.28668	0.28668
Random Effects		
Point Estimate	0.35748	0.35748
Lower Limit	0.24390	0.24390
Upper Limit	0.46139	0.46139
Q Value	2130.72406	2130.72406

Table B.6 Results of Duval and Tweedie's Trim and Fill for Conscientiousness

	Observed values	**Adjusted values**
Studies trimmed	0	0
Fixed Effects		
Point Estimate	0.28730	0.28730
Lower Limit	0.28180	0.28180
Upper Limit	0.29277	0.29277
Random Effects		
Point Estimate	0.27287	0.27287
Lower Limit	0.23930	0.23930
Upper Limit	0.30579	0.30579
Q Value	85.38295	85.38295

Table B.7 Results of Duval and Tweedie's Trim and Fill for Openness

	Observed values	**Adjusted values**
Studies trimmed	0	0
Fixed Effects		
Point Estimate	0.31225	0.31225
Lower Limit	0.30684	0.30684
Upper Limit	0.31763	0.31763
Random Effects		
Point Estimate	0.29558	0.29558
Lower Limit	0.17784	0.17784
Upper Limit	0.40499	0.40499
Q Value	2124.68335	2124.68335

Table B.8 Results of Duval and Tweedie's Trim and Fill for Neuroticism

	Observed values	**Adjusted values**
Studies trimmed	0	0
Fixed Effects		
Point Estimate	0.20411	0.20411
Lower Limit	0.19838	0.19838
Upper Limit	0.20983	0.20983
Random Effects		
Point Estimate	0.30752	0.30752
Lower Limit	0.21169	0.21169
Upper Limit	0.39749	0.39749
Q Value	1336.52488	1336.52488

B.8 Homogeneity

Table B.9 Results of homogeneity test for Agreeableness

	Fixed Model	Random Model
Effect size and 95% interval		
Number of Studies	11	11
Point Estimate	0.234	0.258
Lower Limit	0.228	0.187
Upper Limit	0.239	0.327
Test of null (2-Tail)		
Z-value	78.065	6.687
P-value	0.000	0.000
Heterogeneity		
Q-value	608.981	-
df (Q)	10	-
P-value	0.000	-
I-squared	98.358	-
Tau-squared		
Tau Squared	0.010	-
Standard Error	0.013	-
Variance	0.000	-
Tau	0.102	-

Table B.10 Results of homogeneity test for Extraversion

	Fixed Model	Random Model
Effect size and 95% interval		
Number of Studies	11	11
Point Estimate	0.281	0.357
Lower Limit	0.276	0.244
Upper Limit	0.287	0.461
Test of null (2-Tail)		
Z-value	94.767	5.861
P-value	0.000	0.000
Heterogeneity		
Q-value	2130.724	-
df (Q)	10	-
P-value	0.000	-
I-squared	99.531	-
Tau-squared		
Tau Squared	0.037	-
Standard Error	0.047	-
Variance	0.002	-
Tau	0.192	-

Table B.11 Results of homogeneity test for Conscientiousness

	Fixed Model	Random Model
Effect size and 95% interval		
Number of Studies	11	11
Point Estimate	0.287	0.273
Lower Limit	0.282	0.239
Upper Limit	0.293	0.306
Test of null (2-Tail)		
Z-value	96.947	15.270
P-value	0.000	0.000
Heterogeneity		
Q-value	85.383	-
df (Q)	10	-
P-value	0.000	-
I-squared	88.288	-
Tau-squared		
Tau Squared	0.001	-
Standard Error	0.002	-
Variance	0.000	-
Tau	0.036	-

Table B.12 Results of homogeneity test for Openness

	Fixed Model	Random Model
Effect size and 95% interval		
Number of Studies	11	11
Point Estimate	0.312	0.296
Lower Limit	0.307	0.178
Upper Limit	0.318	0.405
Test of null (2-Tail)		
Z-value	105.937	4.780
P-value	0.000	0.000
Heterogeneity		
Q-value	2124.683	-
df (Q)	10	-
P-value	0.000	-
I-squared	99.529	-
Tau-squared		
Tau Squared	0.037	-
Standard Error	0.047	-
Variance	0.002	-
Tau	0.192	-

Table B.13 Results of homogeneity test for Neuroticism

	Fixed Model	Random Model
Effect size and 95% interval		
Number of Studies	11	11
Point Estimate	0.204	0.308
Lower Limit	0.198	0.212
Upper Limit	0.210	0.397
Test of null (2-Tail)		
Z-value	67.892	6.055
P-value	0.000	0.000
Heterogeneity		
Q-value	1336.524	-
df (Q)	10	-
P-value	0.000	-
I-squared	99.252	-
Tau-squared		
Tau Squared	0.023	-
Standard Error	0.029	-
Variance	0.001	-
Tau	0.152	-

B.9 Meta-Regression Results

We have divided this section into six sub-sections (into the six moderators) where we included the screenshots from the results of our meta-regressions that we performed for each personality traits, along with the scatterplot of the studies by the very same moderator (with given regression line and confidence interval in the figure).

B.9.1 Moderator 1: Social Media Platform

Set	Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value	
	Intercept	0,2644	0,0553	0,1561	0,3727	4,79	0,0000	
Soc. Media Platform	**Soc. Media Platform:**	-0,0851	0,1101	-0,3008	0,1307	-0,77	0,4397	Q=11,72, df=3, p=0,0084
	Soc. Media Platform:	0,7086	0,2229	0,2718	1,1453	3,18	0,0015	
	Soc. Media Platform: Twitter	-0,0282	0,0907	-0,2060	0,1496	-0,31	0,7560	

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
$Q = 11{,}72$, $df = 3$, $p = 0{,}0084$
Goodness of fit: Test that unexplained variance is zero
$Tau^2 = 0{,}0109$, $Tau = 0{,}1044$, $I^2 = 98{,}80\%$, $Q = 585{,}37$, $df = 7$, $p = 0{,}0000$

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
$Tau^2 = 0{,}0104$, $Tau = 0{,}1020$, $I^2 = 98{,}36\%$, $Q = 608{,}98$, $df = 10$, $p = 0{,}0000$
Proportion of total between-study variance explained by Model 1
R^2 analog = 0,00 (computed value is -0,05)

Number of studies in the analysis **11**

Fig. B.16 Meta-regression results for Agreeableness with moderator Social Media Platform

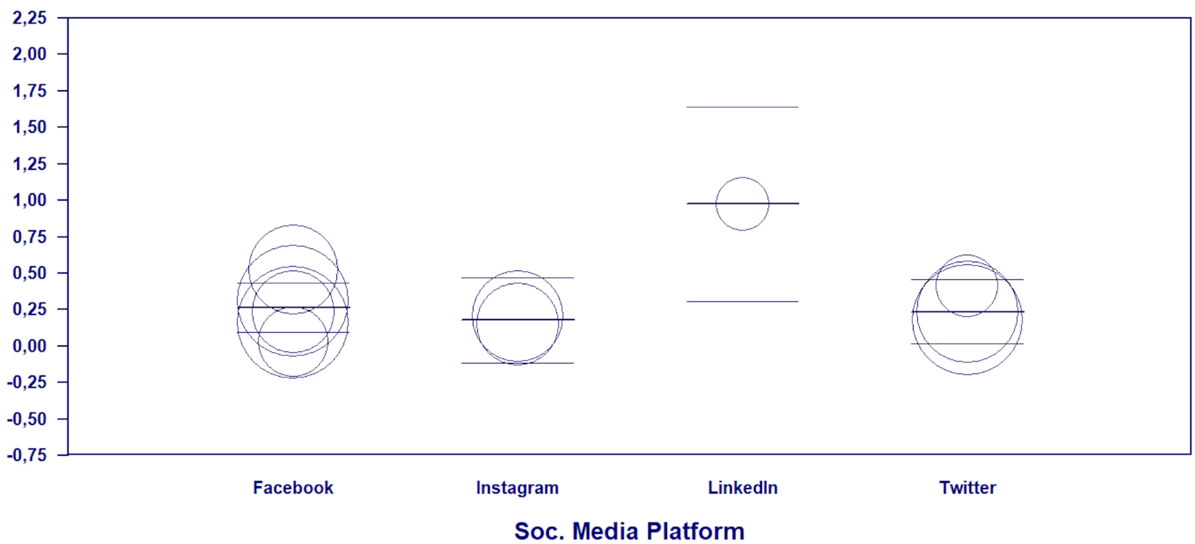

Fig. B.17 Scatterplot for Agreeableness for studies by Social Media Platform

Set	Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value	
	Intercept	0,3354	0,0946	0,1500	0,5208	3,55	0,0004	
Soc. Media Platform	Soc. Media Platform:	-0,0791	0,1794	-0,4306	0,2725	-0,44	0,6593	Q=19,04, df=3, p=0,0003
	Soc. Media Platform:	1,1921	0,2894	0,6248	1,7594	4,12	0,0000	
	Soc. Media Platform: Twitter	-0,0482	0,1558	-0,3536	0,2572	-0,31	0,7569	

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 19,04, df = 3, p = 0,0003

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0391, Tau = 0,1978, I^2 = 99,66%, Q = 2083,19, df = 7, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0369, Tau = 0,1920, I^2 = 99,53%, Q = 2130,72, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,00 (computed value is -0,06)

Number of studies in the analysis 11

Fig. B.18 Meta-regression results for Extraversion with moderator Social Media Platform

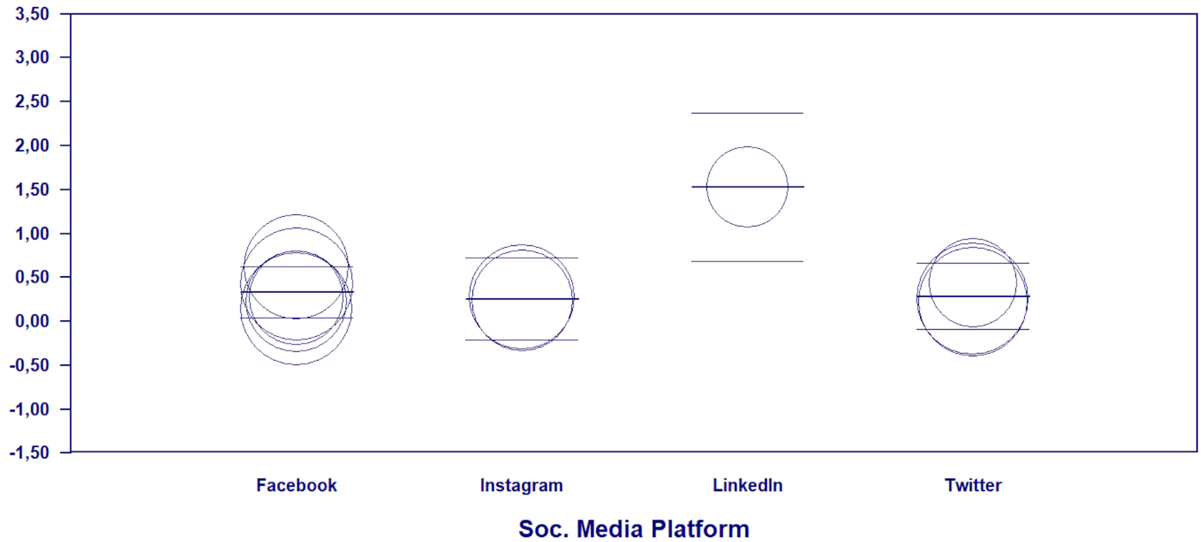

Fig. B.19 Scatterplot for Extraversion for studies by Social Media Platform

Set	Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value	
	Intercept	0,3099	0,0151	0,2803	0,3396	20,47	0,0000	
Soc. Media Platform	Soc. Media Platform:	-0,0978	0,0630	-0,2214	0,0258	-1,55	0,1208	
	Soc. Media Platform:	0,6630	0,1908	0,2890	1,0371	3,47	0,0005	Q=31,13, df=3, p=0,0000
	Soc. Media Platform: Twitter	-0,1185	0,0291	-0,1755	-0,0615	-4,08	0,0000	

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 31,13, df = 3, p = 0,0000

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0005, Tau = 0,0218, I^2 = 78,25%, Q = 32,19, df = 7, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0013, Tau = 0,0362, I^2 = 88,29%, Q = 85,38, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,64

Number of studies in the analysis **11**

Fig. B.20 Meta-regression results for Conscientiousness with moderator Social Media Platform

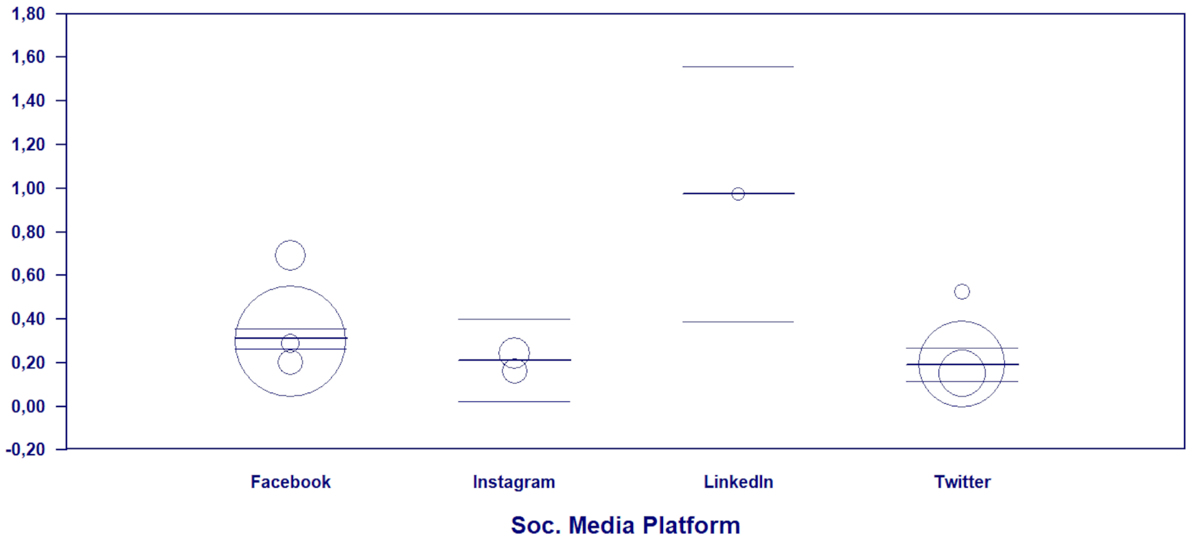

Fig. B.21 Scatterplot for Conscientiousness for studies by Social Media Platform

Set	Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value	
	Intercept	0,3675	0,0936	0,1841	0,5510	3,93	0,0001	
Soc. Media Platform	Soc. Media Platform:	-0,1507	0,1775	-0,4986	0,1973	-0,85	0,3961	Q=1,28, df=3, p=0,7341
	Soc. Media Platform:	0,0443	0,2875	-0,5192	0,6078	0,15	0,8776	
	Soc. Media Platform: Twitter	-0,1337	0,1542	-0,4358	0,1685	-0,87	0,3860	

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 1,28, df = 3, p = 0,7341

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0382, Tau = 0,1954, I^2 = 99,66%, Q = 2033,69, df = 7, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0368, Tau = 0,1917, I^2 = 99,53%, Q = 2124,68, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,00 (computed value is -0,04)

Number of studies in the analysis 11

Fig. B.22 Meta-regression results for Openness with moderator Social Media Platform

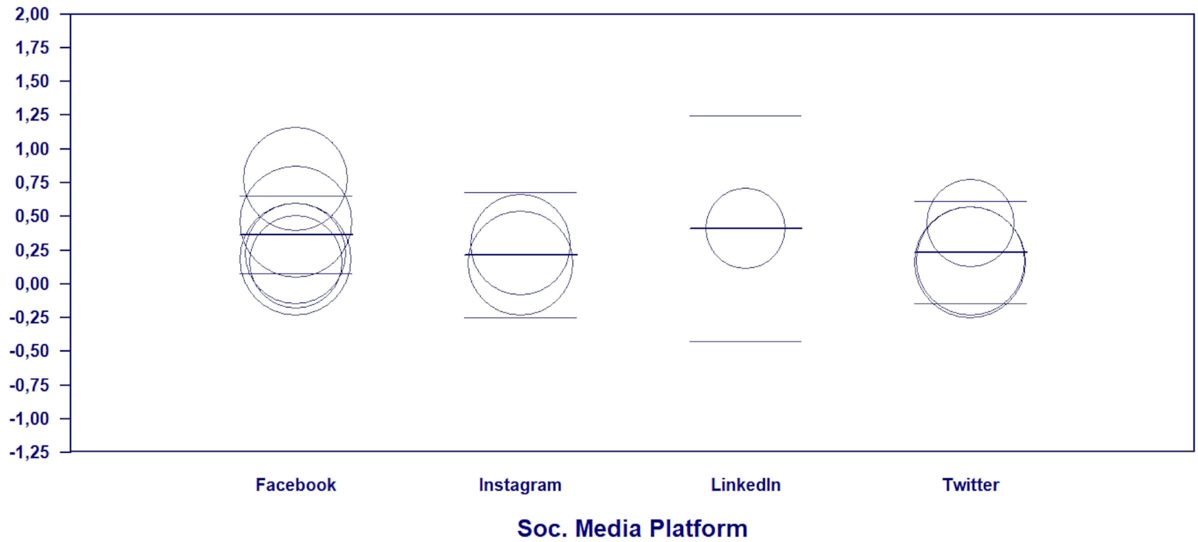

Fig. B.23 Scatterplot for Openness for studies by Social Media Platform

Set	Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value	
	Intercept	0,2473	0,0767	0,0970	0,3975	3,22	0,0013	
Soc. Media Platform	Soc. Media Platform:	-0,0726	0,1470	-0,3606	0,2155	-0,49	0,6215	
	Soc. Media Platform:	1,2250	0,2564	0,7224	1,7275	4,78	0,0000	Q=24,46, df=3, p=0,0000
	Soc. Media Platform: Twitter	0,0930	0,1263	-0,1546	0,3406	0,74	0,4615	

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 24,46, df = 3, p = 0,0000
Goodness of fit: Test that unexplained variance is zero
Tau² = 0,0241, Tau = 0,1554, I² = 99,46%, Q = 1288,63, df = 7, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
Tau² = 0,0231, Tau = 0,1518, I² = 99,25%, Q = 1336,52, df = 10, p = 0,0000
Proportion of total between-study variance explained by Model 1
R² analog = 0,00 (computed value is -0,05)

Number of studies in the analysis **11**

Fig. B.24 Meta-regression results for Neuroticism with moderator Social Media Platform

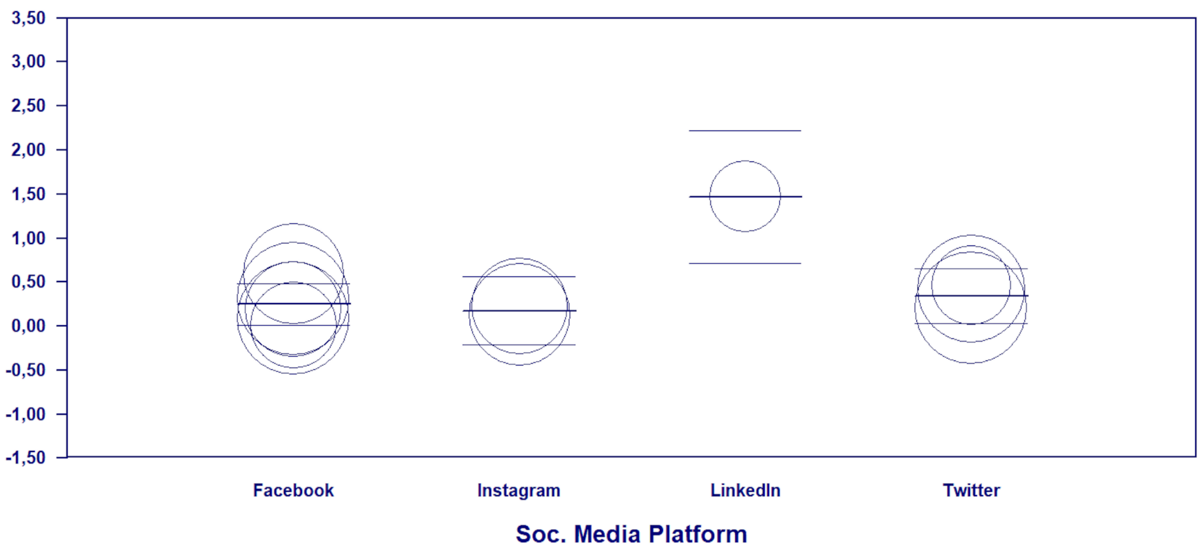

Fig. B.25 Scatterplot for Neuroticism for studies by Social Media Platform

B.9.1 Moderator 2: Single vs. Multiple

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,6404	0,1093	0,4262	0,8546	5,86	0,0000
SINGLE VS MULTIPLE: Single	-0,4284	0,1165	-0,6568	-0,1999	-3,68	0,0002

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 13,51, df = 1, p = 0,0002

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0101, Tau = 0,1003, I^2 = 98,46%, Q = 585,34, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0104, Tau = 0,1020, I^2 = 98,36%, Q = 608,98, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,03

Number of studies in the analysis **11**

Fig. B.26 Meta-regression results for Agreeableness with moderator Single vs. Multiple

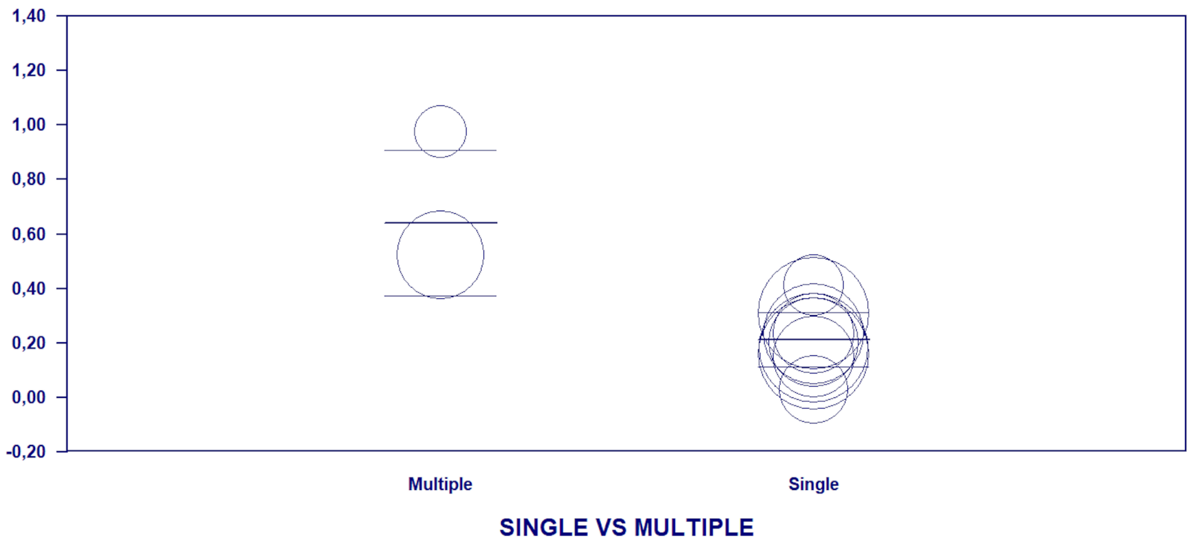

Fig. B.27 Scatterplot for Agreeableness for studies by Single vs. Multiple

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,2704	0,0688	0,1356	0,4053	3,93	0,0001
SINGLE VS MULTIPLE:	0,6849	0,1773	0,3374	1,0323	3,86	0,0001

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 14,93, df = 1, p = 0,0001

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0363, Tau = 0,1906, I^2 = 99,57%, Q = 2089,67, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0369, Tau = 0,1920, I^2 = 99,53%, Q = 2130,72, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,01

Number of studies in the analysis 11

Fig. B.28 Meta-regression results for Extraversion with moderator Single vs. Multiple

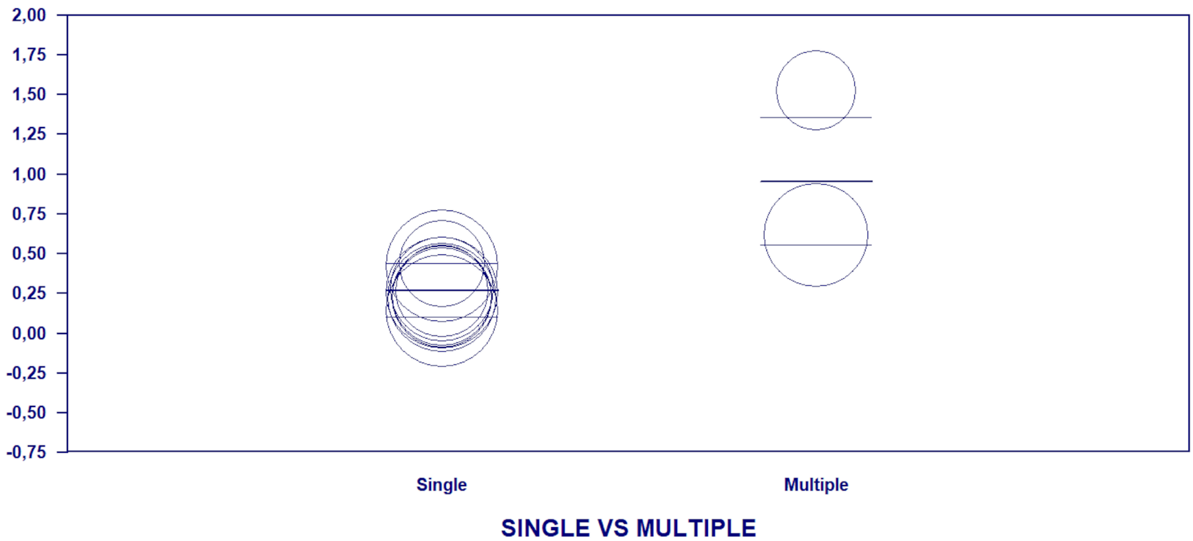

Fig. B.29 Scatterplot for Extraversion for studies by Single vs. Multiple

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,2602	0,0148	0,2312	0,2891	17,62	0,0000
SINGLE VS MULTIPLE:	0,4769	0,0771	0,3259	0,6279	6,19	0,0000

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 38,31, df = 1, p = 0,0000

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0007, Tau = 0,0262, I^2 = 81,42%, Q = 48,43, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0013, Tau = 0,0362, I^2 = 88,29%, Q = 85,38, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,47

Number of studies in the analysis 11

Fig. B.30 Meta-regression results for Conscientiousness with moderator Single vs. Multiple

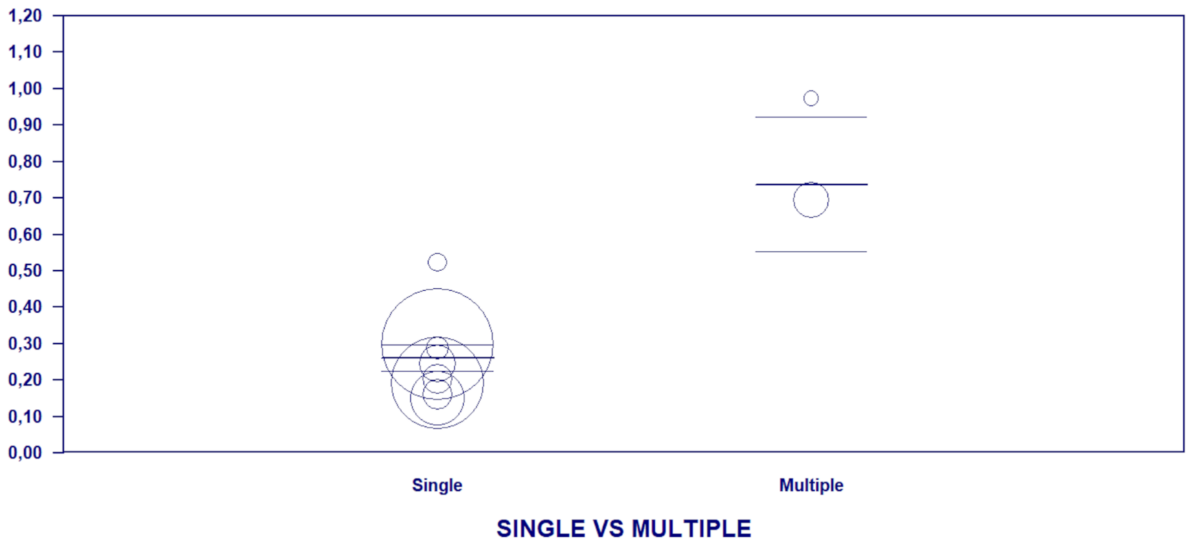

Fig. B.31 Scatterplot for Conscientiousness for studies by Single vs. Multiple

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,2450	0,0689	0,1101	0,3800	3,56	0,0004
SINGLE VS MULTIPLE:	0,3955	0,1774	0,0478	0,7432	2,23	0,0258

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 4,97, df = 1, p = 0,0258
Goodness of fit: Test that unexplained variance is zero
Tau^2 = 0,0364, Tau = 0,1908, I^2 = 99,57%, Q = 2094,02, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
Tau^2 = 0,0368, Tau = 0,1917, I^2 = 99,53%, Q = 2124,68, df = 10, p = 0,0000
Proportion of total between-study variance explained by Model 1
R^2 analog = 0,01

Number of studies in the analysis **11**

Fig. B.32 Meta-regression results for Openness with moderator Single vs. Multiple

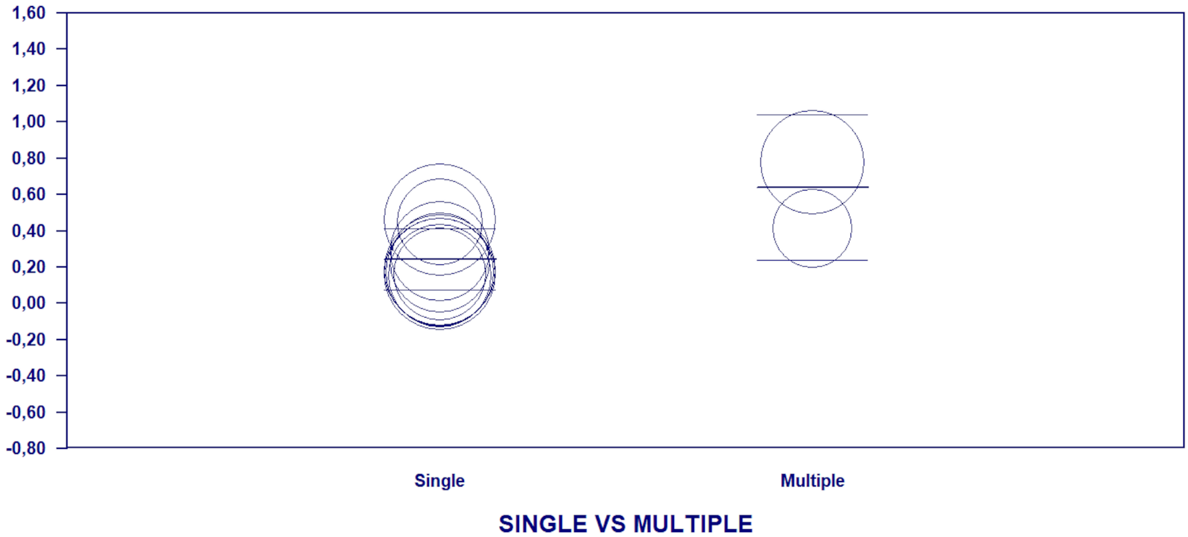

Fig. B.33 Scatterplot for Openness for studies by Single vs. Multiple

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,2248	0,0558	0,1154	0,3343	4,03	0,0001
SINGLE VS MULTIPLE:	0,6552	0,1489	0,3633	0,9470	4,40	0,0000

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 19,35, df = 1, p = 0,0000

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0223, Tau = 0,1493, I^2 = 99,30%, Q = 1286,15, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0231, Tau = 0,1518, I^2 = 99,25%, Q = 1336,52, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,03

Number of studies in the analysis 11

Fig. B.34 Meta-regression results for Neuroticism with moderator Single vs. Multiple

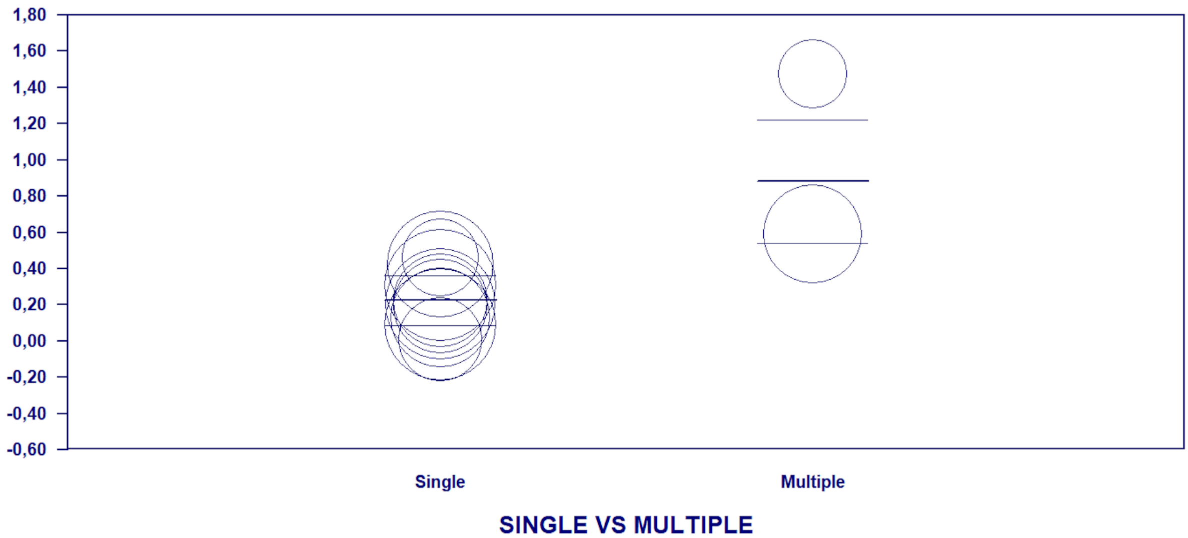

Fig. B.35 Scatterplot for Neuroticism for studies by Single vs. Multiple

B.9.1 Moderator 3: Demographic

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,5230	0,1278	0,2725	0,7735	4,09	0,0000
Demographic ?: No	-0,2844	0,1340	-0,5469	-0,0219	-2,12	0,0337

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 4,51, df = 1, p = 0,0337
Goodness of fit: Test that unexplained variance is zero
Tau^2 = 0,0102, Tau = 0,1012, I^2 = 98,49%, Q = 595,64, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
Tau^2 = 0,0104, Tau = 0,1020, I^2 = 98,36%, Q = 608,98, df = 10, p = 0,0000
Proportion of total between-study variance explained by Model 1
R^2 analog = 0,02

Number of studies in the analysis 11

Fig. B.36 Meta-regression results for Agreeableness with moderator Demographic

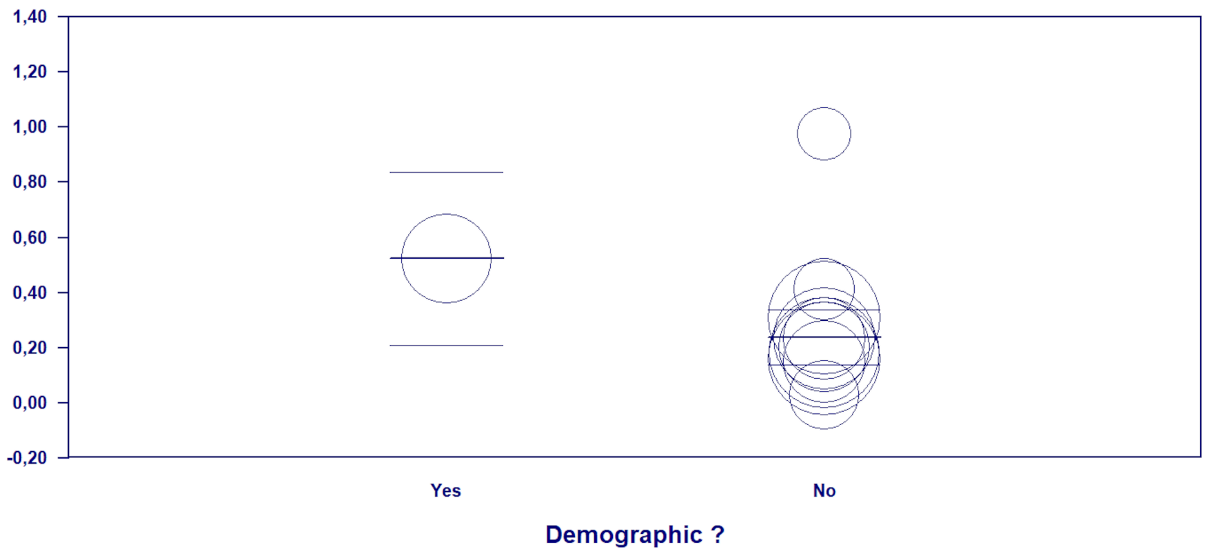

Fig. B.37 Scatterplot for Agreeableness for studies by Demographic

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,6184	0,2069	0,2128	1,0240	2,99	0,0028
Demographic ?: No	-0,2701	0,2175	-0,6964	0,1562	-1,24	0,2143

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 1,54, df = 1, p = 0,2143

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0367, Tau = 0,1916, I^2 = 99,57%, Q = 2112,90, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0369, Tau = 0,1920, I^2 = 99,53%, Q = 2130,72, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,00

Number of studies in the analysis 11

Fig. B.38 Meta-regression results for Extraversion with moderator Demographic

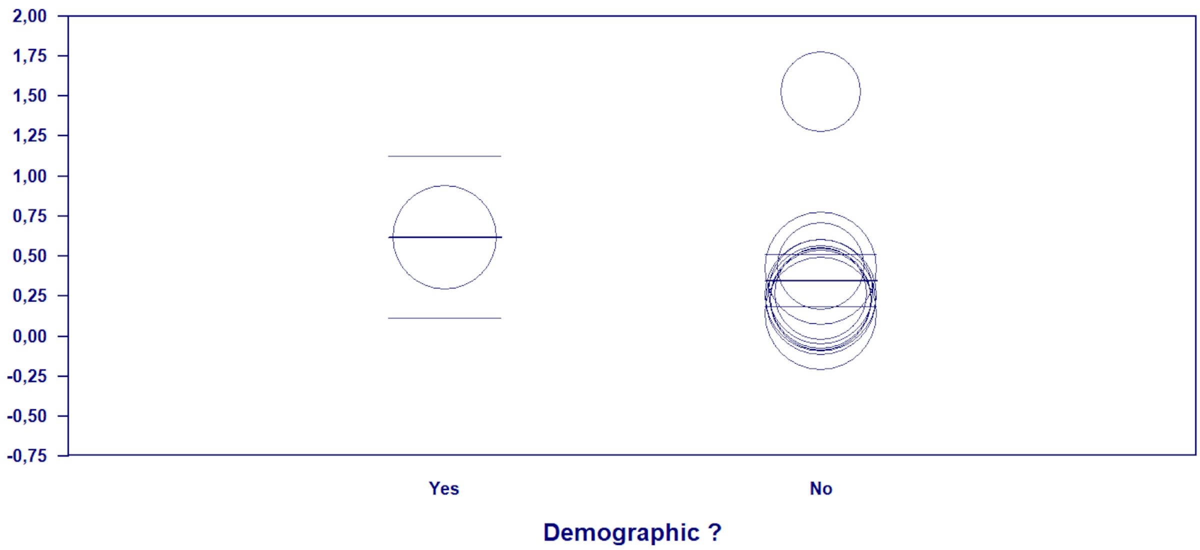

Fig. B.39 Scatterplot for Extraversion for studies by Demographic

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,6931	0,0835	0,5294	0,8569	8,30	0,0000
Demographic ?: No	-0,4306	0,0851	-0,5973	-0,2638	-5,06	0,0000

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 25,61, df = 1, p = 0,0000
Goodness of fit: Test that unexplained variance is zero
Tau^2 = 0,0009, Tau = 0,0297, I^2 = 84,86%, Q = 59,43, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
Tau^2 = 0,0013, Tau = 0,0362, I^2 = 88,29%, Q = 85,38, df = 10, p = 0,0000
Proportion of total between-study variance explained by Model 1
R^2 analog = 0,33

Number of studies in the analysis **11**

Fig. B.40 Meta-regression results for Conscientiousness with moderator Demographic

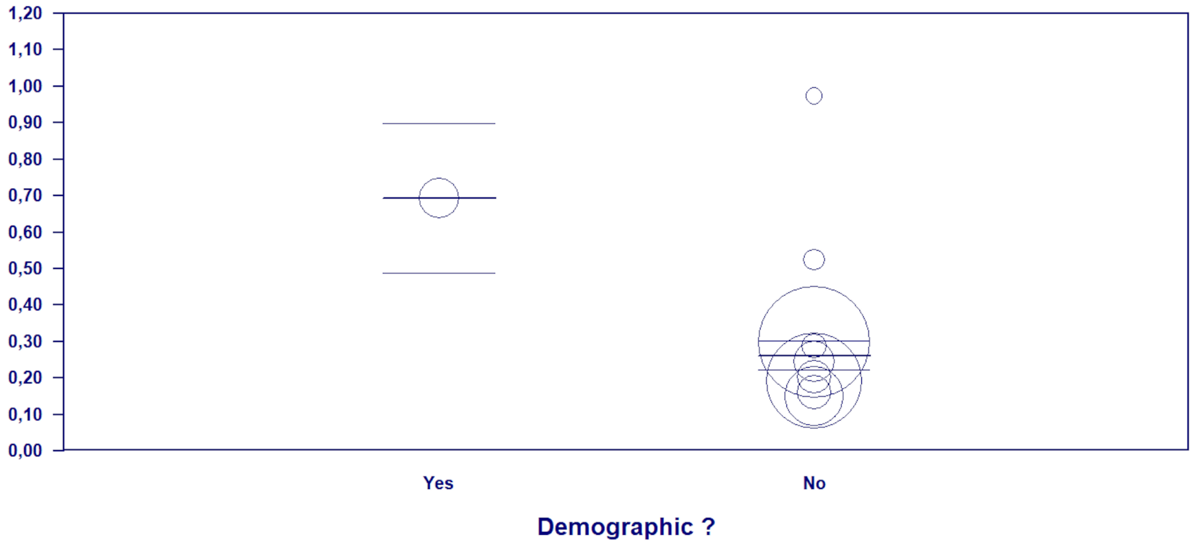

Fig. B.41 Scatterplot for Conscientiousness for studies by Demographic

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,7753	0,2060	0,3715	1,1791	3,76	0,0002
Demographic ?: No	-0,5200	0,2165	-0,9444	-0,0956	-2,40	0,0163

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 5,77, df = 1, p = 0,0163

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0363, Tau = 0,1906, I^2 = 99,57%, Q = 2091,09, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0368, Tau = 0,1917, I^2 = 99,53%, Q = 2124,68, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,01

Number of studies in the analysis 11

Fig. B.42 Meta-regression results for Openness with moderator Demographic

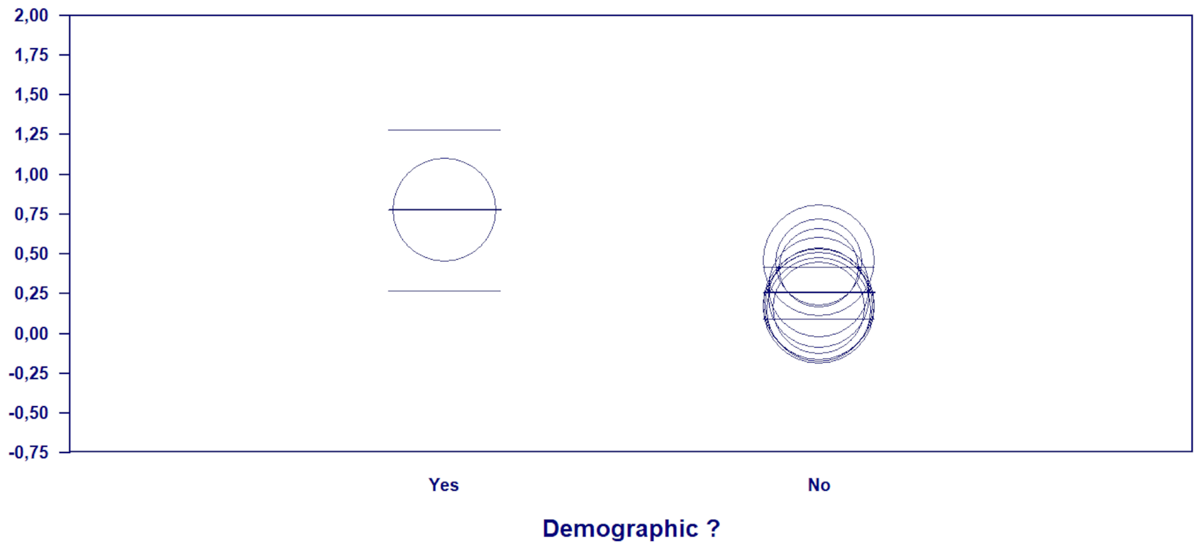

Fig. B.43 Scatterplot for Openness for studies by Demographic

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,5901	0,1698	0,2573	0,9230	3,47	0,0005
Demographic ?: No	-0,3011	0,1785	-0,6509	0,0487	-1,69	0,0916

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 2,85, df = 1, p = 0,0916

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0228, Tau = 0,1508, I^2 = 99,31%, Q = 1312,42, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0231, Tau = 0,1518, I^2 = 99,25%, Q = 1336,52, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,01

Number of studies in the analysis **11**

Fig. B.44 Meta-regression results for Neuroticism with moderator Demographic

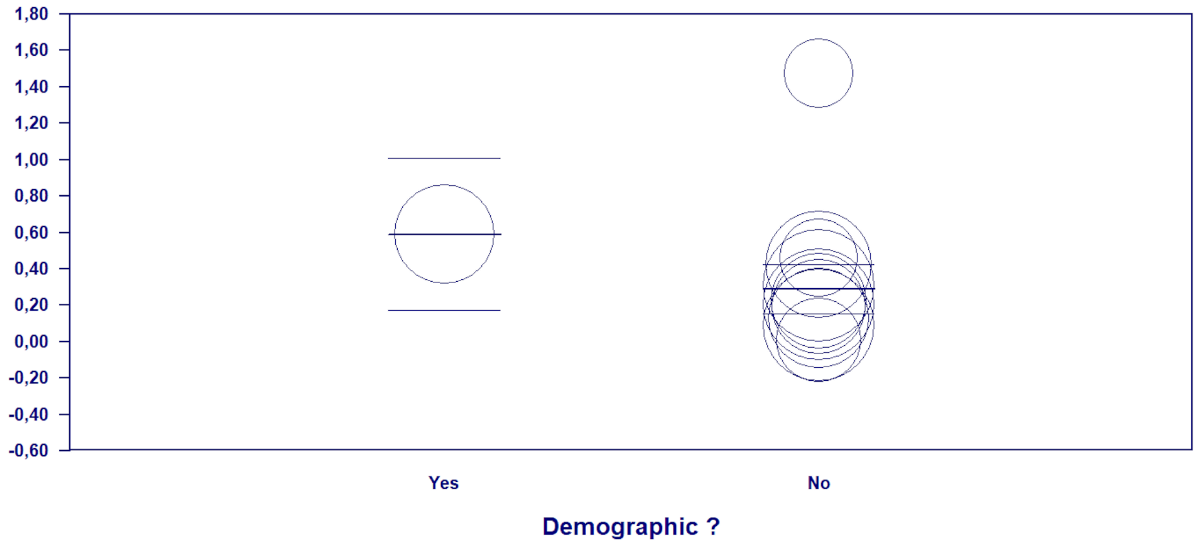

Fig. B.45 Scatterplot for Neuroticism for studies by Demographic

B.9.1 Moderator 4: Activity

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,3456	0,0380	0,2711	0,4201	9,09	0,0000
ACTIVITY ?: No	-0,1612	0,0483	-0,2559	-0,0665	-3,34	0,0008

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 11,14, df = 1, p = 0,0008

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0026, Tau = 0,0512, I^2 = 69,07%, Q = 29,09, df = 9, p = 0,0006

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0104, Tau = 0,1020, I^2 = 98,36%, Q = 608,98, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,75

Number of studies in the analysis **11**

Fig. B.46 Meta-regression results for Agreeableness with moderator Activity

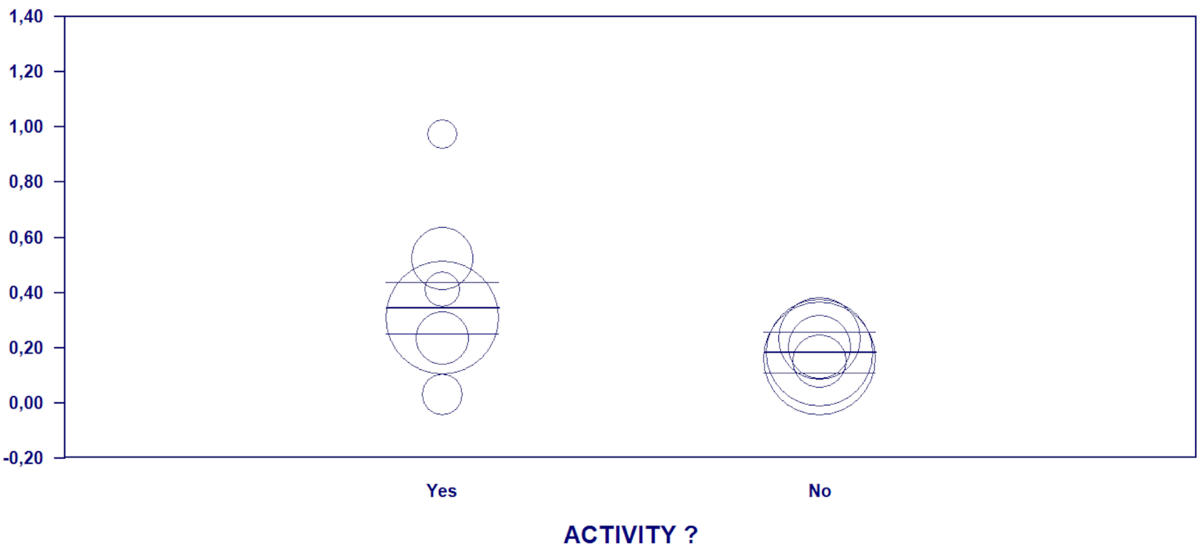

Fig. B.47 Scatterplot for Agreeableness for studies by Activity

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,4920	0,0582	0,3779	0,6061	8,45	0,0000
ACTIVITY ?: No	-0,2714	0,0775	-0,4233	-0,1194	-3,50	0,0005

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 12,26, df = 1, p = 0,0005

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0104, Tau = 0,1018, I^2 = 89,83%, Q = 88,50, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0369, Tau = 0,1920, I^2 = 99,53%, Q = 2130,72, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,72

Number of studies in the analysis **11**

Fig. B.48 Meta-regression results for Extraversion with moderator Activity

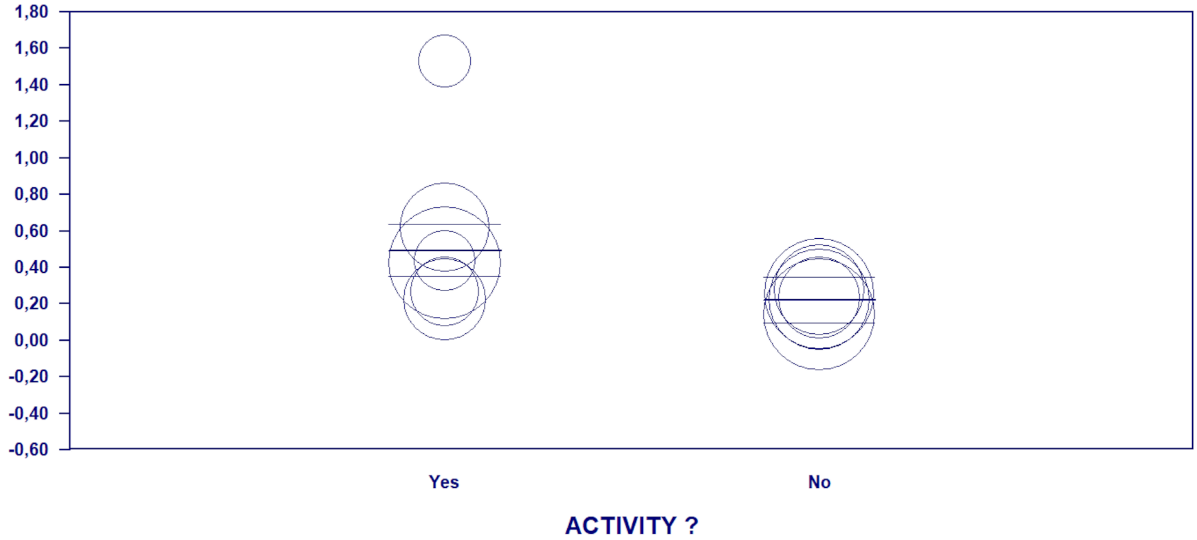

Fig. B.49 Scatterplot for Extraversion for studies by Activity

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,4315	0,0570	0,3199	0,5432	7,58	0,0000
ACTIVITY ?: No	-0,2163	0,0756	-0,3644	-0,0681	-2,86	0,0042

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 8,18, df = 1, p = 0,0042

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0097, Tau = 0,0984, I^2 = 89,18%, Q = 83,18, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0013, Tau = 0,0362, I^2 = 88,29%, Q = 85,38, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,00 (computed value is -6,39)

Number of studies in the analysis 11

Fig. B.50 Meta-regression results for Conscientiousness with moderator Activity

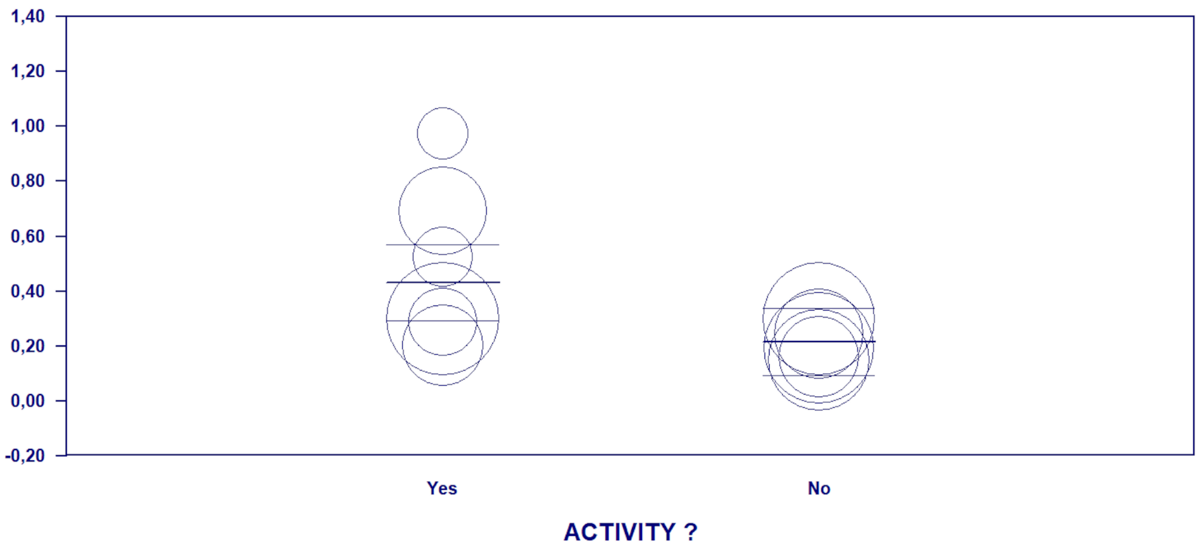

Fig. B.51 Scatterplot for Conscientiousness for studies by Activity

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,4585	0,0387	0,3826	0,5344	11,84	0,0000
ACTIVITY ?: No	-0,2797	0,0493	-0,3762	-0,1831	-5,68	0,0000

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 32,21, df = 1, p = 0,0000

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0028, Tau = 0,0527, I^2 = 70,28%, Q = 30,29, df = 9, p = 0,0004

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0368, Tau = 0,1917, I^2 = 99,53%, Q = 2124,68, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,92

Number of studies in the analysis 11

Fig. B.52 Meta-regression results for Openness with moderator Activity

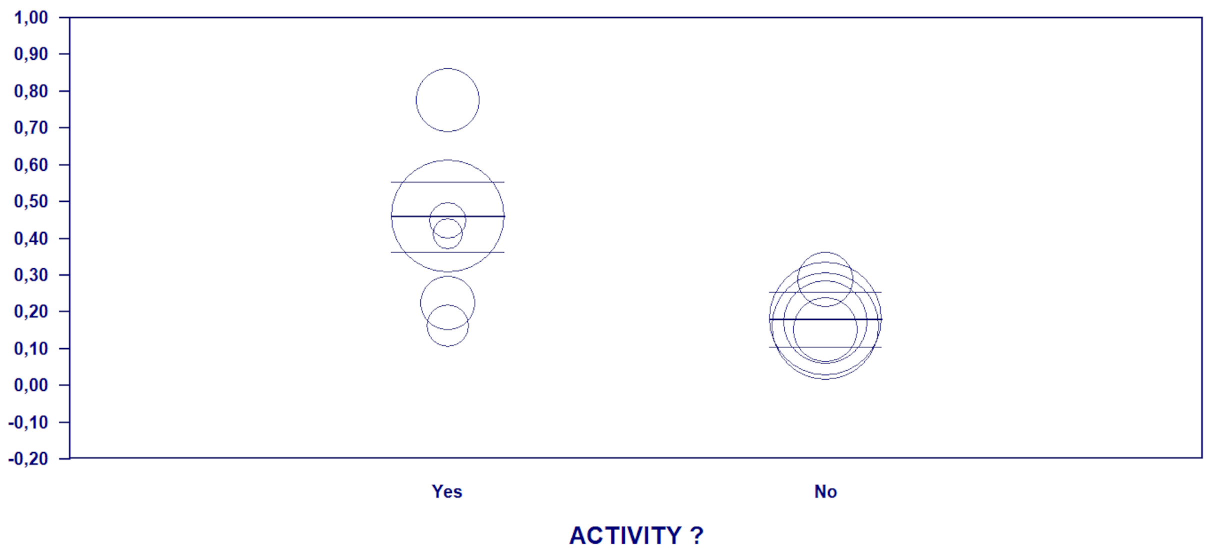

Fig. B.53 Scatterplot for Openness for studies by Activity

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,4254	0,0689	0,2904	0,5604	6,17	0,0000
ACTIVITY ?: No	-0,2128	0,0937	-0,3965	-0,0291	-2,27	0,0232

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 5,15, df = 1, p = 0,0232

Goodness of fit: Test that unexplained variance is zero

Tau² = 0,0172, Tau = 0,1313, I^2 = 93,62%, Q = 141,15, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau² = 0,0231, Tau = 0,1518, I^2 = 99,25%, Q = 1336,52, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,25

Number of studies in the analysis **11**

Fig. B.54 Meta-regression results for Neuroticism with moderator Activity

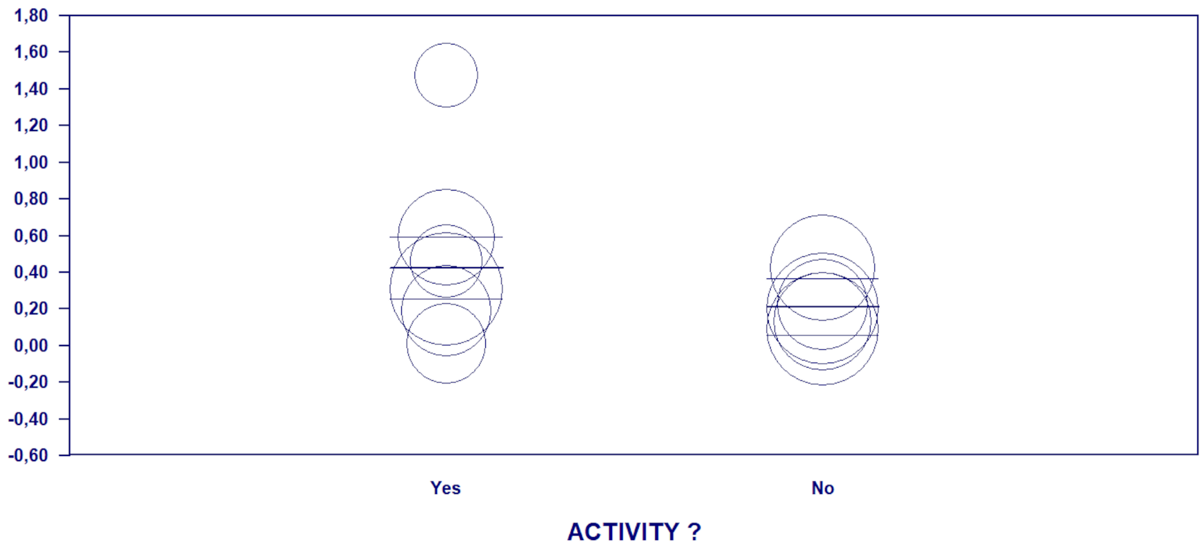

Fig. B.55 Scatterplot for Neuroticism for studies by Activity

B.9.1 Moderator 5: Language

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,2814	0,0474	0,1885	0,3743	5,94	0,0000
Language ?: No	-0,0446	0,0627	-0,1674	0,0783	-0,71	0,4771

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 0,51, df = 1, p = 0,4771

Goodness of fit: Test that unexplained variance is zero

Tau² = 0,0058, Tau = 0,0759, I² = 83,17%, Q = 53,47, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau² = 0,0104, Tau = 0,1020, I² = 98,36%, Q = 608,98, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R² analog = 0,45

Number of studies in the analysis **11**

Fig. B.56 Meta-regression results for Agreeableness with moderator Language

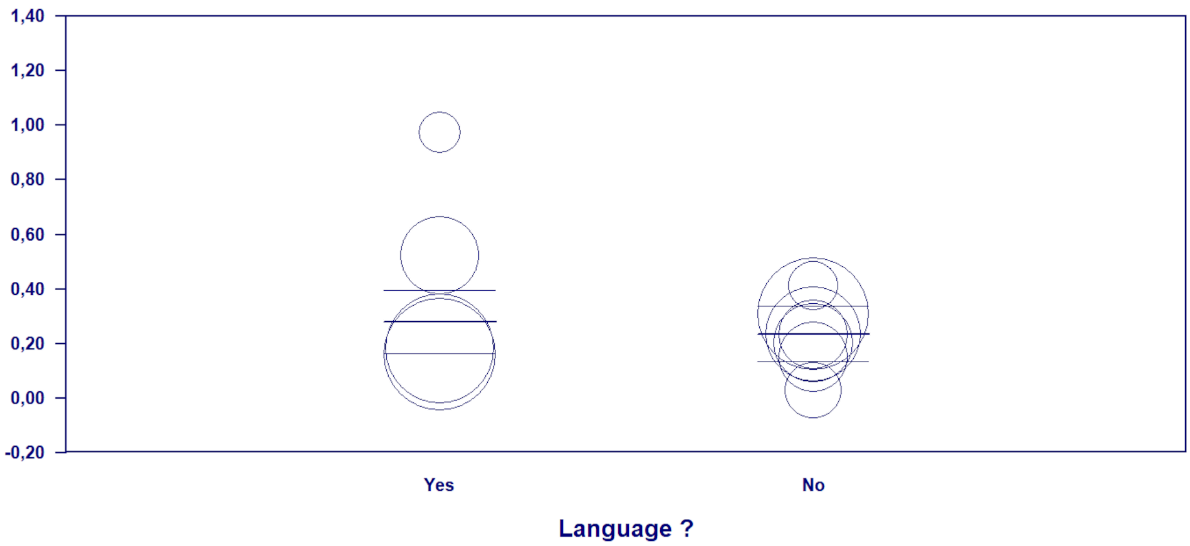

Fig. B.57 Scatterplot for Agreeableness for studies by Language

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,4508	0,0786	0,2968	0,6048	5,74	0,0000
Language ?: No	-0,1542	0,1000	-0,3502	0,0418	-1,54	0,1231

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 2,38, df = 1, p = 0,1231
Goodness of fit: Test that unexplained variance is zero
Tau^2 = 0,0191, Tau = 0,1380, I^2 = 94,23%, Q = 156,00, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
Tau^2 = 0,0369, Tau = 0,1920, I^2 = 99,53%, Q = 2130,72, df = 10, p = 0,0000
Proportion of total between-study variance explained by Model 1
R^2 analog = 0,48

Number of studies in the analysis 11

Fig. B.58 Meta-regression results for Extraversion with moderator Language

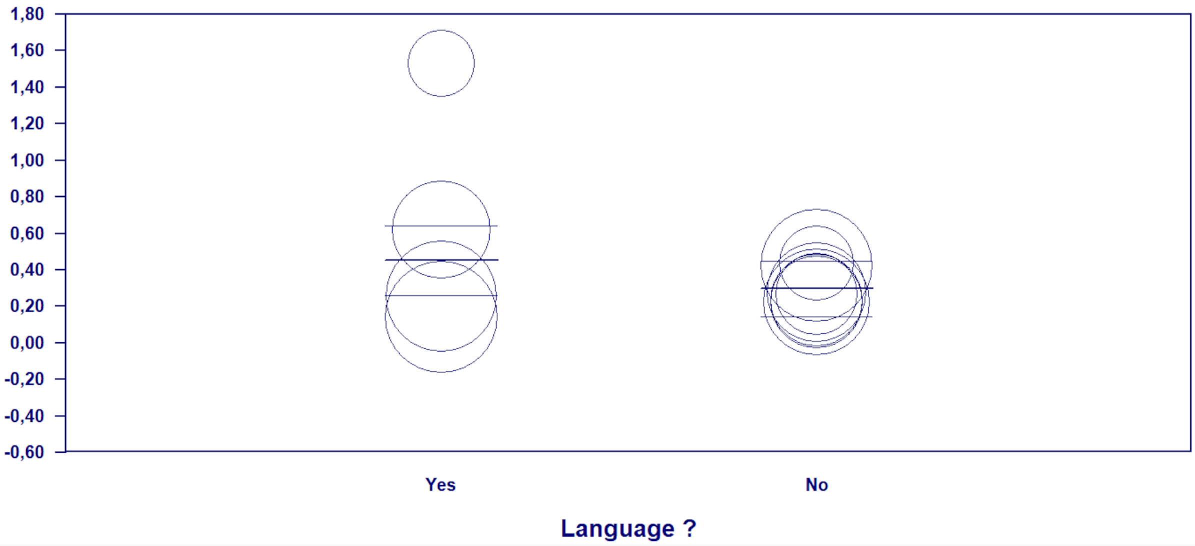

Fig. B.59 Scatterplot for Extraversion for studies by Language

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,4012	0,0594	0,2848	0,5176	6,76	0,0000
Language ?: No	-0,1543	0,0769	-0,3051	-0,0036	-2,01	0,0448

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 4,03, df = 1, p = 0,0448
Goodness of fit: Test that unexplained variance is zero
$Tau^2 = 0,0099$, Tau = 0,0994, $I^2 = 89,43\%$, Q = 85,19, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
$Tau^2 = 0,0013$, Tau = 0,0362, $I^2 = 88,29\%$, Q = 85,38, df = 10, p = 0,0000
Proportion of total between-study variance explained by Model 1
R^2 analog = 0,00 (computed value is -6,54)

Number of studies in the analysis 11

Fig. B.60 Meta-regression results for Conscientiousness with moderator Language

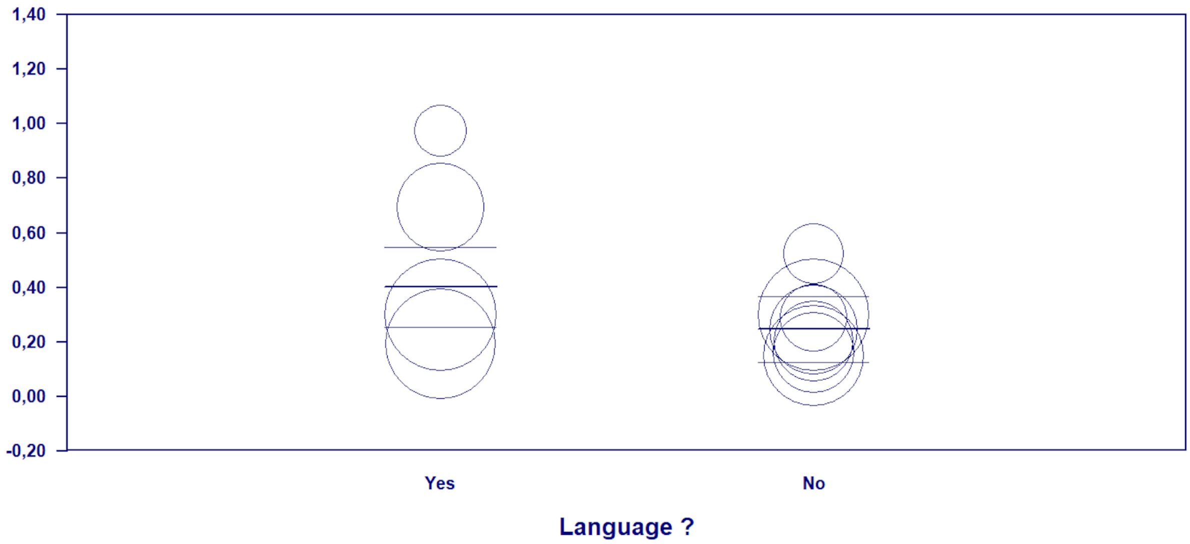

Fig. B.61 Scatterplot for Conscientiousness for studies by Language

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,3400	0,0714	0,2000	0,4799	4,76	0,0000
Language ?: No	-0,0664	0,0913	-0,2454	0,1126	-0,73	0,4672

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 0,53, df = 1, p = 0,4672

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0152, Tau = 0,1235, I^2 = 92,89%, Q = 126,66, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0368, Tau = 0,1917, I^2 = 99,53%, Q = 2124,68, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,59

Number of studies in the analysis 11

Fig. B.62 Meta-regression results for Openness with moderator Language

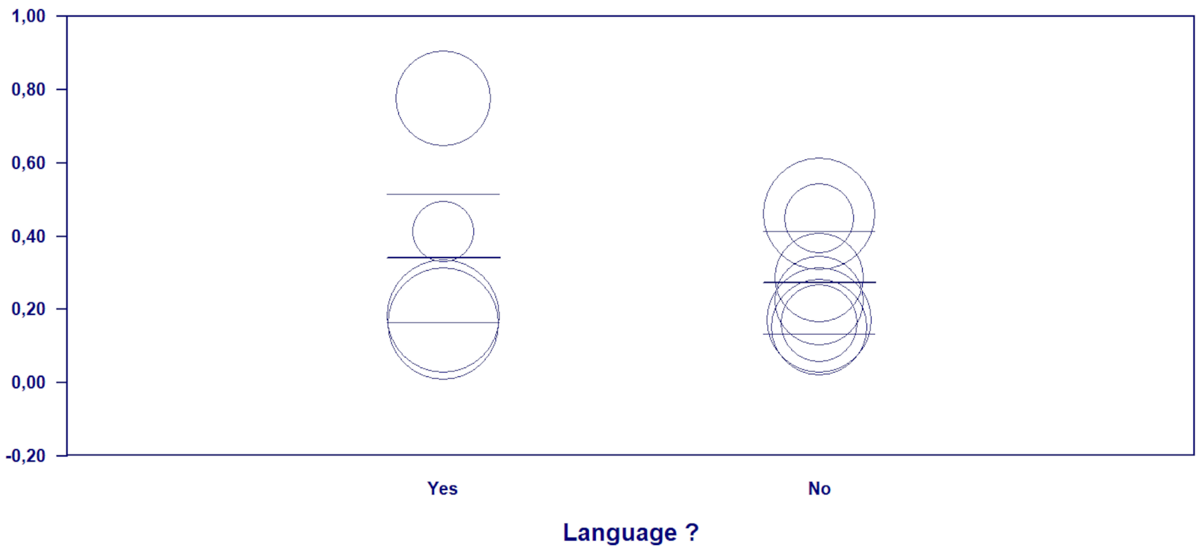

Fig. B.63 Scatterplot for Openness for studies by Language

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,3991	0,0765	0,2491	0,5491	5,21	0,0000
Language ?: No	-0,1424	0,0975	-0,3336	0,0487	-1,46	0,1442

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 2,13, df = 1, p = 0,1442

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0179, Tau = 0,1339, I^2 = 93,89%, Q = 147,37, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0231, Tau = 0,1518, I^2 = 99,25%, Q = 1336,52, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,22

Number of studies in the analysis 11

Fig. B.64 Meta-regression results for Neuroticism with moderator Language

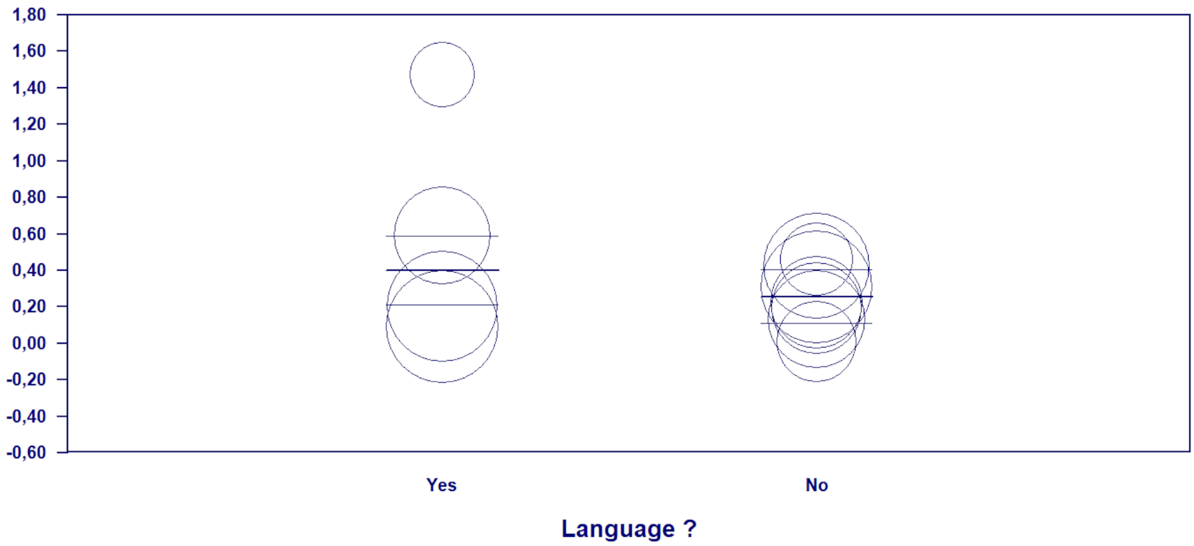

Fig. B.65 Scatterplot for Neuroticism for studies by Language

B.9.1 Moderator 6: Pictures

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,2018	0,0725	0,0597	0,3438	2,78	0,0054
Pictures ?: No	0,0876	0,0857	-0,0803	0,2556	1,02	0,3064

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 1,05, df = 1, p = 0,3064
Goodness of fit: Test that unexplained variance is zero
Tau^2 = 0,0105, Tau = 0,1027, I^2 = 98,52%, Q = 608,55, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
Tau^2 = 0,0104, Tau = 0,1020, I^2 = 98,36%, Q = 608,98, df = 10, p = 0,0000
Proportion of total between-study variance explained by Model 1
R^2 analog = 0,00 (computed value is -0,01)

Number of studies in the analysis 11

Fig. B.66 Meta-regression results for Agreeableness with moderator Pictures

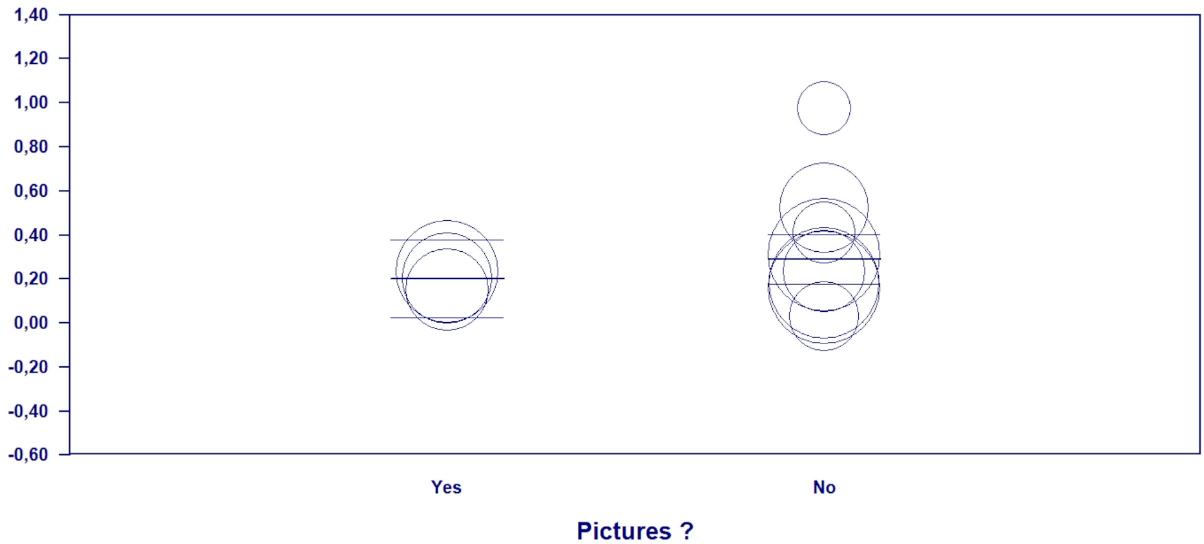

Fig. B.67 Scatterplot for Agreeableness for studies by Pictures

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,2446	0,1194	0,0105	0,4786	2,05	0,0406
Pictures ?: No	0,1823	0,1416	-0,0952	0,4597	1,29	0,1979

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 1,66, df = 1, p = 0,1979

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0373, Tau = 0,1931, I^2 = 99,58%, Q = 2128,87, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0369, Tau = 0,1920, I^2 = 99,53%, Q = 2130,72, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,00 (computed value is -0,01)

Number of studies in the analysis 11

Fig. B.68 Meta-regression results for Extraversion with moderator Pictures

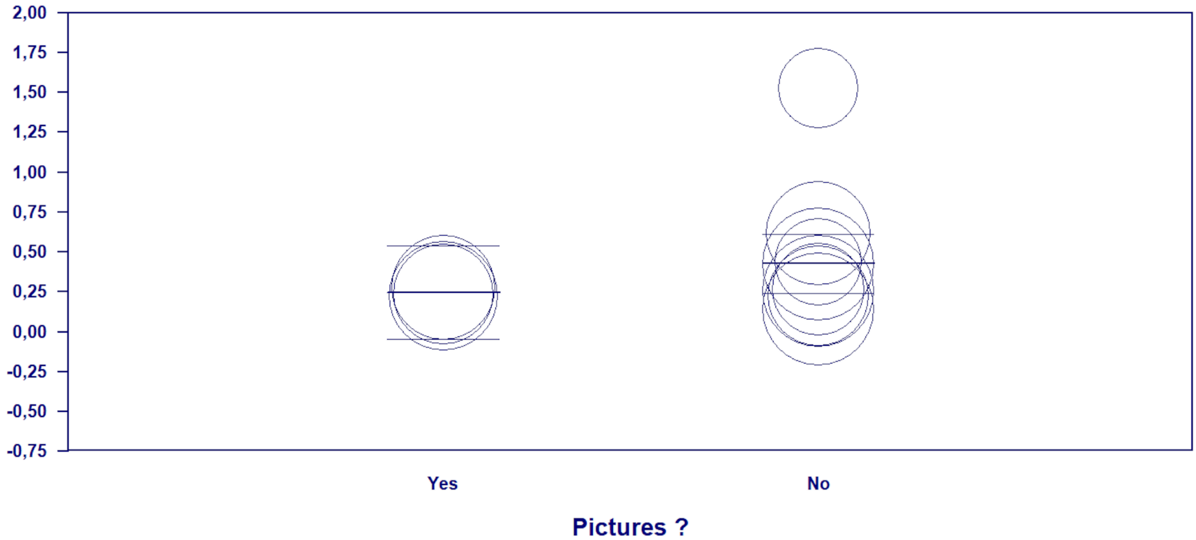

Fig. B.69 Scatterplot for Extraversion for studies by Pictures

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,1787	0,0434	0,0938	0,2637	4,12	0,0000
Pictures ?: No	0,1202	0,0474	0,0272	0,2131	2,53	0,0113

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 6,42, df = 1, p = 0,0113

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0012, Tau = 0,0341, I^2 = 87,99%, Q = 74,96, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0013, Tau = 0,0362, I^2 = 88,29%, Q = 85,38, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,11

Number of studies in the analysis 11

Fig. B.70 Meta-regression results for Conscientiousness with moderator Pictures

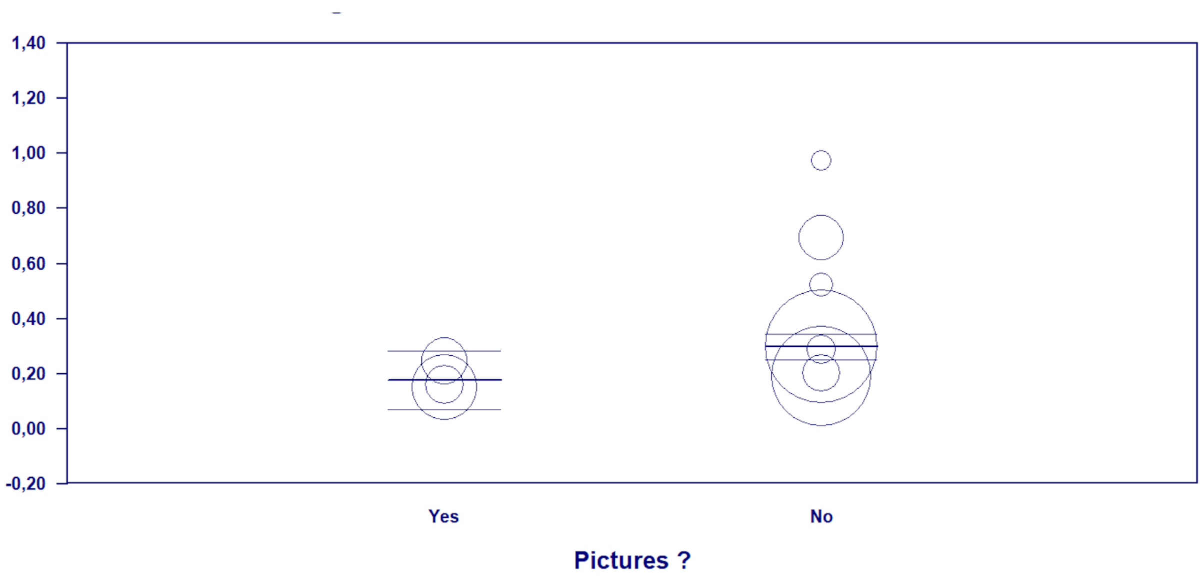

Fig. B.71 Scatterplot for Conscientiousness for studies by Pictures

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,2005	0,1190	-0,0327	0,4337	1,69	0,0919
Pictures ?: No	0,1464	0,1410	-0,1300	0,4228	1,04	0,2993

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 1,08, df = 1, p = 0,2993

Goodness of fit: Test that unexplained variance is zero

Tau^2 = 0,0370, Tau = 0,1922, I^2 = 99,57%, Q = 2110,77, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

Tau^2 = 0,0368, Tau = 0,1917, I^2 = 99,53%, Q = 2124,68, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,00 (computed value is -0,01)

Number of studies in the analysis 11

Fig. B.72 Meta-regression results for Openness with moderator Pictures

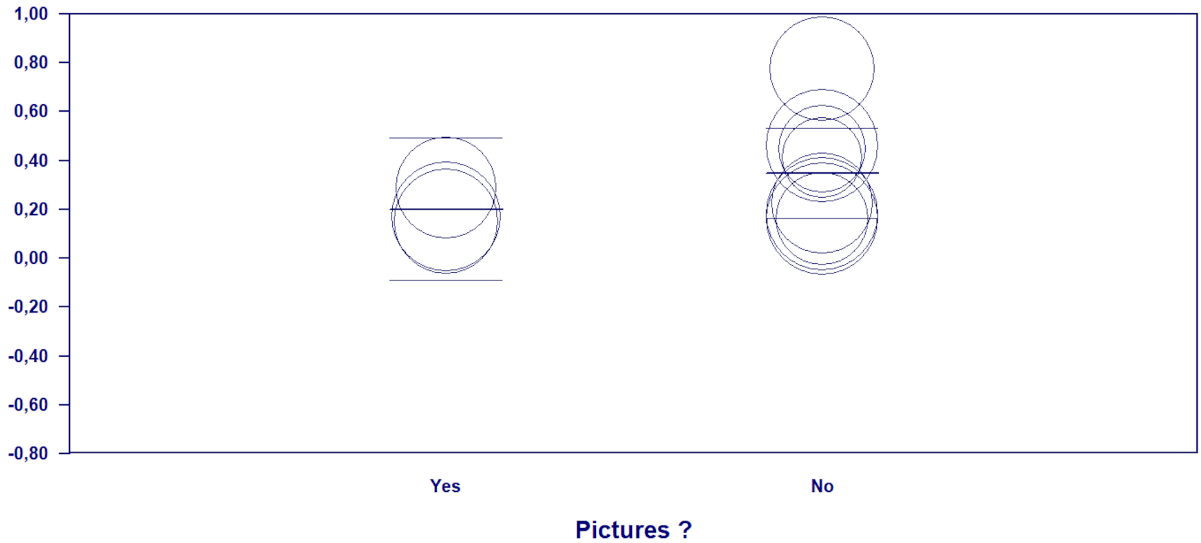

Fig. B.73 Scatterplot for Openness for studies by Pictures

Covariate	Coefficient	Standard Error	95% Lower	95% Upper	Z-value	2-sided P-value
Intercept	0,2678	0,0976	0,0764	0,4592	2,74	0,0061
Pictures ?: No	0,0706	0,1159	-0,1565	0,2978	0,61	0,5421

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

Q = 0,37, df = 1, p = 0,5421

Goodness of fit: Test that unexplained variance is zero

$Tau^2 = 0,0232$, Tau = 0,1522, $I^2 = 99,32\%$, Q = 1327,06, df = 9, p = 0,0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

$Tau^2 = 0,0231$, Tau = 0,1518, $I^2 = 99,25\%$, Q = 1336,52, df = 10, p = 0,0000

Proportion of total between-study variance explained by Model 1

R^2 analog = 0,00 (computed value is -0,01)

Number of studies in the analysis **11**

Fig. B.74 Meta-regression results for Neuroticism with moderator Pictures

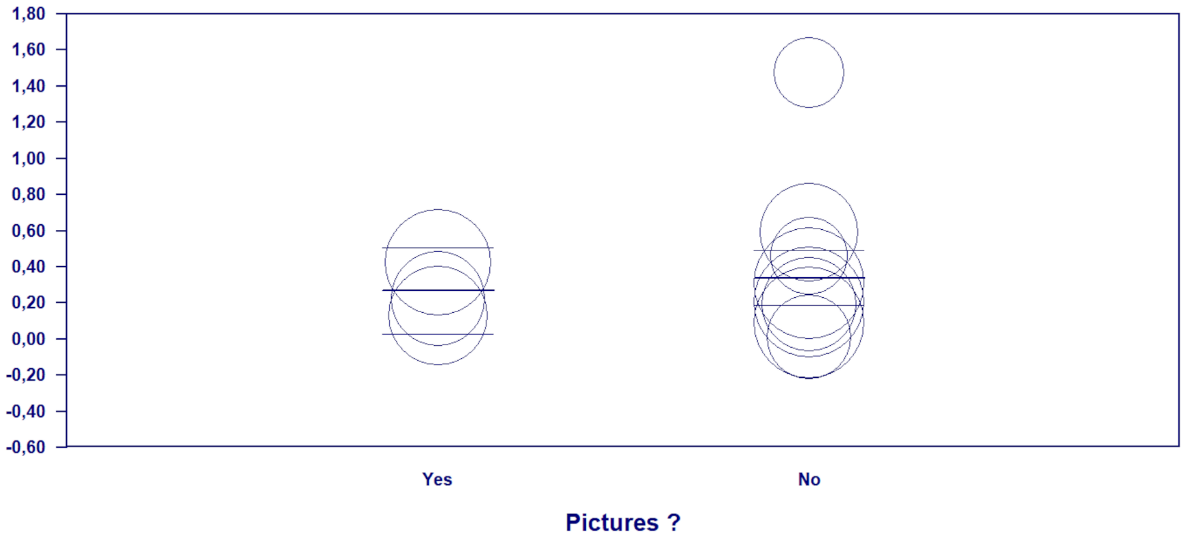

Fig. B.75 Scatterplot for Neuroticism for studies by Pictures

APPENDIX C

This appendix provides two tables that are used to obtain values in the meta-analysis.

	p-value (Two-tailed)								
df	.80	.50	.20	.10	.05	.02	.01	.002	.001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	318.309	636.621
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Table C.1 Two-tailed *t*-values from the *t*-distribution by *df* and *p*-value (Lipsey and Wilson 2001)

v	α = 0.100	0.050	0.025	0.010	0.005	0.001
1	2.7055	3.8415	5.0239	6.6349	7.8794	10.8276
2	4.6052	5.9915	7.3778	9.2103	10.5966	13.8155
3	6.2514	7.8147	9.3484	11.3449	12.8382	16.2662
4	7.7794	9.4877	11.1433	13.2767	14.8603	18.4668
5	9.2364	11.0705	12.8325	15.0863	16.7496	20.5150
6	10.6446	12.5916	14.4494	16.8119	18.5476	22.4577
7	12.0170	14.0671	16.0128	18.4753	20.2777	24.3219
8	13.3616	15.5073	17.5345	20.0902	21.9550	26.1245
9	14.6837	16.9190	19.0228	21.6660	23.5894	27.8772
10	15.9872	18.3070	20.4832	23.2093	25.1882	29.5883
11	17.2750	19.6751	21.9200	24.7250	26.7568	31.2641
12	18.5493	21.0261	23.3367	26.2170	28.2995	32.9095
13	19.8119	22.3620	24.7356	27.6882	29.8195	34.5282
14	21.0641	23.6848	26.1189	29.1412	31.3193	36.1233
15	22.3071	24.9958	27.4884	30.5779	32.8013	37.6973
16	23.5418	26.2962	28.8454	31.9999	34.2672	39.2524
17	24.7690	27.5871	30.1910	33.4087	35.7185	40.7902
18	25.9894	28.8693	31.5264	34.8053	37.1565	42.3124
19	27.2036	30.1435	32.8523	36.1909	38.5823	43.8202
20	28.4120	31.4104	34.1696	37.5662	39.9968	45.3147
21	29.6151	32.6706	35.4789	38.9322	41.4011	46.7970
22	30.8133	33.9244	36.7807	40.2894	42.7957	48.2679
23	32.0069	35.1725	38.0756	41.6384	44.1813	49.7282
24	33.1962	36.4150	39.3641	42.9798	45.5585	51.1786
25	34.3816	37.6525	40.6465	44.3141	46.9279	52.6197
26	35.5632	38.8851	41.9232	45.6417	48.2899	54.0520
27	36.7412	40.1133	43.1945	46.9629	49.6449	55.4760
28	37.9159	41.3371	44.4608	48.2782	50.9934	56.8923
29	39.0875	42.5570	45.7223	49.5879	52.3356	58.3012
30	40.2560	43.7730	46.9792	50.8922	53.6720	59.7031
31	41.4217	44.9853	48.2319	52.1914	55.0027	61.0983
63	77.7454	82.5287	86.8296	92.0100	95.6493	103.4424
127	147.8048	154.3015	160.0858	166.9874	171.7961	181.9930
255	284.3359	293.2478	301.1250	310.4574	316.9194	330.5197
511	552.3739	564.6961	575.5298	588.2978	597.0978	615.5149
1023	1081.3794	1098.5208	1113.5334	1131.1587	1143.2653	1168.4972

Table C.2 Chi-square critical values (Lipsey and Wilson 2001)