Gene-Environment Interplay for Well-Being

van de Weijer, Margot Petra

2023

DOI (link to publisher)
10.5463/thesis.236

document version
Publisher's PDF, also known as Version of record

GENE-ENVIRONMENT INTERPLAY FOR WELL-BEING

Margot P. van de Weijer

Gene-Environment Interplay for Well-Being

Margot P. van de Weijer

Paranymphs:	Lianne P. de Vries Zenab Tamimy
Funding:	This work is supported by an ERC Consolidator grant (WELL-BEING 771057 PI Bartels)
ISBN:	978-94-6483-133-7
Cover design:	Linda Jasmijn van de Weijer | lindavandeweijer99@gmail.com
Lay-out:	Publiss | www.publiss.nl
Print:	Ridderprint | www.ridderprint.nl

VRIJE UNIVERSITEIT

Gene-Environment Interplay for Well-Being

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Gedrags- en Bewegingswetenschappen
op woensdag 5 juli 2023 om 13.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Margot Petra van de Weijer

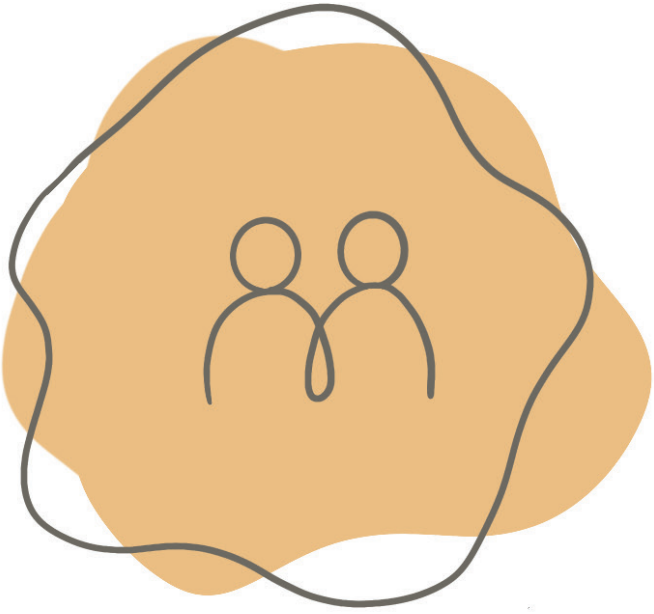
geboren te Amsterdam

Promotor:	prof.dr. M. Bartels
Copromotoren:	dr. B.M.L. Baselmans dr. H.M. Pelt
promotiecommissie:	prof.dr. E.M. Sijbrandij prof.dr. G.W.J.M. Stevens prof.dr.mr. C.H. Vinkers dr. M. Helbich dr. Y.E. Willems dr. B. Sauce Silva

Aan mijn ouders, Teus en Anna

Voor de genen, maar vooral voor de liefdevolle omgeving

TABLE OF CONTENTS



CHAPTER 1

General Introduction

Based on:

Van de Weijer, M.P., de Vries, L.P., & Bartels, M. (2022). Happiness and Wellbeing; the value and findings from genetic studies. In Tarnoki, A., Tarnoki, D. Harris, J. & Segal, S. (2022) *Twin Research for Everyone*. Academic Press.

and

Bartels, M., Nes, R. B., Armitage, J. M., **van de Weijer, M. P.**, de Vries, L. P., & Haworth, C. (2022). Exploring the biological basis for happiness. In Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2022*). World Happiness Report 2022.* New York: Sustainable Development Solutions Network.

and

Van de Weijer, M.P., Baselmans, B.M.L., van der Deijl, W., & Bartels, M. (2018). A growing sense of well-being: a literature review on the complex framework well-being. DOI: 10.31234/osf.io/3rmx9.

1. INTRODUCTION

There is a growing global recognition of well-being as an important public policy goal, with more and more consideration of well-being to inform governmental decision-making[1–3]. Similarly, well-being is becoming an increasingly important research topic across different disciplines, including the field of behavior genetics. Within this field, and in my dissertation, we try to answer questions about the contribution of genetic and environmental factors to individual differences in well-being. Additionally, we are interested in finding out more about the dynamic interplay between these genetic and environmental factors.

Importantly, the term 'well-being' embodies a multitude of concepts with varying meanings depending on context and discipline. Here, we focus on the meaning of well-being as employed in psychology and social sciences. It is important, though, to first remark its philosophical origin. I start this introduction by examining how philosophical ideas on happiness developed into psychological constructs over time, and how they have shaped modern day well-being definitions (**Section 2**). Next, I discuss how previous behavior genetic research has helped us advance our understanding of well-being. I discuss how, by partitioning the variance of well-being into genetic and environmental sources of variation, twin studies enable us to interpret causes of individual differences in well-being (**Section 3**). Furthermore, I summarize how knowledge gained from twin studies on well-being have fueled follow-up in-depth analyses in both genetic and environmental directions (**Sections 4 and 5**). In the final section of the introduction, I discuss how the work presenting in this dissertation builds upon the literature to progress in the field (**Section 6)**.

2. WELL-BEING AND ITS PHILOSOPHICAL ROOTS.

For centuries, people have asked themselves questions about the nature of well-being and happiness. This can be traced back to ancient Greek philosophers, such as Aristotle and Socrates, who already wondered about the prerequisites for living a satisfactory life[4]. Traditionally, well-being was divided in hedonic and eudaimonic well-being.

2.1 Hedonism.

"Pleasure is our first and kindred good. It is the starting point of every choice and of every aversion, and to it we come back, inasmuch as we make feeling the rule by which to judge of every good thing."

- Letter to Menoeceus, Epicurus

Ancient hedonism is centered around pleasure, or how good a person feels about his or her life[5]. From this perspective, well-being is about balancing pleasure and pain, that is: how to maximize pleasure and minimize pain. When examining ancient hedonistic thinkers, the element of pleasure is always prominent. Aristippus (c. 435 – c. 356 BCE), one of Socrates' students, was the founder of the Cyrenaic school of Philosophy, a school that taught pleasure was the ultimate goal of human life, and that the pursuit of pleasure was the purpose of human existence. They are therefore considered as one of the first to teach the hedonistic line of thought. Around the same time, Democritus (c. 460 – c. 370 BCE) also devoted his time to the hedonism. Democritus' line of thought can be characterized as a type of enlightened hedonism: the good was held to an internal state of mind[6], indicating that the well-being of an individual can be ascribed to a person's cast of mind, instead of (only) to external factors.

More recent examples of hedonists are Jeremy Bentham (1748-1832) and John Stuart Mill (1806-1873). Their philosophies can be defined as hedonistic utilitarianism, according to which we ought to maximize our sum-total of well-being[7]. According to Bentham's narrow hedonism, different pains and pleasures possess different values, and their sum determines a person's hedonic level, their level of well-being. The two most fundamental aspects in this theory are duration and intensity: these factors determine the value of an individuals' pleasures and pains[8]. That is to say, the higher the intensity, and the longer the duration, the higher the value of a pain or pleasure. However, according to Mill, this form of hedonism lacks a dimension: quality. His objection to Bentham is that "*It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied*"[9]. This implies that well-being is not a mere summation of quantities of pleasure, but that qualitatively better pleasures contribute more to well-being.

2.2 From hedonism to subjective well-being

Comparing these 19th century philosophers with their ancient counterparts, we see a more careful and detailed analysis of the concept of hedonic happiness. In modern-day behavioral and social sciences, the term hedonic well-being is less frequently mentioned. However, this does not mean that the hedonistic line of thought is unpopular among contemporary social scientists. Rather, we observe a shift in terminology: contemporary scientists prefer to use the terms subjective well-being (SWB) or happiness rather than pleasure and hedonism. A likely reason for this shift is that hedonism is a philosophical concept that has no clear method of measurement. Therefore, researchers have tried to redefine hedonism into an operational definition.

While many methods have been proposed for measuring and conceptualizing SWB[10], a widely adopted definition is that of Diener[11]. According to this conceptualization, SWB consists of three hallmarks: 1) it is subjective (objective influences are not necessarily part of the construct), 2) it includes positive measures (it is not just the absence of negative factors), and 3) it includes a global assessment of all aspects of a person's life, not just of one or a few domains. Three separate components are used to measure this construct: positive affect, negative affect, and life satisfaction[12]. While the ancient concept of hedonism is not exactly the same as modern SWB, it is very likely that SWB exists as a result of the hedonic line of thought. Conceptually, positive and negative affect, also referred to as the affective/emotional aspect of SWB, are similar to the ancient ideas of pains and pleasures contributing to hedonic levels. Life satisfaction (also referred to as the cognitive component of SWB), defined as a global judgment of one's life, could intuitively be comparable to the overall hedonic level of an individual over their life as a whole, but life satisfaction could also be considered a newer addition to this type of well-being and not strictly a hedonic concept. A person's hedonic state is the overall balance of pleasure and pain experienced at a particular point in time[7]. In contrast, life satisfaction is an evaluation a person makes about their life by their own standards. These two concepts do not necessarily coincide, as a person may be satisfied with states that might not feel good, like in the context of childbirth[13]. Moreover, in his *Conditions of Happiness*, Veenhoven[14] mentions that ancient hedonists equate the evaluation of happiness with a focus on sensory pleasures, while the modern-day concept of happiness more strongly focusses on affective and cognitive pleasures. Taken together, while hedonic levels of well-being are an important aspect of SWB, it does not capture the complete SWB construct.

2.3 Eudaimonism

"Again, our definition accords with the description of the happy man as one who 'lives well' or 'does well'; for it has virtually identified happiness with a form of good life or doing well."

- Aristotle, Nicomachean Ethics, 1098b [4]

Eudaimonia is a Greek word commonly translated as well-being or flourishing. Synonyms for eudaimonia are living well or doing well. Ancient eudaimonic philosophers based their ethical theories on the concept of this eudaimonism[15], and ancient eudaimonism takes well-being to be constituted by virtue and the fulfillment of human capacities. Whereas the hedonic tradition limited the concept of well-being to the balance of pleasure and pain, the eudaimonic tradition takes virtuous activity to be necessary for well-being as well. However, this is not to say that hedonic philosophers ascribed no value to virtue at all. Epicurus, a hedonistic philosopher, believed that acting just was a perquisite for living a pleasant life[15]. Perhaps more characterizing to the eudaimonic tradition of well-being is the principle of self-fulfillment. The most important contributor to, as well as founder of, the eudaimonic line of thought as discussed in this chapter, is Aristotle (c. 384 – c. 322 BCE). Aristotle rejected the hedonistic definition of well-being, describing it as "*vulgar*"[16]. According to Aristotle, well-being can be interpreted as well-living: it is about the actualization of human potential. Virtue, defined as knowledge (practiced over time) about how to live well, is an important aspect of this theory[17]. Therefore, the Aristotelian concept of well-being has more to do with the fulfillment of a person's nature: it aims at reaching one's fullest potential in line with one's deeper principles[18]. Another important contributor to eudaimonic well-being is Cicero (c. 106 – c. 43 B. C.). He believed that eudaimonia requires living well in one's social environment, and that friendships are thus an important aspect of well-being. This is another notable difference between hedonism and eudaimonism, where the former has an individualistic focus, while the latter is more in line with a collectivist point of view. A more modern, famous example of a eudaimonic theory is Maslow's hierarchy of needs, as proposed by Abraham Maslow[19]. This theory describes five different stages of human growth, starting at the most basic level of physiological needs. Every time the need belonging to a particular level is fulfilled, one moves up a stage in the hierarchy. The highest level a person can reach is self-actualization, which is only reached by one in a hundred

people (according to Maslow). In his theory of human motivation[19], Maslow refers to self-actualization in the sense of the Aristotelian tradition: "*It refers to the desire for self-fulfillment, namely, to the tendency for him to become actualized in what he is potentially. This tendency might be phrased as the desire to become more and more what one is, to become everything that one is capable of becoming.*"

2.4 From eudaimonism to psychological well-being

Where the hedonic line of thought has largely been replaced by SWB in the empirical literature, the eudaimonic tradition has gradually shifted towards psychological well-being (PWB). Whilst creating valid measurement methods for SWB was starting to gain popularity amongst the social sciences around the 1970/1980s, valid measurements for PWB seemed to be lacking at that time. Especially the absence of self-actualization within PWB conceptualization was troubling, and gave rise to a new formulation for capturing this construct[20]. This new formulation for PWB, developed by Ryff and colleagues, consists of six core dimensions: Self-Acceptance, Positive Relations with Others, Autonomy, Environmental Mastery, Purpose in Life and Personal Growth. While many other measurement instruments for PWB are available nowadays, these six core dimensions are still widely used to assess PWB. Notably, there are other modern perspectives on eudaimonia, such as Self-Determination theory[21]. However, here, we only present the PWB formulation as proposed by Ryff due to its frequent application in behavioral and social sciences. PWB, as proposed by Ryff, is without doubt a result of eudaimonic thinking: it was in her intention to create a measure that captures the eudaimonic line of thought: "*Indeed, the deeper philosophical roots of the new model of well-being resided in Aristotle's formulation of the highest human good, which in his Nichomachean Ethics he termed eudaimonia*"[22]. Therefore both PWB and eudaimonia are predominantly concerned with the development and self-realization of an individual[20]. However, a difference that can be pointed out between ancient eudaimonism and PWB is that in the Aristotelian tradition, eudaimonia did not just concern subjective experience, but intersubjective experience: a way of being in the world[23]. Ryff's PWB scales, though, still have more focus on the subjective, individualistic values. This is not surprising since Western countries (in which Ryff's scales are often applied) mostly have individualistic values instead of collectivistic ones. The most intersubjective scale is the positive relations with others scale. While this scale does measure the concern someone has for others, it does not place as large emphasis on intersubjective values as the ancient eudaimonic tradition.

3. GENETIC AND ENVIRONMENTAL INFLUENCES ON WELL-BEING

The previous section demonstrated how well-being became a topic of interest within philosophy and (positive) psychology. In the early 2000s, well-being also started attracting interest in the field of behavior genetics. Until that point, little attention had been devoted to the potential role that genetics might play in well-being. One of the first questions the field thus had to answer was if genetic differences between people played a role in individual differences in well-being, and if so, to what extent.

To quantify the contributions of genetic and environmental factors on individual differences, the so-called *classical twin design* (CTD) is most often used. This design relies on the fact that monozygotic (MZ) twins share approximately 100% of their genes, while dizygotic (DZ) twins share on average 50% of their segregating genes. This allows decomposition of the variation of a trait, or covariation between a set of traits, into four potential sources of variation: 1) additive genetic (A) factors, shared 100% by MZ twins and 50% by DZ twins, 2) dominant genetic (D) factors, shared 100% by MZ twins and 25% by DZ twins, 3) common environmental (C) factors, shared completely by both types of twin pairs, and 4) unique environmental (E) factors (and measurement error), completely unshared by both types of twin pairs. These additive genetic factors are also called the *narrow-sense heritability*, or the proportion of variation in well-being that can be accounted for by variation in additive genetic effects. The *broad-sense heritability* includes genetic variation due to both additive and dominant genetic factors.

In 2015, two comprehensive reviews on the causes of individual differences in well-being were published[24,25]. The reviews included studies that partitioned variance in well-being based on the twin(-family) design. Results of these twin-family studies into the genetic and environmental influences on well-being revealed a range of heritability estimates, but when meta-analyses were used to estimate heritability across the studies the meta-analytic results converged on the (narrow-sense) heritability estimate. In a book chapter by Nes and Røysamb, the weighted average heritability, across 13 independent studies including more than 30,000 twins (aged 12-88) from seven different countries, was estimated at 40% (CI: 37%-42%)[25]. Similarly, in a paper by Bartels, the weighted average heritability of well-being, based on a sample size of 55,974 individuals, was 36% (34%–38%), while the weighted average heritability for satisfaction with life was 32% (29%–35%) (n = 47,750)[24]. These similar

results, with overlapping confidence intervals, provide a more robust estimate of the genetic influence on well-being. Both reviews and meta-analyses showed that both genetic and environmental influences are important for individual differences in well-being. The meta-analyses indicate that genetic influences on well-being are mainly additive and that the environmental influences appear to be non-shared.

Since 2015, the twin design has been used in an additional 15 studies that investigate the heritability of well-being using different measures of well-being. **Figures 1.1 and 1.2** summarize the heritability estimates of twin studies in the earlier meta-analyses, and of the recent twin studies on well-being. The heritability estimates of the recent studies on well-being vary somewhat (range: 0.27-0.67) but are mostly in line with the previous meta-analytic estimates. The effect of shared environment is small, but significant in a few studies in younger participants. In contrast to earlier studies, none of the recent studies reported evidence for non-additive genetic effects.

With respect to the stability of the variance decomposition of well-being across the lifespan, a study in a Dutch twin sample[26] investigated the contribution of genetic and environmental factors on well-being and depression across different ages. Genetic factors explained a substantial part of the phenotypic variance in well-being during childhood, adolescence, and adulthood (range 31–47%). In the younger samples, shared environmental influences explained a large part of the variation, but these disappear with age (as twins share less and less of their environments as they age). Regarding the association between well-being and depression, the contribution of genetic factors increased from childhood to adolescence, meaning that environmental factors are important in explaining the relationship between well-being and depressive symptoms in childhood, while in adolescence genetic factors play a larger role. In addition, Røysamb and colleagues examine how the heritability and changeability of well-being fit together by reviewing existing research[27]. Two important conclusions from this work are that 1) genetic factors contribute mostly to stability in well-being, while environmental factors contribute to both short- and long-term change, and that 2) the heritability estimates depend on environmental variation.

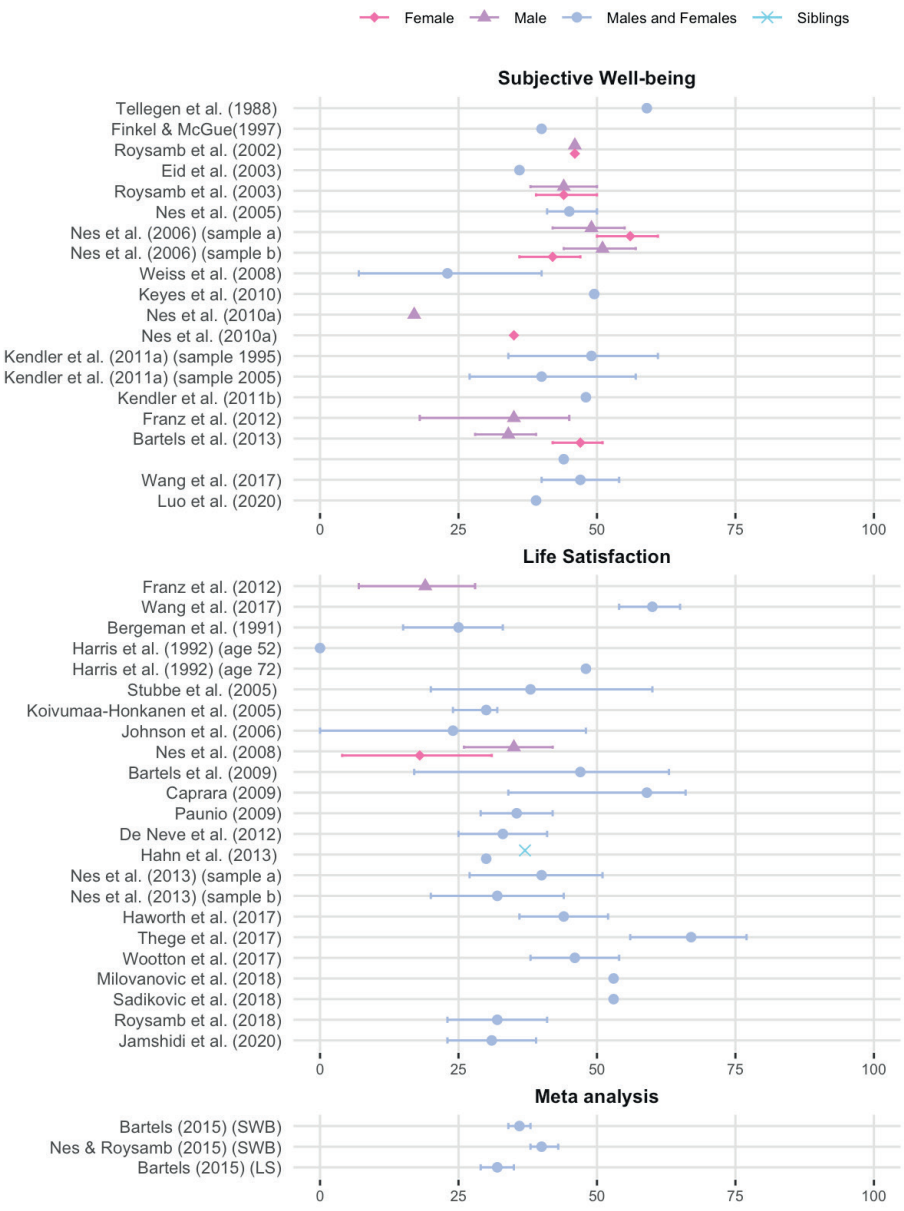


Fig 1.1 Heritability estimates for well-being domains.

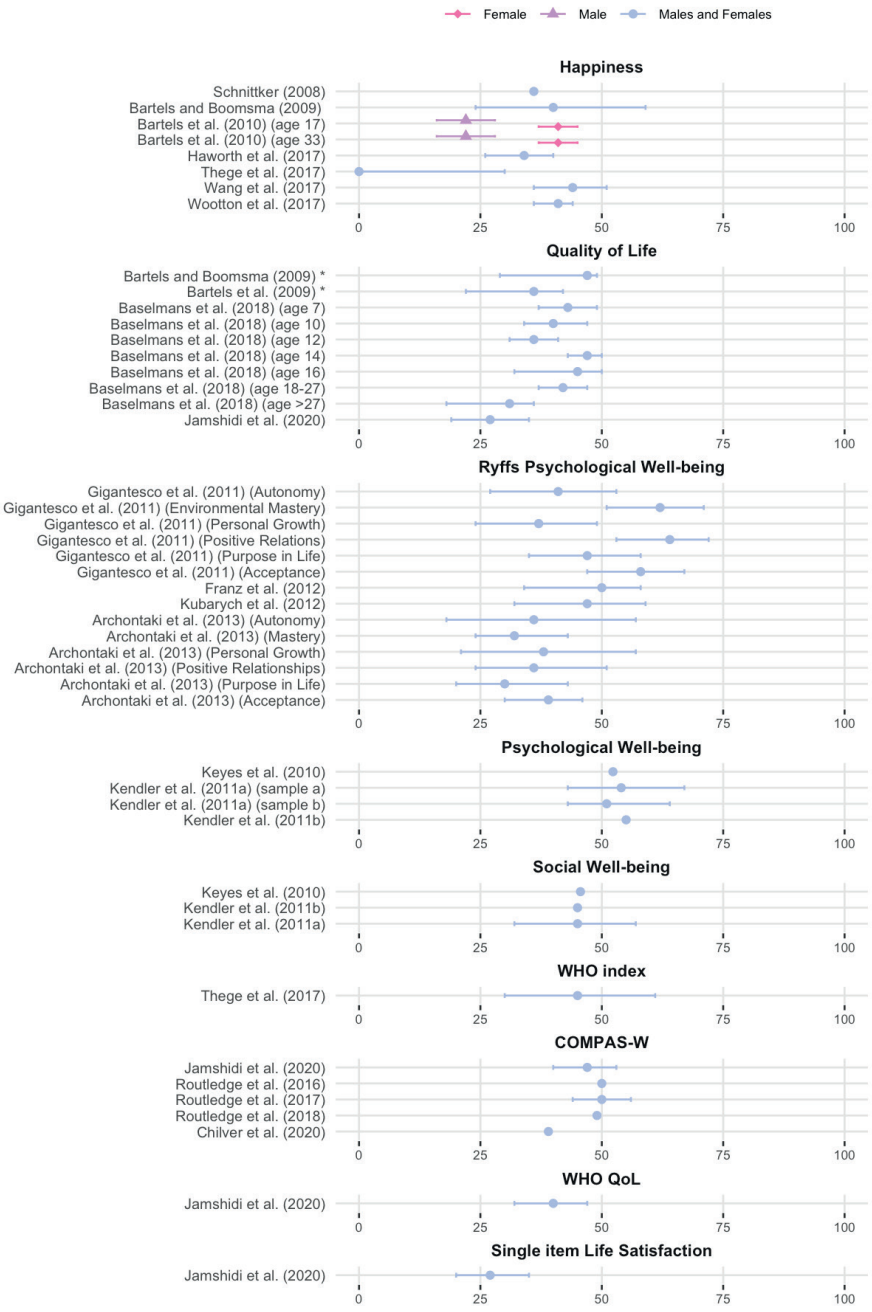


Fig 1.2 Heritability estimates for well-being domains. * = broad sense heritability (includes dominance).

4. MOLECULAR GENETIC FINDINGS FOR WELL-BEING

Results from well-being twin studies have acted as a catalyst for more in-depth analyses of genetic and environmental effects. By revealing that a substantial part (~40%) of the variation in well-being can be attributed to genetic influences, an obvious next step was to try to identify specific genomic regions associated with well-being.

The first reliable molecular evidence for the genetic complexity of well-being came from a method called *Genome-wide Complex Trait Analysis* (GCTA), where the proportion of phenotypic variance explained by all genome-wide SNPs (single nucleotide polymorphisms – DNA sequence variation of a single nucleotide) is estimated by comparing the phenotypic and genetic similarity across a group of unrelated individuals[28]. In a pooled sample of ~11.500 unrelated genotyped Swedish and Dutch participants, well-being was measured using the positive affect subscale of the Center for Epidemiology Studies Depression Scale (CES-D). Based on this approach, it was estimated that 12-18% of the variance in well-being was accounted for by the additive effects of the SNPs measured on genotyping platforms[29].

Next, the development of *genome-wide association studies* (GWASs), allowed for the first identification of specific genetic variants associated with well-being. In a GWAS, millions of genetic variants are measured and associated with a phenotype in a large group of individuals. The association between each genetic variant and an outcome of interest is tested with a strong correction for multiple testing, so that the chance of finding false positives is greatly reduced. The first successful GWAS for well-being (N = 298,420) was performed in 2016. This study led to the identification of 3 genetic variants associated with well-being (defined as life satisfaction and positive affect)[30]. The SNPs had estimated effects in the range of 0.015–0.018 standard deviation per allele. Additionally, high genetic correlations ($r_g > .75$) between life satisfaction, positive affect, neuroticism, and depressive symptoms suggested a common liability, and this common liability can be leveraged to increase the power to identify associated genetic variants. To this end, the largest GWAS for well-being combined these 3 traits and coined them 'the well-being spectrum'. In this study, 304 independent significant variant-phenotype associations were identified for the well-being spectrum, with 148

and 191 associations specific for life satisfaction and positive affect, respectively. Biological annotation revealed evidence for enrichment of genes differentially expressed in the subiculum (part of the hippocampus) and enrichment for GABAergic interneurons. However, even with this progress, the identified variants account for only a small percentage of the variation, meaning that we still have a long road ahead before we completely capture well-being genetics.

These analyses taught us about the genetic complexity of well-being, with likely thousands of variants contributing to variation in the trait. These studies also revealed that each genetic variant only contributes a tiny amount to the variation in well-being, so that we cannot speak of a single "happiness gene" or a few "happiness genes" that assert substantial influence on well-being. However, one way in which we can use the results from these GWASs is by aggregating all small effects into a weighted sum called a *polygenic score* (PGS)[31]. An individual's polygenic score reflects a genetic susceptibility for a trait of interest, in our case well-being. These PGSs can be used in follow-up analyses, for example to examine cross-phenotype overlap[32], or as instruments to examine causality in *Mendelian Randomization*[33], a form of genetic instrumental variable analysis.

5. ENVIRONMENTAL INFLUENCES ON WELL-BEING

While there is substantial genetic influence on variation in well-being, the remaining majority of variance is caused by environmental influences. Again, while twin-and family-studies tell us something about the relative influence of the environment, they do not clarify which environmental influences are important. We can draw a few conclusions from the existing literature on the association between well-being and environmental factors. On the socio-environmental side, it seems that factors associated with social connectedness, such as the quality of social contacts[34] and social support[35] are important for well-being. However, on the more contextual/physical environment side, there is not a lot of consensus on which environmental factors are important. Not only do studies produce contradicting results, but there also seems to be a lack of meta-analytic oversight. This lack of meta-analyses can mostly be explained by the fact that studies use varying designs, making it difficult to directly compare outcomes. There are some overview studies for specific environmental factors from the well-being literature in general, but these studies also fail to present conclusive evidence. For example, Lovell and colleagues

examined the association between exposure to biodiverse environments and well-being and conclude that there is some evidence for a small positive effect, but that much of the evidence is inconclusive[36]. Similarly, Vanaken & Danckaerts[37] and Houlden et al.[38] examined the literature related to the relation between green space exposure and well-being in children and adults, respectively. They both conclude there is limited evidence for a positive effect. Similar to genetic effects, it is likely that effects for these types of environmental factors are small, and that we need large sample sizes to identify them. Unfortunately, even though there is much literature examining the associations between different environmental variables and well-being, it seems we are far from having a complete picture of these environmental influences.

For future research in this area, it is important we continue with large scale investigations into these environmental factors. For example, more homogeneity can be achieved by employing a design that is similar to that used in GWA studies but include multiple environmental factors instead of multiple genetic variants. By performing such *Environment-Wide Association Studie*s, we can study the effect of environmental variables in different populations and geographical levels in a consistent manner. Ni and colleagues already applied such a design for well-being, where they assessed the association between 194 psychosocial and behavioral factors and physical, mental and social well-being in a large Hong Kong sample[39]. They reported that only depressive symptoms, life satisfaction, and happiness were simultaneously associated with these three domains of WB. To develop a full picture of the well-being *exposome* (i.e. the collective of exposures people experience, and how these exposures influence well-being), it is important we continue this progress by studying other types of environmental factors in an environment-wide context, such as the physical and social environment. Moreover, as we have seen in this chapter, there is a considerable genetic influence on well-being. Environmental factors are also partly under genetic control[40], meaning that exposure to certain environments might be driven by genetic factors. Therefore, to fully understand the association between well-being and environmental factors, this gene-environment interplay also needs to be considered. As mentioned earlier, there is a lot of inconsistent results from studies examining the environment in relation to well-being. Part of this inconsistency might be explained by the fact that most studies do not use genetically sensitive designs. Behavior genetic research

can help us elucidate the extent to which covariation between well-being and environmental factors is genetic in nature, for instance using bivariate designs that partition covariance into genetic and environmental sources.

6. THIS DISSERTATION

The previous sections demonstrate that while there has been immense progress in genetic and environmental research on well-being, there are still many interesting questions that need answers. In this dissertation, I examine the way in which our environment impacts our well-being using genetically sensitive designs. Moreover, I investigate the dynamic interplay between genetic and environmental influences.

In the first part of this dissertation, I re-evaluate the way in which well-being was previously examined from a genetic and phenotypic perspective. First, in **Chapter 2**, we use a network approach to examine the well-being spectrum, and compare results to traditional factor analyses. The different approaches present us with only partly overlapping conclusions, which indicates that the use of both can help us gain different, but complementary information. Second, in **Chapter 3**, a systematic review on candidate gene literature for well-being is presented. By re-evaluating these studies in light of more recent, well-powered genome-wide evidence, we conclude that the results from these candidate gene studies cannot be replicated, and that the field should move away from this approach.

The second part of the dissertation focuses on the well-being exposome. In **Chapter 4**, we present an environment-wide association study where we examine associations between 139 objective environmental indicators and well-being. We conclude that, at the neighborhood level, there was especially evidence for associations between well-being and socioeconomic and safety factors. These analyses are followed up in **Chapter 5**, where both subjective and objective environmental indicators are used in poly-environmental scores based on elastic net regression. We find that a poly-environmental score based on subjective environmental indicators predicts approximately half of the environmental variance in well-being, while one based on objective environmental indicators does provide a further contribution to explaining variance in well-being. Moreover, by combing the poly-environmental score with a polygenic score in a UK sample, we find evidence for gene-environment correlation.

In the third part of this dissertation, the focus shifts to an extreme environmental shift that took place during my PhD trajectory: the COVID-19 pandemic. In **Chapter 6**, a bivariate classical twin design is presented where we examine the effect of the first COVID-19 pandemic lockdown on the variance decomposition of Self-Rated Health (SRH). We find that, during this first lockdown, mean SRH increases, but the relative amount of variance that can be accounted for by genetic and environmental factors remains unchanged. In **Chapter 7**, we examine the influence of this first lockdown on Quality of Life (QoL) in a sample of multiples and their family members using the Mendel software. We see decreased levels of QoL during the first lockdown, and a large increase in unique environmental variance, resulting in a decreased heritability.

The last section of my dissertation moves toward more causal interpretations in the context of well-being. First, in **Chapter 8**, we are interested in examining the extent to which genetic factors explain associations between adolescent well-being and different aspects of the social environment. Using twin-difference scores and bivariate twin models, we find evidence for large (73-91%) genetic influence on the associations between well-being and family conflict and functioning, leisure time sport/scouting clubs, and satisfaction with friendships. Finally, in **Chapter 9**, we use four different methods for examining causality in a large UK dataset to assess a potential causal effect of longer education on a range of mental and physical health outcomes. For none of the outcomes, a consistent causal effect was found across all four methods.

We conclude this dissertation in **Chapter 10**, with a summary and general discussion. In this discussion, I reflect upon how the combined work presented in this dissertation contributes to the field. Moreover, I provide some perspective on the exciting ways in which future research can build upon this work and how we can use this type of research to inform policy and improve population outcomes.

PART

Re-Evaluating Well-Being Phenotypes and Genetics

CHAPTER 2

Connecting the dots: Using a network approach to study the well-being spectrum

Submitted as: Van de Weijer, M.P., Landvreugd, A., Pelt, D.H.M., & Bartels, M. (*under review*). Connecting the dots: Using a network approach to study the well-being spectrum.

*supplementary materials accessible at: https://drive.google.com/drive/folders/1u-kQ5ARNRTCjvek_TtvFKMKtoF8xxxud?usp=sharing

ABSTRACT

In the past few years, the network approach has gained both popularity and criticism in the application to psychological constructs. In this paper we used psychometric network approach to study the structure of well-being, and compare the results with results from factor analytic models. First, in a trimming sample of N=1343 participants, we examine potential item redundancy based on associations between satisfaction with life, subjective happiness, quality of life, flourishing, self-rated health, depressive symptoms, neuroticism, and loneliness items. Next, we fit the network in an estimation sample of *N*=759 participants, and examine the performance and accuracy of the network. Lastly, we perform exploratory and confirmatory factor analyses to compare the results and feasibility of both approaches. Our final network consists of a positive cluster including satisfaction with life, subjective happiness, and flourishing items, and a negative cluster including depressive symptoms, loneliness, and neuroticism items. While items belonging to the same well-being measure clustered together, most well-being items were densely connected, re-affirming the complexity of the construct. The factor analyses, on the other hand, suggest six independent but moderately to strongly correlated factors were a better fit for the data than a model with one or two overarching well-being factors. While it is not possible to determine whether the factor or network structure is a better depiction of reality based on these results, we find that both approaches provide us with different but complementary information.

1. INTRODUCTION

Defining and delineating well-being as a construct has proven to be a difficult challenge for the field of positively psychology. There are many different well-being theories, and it is often unclear how these different theories relate to or complement each other. Well-being can be considered an umbrella term for many different more or less connected constructs, which can lead to difficulties when interpreting and comparing results from positive psychological research[10].

Various researchers have made an effort to outline different well-being theories based on existing research[4,41,42]. While it is beyond the scope of the present study to review all existing theories, we provide a brief overview of some of the main theories and their origins. Most of the existing positive psychological theories on well-being originate from philosophical traditions. Lambert, Passmore, and Holder distinguish four (partly overlapping) philosophical traditions that were influential for well-being research: utilitarianism (focus on community well-being and maximizing happiness), virtue philosophy (focus on character strengths), hedonism (focus on maximizing pleasure), and eudaimonism (focus on functioning well and meaning)[4]. An influential contemporary well-being theory combining aspects of all four of these traditions is the PERMA model[43]. The theory postulates that Positive Emotion, Engagement, Relationships, Meaning, and Accomplishment (PERMA) are the building blocks of well-being. While the different PERMA elements correlate with each other[44], they are believed to also independently contribute to overall well-being and can be measured and defined independently using the PERMA profiler[45].

Most contemporary well-being theories tend to be the result of a philosophical tradition. For example, Diener's theory on subjective well-being (SWB) is grounded in the hedonistic tradition of well-being and proposes that SWB is comprised of life satisfaction (cognitive SWB), high levels of positive affect, and low levels of negative affect (emotional SWB)[11]. On the other hand, Ryff's theory on psychological well-being (PWB) is grounded in the eudaimonic tradition of well-being and states that PWB is comprised of six dimensions: autonomy, environmental mastery, personal growth, positive relationships, purpose in life, and self-acceptance[46]. A similar influential theory is self-determination theory (SDT)[21]. Central to SDT is an individual's experience of autonomy, competence, and relatedness, which are

argued to promote well-being. This is slightly different from Ryff's PWB theory, where these dimensions are believed to be components of well-being. Combining aspects from both eudaimonic and hedonic theory, Keyes formulated a theory on flourishing that posits that well-being or mental health is defined by high levels of emotional, psychological, and social well-being, and an absence of psychopathy[47]. The inclusion of social well-being is in line with eudaimonic ideology and is well-supported by well-being literature[48]. While these theories focus on specific aspects of well-being, it is also possible to evaluate well-being in a broader context. Well-being is highly (phenotypically and genetically) correlated to multiple traits, such as depression, neuroticism, loneliness, and self-rated health. In our own work, we collectively refer to these traits as "the well-being spectrum" (WBS)[32].

Importantly, these theories do not necessarily claim to provide a comprehensive well-being framework encompassing the well-being construct in its entirety. Because of this, it is not clear how these different theories combine into one framework: it is unlikely they are all touching upon completely separate domains of the same overarching construct, but it is also unclear to what extent the different well-being constructs overlap. The most common way in which this issue has been studied is through factor analytical methods. In these models, item responses are modelled so that they "load" onto higher-order well-being factors such as SWB and PWB, and the relation between these higher order factors is evaluated by correlating them with each other. These studies provide mixed results in terms of the structure of well-being, with some studies finding single factor solutions[49], and some finding multiple-factor solutions with varying degrees of correlations between these factors[50–52]. Factor analytical methods implicitly assume a top-down (reflective) model in which correlations between indicators are explained by the latent factor. This means that they assume that conditional on the latent factor, residual correlations between the items are zero. Consequently, information on the associations between the different items, independent from them loading on the same higher-order factors, is lost. Therefore, by modeling well-being items as part of an overarching construct, we risk losing important information on the relation between these different components at the item level.

An alternative to approaching well-being as a construct with a latent factor, is approaching well-being as a network of interacting aspects[53]. The main conceptual difference here is that factor analysis is founded on the idea of a common latent

factor (e.g., well-being) that causes the related "symptoms" (e.g., life satisfaction aspects). Contrarily, network theory advocates that symptoms are all part of an interactive system. In this way, network analysis allows us to take a closer look at item-item associations without assuming that their correlations stem from an overarching factor. A network consists of a set of nodes (i.e. symptoms) along with a set of specified ties (edges) linking the nodes[54,55]. The general aim is to characterize the structure of the network and the position of nodes, and to use the network to better understand the examined phenotype. The network is used to reveal which components are the most "central" using the concept of centrality: components with a high degree of centrality are most strongly connected to other items[56], and are therefore thought to be most influential in the network. Following this line of reasoning, components that are central to the network (i.e., high levels of centrality), may serve as targets for the development of prevention and intervention strategies.

A few studies have applied network approaches to well-being phenotypes. In a sample of Chinese adolescents, a network model was applied to the 20-item Chinese version of the engagement, perseverance, optimism, connectedness, and happiness (EPOCH) scale. Being cheerful, being absorbed in current activities, and being optimistic were the most central components of the network, and as suggested by the authors, might serve as useful targets for improving well-being in adolescents[57]. In another study, fourteen well-being items measuring affective-emotional aspects, cognitive-evaluative dimensions, and psychological functioning were used to create a network in a UK sample[58]. Three items related to self-perception and cheerfulness were most central to the network, suggesting that these domains play an important role in influencing other aspects of well-being. Besides prevention and intervention, there are two studies that applied networks to investigate the structure of well-being as a research topic, i.e. to clarify terminology and well-being concepts. One example is a paper by Giuntoli and Vidotto, which estimates a network that included measures of both SWB and PWB in an Italian adult sample (N = 2392)[59]. Based on their findings, the authors conclude that the final network was most in line with Diener's definition of well-being, with life satisfaction, positive and negative experiences, and perceived positive functioning as different, but connected, well-being domains. The second example examines how fluctuations in specific components of well-being are

associated with fluctuations in other components of well-being by estimating a network based on time-series data of well-being (N = 151)[60]. Their analysis suggests that feeling satisfied is not just a component of well-being, but also plays an active role in triggering other related well-being aspects, such as cheerfulness.

Similar to the factor analytic approach, the network approach has its own limitations. A common critique is the validity of the centrality indices in the context of psychological networks. Bringmann and colleagues[61] provide three reasons for why centrality indices might not be suitable for psychological networks. First, the indices were originally developed for social networks (where connections are direct representations of raw data) but these are substantially different from psychological networks (where connections are coefficients derived from a model). Second, some indices (especially closeness and betweenness) have shown to be unstable in psychological networks. One of the reasons why this instability might occur is that the centrality indices are susceptible to which nodes are included the network, and it is unknown which nodes should be included beforehand[62]. Third, there has been little research on the predictive power of centrality indices in psychological networks. Whether centrality indices can thus point to clinically relevant symptoms as targets for intervention is not entirely clear. All in all, while centrality indices seem like a useful advantage of network theory, it is not certain they are as useful for psychological networks as they are for social networks.

To further explore the value of network theory for studying well-being specifically, we estimate a broad network that includes different well-being measures, but also the broader WBS depressive symptoms, neuroticism, self-rated health, and loneliness in a sample of Dutch adults (discovery N=1343, replication N=759). By estimating a broad WBS network, we aim to get better insight into well-being in terms of how clearly delineated or interconnected well-being items from different domains are. In addition, we explore the WBS in classic exploratory and confirmatory factor models. This allows us to compare the results and to evaluate the advantages and disadvantages of both approaches. We conclude the study by considering the added value of network science as a method for answering questions about the nature of the well-being construct.

2. MATERIALS AND METHODS

2.1 Sample

Study participants are voluntarily registered with the Adult Netherlands Twin Register (ANTR)[63]. For the current project we made use of four waves of NTR data collection: 1) the 8th wave of data collection, collected from 2008 to 2010, 2) the 10th wave of data collection, collected from 2012 to 2014, 3) the 13th wave of data collection, collected in 2017-2018, 4) and the 14th wave of data collection, collected in 2019- February 2020. These waves were selected based on the availability of relevant well-being variables. Participants were included if they participated in at least one of these surveys. If data on multiple time-points were available, we selected the most recent time-point.

Since the NTR collects data in multiples and their family members, many individuals are genetically related to each other, meaning that the observations are not entirely independent. To prevent bias due to these dependencies, we selected two samples so that within each sample, all individuals were genetically unrelated to each other. These samples were used as a trimming sample (to check for potential redundant nodes) and an estimation sample (to estimate the network) (see Figure 1). The samples included only participants that had complete data available for all the different traits. In total, the trimming sample included 1343 individuals (63% females, M_{age} = 53.18, SD_{age} = 9.45). The estimation sample included 759 participants (75% females, M_{age} = 45.27, SD_{age} = 11.12).

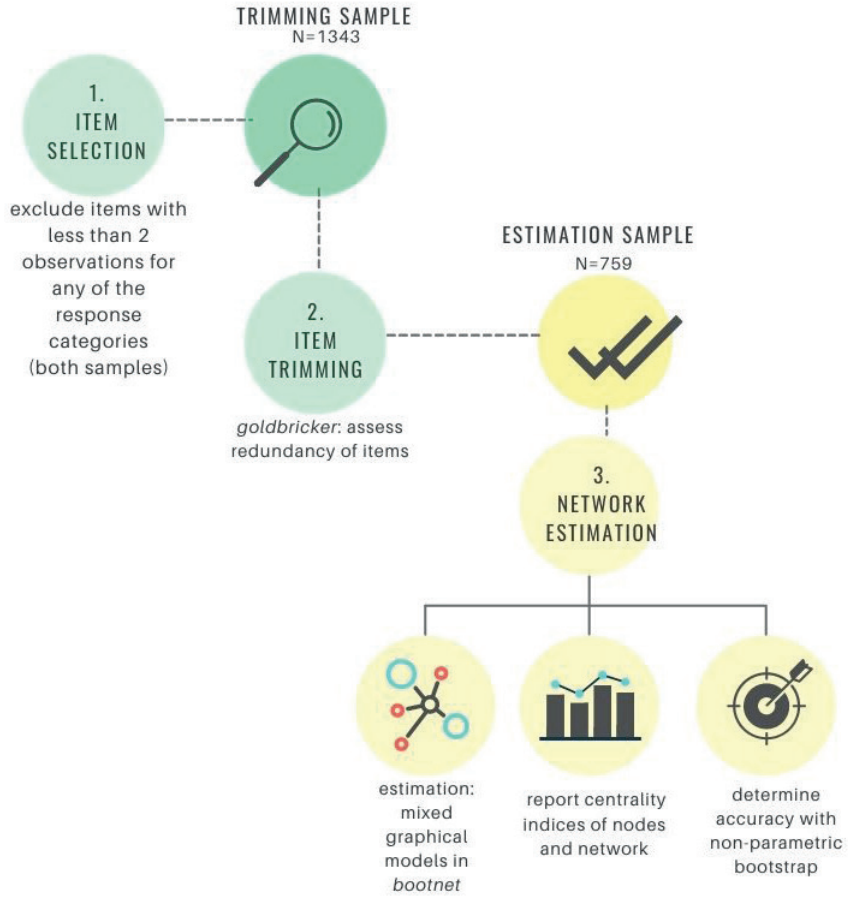


Fig 1. Overview of the analysis plan.

2.2 Measures

To assess the well-being spectrum phenotypes the following standardized instruments were used:

The Subjective Happiness Scale[64]. Four items were rated on a Likert scale from 1 (strongly disagree) to 7 (strongly agree). An example of an item is: "*On the whole, I am a happy person*". We recoded the items so that for all items, a higher score meant higher levels of happiness.

The Satisfaction with Life Scale[12]. Five items were rated on a Likert scale from 1 (strongly disagree) to 7 (strongly agree). An example of an item is: '*My living conditions are excellent*'.

Cantril's ladder[65] was used to assess *Quality of Life* (QoL). Participants were asked '*Where on the scale would you put your life in general?*', with 0 representing the worst possible life and 10 representing the best possible life.

The Short Flourishing Scale[66]. The scale contains 8 items that are rated from 1-7 using a Likert scale. 1 resembles 'strongly disagree' and 7 resembles 'strongly agree'. An example of an item is: '*I am competent and capable in the activities that are important to me*'.

Depressive symptoms. The depressive problems subscale from the adult self-report (ASR) of The Achenbach System of Empirically Based Assessment (ASEBA) was used to assess depressive symptoms[67]. 14 items were rated from 0-2 (0= not true, 1 = somewhat true, 2= very true). An example of an item is: '*I feel worthless or inferior*'.

Loneliness. The three items from the short scale for assessing loneliness in large epidemiological studies were used to assess loneliness[68]. For each item, participants indicated how often they identify with a statement, rated as: 0=almost never, 1=sometimes, or 2=often. An example of a statement is: '*How often do you feel isolated from others?*'.

Neuroticism. The NEO-FFI (NEO Five Factor Inventory) neuroticism subscale was used to assess neuroticism[69]. The subscale consists of 12 items, and each item was rated on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree. An example of an item is '*I often feel tense and jittery*'. Half of the items were reverse-coded so that a higher score indicated higher levels of neuroticism.

Self-rated health. A single item was used to evaluate self-rated health: '*How would you rate your general health?*'[70]. This item was rated on a 5-point scale ranging from 'Bad' to 'Excellent'.

2.3 Statistical analyses

2.3.1 Network analysis

An overview of the different steps of the analysis plan is depicted in Figure 1. Below, we provide more detail on each separate step.

Item selection

Before estimating networks, we examined the distribution of all the items. We excluded ordinal variables having less than 2 observations for any of the observed response categories. The threshold value 2 was chosen because this is a requirement for the non-parametric bootstrap of ordinal items we performed[71] at a later stage.

To estimate the most parsimonious network in the estimation sample, we used the trimming sample to examine item redundancy (i.e., items that are not essential to the network since they correlate highly with other items). The goldbricker function implemented in the networktools R package[72] was used to assess potential item redundancy. With this function, strongly correlated item pairs ($r \geq .7$) that had less than 50% unique combinations with other items (i.e. less than 50% of significantly different correlations with other nodes, $p = .05$) were identified. Next, the net_reduce function was used to choose the more unique node of each redundant pair and remove the redundant one. Based on the network trimming in the trimming sample, we estimated the network without redundant nodes in the estimation sample.

Regularized network estimation

We estimated the WBS network using the estimation sample with all items that remained after the item selection and item trimming phase. We included sex and age as covariates. The network was estimated using the *bootnet* package[55], and visualized using the *qgraph* package[73] in Rstudio[74]. Since mixed variable types (continuous and ordinal) were included in the network, the function *Mixed Graphical Models (MGM)*, which allows for the inclusion of both categorical and continuous data, was chosen as the best regularized estimation method for our data[75]. The model employed by MGM is a pairwise Markov random field (PMRF) model, where nodes are connected by undirected edges, and unconnected

nodes are independent after conditioning on all other variables. Least absolute shrinkage and selection operator (LASSO) regularization with Extended Bayesian Information Criterion (EBIC) model selection was applied to limit the number of spurious edges. The EBIC tuning parameter γ (gamma) controls the level of sparsity (i.e. the likelihood that spurious edges are removed). The parameter typically ranges from 0 to 0.5, where lower values are most sensitive to edge detection with higher risks of false positives, whereas higher values have the highest specificity, with the risk of excluding true edges. To avoid false positives, we set the tuning parameter to the default value (for mgm) of 0.5. The network was plotted using the multidimensional scaling (MDS) function implemented in the *networktools* R package. In MDS plots, the distance between the nodes is reflective of the strength of the association between two nodes, with nodes placed closer together sharing stronger associations.

Centrality and Clustering

We examined the centrality index *strength* (the sum of absolute edge weights connected to each node), which indicates how strongly a node is directly connected to other nodes. The strength centrality measure works optimally in a network with exclusively positive edges as this index does not distinguish between positive and negative edges. We, however, expected, due to the WBS structure, positive (well-being, self-rated health) as well as negative (neuroticism, depression, loneliness) edges. Therefore, we also estimated the *expected influence* (EI) of the nodes[76]. Expected influence (EI) assesses a node's influence while accounting for both negative and positive edges. Nodes with higher EI would play a bigger role in the etiology of well-being. In case of a node with both negative and positive edges, expected influence is a preferable measure over strength, as 1) a node with a comparable number positive and negative edges might have little influence on the overall network since these influences have opposing effects on the network, and 2) a node with a comparable number of stronger positive and negative edges may have little cumulative influence on the network. Based on simulations, EI does indeed seem to outperform strength as in the presence of negative nodes Identifying highly influential nodes in the complicated grief network[76].

To examine the network as a whole, we estimated the *global clustering coefficient* and *local clustering coefficients* of the network. The *global clustering coefficient*

(i.e. transitivity) is an estimate for how often a node's neighbouring nodes are also connected to each other[77]. It reflects the number of closed triads (groups of three nodes that are all connected to each other) over the number of possible triads, with a global clustering coefficient of 0 meaning that none of the triads are closed, and 1 meaning that every triad is closed. A network with a high global clustering coefficient is thus characterized by a highly connected and clustered network structure, while a low global clustering coefficient indicates the network is comprised of numerous weak ties. Next, we calculated *local clustering coefficients* (as implemented in the *qgraph* R package) using Zhang & Horvath's weighted clustering coefficient[78]. The coefficient indicates the likelihood that a node's neighbouring nodes (i.e. the nodes that are connected to a particular node) are also connected. A local clustering coefficient of 1 indicates that the node is at the centre of a fully interlinked cluster, while a coefficient of 0 indicates that a node's neighbouring nodes are not connected at all.

Edge-weight accuracy

Lastly, we examined how accurately we estimated the edge-weighs in our network by using the non-parametric bootstrapping in *bootnet*[55]. Using this method, observations are resampled with replacement to create new plausible datasets where the edge-weights can be re-estimated in. Based on 1000 bootstraps, a 95% confidence interval (CI) around the edge-weights was estimated. These Cis can be used to assess accuracy of the edge-weights, with wider Cis reflecting less accurate edges.

2.3.2 Factor Analysis

For comparison purposes, we used the trimming and estimation samples to run exploratory (EFA) and confirmatory factor analysis (CFA), respectively. First, in EFA, we examined the number of factors to extract from the data using the "parallel" function in the *psych* package in R[79]. We examine how the items load on that number of factors using the 'fa' function, where we use minimum residual factor extraction and oblimin rotation, since we expect the different well-being components to be correlated. We examine potentially redundant items based on their communalities (i.e. the proportion of an item's variance that can be explained by the factors). As a threshold, a communality of over .3 was deemed acceptable.

We use this threshold to compare which items are left out of the network analysis to those left out based on factor analysis.

To enable comparison with the network analyses, we use the network-based trimmed estimation sample for our CFA. In this reduced set of items we assume that all items load on a factor representing their corresponding construct. Using the *lavaan* package[80] we compare the fit of three models: 1) a six factor model with correlated factors. This model contains one factor for each included construct (excluding quality of life and self-rated health since these were removed in the trimming stage) , 2) a higher-order factor model where the six factors load on one second-order "well-being spectrum" factor, and 3) a higher-order factor model where the "positive" traits (satisfaction with life, subjective happiness, and flourishing) load on a positive second-order factor, and all "negative" factors (depression, neuroticism, loneliness) load on a negative second-order factor (where the higher-order factors are allowed to correlate). We compare the fit of the higher-order models to the fit of the six-factor model without higher-order factors using a likelihood ratio test for comparing nested lavaan models.

2.3.3 Comparison of the network model and the factor model

After fitting both models, we compared the results in two ways. First, we examined which items were excluded from the well-being network based on the redundancy, and compared this to the items that were excluded based on communalities in the EFA. Second, we compared the structure of the WBS based on the network approach and the factor analytic approach.

3. RESULTS

3.1 Network Analysis

Item selection

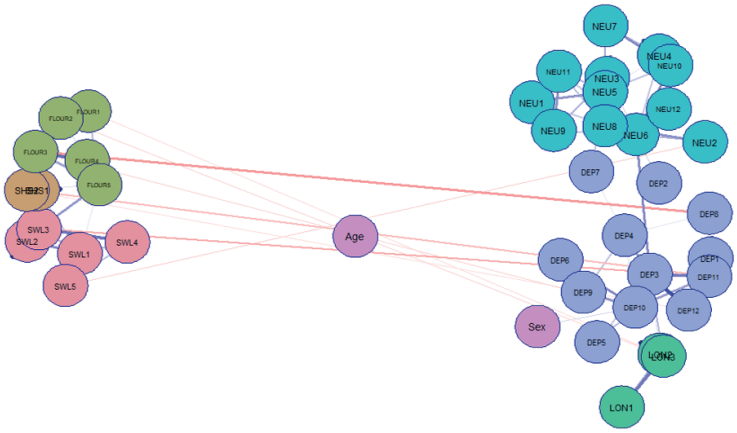
Five items were removed because they did not meet the threshold of at least two observations in each category; a self-rated health item ("*How would you rate your general health?*"), three items from the flourishing scale ("*I am engaged and interested in my daily activities*", "*I actively contribute to the happiness and well-being of others*", and "*People respect me*"), and one depression item ("*I deliberately try to hurt or kill myself*").

Network trimming

Using the goldbricker function four nodes were identified that could be removed. The quality of life item (*"Where on the scale would you put your life in general?"*) was deemed redundant since it was not significantly different from a satisfaction with life item ("*I'm satisfied with my life")* within the context of this network. Two subjective happiness scale items (*"On the whole I am a happy person"* & *"On the whole, I am very happy, I enjoy life come what may and I always make the best of things"*) were excluded because of redundancy in the context of two other subjective happiness items: "*Compared with most of my peers, I am less happy than they are"* and "*On the whole, I am not very happy, although I am not depressed I never seem to be as happy as I could be*", respectively. Lastly, one depression item (*"I feel tired without good reason"*) had a redundant role in the network because of another depression item ("*I do not have much energy*"). After removing these items from the estimation data, 41 items (including covariates) were left for network estimation.

Network structure

The multidimensional scaling (MDS) layout of the well-being network is shown in Figure 2, where the distance between the nodes is reflective of how strongly the nodes are correlated[81]. A visual inspection of this graph reveals two clusters: a depression, loneliness, and neuroticism cluster reflecting the more negative aspects of the WBS, and a cluster of the different well-being measures, reflecting the positive aspects of the WBS. Supplementary Table 1 provides the partial correlation matrix that underlies the network depicted in Figure 2 (with item descriptions in Supplementary Table 2). The positive and negative cluster are mostly connected through depression nodes connecting to different well-being nodes. While loneliness and neuroticism are also directly connected to flourishing and satisfaction with life, respectively, they are mostly indirectly connected to well-being items through depression nodes. Additionally, we see that items that belong to the same questionnaire tend to cluster together. While not immediately obvious from the graph, age was not connected to any of the variables.



Satisfaction with life

1. My life is going more or less as I wished.
2. My living conditions are excellent.
3. I'm satisfied with my life.
4. Until now I've always gotten the most important things I wanted in life.
5. If I had to live my life again, I would do more or less the same.

Subjective happiness

1. Compared with most of my peers, I am less happy than they are (RV).
2. On the whole, I am not very happy, although I am not depressed I never seem to be as happy as I could be (RV).

Flourishing

1. I lead a purposeful and meaningful life.
2. My social relationships are supportive and rewarding.
3. I am competent and capable in the activities that are important to me.
4. I am a good person and live a good life.
5. I am optimistic about my future.

Depressive symptoms

1. I cry a lot
2. I do not eat as well as I should
3. I feel worthless or inferior
4. I feel very guilty.
5. There is very little that I enjoy
6. I sleep more than most other people during day and/or night
7. I have trouble making decisions
8. I think about killing myself
9. I have trouble sleeping
10. I do not have much energy
11. I am unhappy, sad, or depressed
12. I feel that I cannot succeed

Covariates

1. Age
2. Sex

Loneliness

1. How often do you feel that you lack companionship?
2. How often do you feel left out?
3. How often do you feel isolated from others?

Neuroticism

1. I am not a worrier (RV).
2. I often feel inferior to others.
3. When I am under a great deal of stress, I sometimes feel like I am going to pieces.
4. I rarely feel lonely or blue (RV).
5. I often feel tense and jittery.
6. I sometimes feel completely worthless.
7. I rarely feel fearful or anxious (RV).
8. I often get angry at the way people treat me.
9. Too often, when things go wrong, I get discouraged and feel like giving up.
10. I am seldom sad or depressed (RV).
11. I often feel helpless and want someone else to solve my problems.
12. At times I am so ashamed I just want to hide.

Fig 2. Multidimensional scaling layout network of the well-being spectrum. Blue lines indicate positive associations, red lines indicate negative associations. RV= reverse coded before the analyses.

Centrality and Clustering

Standardized centrality indices for each item are depicted in Figure 3. Lower, negative Z-scores indicate nodes with the least strength, while higher, positive Z-scores indicate nodes with the highest strength. The nodes that scored relatively high on strength were SWL3 (*'I am satisfied with my life'*), NEU6 (*'I sometimes feel completely worthless'*), DEP3 (*'I feel worthless or inferior'*), and DEP11 (*"I am unhappy, sad, or depressed"*). Sex, which was included as a covariate, scored the lowest. We estimated both strength and expected influence since our network contained both negative and positive nodes (which is not taken into account by the strength index). However, because there were not many negative edges in the network, the results for strength and expected influence are very similar. The global clustering coefficient of the entire network was .32, and the local clustering coefficients ranged from 0 to .38 (see Supplementary Table 3).

Edge weight accuracy

Supplementary Figure 1 contains the bootstrapped CIs of the edge-weights (edge labels were left out for readability) that were estimated to examine the edge-weight accuracy. On the y-axis are all edges in the network, and on the x-axis it shows the strength of the edge-weights, with red dots as point estimates, the black dots the bootstrap means, and the grey area as 95% confidence intervals. Overall, we find relatively large CIs. These CIs do not reflect whether or not an edge should have been set at zero, but rather the accuracy of the estimated edge-weights, indicating that the strength of the edges should be interpreted with caution.

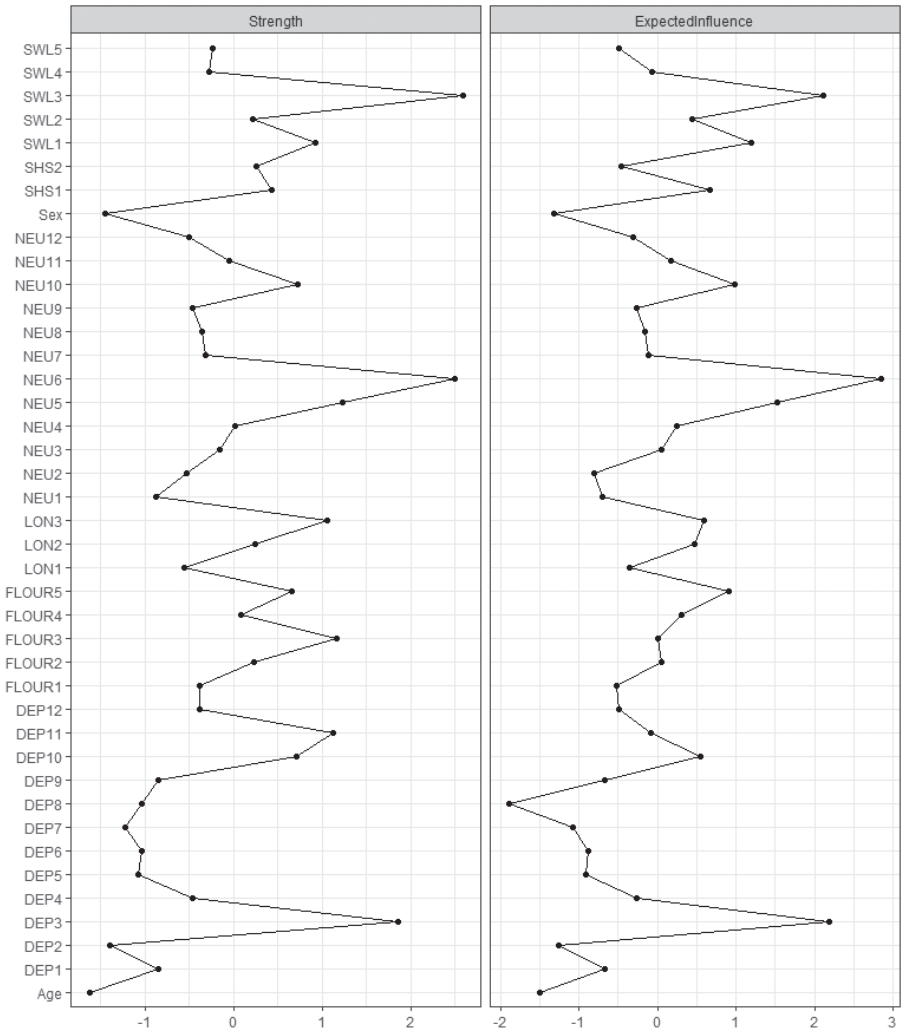


Fig 3. Centrality indices of all nodes (see Figure 2 for item descriptions).

3.2 Factor Analysis

We exploratively examined the factor structure of our items using EFA in our trimming sample. Parallel analysis suggested an nine-factor solution. A closer examination of the extracted factor solution indicated that only two items had factor loadings >.3 on the last two factors, and that both these factors explained only 1% of the variance. Therefore, a seven-factor solution seemed a more sensible

solution (variance explained = 47%). Within this solution (see Supplementary Table 4), items from three scales loaded exclusively on their own intended factors: neuroticism (factor 1), flourishing (factor 2), and loneliness (factor 5). Factor 3 was a composite of quality of life, self-rated health, SWL, and two SHS items. Three of the four SHS items additionally loaded on Factor 7. The 4th factor included 7 depression items, whereas the 6th factor included 3 other depression items and additionally, self-rated health loaded on this factor. Moreover, in the seven-factor solution, two neuroticism items and six depression items had a communality lower than .3 (see Supplementary Table 5).

In our confirmatory factor analysis, we compared the fit of three models: a model with six correlated factors corresponding to the six well-being constructs, a model where we include one higher order well-being spectrum factor on which the six latent factors load, and a model where we include one higher order factor for positive traits, and one for the negative traits. We find that both the model with the single higher-order factor and the two higher-order factors fit the data significantly worse than the six-factor model without higher-order factors (see Table 1 for fit indices). In the six-factor model, all correlations between the latent factors were of medium to high strength, with absolute correlations ranging between .52 and .86 . An overview of the correlations between all factors can be found in Supplementary Table 6.

Table 1

Model fit comparisons

	df	AIC	BIC	CFI	RMSEA	χ^2	$\Delta\chi^2$	Δ *df*	*p*-value
Six factor model	687	58061	58490	.915	.046	1789.5			
One higher-order factor	696	58171	58559	.906	.049	1917.9	128.46	9	$<2.2x10^{-16}$
Two correlated higher-order factors	695	58462	58462	.914	.047	1814.9	25.459	8	.001

Note. We compare the six-factor model to the two other models.

3.3 Comparison of the network model and the factor model

There are two comparisons we can make between our network and factor model. First, we used the trimming sample to exclude possibly redundant items based on item-item correlations in the network analysis, and to examine which items were badly captured by the factors in the EFA. Thus, for the former, we exclude

items based on the fact that they are almost fully captured by other items (i.e., redundant) in the network. In contrast, the latter excludes items that do not seem to fit in with the rest of the items based on the common variance between items. Unsurprisingly, these two opposing strategies lead to very different outcomes: whereas the network trimming leads to the exclusion of mostly well-being items, the factor strategy leads to the exclusion of neuroticism and depression items. Since redundancy in network analysis is based on high correlations between different items, we find that items deemed redundant in the network analysis have relatively high communalities in the factor analysis (between .469 and .750).

Second, we used the trimmed sample (with item exclusion based on the network trimming) to examine the structure of the WBS based on network estimation and CFA. The network analysis indicates that the different items are clustered within their own construct but were simultaneously highly interconnected across different constructs. Moreover, we found that the more items belonging to more positive phenotypes clustered on one side of the network, while the items belonging to more negative phenotypes clustered on the other side of the network. For the factor analysis, on the other hand, we found that six separate factors corresponding to the separate constructs were a better fit to the data than a factor model with one or two higher-order factors.

4. DISCUSSION

In the present study, we set out to study the well-being spectrum from a network perspective to get better insight into the construct itself and evaluate the added value of network psychometrics (compared to traditional factor analyses). The final network suggests two clusters, with on the one hand the "negative" spectrum items of depression, loneliness, and neuroticism, and on the other hand the "positive" spectrum items from the different well-being measures. On the contrary, confirmatory factor analyses suggests that a separate factor for each construct is a better fit for the data than a model including higher-order factors representing the more "positive" and "negative" traits.

Before estimating the network, we excluded items that could be classified as redundant nodes (when 2 items correlate > .7 and <50% significantly different correlations with other items). This led to the exclusion of the quality of life

item, two subjective happiness items and one depression item. Importantly, the exclusion of these items does not mean they are unrelated to the network, but rather that the item is redundant for this specific network because there is another item that plays a similar role in the network. For example, one of the satisfaction with life items was highly correlated with the excluded quality of life item (r = .73) and additionally correlated similarly with other nodes in the network. Since this satisfaction with life item was more unique to the network than the quality of life item (in terms of its redundancy statistics with other nodes), this item was retained instead of the quality of life item. This does not mean that quality of life should be disregarded with respect to well-being, but rather that it correlates similarly with the rest of the network as satisfaction with life. In the factor analysis, items are excluded using an opposite strategy, namely using the proportion of an item's variance that can be explained by the factors (i.e. commonality). Unsurprisingly, this led to the exclusion of different items: 3 depression items and 2 neuroticism items had a commonality lower than .3.

We estimated the final WBS network in the estimation sample. Visual inspection of the network suggests the presence of two smaller networks, connected through a few items. One cluster consisted of positive items for subjective happiness, satisfaction with life and flourishing, while the other, more negative, cluster included depressive symptoms, neuroticism, loneliness, and sex. The positive and negative cluster were predominantly connected by edges between multiple depression items and multiple well-being items, and not by one or two "bridge items" (see Supplementary Tables 1-2). Age was not connected to other nodes in the network, i.e., it was independent after conditioning on all other variables, indicating that the network structure is independent of the age of the participant in our adult sample. Since our current sample only included adults, it would be interesting would be to repeat our current efforts in a sample of children/adolescents or in a sample of older adults to see if age does affect the network in such a sample.

With respect to the well-being items, we see that items belonging to the same measurement instrument tend to cluster together, but there are also several connections between well-being items from different instruments. On the one hand, the clustering of well-being items belonging to the same measurement instrument (i.e. flourishing, satisfaction with life, and subjective happiness) is in

line with theories such as Keyes' theory on flourishing or Diener's theory on SWB that distinguish different well-being domains such as cognitive well-being and psychosocial well-being[11,47]. On the other hand, it also becomes clear that all well-being items are highly clustered and interconnected, suggesting that the different domains are not as clearly delineated as may be claimed by different theories/ factor analytic studies. Taken together, these results are very similar to previous findings for the WBS where we found high phenotypic and genetic correlations between WBS measures, and additionally identified a genomic factor model where positive and negative traits loaded on separate, but highly correlated factors[32]. This is also similar to findings by Giuntoli and Vidotto, who conclude based on their network analyses that different SWB and flourishing components are closely related constructs[59]. In addition, it is in line with other studies emphasizing that different well-being phenotypes are highly interconnected[49,82,83].

We also ran exploratory and confirmatory factor analyses to examine the structure of the well-being spectrum. The exploratory factor analyses showed that some factors clearly represented a distinct well-being construct, while others were more mixed. This is in line with our network results where most items clustered within their own construct but were also interconnected across different constructs. Moreover, our confirmatory analyses indicated that six independent but moderately to strongly correlated factors were a better fit for the data than a model with one or two overarching well-being factors. In contrast to the confirmatory factor analysis, our network approach suggests that the well-being items form a connected system consisting of a positive and negative side. While it is not possible to determine whether the factor or network structure is a better depiction of reality based on these results, it is interesting to see how both approaches provide us with different but complementary information.

One of the advantages of network psychometrics is the possibility to examine nodes in terms of their individual strength in the network. Four items scored relatively high compared to all other nodes: SWL3 ('*I am satisfied with my life*'), NEU6 ('*I sometimes feel completely worthless*'), DEP3 ('*I feel worthless or inferior*'), and DEP11 ("*I am unhappy, sad, or depressed*").This indicates that these nodes have stronger connections to other nodes in the network. A potential interpretation is that the most central nodes reflect the items that are most representative of the WBS. Examining these items in the factor analytic context, these are also items

with high communalities compared to the rest of the items (see Supplementary Table 5). Thus, to get a general idea of an individual's well-being with a limited resources, one might benefit most from examining these items. Additionally, the four nodes that were removed due to redundancy are also interesting for follow-up analyses: these items were removed because they correlated strongly with one or more other items, indicating that they might also be central to the WBS. That is, redundant items do not contribute unique information to the network with regard to well-being when taking the other items into account, which means that they do carry substantial information common to one or more items in network, and thus might be a target for use with limited resources.

Besides being subjected to the common critiques on network analysis mentioned in the introduction[61,62], these findings should be interpreted in the context of a number of other limitations as well. First, we found that items that belong to the same questionnaire tend to cluster together. While this partly reflects these items successfully capturing a particular phenotype, this likely also reflects participants answering questions belonging to the same measurement instrument more similarly than questions from different instruments as these items were presented with the same response format[84]. The response format (e.g., scale, wording) of the different questionnaires is not always the same, potentially leading to clustering. Second, items in a questionnaire are often designed to load on one certain scale and not on another scale. This is accomplished by means of factor analysis, where highly orthogonal factors are created. This artificial way of designing items could be interfering with the actual underlying correlations between the items of different scales. Third, we were limited by the well-being items that were previously collected in our sample. For example, we did not include Ryff's different scales or items corresponding to Keyes' social well-being domain. This relates to the boundary specification problem, referring to the difficulty of deciding which nodes should be included when estimating a network[62]. Ideally, all possible nodes should be included in a network, but it is impossible to know the boundary of the theoretical network, and so we are bounded by what we have measured. Importantly, this is not unique to network analyses, but also applies to factor analytic studies. Nevertheless, for future research, it would be interesting to estimate an even broader well-being network that includes items based on these different theories.

To conclude, the results described in this study support previous research on the WBS that links different items assessing well-being, depression, neuroticism, and loneliness form two highly interconnected positive and negative clusters[32]. Additionally, we identify four nodes most central to the network: one satisfaction with life item, one neuroticism item, and two depression items. This suggests that to get a general sense of the WBS, these items would serve as the most informative items to evaluate. When comparing the network results to the factor analytic results, we find only partly overlapping conclusions. Nevertheless, taking a network perspective re-affirmed prior research that demonstrates the complex interconnectivity of different well-being (related) phenotypes. While several items definitely cluster with other items within the same construct, the network results also reject the view of clearly delineated well-being domains. To develop a more complete picture of well-being, including hedonic and eudaimonic aspects in a network context, additional studies are needed that include more well-being measures that measure these different aspects of well-being.

CHAPTER 3

A re-evaluation of candidate gene studies for well-being in light of genome-wide evidence

Published as: Van De Weijer, M. P., Pelt, D. H. M., De Vries, L. P., Baselmans, B. M. L., & Bartels, M. (2022). A Re-evaluation of Candidate Gene Studies for Well-Being in Light of Genome-Wide Evidence. *Journal of Happiness Studies*, 1-23.

*supplementary materials accessible at: https://doi.org/10.1007/s10902-022-00538-x

ABSTRACT

Ever since twin-family studies found that a substantial amount (±40%) of the variation in well-being can be explained by genetic variation, several candidate genes have been proposed explaining this variation. However, these candidate gene and candidate gene-by-environment interaction studies have been surrounded by controversy regarding the validity and replication of their results. In the present study, we review the existing candidate gene literature for well-being. First, we perform a systematic literature search that results in the inclusion of 41 studies. After describing the results of the included studies, we evaluated the included candidate polymorphisms by 1) looking up the results for the studied candidate SNPs in a large well-being genome-wide association study, 2) performing association analyses in UK biobank (UKB) data for the candidate variable number tandem repeats (VNTR) and the APOE ε4 allele, and 3) studying possible candidate interactions with positive and negative environmental moderators using UKB data. We find no support for any of the candidate genes or candidate gene-environment interactions for well-being, with the exception of two SNPs that were chosen based on genome-wide evidence. While the generalizability of our findings is limited by our phenotype and environment definitions, we strongly advise well-being researchers to abandon the candidate gene approach in the field of well-being and move toward genome-wide approaches.

1. INTRODUCTION

Ever since it was discovered that well-being or happiness, is a heritable trait[24,25,85], well-being researchers have aspired to find the genetic variants responsible for variation in well-being. In 1996, Lykken and Tellegen reported on the first twin analyses and estimated that 44% to 52% of the variance in well-being is associated with genetic variation[86]. In the same year, Hamer predicted that about 10-20 genomic loci would be involved in explaining the heritability of happiness and he proposed a strategy for finding 'happiness genes' by association analysis using loci chosen on the basis of function[85]. Based on earlier biological findings, this quest started with a focus on these so-called *candidate genes* that were hypothesized to hold some sort of biological function important for the biological correlates of well-being. With this in mind several candidate genes for well-being have been investigated.

To illustrate, a polymorphism that was deemed a candidate gene for influencing well-being was the 5-HTTLPR-Variable Number Tandem Repeat (VNTR, a pattern of one or more nucleotides that is repeated, with the number of repeats varying across individuals). The VNTR is located in the promotor region of the *SLC6A4* gene that codes for serotonin (5-HT) transporters, a neurotransmitter commonly implicated in mood disorders and emotional processing[87,88]. The 5-HTTLPR polymorphism was first studied in relation to well-being in a sample of 2574 Americans[89]. It was found that individuals with the short version of 5-HTTLPR reported higher levels of life satisfaction than individuals with the longer version, a finding that failed replication in a study a year later[90]. Since then, several studies have examined the association between the 5-HTTLPR VNTR and well-being, producing mixed results[91–96].

While the rationale behind most candidate genes seems reasonable, a large problem of the candidate gene literature in general is that results are mixed and do not seem to replicate[97]. On of the proposed reasons for the lack of replication is that, in retrospect, effect sizes of individual genetic variants are very small [98]. Therefore, the samples used in the candidate gene studies in general are too small (ranging from less than a hundred to a couple of thousand individuals), leading to many false positive findings[97,99–101]. Another reason for why candidate gene studies like these were producing mixed effects is that the effects of these genes might

depend on the environment. As a result, candidate gene-environment interaction studies started examining the interactions between the genetic polymorphisms and environmental exposures on well-being (i.e., gene-environment interaction). Many of these hypothesized interactions are based on the "*differential susceptibility hypothesis*". This hypothesis states that individuals who are most vulnerable to adversity/negative environments are also most likely to benefit from supportive/positive environments[102]. Candidate gene-environment interaction studies thus examine if carriers of one or two alleles of a particular gene are more adversely affected by negative environments, or more positively affected by positive environments, than non-carriers. For example, Sheffer-Matan and colleagues found that only individuals with the 5-HTTLPR short allele(s) were happier when they perceived higher social support from their friends[103].

Most of the studies mentioned above focus on a definition of well-being or happiness that is most in line with a person's subjective evaluation of their life and well-being, also referred to as *subjective* well-being. (SWB). Another major well-being definition and line of research is *psychological well-being* (PWB). One of the most influential theories in this respect is Ryff's theory on PWB, which states that PWB is comprised of the six dimensions: autonomy, environmental mastery, personal growth, positive relationships, purpose in life, and self-acceptance[46]. Importantly, many different well-being definitions exist that focus on SWB, PWB, or a combination of both. While it is beyond the scope of this study to provide an extensive overview of these theories, many well formulated reviews exist[4,104,105]. With respect to genetic studies on well-being, the focus has predominantly been on a subjective well-being definition, since genetic studies for wellbeing leverage very large available samples with DNA information that often have not directly been designed for wellbeing research but contain wellbeing assessments anyway.

To more systematically search for genetic variants for complex traits, the so-called Genome-Wide Association (GWA) study approach was introduced[106]. In a GWA study, several millions of single nucleotide polymorphisms (SNPs) are studied in relation to the outcome measure in a hypothesis-free fashion. Using the GWA design, it was quickly discovered that most behavioral/psychological traits are influenced by hundreds to thousands of genetic variants, with most of them carrying tiny effects[107]. As a consequence, to be able to detect these small effects, performing reliable GWA studies requires large sample sizes, often

ranging from a few hundred-thousand to millions of study participants. In the context of well-being, the first genome-wide *hits* were identified in 2016, in a GWA study examining subjective well-being data from almost 300.000 individuals[30]. Since then, two more GWA studies have been performed for well-being, both of them examining well-being in the context of a *well-being spectrum* consisting of the highly genetically correlated traits subjective well-being, depressive symptoms, and neuroticism[108,109]. By jointly analyzing these traits, Turley and colleagues [109] identified 49 genetic variants associated with subjective well-being (*N*=354,462). Baselmans and colleagues also jointly analyzed these traits in a multivariate fashion resulting in 304 hits, and additionally generated trait-specific estimates for each SNP, and identified 148 and 191 significant hits for life satisfaction and positive affect, respectively (N_{obs} = 2,370,390)[108]. These results reflect a linear positive relation between sample size and the number of hits identified, an effect which has also been demonstrated empirically[110].

In light of the results that emerged from GWA studies, several researchers started to re-evaluate previous evidence from candidate gene and candidate gene-interaction studies for different traits. In this way, it was found that data from a large population-based sample did not support previous major candidate genes for depression[97]. This includes the 5-HTTPLR gene, studied >500 times as a candidate gene for depression. Similarly, in a study examining historical candidate genes for schizophrenia in light of results from a large genomic study, no robust evidence was found for the role of the proposed candidate genes[111]. Like the aforementioned studies for depression and schizophrenia, the GWAS findings for well-being allow for a re-evaluation of the role of candidate genes for well-being. For the present study, we scan the existing literature for candidate gene studies on well-being and summarize the outcomes of these studies. Second, we look up the studied SNPs in the most recent large GWA study for well-being. Lastly, we examine potential associations of four frequently studied VNTRs (*SLC6A3, DRD4, SLC6A4* (a.k.a. 5HTTLPR)*,* and *MAOA)* and the APOE ε4 allele, with well-being in a large sample from the UK Biobank. In line with the differential susceptibility hypothesis, we also examine potential interactions with positive and negative environmental moderators. In this way, we re-evaluate the role of these candidate genes to explain differences in well-being. With this information we aim to inform the field on pursuing or abandoning (relative expensive) candidate-gene based research approaches.

2. MATERIALS AND METHODS

2.1 Systematic literature search

Articles were retrieved from *PubMed* (http://www.ncbi.nlm.nih.gov/pubmed) and *Web of Science* (http://apps.webofknowledge.com) through a computerized literature search. A literature search was conducted for studies published up to January 28, 2022. The following search terms were used: "well-being" or "wellbeing" or "well being" or "quality of life" or "satisfaction with life" or "life satisfaction" or "happiness" or "positive affect" or "flourishing" or "meaning in life" or "purpose in life" or "Ryff*" or "PERMA" or "eudai*" or "eudem", and "genes" or "gene" or "genetics" or "polymorphism". Studies were included if they 1) examined association(s) between some measure of (mental) well-being and one or more candidate genes (not GWA studies), 2) were peer-reviewed, 3) published in English, and 4) examined these associations in a non-patient/non-clinical human population. Importantly, we only included studies that aimed to examine well-being as a phenotype, and not well-being-related phenotypes such as depressive symptoms.

2.2 SNP look-up

For our SNP look-up, we used summary statistics from Baselmans et al.[108]. Details on this genome-wide association meta-analysis (GWAMA) can be found in the original paper. Briefly, this study performed multivariate GWAMA for four genetically highly related traits: positive affect, life satisfaction, neuroticism, and depressive symptoms, collectively referred to as the well-being spectrum (*N* observations=2,370,390). The study performed univariate meta-analyses for all traits separately, as well as multivariate analyses where the traits were combined, resulting in 304 significant independent hits. For each candidate gene study identified through our systematic literature search (independent of the outcome of that candidate gene study), we looked up the candidate SNPs in the *N*-weighted GWAMA summary statistics for: 1) life satisfaction, 2) positive affect, and 3) the well-being spectrum composite score. We report the p-values of each of these candidate SNPs in the GWAS summary statistics and compare it to the p-values of the original studies.

2.3 UK Biobank (UKB)

We used data from the UKB to test for potential associations between widely studied VNTRs, APOE ε4, and well-being. The UKB is a UK cohort study with genetic and phenotypic data on approximately 500,000 individuals aged between 40 and 69 years old at recruitment[112]. We included a subset of participants with available well-being data. Well-being was approximated using a happiness question: "In general how happy are you?". This question was answered by 214,357 participants (on four instances) from the initial touchscreen interview (UKB data-field 4526), and by 157,335 participants who completed an online follow-up questionnaire (UKB data-field 20458). If a participant had data available for multiple instances, we selected the last time-point. Participants could answer the question on scale from 1 to 6 ranging from Extremely happy (1) to Extremely unhappy (6). We reverse-coded the item so that a higher score on the scale reflected a higher level of happiness. To limit bias due to population stratification, we reduced our sample to individuals of Caucasian British ancestry (based on self-report, UKB data-field 22006). In total, this led to a sample size of 226,842 individuals with happiness data.

2.4 VNTR Association Analyses

VNTR data are available in UKB for four highly studied candidate VNTRs in psychiatric genetics, located in *SLC6A3, DRD4, SLC6A4* (*5HTTLPR*)*,* and *MAOA*. Additionally, the moderating SNP rs25531 in *SLC6A4* was imputed to the UKB and included in the present study. These VNTRs (and modifying SNP) were imputed previously in the UKB sample using the Family Transitions Project (FTP), the Center for Antisocial Drug Dependence (CADD), and the Genetics of Antisocial Drug Dependence (GADD) studies as reference panels and show good imputation accuracy (>0.96 for all four VNTR variants)[113]. *SLC6A3, DRD4,* and *SLC6A4* were imputed as bi-allelic short/long alleles, while the *MAOA* was imputed as bi-allelic risk/wild-type. It is the largest sample for which these VNTRs are available, and the data has have been used to study potential associations between depression and these candidate VNTRs[97]. We analyzed additive associations between happiness and the four VNTRs imputed to UKB using linear association analysis in plink[114]. Age, sex, genotyping batch, and the first 25 ancestry-informative principal components (PCs) were included as covariates. Since we repeated the analysis six times, once for each VNTR, once for the moderating SNP, and once for APOE, we employed a Bonferroni corrected significance threshold of α=0.05/6 = .008.

2.5 APOE ε4

UKB data was used to test whether the presence of the *APOE ε4* allele was associated with happiness. SNP data for rs429358 and rs7412 was used to determine *APOE* genotypes (*APOE* ε4 present/not present). We tested for association using linear regression models in R, including age, sex, genotyping batch, and the first 25 ancestry-informative principal components (PCs) as covariates.

2.6 Interaction studies

A subset of the articles identified in our systematic literature search examined gene-environment interactions within the differential susceptibility framework[91,94,103,115–118]. These studies were performed for *APOE ε4*, the *MAOA* VNTR, *OXTR*, and the *5-HTTLPR* gene (see results and Online Resource Table 1). We tested for interaction with both positive and negative environmental moderators for the VNTRs and *APOE* genes using UKB data. In line with Border and colleagues[97], we included childhood trauma, adult trauma, and recent trauma as negative environmental moderators. As positive moderators we included frequency of friends/family visits, and ableness to confide. Details on these variables can be found in Online Resource Table 2.

Regression analyses where we tested for interactions between *SLC6A3, DRD4, SLC6A4* (5-HTTLPR)*, MAOA* VNTRs, and the rs25531 SNP in *SLC6A4* and our positive and negative moderators were performed in plink. We tested for interactions between our moderators and the presence/absence of the *APOE ε4* allele in R. Happiness, age, and continuous moderators were standardized prior to the analyses. Age, sex, the first 25 ancestry informative PCs, all covariate-by-polymorphism interaction terms, and all covariate-by-moderator interaction terms were included as covariates [119]. To test for significance, a Bonferroni corrected significance threshold of α=0.05/(6 polymorphisms x 3 moderators=)18 = .003 was used.

3. RESULTS

3.1 Identified literature

A PRISMA flow diagram[120] of our search process is depicted in Figure 1. Of the 11,400 studies identified in our literature search, 41 were included in the current study. Table 1 provides an overview of these studies and the genetic polymorphisms

that were examined. More details on the individual studies can be found in Online Resource Table 3 and 4. Of the 41 included studies, 16 examined the effect of one or more candidate SNPs on a well-being outcome, 1 examined a candidate haplotype polymorphism, 19 examined the effect of a candidate VNTR on a well-being outcome, 3 examined both SNPs and VNTRs, and 2 examined the association between the *APOE ε4* allele and well-being. Some of these studies examined main effects while others also examined interaction effects. The reasons behind studying these genes (and interactions) as candidates in the context of well-being (as stated by the original studies) are listed in Online Resource Table 1.

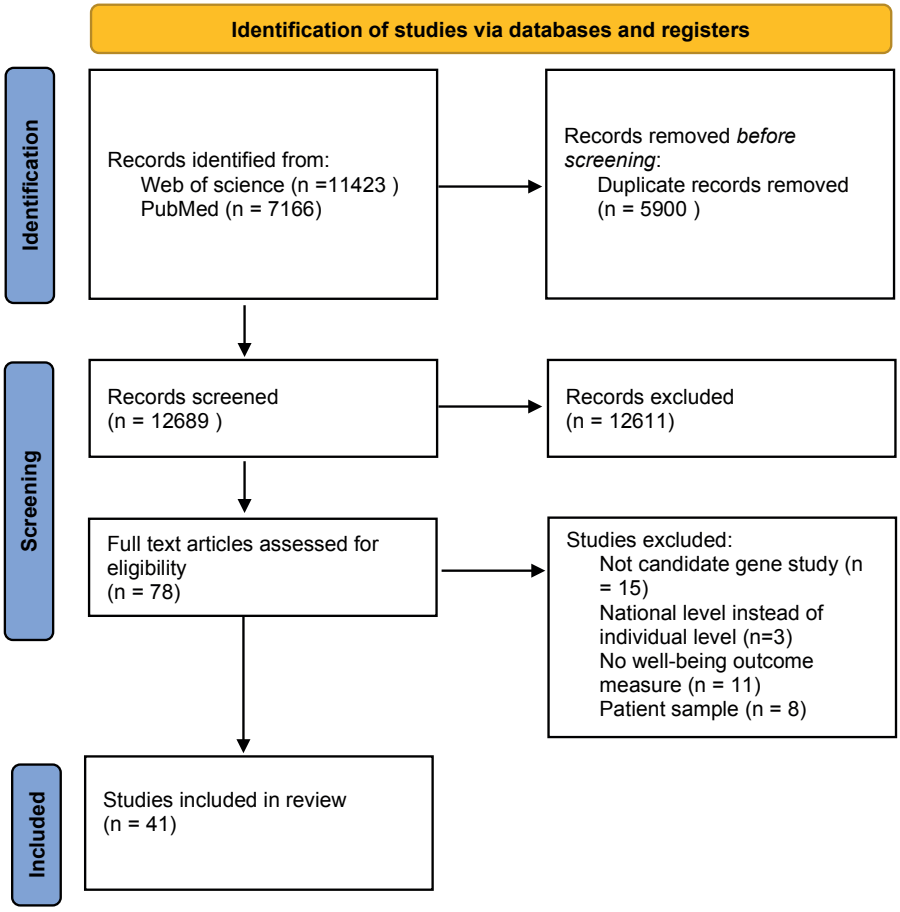


Figure 1. PRISMA flow diagram of the conducted literature search

Table 1

Overview of candidate genes studied in relation to well-being

Gene	Papers	Significant*
VEGF; vascular endothelial growth factor	[121]	no
IGF-1; insulin-like growth factor 1	[121]	no
OXTR; oxytocin receptor	[122]	no
	[123]	yes
	[115]	no
	[124]	no
OXT; oxytocin	[125]	yes
COMT; catechol-O-methyltransferase	[126]	yes
	[127]	yes
	[128]	yes
	[129]	no
RAPGEF6; rap guanine nucleotide exchange factor 6	[130]	no
	[131]	no
AR CAG repeat: androgen receptor	[132]	no
CSE1L; chromosome Segregation 1 Like	[130]	no
NMUR2; neuromedin U Receptor 2	[130]	yes
CNR1; cannabinoid receptor 1	[133]	yes
	[134]	no
HTR2A; the serotonin 2a receptor gene	[135]	no
Y-DNA haplogroup *D-M55*	[136]	yes
CRHR1; corticotropin-releasing hormone receptor 1	[137]	yes
DIO2; iodothyronine Deiodinase 2	[138]	no
KSR; kinase suppressor of RAS	[131]	no
LOC105377703	[131]	no
CYP19A1; cytochrome P450 family 19 subfamily A member 1	[139]	yes
PER3; Period Circadian Regulator 3 (VNTR)	[140]	no
MAOA; monoamine oxidase A (VNTR)	[141]	yes
	[142]	yes
	[103]	no
	[143]	no

Gene	Papers	Significant*
5-HTTLPR; serotonin transporter (VNTR)	89	yes
	90	no
	91	no
	94	no
	95	no
	144	yes
	116	no
	130	no
	92	yes
	145	no
	146	no
	137	no
	129	yes
	96	no
	147	no
	103	no
	117	no
	148	no
	149	no
APOE ε4; e4 allele apolipoprotein E gene	118	yes

3.2 Summary of results from candidate literature

3.2.1 SNPs

An overview of the candidate gene studies that examined associations between well-being and on one or more SNPs is presented in Table 2. Some of these candidate gene studies also included interaction effects, which are discussed in a later section. Candidate gene studies for *VEGF* SNPs, *IGF-1* SNPs, *OXTR* rs2254298, *OXTR* rs2228485, *OXTR* rs2268498, *RAPGEF6* rs3756290, *DIO2* Thr92Ala rs225014, *KSR2* rs7973260, *HTR2A* rs6311, and *LOC105377703* rs4481363 did not find significant associations with well-being. One study found a significant association between *OXTR* rs53576 and well-being in adults[123], but this result was not replicated in adolescents or in other studies. Besides this SNP, six other candidate SNPs were reported to be significantly associated with well-being: *OXT* rs4813625[125], *COMT* rs4680[126–128], *NMUR2* rs4958581[130], *CNR1* rs806377[130], *CRHR1* rs878886[137], and *CYP19A1* rs700518[139] (without replication efforts). One study that is not mentioned in Table 2 is a study that examined the Y-DNA haplogroup DM55, a genetic polymorphism unique to Japan. This study[136] found an association between subjective happiness and DM55,

where mean happiness was higher in females than in non-carrier males, but the differences between females and carrier males, and between carrier and non-carrier males were not significant. Since this haplotype is unique to Japan, we were unable to evaluate this study in light of the GWAS results or UKB dataset.

Table 2

P-values of candidate polymorphisms in GWAS

Gene; polymorphism	Papers	paper p-value	PA GWAS p-value	LS GWAS p-value	3-WBS GWAS p-value
VEGF; rs699947	[121]	.30	.93	.23	.38
VEGF; rs833068		.45	.87	.69	.53
VEGF; rs3024994		.18	.65	.27	.40
VEGF; rs2146323		.49	.45	.50	.58
VEGF; rs3025033		.66	.31	.74	-
VEGF; rs3025035		.07	.31	.52	-
IGF-1; rs2288377		.30	.78	.44	-
IGF-1; rs35767		.50	.84	.64	.23
IGF-1; rs35765		.93	.63	.98	.40
IGF-1; rs7965399		.11	.94	.10	.62
OXTR; rs53576	[122]	.21	.68	.06	.56
	[123] (adults)	**.045**			
	[123] (adolescents)	.99			
	[115]	n.s.			
	[124]	.80			
OXTR; rs2254298	[123] (adults)	.07	.04	.70	.27
	[123] (adolescents)	.44			
	[124]	.98			
OXTR; rs2228485	[123] (adults)	.35	.03	.39	.26
	[123] (adolescents)	.57			
OXTR; rs2268498	[124]	.58	.36	.01	.34
OXT; rs4813625	[125]*	**.02**	.28	.46	-
COMT; Val158Met (rs4680)	[126]	**.01**	.33	.84	.91
	[127]	**.01**			
	[128]	**.02**			
RAPGEF6; rs3756290	[130]	.64	.0002	.04	.10
	[131]	.82			
CSE1L; rs2075677	[130]	.58	**$2.28x10^{-13}$**	.002	**$2.54x10^{-9}$**
NMUR2; rs4958581	[130]	**.01**	.0002	.16	-
CNR1; rs806377	[133]	**<.05**	.66	.29	.98
HTR2A; r6311	[135]	.59	-	.60	-
CRHR1; rs878886	[137]	**.004**	.38	.62	-
DIO2; Thr92Ala; rs225014	[138]	n.s.	.43	.52	.81
KSR2;rs7973260	[131]	0.72	.54	.45	$8.54x10^{-5}$
LOC105377703-rs4481363	[131]	0.94	.03	.02	**$1.28x10^{-13}$**
CYP19A1 Val80; rs700518	[139]*	**<.001**	.34	.55	.56

Note. n.s. = non-significant (unreported) p-value. "-" indicates that the SNP was not examined in the relevant GWAMA. P-values indicated in bold are significant according to the original study.

* multiple measures of well-being were used, we report the most significant one

3.2.2 VNTRs

Across the candidate gene literature identified through our systematic literature search, 3 VNTRs were studied in relation to well-being: the *MAOA* VNTR, the *5-HTTLPR* VNTR and the *PER3* VNTR (see Table 1). For 5-HTTLPR, 4 studies found statistically significant associations with well-being, while 13 studies did not find significant associations (see Table 1). Four studies examined the relation between well-being and the VNTR region in the *MAOA* gene. Gureev and colleagues[142] found an association between this VNTR and subjective well-being in men, while Chen and colleagues[141] found an association between happiness and the *MAOA* VNTR in women, but not men. Sheffer-Matan and colleagues[103] did find a significant interaction between *MAOA* and social support from friends, but did not find a significant main effect for *MAOA* on happiness. Lu and colleagues also did not find a significant main effect for MAOA on subjective well-being[143]. Lastly, Lázár and colleagues[140] examined if there was an association between a VNTR in the PERIOD3 (*PER3*) region and (positive and negative) affect, but did not find a significant effect of genotype on affect.

3.2.3 APOE ε4

Two studies examined associations between the *APOE ε4* allele and well-being. Blazer and colleagues[149] examined associations between the ε4 allele and five parameters of quality of life (including a measure of mental quality of life, measured based on a combination of life satisfaction and depression items) in individuals with good quality of life, but did not find any significant association. Martin and colleagues[118] examined whether centenarians carrying the *APOE ε4* allele scored lower on positive affect than centenarians without the *APOE ε4* allele. They found that carriers scored significantly higher on positive affect than non-carriers.

3.2.4 Interaction studies

Eighteen of the included studies examined interaction effects with candidates genes. Details on the interaction studies can be found in Table 3. Across these 18 studies, 28 interactions were studied for 9 candidate genes: *OXTR* (2 studies), *5-HTTLPR* (10 studies), *MAOA* (1 study), *AR* (1 study), *COMT* (1 study), *APOE ε4* (1 study), *CNR1* (1 study), *HTR2A* (1 study), and *CYP19A1* (1 study).

Twelve of the 28 studied interactions were statistically significant. Eight interactions with *5-HTTLPR* significantly predicted various measures of well-being: two-way interactions with positive parenting[94], sleep quality[95], life events[116], *BDNF*[147], daily

events[117], social support[103], sleep quality[96], and a three-way interaction with early life stress and age[91]. One significant interaction was found for the *COMT* gene: an interaction with age[129]. The remaining three interactions were an interaction between the *MAOA* gene and social support[103], an interaction between *CNR1* and culture[134], and an interaction between *CYP19A1* and gender[139]. Only for the latter interaction (between *CYP19A1* and gender to predict cognitive well-being), both a significant main effect for genotype and a significant interaction effect was found[139].

Table 3

Interactions examined in candidate gene studies

Interaction	Predicted	Paper	Main effect genotype	Interaction
OXTR x positive parenting	Positive affect	115	no	yes
OXTR x prosocial spending	Positive affect	124	no	no
5-HTTLPR x early life stress x age	Evaluative well-being	91	no	yes
5-HTTLPR x early life stress x age	Affective well-being	91	no	no
5-HTTLPR x positive parenting	Positive affect	94	no	yes
5-HTTLPR x sleep quality	Positive affect	95	no	yes
5-HTTLPR x life events	Life satisfaction	116	no	yes
5-HTTLPR x BDNF	Well-being	147	no	yes
5-HTTLPR x daily events	Positive affect	117	no	yes
5-HTTLPR x social support	Happiness	103	no	yes
5-HTTLPR x age	Affective well-being	129	yes	no
5-HTTLPR x sleep quality	Positive affect	96	n.r.	yes
5-HTTLPR x parents' relationship quality	Subjective happiness	148	no	no
5-HTTLPR x parental violence	Subjective happiness	148	no	no
5-HTTLPR x parental attention	Subjective happiness	148	no	no
5-HTTLPR x childhood income	Subjective happiness	148	no	no
MAOA x social support	Happiness	103	no	yes
AR CAG repeat x time	Psychological well-being	132	n.r.	no
COMT x gender	General well-being	128	yes	no
APOE e4 x proximal events	Positive affect	118	yes	no
APOE e4 x distal events	Positive affect	118	yes	no
APOE e4 x engaged lifestyle	Positive affect	118	yes	no
CNR1 x culture (Japan, Canada)	Subjective happiness	134	no	yes
HTR2A x country (Japan, US)	Subjective happiness	135	no	no
CYP19A1 Val80 x gender	Cognitive well-being (life satisfaction)	139	yes	yes
CYP19A1 Val80 x gender	Affective well-being	139	no	no
CYP19A1 Val80 x gender	Psychological well-being (flourishing)	139	no	no
CYP19A1 Val80 x gender	General subjective well-being	139	yes	no

Note. n.r.=not reported

3.3 Evaluation of results from candidate literature

3.3.1 SNP look-up

For all candidate gene studies identified through our literature search that examined individual SNPs, we looked up the relevant SNPs in summary statistics from the GWA meta-analyses for life satisfaction, positive affect, and the well-being spectrum from Baselmans and colleagues[108]. Table 2 lists these SNPs, the p-values in the original studies (rounded to 2 decimals), and the p-values in these GWA studies. When a "-" is presented instead of a p-value, it means the relevant SNP was not present in the GWAS summary statistics. Figures 2-4 depict Manhattan plots for life satisfaction, positive affect, and the well-being spectrum with the candidate SNPs highlighted. Two SNPs were significant at a genome-wide level (p=5x10^{-8}): *CSE1L*- rs2075677 & *LOC105377703*-rs4481363. Importantly, in the candidate gene study where these SNPs were examined[130], the SNPs were selected based on evidence from an earlier genome-wide association study[30]. None of the other SNPs, and thus candidate genes, were significantly associated with life satisfaction, positive affect or the well-being spectrum composite score.

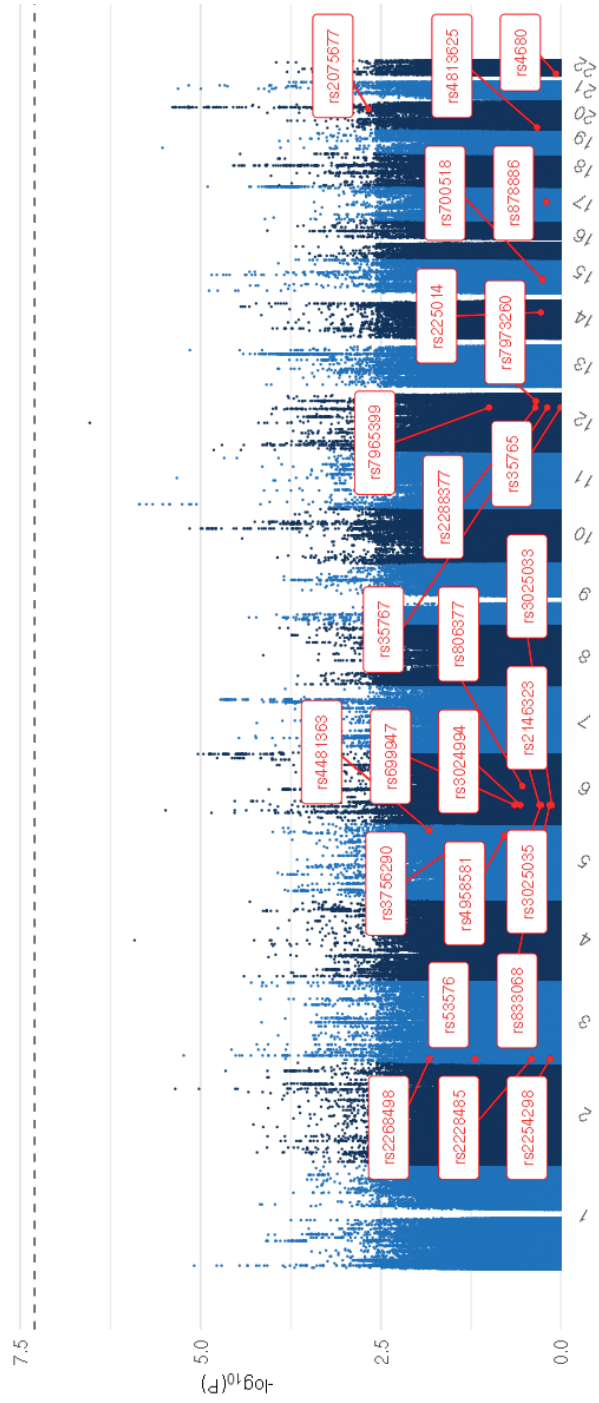


Figure 2. Life satisfaction GWAMA Manhattan plot with highlighted candidate SNPs. The dotted line represents the Bonferroni corrected significance threshold.

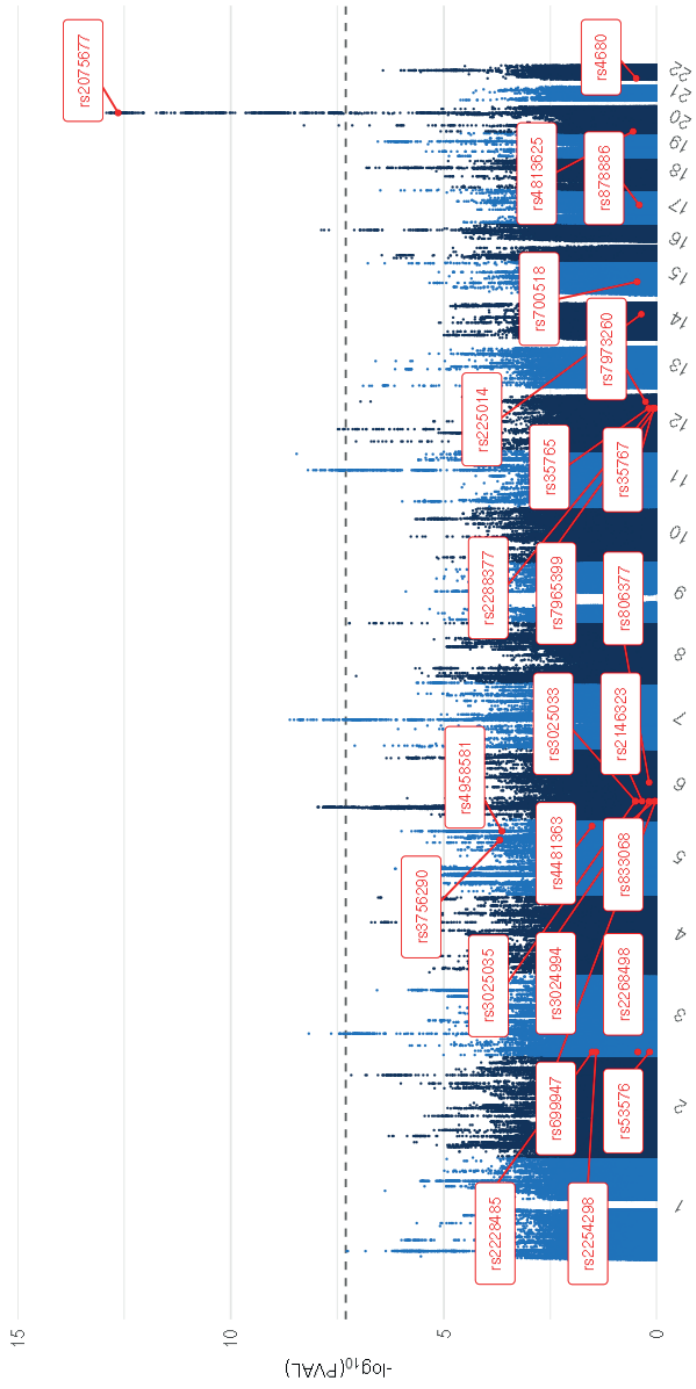


Figure 3. Positive affect GWAMA Manhattan plot with highlighted candidate SNPs. The dotted line represents the Bonferroni corrected significance threshold

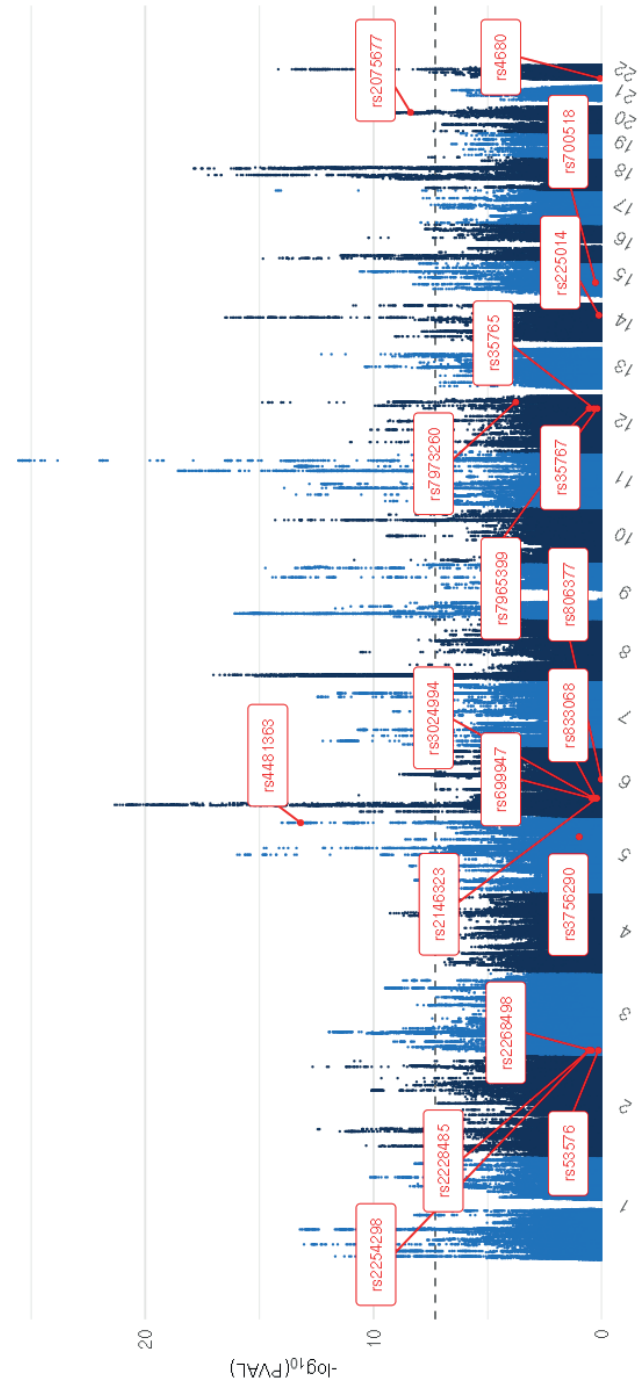


Figure 4. Well-being spectrum GWAMA Manhattan plot with highlighted candidate SNPs. The dotted line represents the Bonferroni corrected significance threshold

3.3.2 VNTR association analyses UKB

Table 4

Results additive association analysis VNTRs UKB

VNTR region	Effect allele	β (SE)	p-value
DAT1	Short allele	.002 (.003)	.33
DRD4	Long allele	.002 (.003)	.56
SLC6A4: rs25531	G	-.002 (.004)	.57
SLC6A4: 5HTTLPR	Short allele	-.002 (.002)	.35
MAOA	Risk allele	-.003 (.002)	.14

Using data from UKB, we analyzed if there was an association between happiness and four commonly studied VNTRs (including the *MAOA* and *5-HTTLPR* VNTR), and a moderating SNP in the *5-HTTLPR* region. Results from our association analysis can be found in Table 4. None of the VNTRs or the moderating SNP were significantly associated with happiness (all $p > .008$).

3.2.3 APOE ε4 association analysis

In the present study, the *APOE* genotype distribution (ε2/ε2: 0.6%, ε2/ε3: 12.4%, ε3/ε3: 58.5%, ε2/ε4: 2.5%, ε3/ε4: 23.6%, and ε4/ε4: 2.3%) was comparable to that of other studies[149,150]. There was no mean difference in happiness between individuals with the *APOE ε4* allele (*M*=4.53, *SD*=.75), and individuals without the *APOE* ε4 allele (*M*=4.53, *SD*=.76) (*t*=-.41, *p*=.685). We did not find a significant association between *APOE ε4* allele presence and happiness (*β*=.0004, *SE*=.003, *p*=.899).

3.3.4 Interaction Analyses UKB

Using UKB data, we tested for interactions of three negative environmental moderators (childhood trauma, adult trauma, and recent trauma) and two positive environmental moderators (frequency of friends/family visits and ableness to confide) with *SLC6A3, DRD4, SLC6A4* (5HTTLPR)*, MAOA* VNTRs, the rs25531 SNP in *SLC6A4*, and the *APOE e4* allele. The results are shown in Table 5. While all of the environmental moderators had a significant main effect on well-being (p-values ranged between 5.48×10^{-309} and 2.28×10^{-15}), none of the polymorphisms or interactions between moderator and polymorphism were significant.

Table 5

Results interaction analyses

		Main effect moderator		Main effect polymorphism		Interaction	
Polymorphism	**Moderator**	**β (SE)**	**p**	**β (SE)**	**p**	**β (SE)**	**p**
DAT1	Childhood Trauma	-.316 (.02)	**$2.94x10^{-43}$**	.001 (.02)	.955	.017 (.01)	.112
	Adult Trauma	-.272 (.02)	**$8.76x10^{-42}$**	.001 (.02)	.935	.004 (.01)	.672
	Recent Trauma	-.273 (.03)	**$5.01x10^{-17}$**	.001 (.02)	.960	.014 (.02)	.371
	Family/Friend visits	.064 (.01)	**$1.97x10^{-19}$**	.005 (.01)	.689	-.003 (.003)	.438
	Able to confide	.261 (.01)	**$2.33x10^{-291}$**	-.011 (.01)	.301	.001 (.003)	.680
DRD4	Childhood Trauma	-.307 (.02)	**$1.42x10^{-41}$**	.010 (.02)	.537	.001 (.01)	.942
	Adult Trauma	-.266 (.02)	**$1.35x10^{-40}$**	.013 (.02)	.445	-.011 (.01)	.277
	Recent Trauma	-.261 (.03)	**$9.71x10^{-16}$**	.004 (.02)	.827	-.015 (.02)	.364
	Family/Friend visits	.060 (.01)	**$1.30x10^{-17}$**	-.009 (.01)	.464	.005 (.004)	.152
	Able to confide	.259 (.007)	**$8.21x10^{-289}$**	-.009 (.01)	.444	.007 (.004)	.060
SLC6A4:	Childhood Trauma	-.310 (.02)	**$6.26x10^{-44}$**	.003 (.03)	.900	.024 (.02)	.184
rs25531	Adult Trauma	-.272 (.02)	**$1.30x10^{-43}$**	-.003 (.03)	.911	.009 (.02)	.541
	Recent Trauma	-.267 (.03)	**$4.98x10^{-17}$**	.008 (.03)	.778	.003 (.03)	.924
	Family/Friend visits	.062 (.007)	**$4.98x10^{-19}$**	-.004 (.02)	.849	.007 (.006)	.236
	Able to confide	.263 (.007)	**$5.84x10^{-309}$**	-.011 (.02)	.550	-.007 (.006)	.241
SLC6A4:	Childhood Trauma	-.302 (.024)	**$3.76x10^{-37}$**	.002 (.01)	.866	-.005 (.01)	.594
5HTTLPR	Adult Trauma	-.266 (.02)	**$1.26x10^{-37}$**	.007 (.01)	.645	-.005 (.01)	.552
	Recent Trauma	-.270 (.03)	**$2.28x10^{-15}$**	.002 (.01)	.875	.003 (.01)	.810
	Family/Friend visits	.065 (.007)	**$1.33x10^{-18}$**	.006 (.010)	.531	-.003 (.003)	.372
	Able to confide	.263 (.007)	**$5.03x10^{-274}$**	.005 (.010)	.585	-.001 (.003)	.646
MAOA	Childhood Trauma	-.312 (.02)	**$3.56x10^{-42}$**	-.010 (.01)	.404	.007 (.008)	.390
	Adult Trauma	-.267 (.02)	**$1.94x10^{-40}$**	-.010 (.01)	.691	-.004 (.007)	.560
	Recent Trauma	-.268 (.03)	**$3.81x10^{-16}$**	-.011 (.01)	.365	.001 (.01)	.968
	Family/Friend visits	.061 (.007)	**$2.09x10^{-17}$**	-.004 (.009)	.664	.003 (.003)	.230
	Able to confide	.263 (.007)	**$1.09x10^{-294}$**	-.0004 (.01)	.960	-.002 (.003)	.466
APOE e4	Childhood Trauma	-.309 (.02)	**$<2x10^{-16}$**	.027 (.02)	.217	.013 (.02)	.385
	Adult Trauma	-.271 (.02)	**$<2x10^{-16}$**	.029 (.02)	.201	.002 (.01)	.880
	Recent Trauma	-.267 (.03)	**$<2x10^{-16}$**	.029 (.02)	.180	.010 (.02)	.647
	Family/Friend visits	.061 (.007)	**$<2x10^{-16}$**	.004 (.02)	.791	.005 (.005)	.267
	Able to confide	.255 (.007)	**$<2x10^{-16}$**	.004 (.02)	.791	-.003 (.005)	.457

Note. Values in bold are significant.

4. DISCUSSION

This study set out by reviewing the candidate gene literature for well-being. To this end, we performed 1) a systematic literature search to identify all the well-being candidate gene literature, 2) a look-up of the studied genomic locations in the largest well-being GWA study, 3) association analyses for commonly studied VNTRs and *APOE* with well-being in UKB data, and 4) association analyses of interactions between negative and positive environmental moderators and the VNTRs and *APOE* in relation to well-being.

In total, 41 studies were included in the present review. Nineteen of these studies examined candidate SNPs in relation to well-being. With sample sizes ranging from less to a hundred to a few thousand, the results from these studies were mixed. Additionally, 20 studies examined potential associations between different VNTRs (*5-HTTLPR*, *MAOA* & *PER3*) and well-being, also producing mixed results. A look up of these SNPs in the GWAS by Baselmans and colleagues[108] revealed no significant associations with life satisfaction, positive affect, or a 3-trait well-being spectrum, with the exception of 2 SNPs across 2 candidate gene studies. In these 2 candidate studies, these SNPs were not significant, but were selected because they were significant in an earlier GWA study [30]. Next, our own association analyses between 5 commonly studied VNTRs (including *5-HTTLPR* & *MAOA*) in over 200,000 individuals of the UKB did not result in significant results. While we were not able to study the association between *PER3* and well-being, this gene was not significantly associated with well-being in the original candidate gene study[140]. Similarly, we failed to identify a significant association between the *APOE ε4* allele and well-being in our UKB analyses. Lastly, 18 of the included studies examined the potential effects of interactions between environmental moderators and genetic polymorphisms. Most often, these studies are based on the differential susceptibility hypothesis stating that individuals who are most vulnerable to adversity/negative environments are also most likely to benefit from supportive/positive environments[102]. To this end, we examined interactions between three negative and two positive environmental moderators and the included VNTRs and the *APOE ε4* allele. None of the interactions significantly predicted well-being in the UKB sample.

Taken together, these results indicate that the candidate gene approach is largely unsuitable for studying both main genotypic effects and gene-environment

interactions in the context of a polygenic complex trait like well-being. While well-being is a heritable trait and many genetic polymorphisms have been associated with well-being in a genome-wide context, individual genetic effects are extremely small, meaning that extremely large sample sizes are required to detect them. This is even more so the case for interaction effects, which are harder to detect than main effects, increasing the required sample size even further[151]. Most candidate gene studies up until now have employed sample sizes too small to detect these effects, ranging from less than 100 to a couple of 1000 individuals (Online Resource Table 3) (average *N*=774). In a study by Okbay and Rietveld[152], Bayesian power analyses indicated that in a scenario of an expected effect size of R^2 =.01 (which is much larger than we would expect for a single variant for well-being) and a sample size of *N*=1000 (and a prior belief in the association of 1%), the power of the test is only 17%. Moreover, the posterior belief in a significant association was still only 3%.

Moreover, the genetics and biology of well-being are too complex to easily form hypotheses on potentially relevant genetic polymorphisms, leading to a lack of support for popular hypotheses such as the *5-HTTLPR* hypothesis. We therefore strongly encourage researchers in the well-being field interested in genetic (and gene-environment) effects to abandon the candidate gene approach and to take on the hypothesis-free GWA approach or use the summary statistics for follow-up analyses.

These summary statistics can be used to calculate so called polygenic scores (PGS): quantitative measures that summarize the estimated effect of many genetic variants on an individual's phenotype, typically calculated as a weighted sum of trait-associated alleles. For example, using summary statistics from the same well-being GWAS as used in this study, Jamshidi and colleagues created PGS to predict different (subjective and psychological) well-being measures[153]. While they found an indication for differences in predictive power across different measurement instruments, none of these differences were statistically significant. Moreover, Patel and colleagues used a PGS for well-being, based on GWAS summary statistics from Turley et al.[109] to study the association between subjective well-being and self-employment. They found that the genetic predisposition for well-being (in the form of this PGS) is positively associated with the likelihood of self-employment and earnings. By using a genetic instrument to examine the consequences of

well-being on self-employment, the study extends existing literature that mainly focused on potential benefits of self-employment for well-being[154]. Furthermore, the summary statistics of these large GWAS studies can be used to study direction of causation in a Mendelian Randomization framework. For example, using this approach de Vries and colleagues[155] report causal relations from well-being to resilience, and Zhou and colleagues[156] report bidirectional causal associations of insomnia with depressive symptoms and subjective well-being.

Our findings are prone to several limitations. First, we relied on a broad definition of well-being that was not limited to one specific well-being construct. We included candidate gene studies that used various measures of both psychological and subjective well-being. However, almost all included studies used a subjective well-being outcome measure for their analyses. However, effect sizes for psychological well-being (in the form of meaning in life) for the only GWAS on this topic show effect sizes in the same range as for subjective well-being[157]. Therefore, we do not expect large effects for individual genetic variants for psychological well-being, leading to the same complication for candidate gene studies on this definition of well-being.

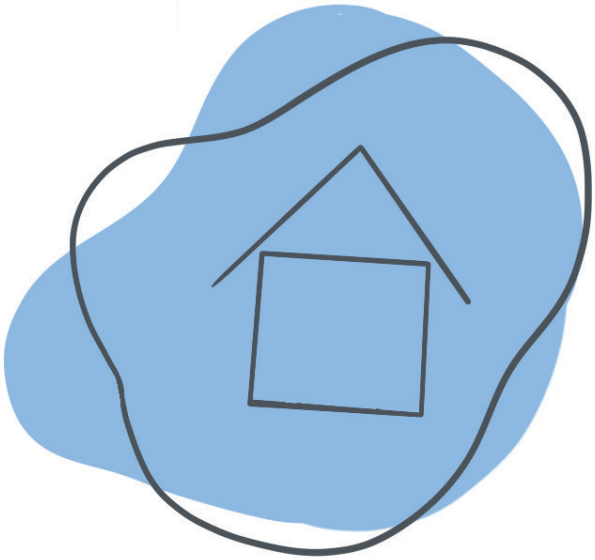
For our SNP look up, we examined results from the Baselmans et al. GWAS[108], including results for positive affect (including happiness measures), life satisfaction, and the well-being spectrum. For our VNTR/*APOE* analyses, we used a UKB measure of happiness. Since our own well-being definitions were not always the same as the constructs used in the different candidate gene studies, we assume that the genetic architecture of different well-being constructs is largely similar, which is confirmed in earlier work reporting high genetic correlations between measures of subjective and psychological well-being[82,157]. Second, while we included different positive and negative environmental moderators in our interaction analyses to test the differential susceptibility hypothesis, they are not identical to the measures used in the included candidate gene-environment studies. It may be the case that we would have found different results if we included different environmental moderators, but given the extremely small effect sizes of significant SNPs, and the abundance of literature showing no evidence for candidate gene-by-environment interactions[101,158], we believe it is unlikely that strong gene-environment effects can be found for individual SNPs. Additionally, the GWA results were based on individuals from European ancestry and the VNTR/APOE analyses were performed

on UK participants. There are currently no large-scale genome-wide studies on the genetics of well-being in non-Caucasian individuals, limiting our ability to draw conclusions on those populations. For two unrelated phenotypes, height and BMI, a substantial genetic correlation was found between European and non-European samples[159]. While this does not necessarily generalize to well-being, it does give a first indication that a proportion of GWAS findings in Europeans are likely applicable to non-Europeans.

While the generalizability of our findings is limited by our phenotype and environment definitions, the strength of this study is that the analyses were performed in a much larger sample than those of the included candidate gene studies. In order to continue the progress made in the area of well-being genetics, we advise to abandon the candidate gene approach and move toward well-powered genome-wide approaches, in line with conclusions from earlier work reviewing candidate gene studies for other phenotypes[97,100]. In the context of gene-environment research, it is unlikely that any individual SNP or gene will have a strong interaction effect with an environmental moderator. Instead of focusing on specific candidate SNPs or candidate genes in gene-environment research, an alternative is to look at the joint effect of many well-being associated SNPs, for instance in the form of polygenic scores. These scores are based on GWA summary statistics and reflect an individual's genetic propensity for a trait of interest. In this way, we might be able to investigate whether the effect of environmental factors is different for people with a different genetic susceptibly – measured across the whole genome rather than a single SNP – for well-being. Moving toward these data-driven approaches will allow us to not only learn more about the biology and genetics of well-being, but will also help us to better understand individual differences in both well-being itself and differences in how people are impacted by environmental factors.

PART II

The Well-Being Exposome



CHAPTER 4

Expanding the environmental scope: An environment-wide association study for mental well-being

Published as: Van De Weijer, M. P., Baselmans, B. M. L., Hottenga, J. J., Dolan, C. V., Willemsen, G., & Bartels, M. (2022). Expanding the environmental scope: an environment-wide association study for mental well-being. *Journal of exposure science & environmental epidemiology,* 32(2), 195-204

*supplementary materials accessible at: https://doi.org/10.1038/s41370-021-00346-0

ABSTRACT

Identifying modifiable factors associated with well-being is of increased interest for public policy guidance. Developments in record linkage make it possible to identify what contributes to well-being from a myriad of factors. To this end, we link two large-scale data resources; the Geoscience and Health Cohort Consortium, a collection of geo-data, and the Netherlands Twin Register, which holds population-based well-being data. We perform an Environment-Wide Association Study (EnWAS), where we examine 139 neighbourhood-level environmental exposures in relation to well-being. First, we performed a generalized estimation equation regression (*N* = 11,975) to test for the effects of environmental exposures on well-being. Second, to account for multicollinearity amongst exposures, we performed principal component regression. Finally, using a genetically informative design, we examined whether environmental exposure is driven by genetic predisposition for well-being. We identified 21 environmental factors that were associated with well-being in the domains: housing stock, income, core neighbourhood characteristics, livability, and socioeconomic status. Of these associations, socioeconomic status and safety are indicated as the most important factors to explain differences in well-being. No evidence of gene-environment correlation was found. These observed associations, especially neighbourhood safety, could be informative for policy makers and provide public policy guidance to improve well-being. Our results show that linking databases is a fruitful exercise to identify determinants of mental health that would remain unknown by a more unilateral approach.

1. INTRODUCTION

Demographic factors are widely recognized as important for people's functioning and mental health. For example, urbanization, i.e. the movement of population from rural to more urbanized areas, is accompanied by both beneficial and detrimental effects on mental health. Urbanization is often associated with economic growth and prosperity[160,161], and comes with better infrastructure and better access to health care services[162]. Mental disorders, though, are more prevalent in more urbanized areas[163,164] for example due to less access to green space[165], increased social stress[166], and less (perceived) neighbourhood safety[167,168]. Moreover, genetic factors influence where people prefer to live and how their environment impacts them. For instance, research into urbanization and schizophrenia showed that individuals with increased genetic predisposition for schizophrenia tend to live in more urbanized areas. While it was previously assumed that the higher schizophrenia prevalence was explained by increased environmental stress in urbanized areas, this study revealed that part of why schizophrenia is more prevalent in cities is because of an increased genetic predisposition[169].

Recent developments in data sharing and linkage are transforming the way we approach mental health topics and its possible correlates. One of the developments that makes it possible to identify what contributes to mental health and human functioning from a myriad of factors is record linkage. By linking large data resources that contain different types of information, novel, otherwise invisible patterns can be uncovered. A well-known example in is the UK Biobank[170]. By linking genetic (and biological, phenotypic) data to existing health records, great advances have been made in identifying risk factors for disorders such as schizophrenia and depression[171,172]. Record linkage is becoming increasingly accessible for researchers across different disciplines and countries. For example, in the Netherlands, data on households, job benefits, education, crime, and more is available on a population-based scale[173]. This type of data can, under certain conditions and strict privacy regulations, be linked to patient data[174], environmental data[175], and other data resources in the country[63,176].

In this paper we illustrate the potential of record linkage to better understand complex human traits to inform prevention, intervention, and policy by investigating environmental factors that potentially influence well-being. Most

existing research on environmental effects for well-being to date follows a pick and choose approach[34,177,178], which could result in selective reporting or overestimation of effects. To overcome these limitations we propose a data-driven design, an *Environment-Wide Association Study*[179] (EnWAS). This approach is based on the principles of a genome-wide association study (GWAS), where each genetic marker in the genome is systematically tested for association with the phenotype of interest. Instead of genetic markers, EnWAS systematically associates environmental variables while reducing the chance of spurious findings by accounting for multiple testing. This data-driven approach is of particular interest given the lack of theoretical inclusion models and was recently successfully applied to examine behavioural patterns, psychosocial factors, mental and physical health conditions, access to and utilization of health care, and anthropometrics with physical, mental and social well-being[39]. From a broad range of psychosocial factors, 3 factors were associated with well-being: depressive symptoms, life satisfaction, and happiness. While this study provides us with valuable information of psychosocial associations with well-being, it did not explore physical environmental factors such as neighbourhood characteristics, in relation to well-being. Given that many governmental decisions and prevention and intervention policies are enrolled at a neighbourhood level it is very important to get an indication of the effect of neighbourhood-level characteristics on person-level well- being.

In order to examine environmental variables associated with well-being, we applied EnWAS by linking well-being data from the population based Netherlands Twin Register (NTR)[63] to environmental data from the Geoscience and Health Cohort Consortium (GECCO)[175]. We examine 139 environmental variables that cover most aspects of people's living environments e.g. land use in terms of build area or green space, and neighbourhood characteristics, such as safety and livability. In addition, given that it is widely accepted that people do not randomly choose where they live[180,181] , that differences in well-being are partly accounted for by genetic differences[24,30], and to overcome possible genetic confounding, we use a genetically informative design. With this design we examine whether environmental exposure is driven by genetic predisposition for well-being. By combining exposome, phenome, and genome data, we aim to extend the limits of traditional approaches to get more comprehensive insight in how well-being can be placed in a broader context[182].

2. MATERIALS AND METHODS

2.1 Sample

This study used well-being data from the Adult sample of the NTR[63,183]. For the current project, we made use of data collected in the 6th wave of data collection (2002/2003), and the 8th wave of data collection (2009/2010). These waves were chosen based on the fact that we collected satisfaction with life data at both these time-points. This resulted in a dataset of 9951 individuals for 2002/2003 and 11,975 individuals for 2009/2010. Sample characteristics can be found in Table 1. Depending on the missing-ness of environmental data per GECCO dataset, the number of individuals per analysis varies slightly across analyses.

Table 1

Sample characteristics

Sample	$n_{individuals}$	n_{males}/ $n_{females}$	Mean age (range)	Mean SWL (SD)
2002/2003 Full Sample	9951	4158/5790*	39.4 (16-85)	26.6(5.26)
2009/2010 Full Sample	11,975	4608/7363*	45.8(16-97)	27.3(5.18)
Polygenic score sample	7527	2602/4926	41.7 (16-90)	27.5(5.18)
2002/2003 PC regression	5655	2603/3052	44.0 (16-85)	26.6(5.22)
2009/2010 PC regression	4922	1702/3219**	48.5(16-97)	27.33(5.25)

Note. SWL= satisfaction with life.

* age was unknown for 2 individuals

** age was unknown for 1 individual

2.2 Well-being data

To quantify well-being, we used the satisfaction with life (SWL) scale[12]. The SWL scale consists of five items measuring satisfaction with life. Each item required a judgment of a given statement pertaining to SWL on a response scale ranging from 1 (strongly disagree) to 7 (strongly agree), summed to create individual sum scores ranging from 7 to 35.

2.3 Environmental exposure data

Environmental data were extracted from the Geoscience and Health Cohort Consortium (GECCO)[175] database. The GECCO database is a centralized collection of longitudinal geo-data on different geospatial levels in the Netherlands. As the

GECCO data were collected in different time frames, we matched the GECCO data to the NTR data available in 2002/2003 and 2009/2010. In total, 1330 postal-code level variables are available within 34 predefined domains in the GECCO database (see Supplementary Tables S1 and S2). The data sources from which GECCO extracted the variables are frequently used government/census data resources, more information on which can be found in Supplementary Table S2. The environmental exposure data that are available in the GECCO database encompass a wide range of environmental domains, including social, physical and demographic variables. We selected variables representative of neighbourhood characteristics, regardless of which environmental domain they encompass. We pre-selected 168 variables based on two criteria: 1) availability at the same time-points as the NTR well-being assessment, and 2) we chose the most representative variables per domain to prevent inclusion of duplicate variables/ variables that were, without inspection of the data, expected not to vary across the Netherlands. Supplementary table S3 provides an overview of these preregistered variables. Ultimately, quality control led to the inclusion of 133 variables grouped in 22 domains (see Supplementary Table S4). Importantly, some GECCO variables were assessed in both 2002/2003 and 2009/2010, and some exclusively at one of the time-points. More specifically, 80 variables were measured exclusively in 2002/2003, 23 variables were measured exclusively in 2009/2010, and 15 variables were measured on both occasions. Four-digit postal codes were used to link the environmental data to individual level well-being. Figure 1 and Table 2 describe the included domains, and Supplementary Table S5 provides descriptive statistics on these exposure variables.

Table 2

Overview of the included environmental domains

Domains	Description
Accessibility	Data on accessibility include the total number of persons and jobs that are accessible within 15, 30, and 45 minutes over the road and on the rail.
Air pollution	Residential exposure to air pollutants was assessed as annual average concentrations of particulate matter with diameters of less than 2.5μm, and between 2.5μm and 10.0μm , PM2.5 absorbance, and annual average concentrations of nitrogen oxides.
Cinema's and movie theatres	Data on the number of cinema's and movie theatres
Facilities care	Data on the number of care-related facilities (e.g. hospitals, care homes)
Facilities culture	Data on the number of cinema's, museums and theatres.
Facilities education	Data on the number of schools/educational locations and students stratified for level of education (see Figure S7 for more information on the Dutch educational system)
Facilities retail outlets	Data on the number of retail outlets
Facilities sport	Data on the number of a variety of sport facilities
Housing benefits	Data on housing benefit receivers, accounting rent, accounting income, the height/sum of housing benefits
Housing stock	Data on number/percentages of houses in the owner occupied sector, and (private and social) rental sector
Income	Data on disposable income per person and household
Core neighbourhood characteristics	Data on core neighbourhood characteristics, e.g. urbanization and mean house value
Land use	Data on number of hectares that are related to specific land use (e.g. traffic, residential area)
Livability	Livability is measured using the "*leefbaarometer"* (LBM total score). Livability is defined as the extent to which the living environment is in line with the conditions and needs of residents. The LBM total score is based on six dimensions. These dimensions are: 1) population composition, 2) social cohesion, 3) public space, 4) safety, 5) level of resources, and 6) housing.
Museums, music theatres and pop podia	Data on number of museums, music theatres and pop podia.
Offices, retail and businesses	Data on number purchased and rented offices, retail and businesses. Data on the area of these buildings and related rental/sale costs are also available.
Primary education	Data on the number of primary schools and the number/ percentages of pupils at these schools (Figure S5).
Secondary education	Data on the number of schools with secondary education and the number/percentages of pupils at these schools (Figure S5)
Socio-economic status scores	Data on socio-economic status scores based on education, income and position in the labour market.
Special education	Data on the number of schools with special education and the number/ percentages of pupils at primary and secondary special schools.
Sport associations	Data on the number of hockey-clubs, baseball clubs, korfball clubs, tennis clubs, rugby clubs, and football clubs.
Transactions and average house prices	Data on the number of transactions, stratified for type of houses. In addition, the data-set includes data on the average house price.

Figure 1. Overview of domains.

2.4 Genetic data

Genotyping was performed on different SNP micro-arrays that were cross-platform imputed using the Genome of the Netherlands (GoNL) reference set[184]. Quality control procedures are described in the Supplementary Methods. Principal component analysis (PCA) was performed to create genomic PCs reflecting ancestry and genotyping batch effects (for details see[181]. In total, genetic data and well-being scores were available for 7527 individuals (see Table 1).

2.5 Analyses

This project was pre-registered at the open science framework (OSF) (https://osf.io/xehkc). Non-pre-registered follow-up analyses are indicated as such throughout the paper.

Pre-registered

2.5.1 Regression analyses

We pre-registered multilevel models to account for potential within-postal code well-being similarity of participants. Supplementary Table S6 summarizes the number of participants per postal code. However, after accessing the data, the intra-class correlation (ICC) for well-being showed that the dependency of the observations within postal code is neglible (.02 for 2002/2003 and .002 for 2009/2010). Therefore, we proceeded our analyses with GEE models, instead of multilevel models. GEE corrects for correlated observations, allowing us to include the full sample (instead of only genetically unrelated individuals). Regression analyses were performed for each environmental predictor, with sex, age, and age-squared as covariates. Familial relatedness was accounted for using an exchangeable conditional covariance matrix based on sandwich-corrected standard errors[185], as implemented in the GEE package in R. Statistical significance was assessed using a Bonferroni-corrected significance threshold of 3.6×10^{-4} (0.05/139). Power to detect associations with different potential effect sizes can be found in the pre-registration.

2.5.2 Polygenic risk score analysis

To assess the role of genetic factors in the associations obtained in the GEE analyses, we performed polygenic score (PGS) prediction analyses. A PGS reflects an individual's genetic liability for a trait of interest, calculated from the effect sizes from GWA summary statistics. The PGSs were computed for the well-being spectrum in NTR participants using the GWA summary statistics (recomputed excluding NTR) from Baselmans et al.[108]. The summary statistics were recomputed using LDpred[186]. These recomputed summary statistics were turned into PGSs using allelic scoring function in PLINK[114]. This function aggregates the number of effect alleles weighted by their effect estimates in each individual to create scores reflecting an individual's genetic liability for a trait. GEE was used to test the association of the well-being spectrum PGSs (independent variable) with significant environmental correlates (dependent variables) from the EnWAS. Age, age-squared, sex, and the first ten genomic PCs were included as covariates.

Additionally, we used the well-being spectrum PGSs to split the sample into septiles to evaluate the potential of stratifying individuals based on a PGS for well-being. The first septile contains participants with the lowest genetic susceptibility for well-being, and the seventh septile contains those with the highest. We calculated the mean well-being and environmental value per septile and compared whether these means differed significantly by examining overlap in confidence intervals.

Non pre-registered

2.5.3 Multicollinearity follow-up

In the univariate analyses the covariates were considered one at a time, thus ignoring the possible correlation between these variables. To illustrate the overlap between the different variables that significantly predict well-being, we visualized the correlations in chord diagrams using the circlize package in R[187]. We plotted the associations separately for the variables from 2002/2003 and 2009/2010, and made separate plots for: a) correlations stronger than .8, and c) correlations stronger than .4.

Next, to accommodate the relative strong correlations between the environmental factors (see Supplementary table S7), we ran a principal component analysis (PCA) of the standardized environmental exposures using the prcomp function from the

stats package in R. We aimed to extract independent principal components (PCs) that explained at least 90% of the environmental data. Next, these uncorrelated PCs were used as independent predictors to predict well-being in an unrelated sample (after the effects of age, age^2, and sex were regressed out). Based on this analysis, we examined how much variance in well-being can be explained by the combined environmental factors.

2.5.5. Socioeconomic status correction

In the exploratory, data-driven approach of our initial pre-registered analyses, we did not correct for socioeconomic status (SES). However, outcomes of the GEE and the principal component analyses suggested a potential role of SES in the associations. Therefore, as none-preregistered follow-up, we repeated the GEE analyses while correcting for SES using two strategies: (1) including the individual's educational attainment to approximate individual SES, and (2) including the GECCO variable "status score of the neighbourhood" as a measurement of neighbourhood SES (see Supplementary Methods for more information).

3. RESULTS

3.1. Regression analyses

In the GEE analyses, 21 of the 139 environmental variables passed the Bonferroni-corrected threshold and thus were found to be associated with well-being (Figure 2-3, and Table 3). These variables were included in the domains: housing stock, income, core neighbourhood characteristics, livability, and SES scores. An overview of all associations can be found in Supplementary Table S8.

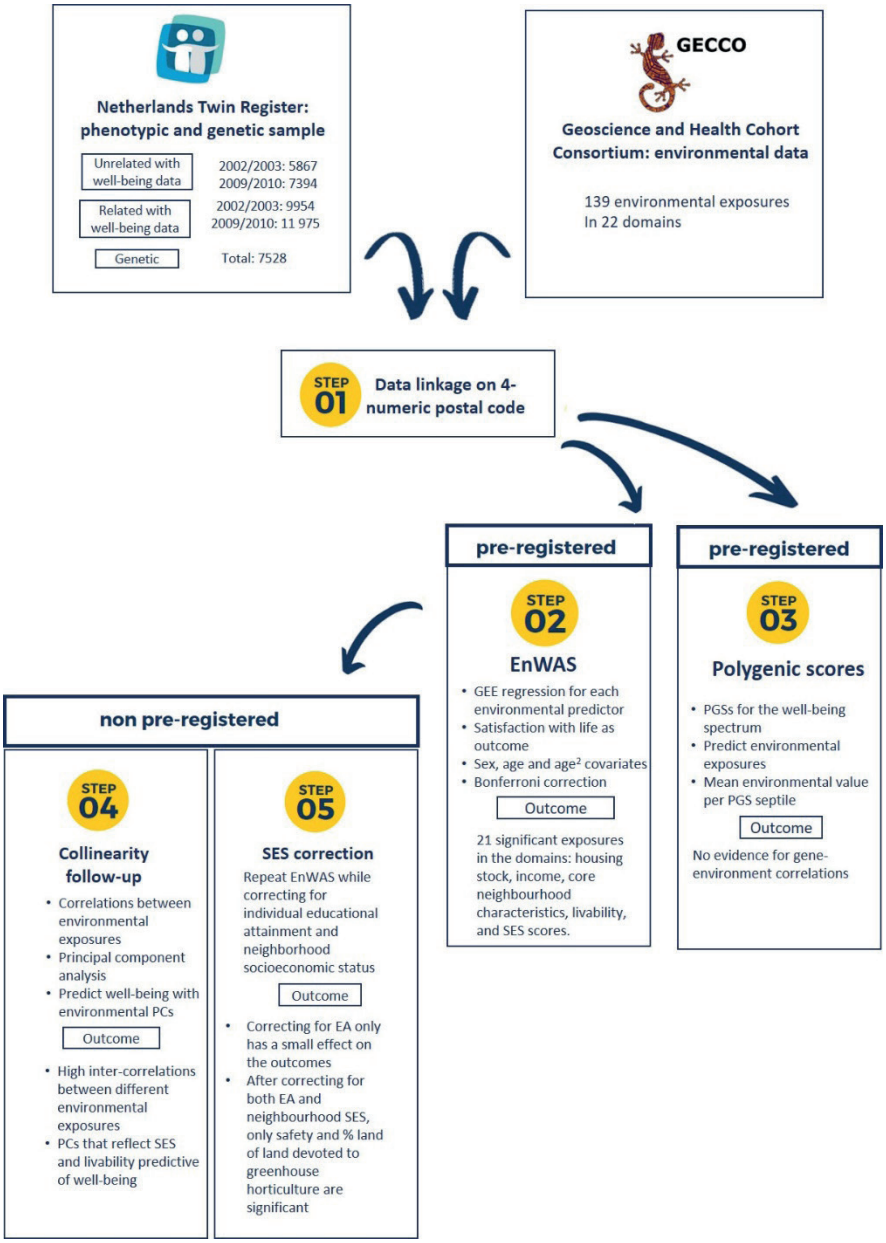


Figure 2. Overview of analyses and results.

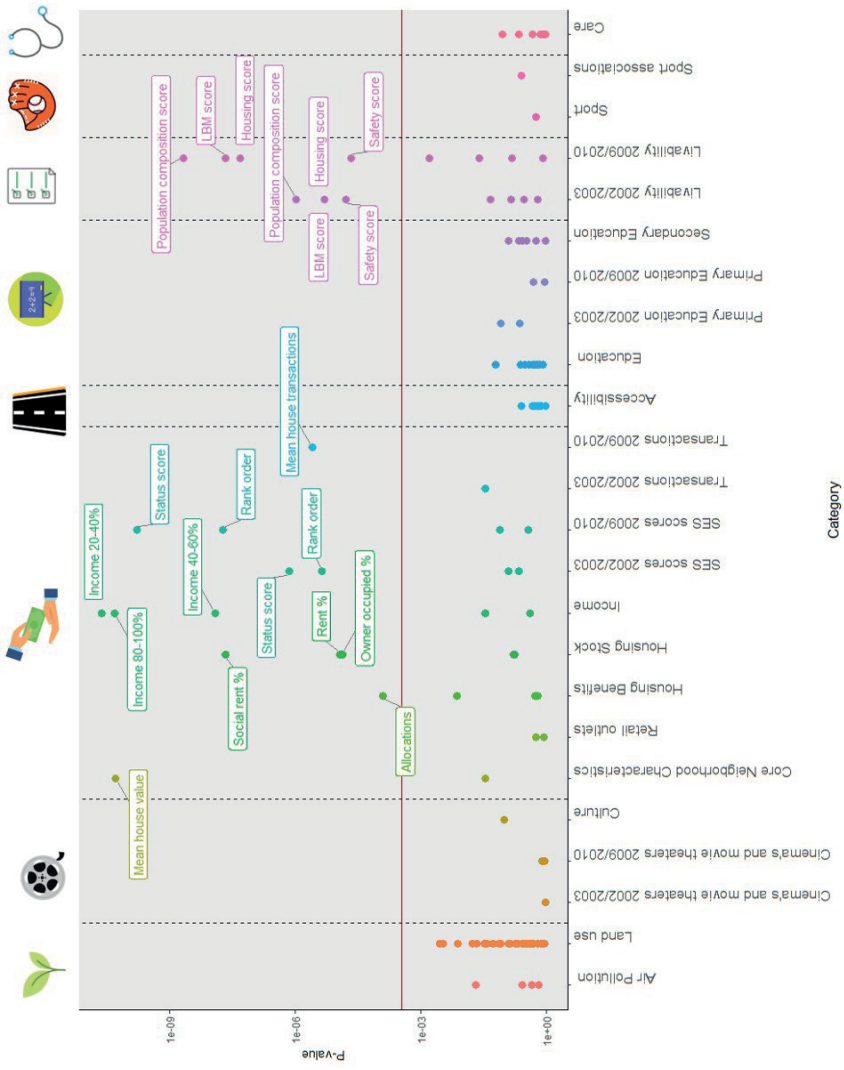


Figure 3. Manhattan plot showing P-values for associations between environmental variables within different categories and well-being.

Table 3

Significant associations with well-being from the generalized estimation equation (GEE) analyses

Domain	Variable	*β* (*SE*) GEE	*P*-value GEE	R^2 GEE
Housing Benefits	Housing benefits (allocations)	-0.045 (.01)	$1.2x10^{-4}$	.002
Housing Stock	Social rental sector %	-0.066 (.01)	$2.14x10^{-8}$	.004
Housing Stock	Rental sector %	-0.051 (.01)	$1.15x10^{-5}$	.003
Housing Stock	Owner occupied %	0.051 (.01)	$1.31x10^{-5}$	.003
Income	Income 80-100%	0.069 (.01)	$4.89x10^{-11}$	.005
Income	Income 20-40%	-0.076 (.01)	$2.37x10^{-11}$	.006
Income	Income 40-60%	-0.059 (.01)	$1.21x10^{-8}$	.003
Core neighbourhood characteristics	Mean house value	0.064 (.01)	$5.21x10^{-11}$	.004
Livability 2002/2003	Population composition	0.057 (.01)	$1.05x10^{-6}$	.003
Livability 2002/2003	Livability (LBM) score	0.053 (.01)	$5.00x10^{-6}$	.003
Livability 2002/2003	Housing score	0.052 (.01)	$5.02x10^{-6}$	.003
Livability 2002/2003	Safety score	0.049 (.01)	$1.60x10^{-5}$	.002
Livability 2009/2010	Livability (LBM) score	0.059 (.01)	$2.19x10^{-8}$	.003
Livability 2009/2010	Housing score	0.057 (.01)	$4.91x10^{-8}$	.003
Livability 2009/2010	Population composition	0.064 (.01)	$2.16x10^{-9}$	.004
Livability 2009/2010	Safety score	0.044 (.01)	$2.12x10^{-5}$	.002
SES scores 2002/2003	Status score	0.057 (.01)	$7.01x10^{-7}$	.003
SES scores 2002/2003	Rank order	-0.052 (.01)	$4.30x10^{-6}$	.003
SES scores 2009/2010	Status score	0.067 (.01)	$1.68x10^{-10}$	.004
SES scores 2009/2010	Rank order	-0.057 (.01)	$1.90x10^{-8}$	.003
Transactions 2009/2010	Mean house transactions	0.051 (.01)	$2.52x10^{-6}$	.003

Note. SES= socioeconomic status, *β*= beta, *SE*= standard error, GEE= generalized estimation equation, R^2= R-squared.

3.2. Polygenic risk score analysis

The well-being spectrum polygenic score predicted well-being in our sample (R^2=.007, *P*= $5.11x10^{-12}$), but it did not predict any of the environmental correlates (Table 4). Additionally, no mean difference between polygenic septiles was observed for any of the variables (see Supplementary Table S9).

Table 4

Associations well-being polygenic score with environmental exposures

Domain	Variable	PGS *β* (*SE*)	PGS *P*-value
Housing Benefits	Housing benefits (allocations)	-0.020 (.02)	0.246
Housing Stock	Social rental sector %	-0.017 (.02)	0.310
Housing Stock	Rental sector %	-0.002 (.02)	0.929
Housing Stock	Owner occupied %	0.002 (.02)	0.929
Income	% people income 5th percentile	0.007 (.02)	0.672
Income	% people income 2th percentile	0.002 (.02)	0.906
Income	% people income 3th percentile	-0.006 (.02)	0.704
Core neighbourhood characteristics	Mean house value	0.009 (.02)	0.560
Livability 2002/2003	Population composition	0.030 (.02)	0.066
Livability 2002/2003	Livability (LBM) score	0.030 (.02)	0.075
Livability 2002/2003	Housing score	0.012 (.02)	0.467
Livability 2002/2003	Safety score	0.011 (.02)	0.517
Livability 2009/2010	Livability (LBM) score	0.015 (.02)	0.313
Livability 2009/2010	Housing score	0.019 (.02)	0.211
Livability 2009/2010	Population composition	0.017 (.02)	0.278
Livability 2009/2010	Safety score	0.008 (.02)	0.614
SES scores 2002/2003	Status score livability	0.017 (.02)	0.289
SES scores 2002/2003	Rank order livability	-0.014 (.02)	0.397
SES scores 2009/2010	Status score livability	-0.011 (.02)	0.477
SES scores 2009/2010	Rank order livability	0.006 (.02)	0.691
Transactions 2009/2010	Mean house transactions	0.016 (.02)	0.334

Note. SES= socioeconomic status, PGS= Polygenic Score, β=beta, SE=standard error.

3.3 Multicollinearity follow-up

Strong correlations (ranging between -1 and .87) were observed between the significant variables from the GEE analyses (Supplementary Table S7). For both time-points, we plotted the variables that were correlated .8 or stronger (Supplementary Figures S1A and S2A), and .4 or stronger (Supplementary Figures S1B and S2B) using chord plots**.** These plots display all associations (above our defined thresholds) between the included variables. The variables are presented in a circle, and whenever a line connects two variables, it indicates they are associated. For both time-points, when we defined the threshold as correlations >.4, we see that all variables are connected to all other variables, creating a

densely connected plot. However, when we increased the threshold to .8, the plots become more organized with only few connections remaining. For the 2002/2003 data, this resulted in a plot with three clusters: 1) a housing cluster with housing score, housing stock owner-occupied, housing stock: rental, and housing stock: social rent, 2) a livability cluster of livability scores, population composition scores, and safety scores, and 3) another livability cluster with status scores and rank order of the neighbourhoods. For the 2009/2010 data, we see two clusters: 1) an SES cluster including two income variables, mean house value, and the status score and rank order of the neighbourhood, and 2) a livability cluster including LBM scores, population composition, and safety.

The PCA extracted 95 and 38 independent PCs for 2002/2003 and 2009/2010, respectively. The first 43 PCs cumulatively explained 90.5% of the 95 environmental variables in the 2002/2003 data, and the first 16 PCs explained 90.7% of the 38 environmental variables in the 2009/2010 data (see Supplementary Table S10). Combined in one linear regression model, these 43 PCs explained 1.45% of the variance in well-being in the 2002/2003 data**.** After correcting for the number of PCs included, this decreased to 0.69% (adjusted R^2). One PC (PC3: β=-0.029, *SE*=0.006, *P*=2.73x10^{-7}) significantly predicted well-being after correcting for multiple testing. For the 2009/2010 data, the 16 PCs explained 1.11% of the variance in well-being, which decreased to 0.79% after correcting for the number of PCs (adjusted R^2). Two PCs significantly predicted well-being (PC1: β=0.0185, *SE*=0.005, *P*=0.0001, PC2: β=-0.0240, *SE*=0.006, *P*=3.4x10^{-5}). Supplementary Table S11 lists the environmental variables with loadings higher than .1 with the significant PCs. For the 2002/2003 data, the PC that significantly negatively predicted well-being was represented by four variables reflecting low-income neighbourhoods . For the 2009/2010 data, one of the PCs (PC1) was indicative of high income and livability, while the other PC (PC2) was indicative of low income and lower livability.

3.5 Analyses with socioeconomic status

Correcting for individual EA had a small effect on the observed associations. After also including the SES of the neighbourhood, only neighbourhood safety and the percentage of land devoted to greenhouse horticulture remained significant.

A summary of all analyses and their results can be found in Figure 2.

4. DISCUSSION

The present study linked two large data-resources in the Netherlands in order to examine potential associations between well-being and a range of environmental factors. Using this environment-wide association approach, we identified 21 environmental factors that were associated with well-being. These factors cluster in the following domains: housing stock, income, core neighbourhood characteristics, livability, and SES. A common theme that emerged is that the identified correlates can be classified as socioeconomic indicators.

An examination of the correlations between these variables reveals that they are not independent. When correcting for individual and neighbourhood SES, only safety and % of land devoted to greenhouse horticulture were significantly associated with well-being, with safer neighbourhoods and neighbourhoods with more greenhouse horticulture showing higher average levels of well-being. A closer examination of the distribution of these two environmental variables in the Netherlands (Supplementary Figures S3-5) revealed that greenhouse horticulture did not show a lot of variation across the country, especially compared to the other associated variables (SES and safety). Therefore, this association should be interpreted with caution. Safety, on the other hand varies widely across the different postal codes. Earlier studies also found associations between psychological health and neighbourhood safety[188–190]. It is furthermore, in line with previous research where well-being was linked to neighbourhood-level SES indicators[191,192]. Moreover, similar results have been found for depression using GECCO data[193]. Importantly, what should be kept in mind when examining the results of this study is that we are examining associations, and not causal effects. For the identified associations, this means two things should be considered. First, there might be some third, mediating factor that explains the associations. Most of the factors assessed in the first round of EnWAS disappeared when we corrected for SES, already suggesting that SES was driving these associations. Secondly, even if there are potential causal associations, we cannot make any statement regarding the direction of the effect.

No effects of genetic differences were observed, indicated by the absence of significant genetic prediction. This indicates that either the genetic predisposition for well-being does not cause individuals to pick certain environments or that we

suffer from a lack of power. Indeed, a post-hoc power analysis (Supplementary Figure 6) indicates that with our current sample size and alpha, we could have detected associations between the well-being polygenic scores and environmental exposures with effect sizes greater than R^2=.002. Thus, associations between the current PRS and the environmental exposures assessed here are likely extremely small. The well-being spectrum polygenic score explains less than a percentage of the variance in well-being itself, and there was no difference in mean well-being between different genetic susceptibility groups. This raises the question of whether a stronger PGS would lead to different results than presented here. Therefore, while any statement on this genetic component is speculative at this moment, we cannot write off the potential role that genetics play in these associations, and encourage future investigations in this area.

From the existing literature, we already knew that the effect of individual genetic variants on well-being is small[29]: 12-18% of the variance in well-being is explained by ~600k genome-wide measured SNPs for complex traits, with GWA study SNP-based heritability estimates around ~5%[30]. Here, we report small environmental effects on well-being. The significant environmental predictors from the EnWAS individually explain only 0.2% to 0.5% of the variance in well-being. Additionally, the PCA showed that the combined effect of the EnWAS variables explains only around ~1% of the variance in well-being. Important to keep in mind while interpreting these effect sizes is the fact that we examined environmental exposures at the postal code level. It is likely that the well-being exposome varies over different geographical levels (e.g. cities, municipalities)[194], measures of well-being, and is differently associated with subjective indicators of the environment[195,196]. Take as an example SES: studies examining the effect of individual-level SES on well-being find estimates as large as 6% explained variance[197,198], which is much larger than our current finding for neighbourhood SES indicators.

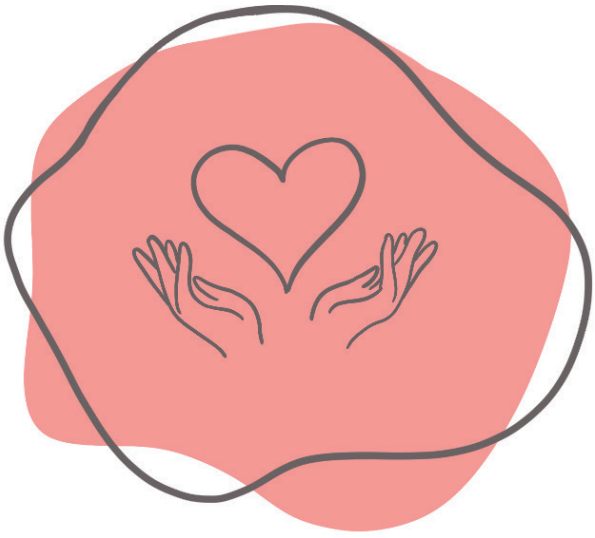
Moreover, we did not, despite our large sample, find any evidence for many previously suggested indicators, such as the presence of green space[199] or air pollution[200]. Different reasons might explain this discrepancy: e.g. the level and country of examination (postal code level in the Netherlands), the use of objective indicators of the environment (instead of subjective experiences), and the measure of well-being we used. Therefore, our findings should be interpreted in the context of this study. Important, though, is that our study investigates the

association between wellbeing and postal code linked variables, e.g. the amount of greenspace in the postal code area. That is a different approach than studying wellbeing in relation to frequency of visiting or enjoying greenspace. In order to develop a full picture of the well-being exposome, it is necessary to take these different aspects into account. Mapping the well-being exposome will also require investigations on different time-points or, optimally, longitudinal investigations tracking the dynamic interplay and direction of causality between environmental factors, biological factors and well-being[201]. For consistency, we decided to assess each variable on the same geospatial scale (PC-4 level). However, this level is likely not the most relevant level for each assessed exposure variable. The methodology used in this project can easily be applied to different levels of analysis (e.g. individual level objective data, individual level subjective data, street level). In this way, we can compare EnWAS results on different levels, offering a replicable means of mapping the well-being exposome. What should additionally be kept in mind is that many studies focus their efforts on one or a few exposures at a time, limiting the potential to study such an exposure in a broader context. This study demonstrates the importance of large, data-driven explorations to get a more adequate image of these intertwined environmental associations.

In the genetics field, small effects are common and combined in polygenic scores that are used for more in-depth analyses. An interesting approach would be to combine environmental effects in "*poly-environmental*" scores. In this way, small environmental effects can be combined and used to predict well-being. An obstacle that needs to be overcome in order to construct these scores is that we need a better understanding of the correlational structure between different environmental factors. In case of polygenic scores, we can correct for correlations between genetic variants based on our knowledge of recombination patterns and linkage disequilibrium[186]. For poly-environmental scores, however, the association between different environmental factors is much more complex and dynamic. By combining small effects in poly-environmental scores, complemented by polygenic scores, it might in the future be possible to develop personalized prevention and intervention strategies for well-being. However, in addition to acquiring better knowledge of the correlational structure of the environment, this will also require more insight into the potential direction of causality of current findings. Another interesting direction for future research that aims to combine

genetic and environmental effects is to compare the well-being of monozygotic twins that are exposed to different living environments. Since monozygotic twins are 100% genetically identical, a difference in well-being between the twins can only be caused by unique environmental experiences. Therefore, by associating monozygotic intra-pair difference scores for well-being with intra-pair difference scores for environmental exposure, it becomes possible to examine the extent to which an association between well-being and an environmental exposure exists independent from genetic and shared environmental factors. In our sample, there was a relatively low number of *complete* monozygotic twin pairs for which we could compute difference scores for both well-being and the environmental exposures (N_{pairs} 2002/2003=389, N_{pairs} 2009/2010=270). As a result, no evidence was found for an association between well-being intra-pair difference scores and any of the environmental exposure intra-pair difference scores (see Supplementary Table S12). Therefore, we encourage other cohorts with larger samples to perform these analyses in order to get a grasp of the potential genetic effects.

To conclude, in this study we combined the strengths of record linkage to understand individual differences in well-being. Taken together, our analyses suggest that, at the postal-code level, the most important predictors of well-being are socioeconomic factors and safety. Moreover, we find that environmental effects are typically small and context dependent, emphasizing the need for large scale linkage efforts and data-driven designs.



CHAPTER 5

Capturing the well-being exposome in poly-environmental scores

In preparation as: Van De Weijer, M. P., Pelt, D.H.M., Baselmans, B.M.L., Ligthart, L., Huider, F., Hottenga, J.-J., Pool, R., & Bartels, M. (i*n preparation*). Capturing the well-being exposome in poly-environmental scores.

*supplementary materials accessible at: https://drive.google.com/drive/folders/19XdAvHMxSW8J3MqLX3r6eSpH8tIZGCaA?usp=sharing

ABSTRACT

In this study we use an aggregated weighted score of environmental effects to study environmental influences on well-being and happiness. To this end, we split a sample of Netherlands Twin Register (NTR) participants into a training (*N*=4857) and test (*N*=2077) sample. In the training sample, we use elastic net regression to estimate effect sizes for associations between life satisfaction and two sets of environmental variables: one based on self-report socioenvironmental data (PES-S) and one based on objective physical environmental data (PES-O). In the test sample, we perform association analyses between different measures of well-being and the two PESs. We find that the PES-S explains ~36% of the variance in well-being, while the PES-O does not significantly contribute to the model. Variance in other well-being measures (i.e. different life satisfaction domains, subjective happiness, quality of life, flourishing, psychological well-being, self-rated health, depressive problems, and loneliness) are explained to varying extents, ranging 6.36% (self-rated health) to 36.66% (loneliness). These predictive values did not change during the COVID-19 pandemic (*N*=3214). Validating the PES-S in the UK Biobank (*N*=40,614), we find that the UKB PES-S explains about ~12% of the variance in happiness. Lastly, we examine if there is any indication for gene-environment correlation (rGE), the phenomenon where ones genetic predisposition influences exposure to the environment, by associating the PESs with polygenic scores (PGS) in a sample of NTR and UK Biobank participants. While the PES and PGS were not correlated in the NTR sample, they were correlated in the larger UKB sample, indicating the potential presence of rGE. We discuss several limitations pertaining to our dataset and reflect on how PESs might be used in future research.

1. INTRODUCTION

Many (socio)environmental exposures have been associated with human well-being. For example, meta-analyses suggest a role for social support[202], green space exposure[38], and socioeconomic status[203], among many other factors. The totality of these environmental exposures can collectively be referred to as the well-being *exposome*[204], which captures all non-genetic exposures influencing variation in well-being from conception onwards. Approximately 60-70% of individual differences in well-being can be traced back to this exposome[24,25]. Complementary to the exposome is the genome, which accounts for the other 30-40% of individual differences in well-being.

While there are many studies examining associations between environmental factors and well-being, they mostly follow a "pick-and-choose" approach, where potential risk factors are selected beforehand based on existing hypotheses, which can result in selective reporting (cherry picking) and overestimation of effects (publication bias). To overcome these problems, we recently conducted an environment-wide association study (EnWAS)[205] to get a better hold of the various environmental influences on well-being. In this EnWAS we were interested in examining individual environmental effects in a hypothesis-free fashion, similar to the genome-wide association approach. In so-called genome-wide association studies (GWAS), millions of genetic variants are associated with an outcome without a priori hypothesizing which variants might be important for the trait of interest. By examining associations between well-being and 139 objective environmental indicators (e.g., green space, livability) in our EnWAS, we found that neighborhood variables related to socioeconomic status and safety were significantly associated with individual-level well-being.

One of the ways in which GWAS findings are applied in both scientific and clinical contexts is by combining the resulting effect sizes (reflecting the strength of associations between individual SNPs and the outcome of interest) into so-called polygenic scores (PGS)[31]. These scores, constructed by aggregating an individual's effect alleles weighted by the respective GWAS effect sizes, reflect an individual's genetic propensity or risk for a certain outcome. PGSs can be used for disease risk estimation, cross-phenotype prediction, and for answering specific research questions such as if the effect of individual environmental exposures varies over

strata of individuals with varying genetic propensities for a specific outcome[31]. In previous research, well-being polygenic scores have been used to predict not only well-being measures[32], but several other phenotypes such as loneliness[206], childhood psychopathology[207], and brain morpholology[208]. Similarly, we can take our investigation of the exposome one step further by summing individual environmental risk across multiple environmental factors into poly-environmental or poly-exposure scores (PESs).

One of the potential difficulties in constructing PESs is that environmental factors are correlated, and that we should take these correlations into account in some way so that we do not overestimate the effect of the environment. When constructing PGSs, it is possible to correct for correlations between genetic variants based on our knowledge of chromosomal recombination patterns and linkage disequilibrium, i.e. the non-random association of alleles at different loci within a population[186]. However, correlations between environmental factors are much more complex and dynamic, complicating the construction of PESs. Existing studies using PESs have focused mainly on disease outcomes such as schizophrenia and psychosis[209–213], where the scores are used to identify at-risk individuals for these outcomes. Overall, these scores seem to be able to explain 10-20% of the variation in case-control status. The manner in which the environmental factors are combined in the calculation of these PESs varies considerably, with some simply summing estimates from systematic reviews or meta-analyses[209,214,215] (without correcting for potential correlations), and others using training datasets where different prediction techniques are used to calculate weights that are used for weighing the estimates[210,216] in a test sample. An important distinction between these two approaches is that the latter takes correlations between predictors into account by weighting the different exposures, while the former does not.

Similar to PGSs, PESs could be used for well-being and cross-phenotype prediction, and for stratifying individuals based on environmental (instead of genetic) 'risk or protection' (i.e. which individuals live in environments that stimulate well-being). Moreover, it allows us to broaden our understanding of the interplay between the genome and exposome with the use of research designs that combine PGSs and PESs. For example, it presents us with a new opportunity to study gene-environment correlation (rGE), the phenomenon where ones genetic predisposition influences exposure to the environment[217]. There are three types of gene-environment

correlation: 1) active, where a person's heritable traits cause them to actively select certain types of environment, 2) evocative, where a person's heritable traits elicit a reaction in other people, which in turn influences one's environment, and 3) passive, where genotype and environment become correlated because a child inherits both genes and a familial environment from their parents[218]. By examining if the PESs and PGSs are associated, we are provided with a first indication of potential rGE, but not necessarily of which type of rGE.

Therefore, the objective of the current study is to explore the potential of PESs in the context of well-being. First, we calculate effect size estimates for two sets of exposome variables (one including objective environmental indicators and one including subjective environmental evaluations) in predicting life satisfaction scores in a Dutch training set using elastic net regression. Next, we use these effect sizes as weights to construct PESs in an independent test set and predict life satisfaction and several other well-being outcomes in this test set. Based on our previous finding that environmental variance for well-being increased during the pandemic[219], we additionally examine if the predictive power of the PES changes from before to during the COVID-19 pandemic. Lastly, we examine if the well-being PES is associated with a well-being PGS to assess potential gene-environment correlation. We compare these findings with a follow-up analysis in the UK Biobank, where we create a similar PES based on socioenvironmental variables and correlate it with the well-being PGS.

2. MATERIALS AND METHODS

This study was pre-registered at the open science framework (OSF) (https://osf.io/5x8kf). Additional follow-up analyses are indicated as "non-preregistered".

2.1 Sample

We used a sample of participants from the Netherlands Twin Register (NTR)[63]. We included multiples and family members who filled out the most recent wave of survey data collection (wave 14 collected in 2019-2022). We use two separate datasets: 1) the first part of the data collection just prior to the COVID-19 pandemic in the Netherlands, collected between June 2019 and February 2020, and 2) the second part of the data collection, collected between February 2020 and May 2022,

during the COVID-19 pandemic. In both samples, we remove genetic relatedness by randomly including one individual from genetically related family members, leading to final sample of N_{max}=6092 individuals with satisfaction with life data in the pre-pandemic sample, and N_{max}=3214 individuals with satisfaction with life data in the pandemic sample. Sample sizes vary across analyses and outcomes based on missingness. Sample characteristics can be found in Table 1.

Table 1

Sample descriptives pre-registered NTR analyses

	NTR			
	full sample* (pre-pandemic)	**training sample****	**test sample****	**pandemic sample**
N males/females	1779/4313	1425/3432	608/1469	886/2328
M (SD) age	49.94 (12.22)	50.01 (12.19)	49.98 (12.39)	35.18 (13.32)
Age range	16-88	16-88	18-86	18-88
M (SD) life satisfaction	27.50 (5.11)	27.50 (5.14)	27.49 (5.04)	27.08 (5.48)
M (SD) family satisfaction	4.66 (.78)	4.66 (.78)	4.65 (.78)	4.74 (.88)
M (SD) financial satisfaction	4.61 (.80)	4.61 (.81)	4.60 (.79)	4.53 (.88)
M (SD) work satisfaction	4.57 (.85)	4.58 (.84)	4.55 (.87)	4.51 (.92)
M (SD) health satisfaction	4.56 (.88)	4.56 (.87)	4.56 (.90)	4.66 (.94)
M (SD) friendship satisfaction	4.71 (.75)	4.72 (.74)	4.68 (.78)	4.81 (.83)
Subjective Happiness	22.81 (4.21)	22.79 (4.22)	22.87 (4.16)	22.10 (4.78)
Quality of Life	7.84 (1.06)	7.83 (1.07)	7.86 (1.03)	7.65 (1.22)
Flourishing	46.18 (5.70)	46.18 (5.70)	46.18 (5.70)	46.02 (6.04)
Psychological well-being	20.73 (5.71)	20.76 (5.68)	20.65 (5.78)	20.47 (5.83)
Self-Rated Health	3.95 (.70)	3.95 (.70)	3.95 (.69)	4.04 (.71)
Depressive Problems	3.64 (3.64)	3.62 (3.64)	3.69 (3.65)	4.56 (4.37)
Loneliness	4.04 (1.35)	4.02 (1.34)	4.10 (1.37)	4.44 (1.51)

*full sample indicates the full set of individuals with non-missing SWL data

**actual sample size per analysis is lower depending on the amount of missingness per PES type

2.2 Measures

Outcome measures

Satisfaction with life was assessed using the 5-item satisfaction with life (SWL) scale[12]. Individual items are scored on a scale from 1 to 7, with higher scores indicating higher levels of satisfaction with life. The item responses are combined into a sum-score ranging from 7-35 (Cronbach's α=.86).

Domain-specific satisfaction items were assessed *family satisfaction, financial satisfaction, friendship satisfaction, work satisfaction*, and *health satisfaction*. For each domain, participants were asked 'in general, how satisfied are you with[...]?'. The items were coded on a scale from 1 to 6, with 1 indicating extremely unhappy, and 6 indicating extremely happy.

Subjective Happiness was assessed using the 4-item subjective happiness scale (SHS)[64]. The items are rated on a Likert scale ranging from 1-7. We recoded two reverse-coded items so that for all items, a higher score indicated higher levels of happiness. The items were combined into a sum-score ranging from 4-28 (Cronbach's α=.87).

Quality of Life was assessed using the single-item Cantril ladder[65]. Participants are asked to answer the question "Where on the scale would you place your life in general?", on a scale from 0 (indicating the worst possible life) to 10 (representing the best possible life).

Flourishing was assessed using the 8-item Short Flourishing Scale[66]. We combined the individual items, which are rated on a Likert scale ranging from 1-7, into a sum-score, where higher values indicate higher levels of flourishing (Cronbach's α=.90).

Psychological well-being was assessed using the psychological well-being (PWB) subscale of the mental health continuum short form[220]. This subscale consists of 6 items rated on a scale from 0 to 5. Item responses were summed to create scores ranging from 0 to 30, with higher scores indicating higher levels of PWB (Cronbach's α=.92).

Self-rated health was assessed with a single item "How would you rate your health in general?"[70]. The item was rated on a 5-point scale ranging from 'Bad'(1) to 'Excellent'(5).

Loneliness was assessed using the 3-item short scale for assessing loneliness in large epidemiological studies[68]. For each item, participants were asked to indicate how often they identified with a statement (e.g. "how often do you feel isolated from others?"), rated as 0= almost never, 1=sometimes, 2=often. The items were summed to obtain a loneliness sum-score (Cronbach's α=.79).

Depressive problems were assessed with the ASR DSM Depressive Problems scale[221]. The scale consists of 14 items, where each item is rated from 0 = not

true, 1 = somewhat true, to 2 = very true. The items were summed to obtain a depressive problems sum-score where higher values indicate higher levels of depressive problems (Cronbach's α=.82).

Predictors

We include two sets of environmental exposures:

1. Objective Environmental Exposures: we used a set of 69 objective environmental measures obtained from the Geoscience and Health Cohort Consortium (GECCO)[175], linked to the NTR data based on 4-numeric postal code. We previously included an overlapping set of variables in an EnWAS[205]. The variables reflect aspects of the physical environment (i.e. amount of traffic in the area), culture, socioeconomic status, accessibility, education, liveability, care, and sports. The included variables reflect time-points between 2017 and 2020. A complete overview and descriptives of these variables can be found in Supplementary Table 1.

2. Subjective Environmental Exposures: A set of 21 subjective socio-environmental indicators included in the same NTR survey wave as the well-being data. These variables reflect participants' subjective evaluations on their relationships, life events, social support, leisure time activities, education, stress, and online and offline social contact. A complete overview and descriptives of these variables can be found in Supplementary Table 2.

All continuous variables were standardized to Z-scores prior to analyses.

Postal code linkage

The GECCO and NTR data were linked based on self-report postal code data from NTR participants that gave permission for data linkage. When participants register for the NTR, they are asked to provide their address. Participants are asked to contact us when they move so that their address can be updated in our administrative database. In addition, addresses are regularly updated by cross-checking with the Dutch Personal Records Database (PRD) (in Dutch: Basis Registratie Personen, BRP), and by enquiring about postal codes in several survey waves and NTR newsletters. For the current project, we used data from the pre-pandemic and pandemic wave 14 NTR data collection. In the pandemic survey,

we included a question asking participants to report their current four-numeric postal code. For those individuals, we used this self-reported postal code. In the pre-pandemic survey, this item was not included. For a subset of 2623 participants of the pre-pandemic sample, addresses were cross-checked with the PRD in September 2022. For this subset, 85% of known postal codes matched with the postal codes reported in PRD. Since the pre-pandemic sample was collected 2-3 years prior to the PRD check, postal code listed at NTR at the time of data collection was used as the postal code for linkage.

Covariates

As phenotypic covariates, we include sex, age, and age^2. For the analyses that include PGSs, we additionally include genotyping platform dummies and the first 10 genomic principal components (PCs) to control for population structure. Smartpca was used to calculate the PCs, using LD-pruned 1000 Genomes imputed SNPs genotyped on one or more platforms.

2.3 NTR genotype data and polygenic scores

Genotype data were available for *N*=558 participants in the test sample. Genotyping was performed on multiple genotyping platforms: Affymetrix 6.0 (*N*=207), Affymetrix Axiom (*N*=43), Affymetrix Perlegen (*N*=34), Illumina Omni 1M (*N*=4), Illumina 660 Human Beadchip 4 (*N*=48), and Illumina GSA (*N*=222). Prior to imputation, samples were excluded if there was a mismatch in reported and genotyped sex, an abnormal inbreeding value (F value <.10 or >.10), or a call rate below 90%. SNPs were removed if the minor allele frequency (MAF) was below < $1x10^{-6}$, there was a deviation from Hardy-Weinberg Equilibrium (HWE) ($p<1x10^{-6}$), or if the missing rate was >5%. After quality control, the SNPs were aligned with the 1000 Genomes reference panel[222], filtering for SNPs with allele frequencies differences >.20, palindromic SNPs, and DNA strand issues. After merging the platforms into a single dataset, samples with DNA identity-by-descent (IBD) status that did not agree with expected familial relationships and CEU population outliers were excluded. The data was phased using Eagle v2.4.1[223] and imputed to 1000 Genomes[222] and TOPMED[224] reference panels using Minimac3-omp[225], following Michigan imputation protocols. The separate platform VCF files were merged into a single file per chromosome with Bcftools 1.9[226]. Prior to polygenic scoring, the imputed data were converted into best

guess data and filtered to include only ACGT, bi-allelic SNPs, SNPs with MAF >.01, HWE p>1x10^{-5}, and genotype call rate >.98.

Polygenic scores were calculated from summary statistics from a well-being spectrum GWAS[108], excluding NTR participants. Before constructing the scores, effect sizes were re-estimated, taking into account linkage disequilibrium (the correlation between SNPs), with LDpred 0.9[186], using an infinitesimal prior and 1000Genomes phase 3 CEU as a reference panel. The re-estimated effect sizes were used for constructing polygenic scores in PLINK[114].

2.4 Pre-registered Statistical analyses

Elastic net model effect size estimation

We randomly divide our pre-pandemic sample into a training (70%) and test (30%) sample. The training sample is used to estimate effect sizes to use as weights to calculate our PES in the test set. We fit two elastic net regression models in the training sample where we predict life satisfaction scores with 1) the objective environmental indicators (*N*=2152), and 2) the subjective environmental variables (*N*=2297). The phenotypic covariates were included in both models. While we use the same training sample for both elastic net models, the sample size varies slightly based on missingness of the predictors. Elastic net regression is a selection and shrinkage method that combines the penalties from ridge (L2) and lasso (L1) regression to optimally deal with correlated predictors and prevent overfitting. The best tuning parameters alpha (extent to which L1 and L2 are applied) and lambda (the penalty, or shrinkage coefficient) were selected based on the lowest RMSE value from 10-fold cross-validation using the *caret* R package[227].

Poly-environmental score prediction

The non-zero coefficients from the model with the most optimal tuning parameters are used for constructing the objective and subjective PESs for life satisfaction in the test sample. The PESs were calculated by summing the predictor variables weighted by their respective coefficients from the training set elastic net models. The PESs were standardized so that resulting associations reflect the impact of an SD increase in the PESs on the different outcomes. We refer to the PES based on objective indicators as the PES-O, and to the PES based on the subjective indicators

as the PES-S. We include the PESs (and phenotypic covariates) as predictors in two separate, and one combined linear regression models where we predict SWL scores.

Cross phenotype prediction

We assess cross-phenotype overlap in exposome associations by using the PESs based on SWL data as predictors in models where we predict the other well-being(-related) phenotypes: subjective happiness, the specific satisfaction domains, quality of life, flourishing, psychological well-being, self-rated health, loneliness, and depressive problems. Since we run separate models for all 12 phenotypes, we use a corrected significance threshold of α=.05/12= .004.

Prediction during the corona pandemic

Part of our well-being data collection took place during the COVID-19 pandemic. Since the effect of environmental factors might be different during the pandemic, we separately analyzed this part of the sample (*N*=1898) to examine whether the PESs predict SWL and the other well-being related phenotypes similarly or differently during the pandemic. Confidence intervals, calculated using the *psychometric* R package[228] were used to compare the R^2 of the prediction models before the pandemic to the R^2 of the prediction models during the pandemic.

Gene-environment interplay

Finally, in a sub-sample of the test set that has genotype data available (*N*=556), we assessed potential gene-environment correlations. To this end, we include polygenic scores (PGSs) in models predicting the PESs (for a description of how these PGSs were constructed, see "NTR genotype data and polygenic scores" section above).We predict the PESs using well-being spectrum PGSs[108], including all phenotypic and genetic covariates (see covariate section above). If the PGS significantly predicts the PES (α=.05), it indicates the genetic predisposition for well-being is associated with the exposure to environmental influences. This association can reflect passive, evocative or active gene-environmental correlations and indicates that environmental exposure is not (only) a random process.

2.6 Non pre-registered - Follow-up analyses

UK Biobank cross-validation

To examine the predictive power of a socioenvironmental PES on well-being, and possible correlations with a well-being polygenic score to assess potential rGE in a larger sample, we used data from the UK Biobank (UKB)[112]. The UKB is a large cohort study with phenotypic, genetic, and biological data of UK individuals recruited between the ages of 40 and 69. A single item on happiness (UKB ID 4526) was used an outcome measure. Participants were asked to answer the question "In general how happy are you?" on a scale from 1 to 6, 1 indicating extremely happy, and 6 indicating extremely unhappy. We reverse coded the item so that higher scores indicated higher levels of happiness (M=4.47, SD=.69). This item most closely resembles subjective happiness from the included NTR well-being measures. We selected socioenvironmental variables that most closely resembled the ones that were selected in the NTR data, resulting in a selection of N=15 variables related to mobile phone use, time spend exercising, leisure time activities, educational attainment, life events, and social support (see Supplementary Table 3 for details on the included predictor variables). The sample consisted of N=40,614 individuals with non-missing phenotype data. Similar to the NTR analyses, we split the sample in a training (70%) and test (30%) sample, and standardized all continuous predictors. The training sample was used to calculate elastic net effect size estimates, which were combined into PESs in the test sample. Age, age^2, and sex were included as covariates in both the training and test stage.

We additionally tested for gene-environment interplay by correlating the PESs with well-being spectrum PGSs. Details on genotyping procedures in UKB can be found elsewhere[112]. To account for linkage disequilibrium, the well-being spectrum summary statistics[108] (excluding UK participants) were reanalyzed with SBLUP[229], using a reference sample of 10,000 random unrelated UKB participants. Next, the re-estimated effect sizes were used to generate PGSs in PLINK[114]. These PGSs were used to predict the PESs, using the first 10 genomic PCs, batch, age, age^2, and sex as covariates.

3. RESULTS

Elastic net model effect size estimation and poly-environmental score prediction

The best model for the objective environmental indicators, selected through 10-fold cross validation, used an α penalty of 1, indicating that the L1 (lasso) penalty function was used (i.e. regression coefficients are shrunk toward zero). The λ parameter, which controls the weighting of the sum of both penalties, was set to 0.07, indicating that the penalty is weighted down substantially (indicating that the penalty is applied to a much lesser extent than it would have with full regularization). In the final model, 40 variables were set to zero, leading to the inclusion of 29 variables with non-zero coefficients in the PES-O (see Supplementary Table 4). When the PES-O were used to predict SWL scores in the test sample (*N*=949), it explained 0.4% of the variance (*p*=.052) in SWL scores in the test data after adjusting for sex, age, and age^2.

The most optimal model for the subjective environmental indicators, selected through 10-fold cross validation, used an α penalty of .55, indicating that about equal weight was given to the ridge and lasso penalties. The λ parameter was set to .05, again indicating that the penalty is weighted down substantially. In the model, 4 coefficients were set to zero, leading to 17 variables with non-zero coefficients that were used for constructing the PES-S (see Supplementary Table 4). The PES-S explained 37.16% of the variance in SWL scores in the test data (*N*=1155), after adjusting for sex, age, and age^2. Combined in one model (*N*=722), the two PESs explained 35.38% of the variation in SWL. Only the PES-S (β=3.05, *SE*=.16, $p<2x10^{-16}$), and not the PES-O (β=.07, *SE*=.15, *p*=.65), significantly predicted SWL in the combined model. The two PESs were uncorrelated (*r*=.07, *p*=.06).

Cross phenotype prediction

To assess the extent to which environmental predictors overlap between SWL and other well-being related phenotypes, we used the two PESs to predict 12 other phenotypes. The amount of variance explained by the PESs in the other outcomes ranged from 6.36% (self-rated health) to 36.66% (loneliness). For all outcomes, only the PES-S (and not the PES-O) was a significant predictor (see Table 2).

Prediction during the COVID pandemic

We used a sample of individuals who filled out the survey during the COVID-19 pandemic to assess if the PESs predicted SWL and other well-being related phenotypes equally well during the pandemic. Using SWL itself as the outcome, the variance explained by the prediction model including the two PESs did not change significantly (R^2=34.33%, CI= 30.84-37.82%), indicating that the included environmental factors to not predict well-being to a lesser extent in these changed environmental circumstances. For all other phenotypes, the amount of variance explained by the two environmental scores combined was also similar (see Figure 1 and Table 2). The PES-S was a significant predictor for all phenotypes during the pandemic, whereas the PES-O did not predict any of the phenotypes.

Gene-environment interplay

We computed well-being spectrum polygenic scores for the subset of participants in the test set that had genotype data available (*N*=558). The polygenic score did not significantly predict well-being (β=.44, *SE*=.23, *p*=.06). Similarly, the PGS did not significantly predict either the PES-S (β=.11, *SE*=.06, *p*=.07, *N*=335), or the PES-O (β=.-.07, *SE*=.05, *p*=.22, *N*=353).

UK Biobank follow-up

In the most optimal training set prediction model (α=.1, λ=.0004), none of the variables were set to zero, leading to 15 variables with non-zero coefficients used for constructing the UKB PES-S in the test sample (see Supplementary Table 5 for the elastic net estimates). The PES explained 12.41% of the variance in happiness scores in the test data (β=.24, *SE*=.01, $p<2x10^{-16}$), after adjusting for sex, age, and age^2.

The PGS significantly predicted happiness (β=.05, *SE*=.01, $p=<2x10^{-16}$), but only explained a small amount of the variance (.04%). Moreover, the PGS also significantly predicted the PES (β=.07, *SE*=.01, $p=4.05x10^{-14}$), similarly explaining .04% of the variance.

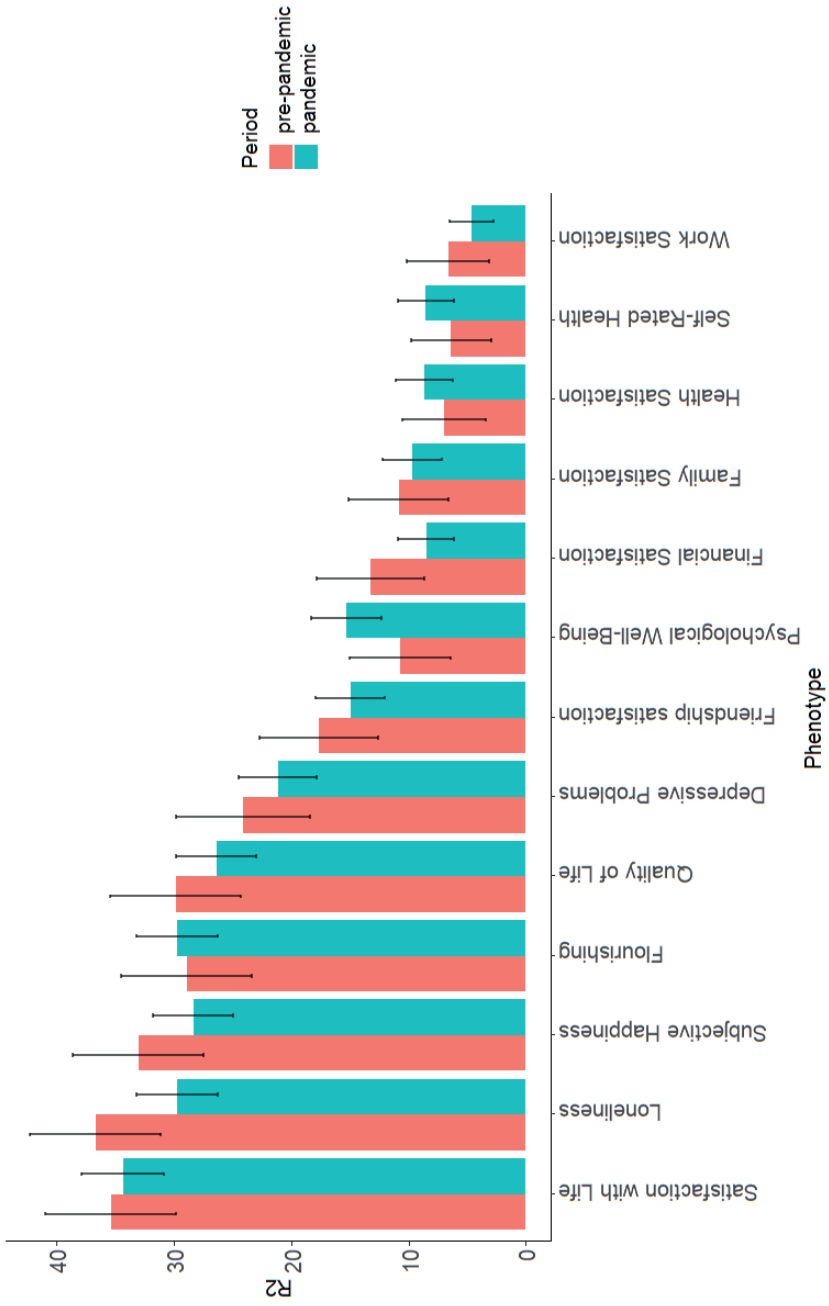


Figure 1. Variance explained by the combined PES-S and PES-O model on outcomes before and during the pandemic (for individual regression coefficients see Table 2).

Table 2

Results from the prediction models including both the PES-S and PES-O

	Pre-pandemic						Pandemic					
	PES-O		PES-S				PES-O		PES-S			
Outcome	**β (SE)**	***p***	**β (SE)**	***p***	***N***	**R^2**	**β (SE)**	***p***	**β (SE)**	***p***	***N***	**R^2**
Satisfaction with life	.07 (.15)	.65	3.05 (.15)	**<2x10^{-16}**	722	35.38%	.13 (.10)	.20	3.22 (.10)	**<2x10^{-16}**	1859	34.33%
Subjective Happiness	-.08 (.12)	.54	2.38 (.13)	**<2x10^{-16}**	722	33.04%	.03 (.09)	.78	2.59 (.10)	**<2x10^{-16}**	1858	28.36%
Quality of Life	.005 (.03)	.89	.59 (.03)	**<2x10^{-16}**	720	29.83%	.04 (.02)	.09	.64 (.02)	**<2x10^{-16}**	1882	26.38%
Flourishing	.11 (.17)	.49	2.89 (.17)	**<2x10^{-16}**	716	28.90%	.09 (.11)	.44	3.25 (.12)	**<2x10^{-16}**	1857	29.75%
Psychological Well-Being	-.08 (.21)	.70	1.98 (.21)	**<2x10^{-16}**	695	10.68%	.04 (.12)	.72	2.27 (.12)	**<2x10^{-16}**	1837	15.29%
Family Satisfaction	.01 (.03)	.60	.24 (.03)	**<2x10^{-16}**	719	10.81%	-.04 (.02)	.03	.27 (.02)	**<2x10^{-16}**	1878	9.65%
Financial Satisfaction	.01 (.03)	.65	.29 (.03)	**<2x10^{-16}**	718	13.21%	.01 (.02)	.47	.27 (.02)	**<2x10^{-16}**	1878	8.50%
Work Satisfaction	.004 (.03)	.90	.22 (.03)	**1.15x10^{-11}**	699	6.63%	.03 (.02)	.12	.20 (.02)	**<2x10^{-16}**	1861	4.61%
Health Satisfaction	.01 (.03)	.82	.25 (.03)	**1.99x10^{-13}**	721	7.00%	.05 (.02)	.02	.27 (.02)	**<2x10^{-16}**	1882	8.66%
Friendship Satisfaction	-.003 (.02)	.92	.31 (.02)	**<2x10^{-16}**	720	17.65%	.01 (.02)	.56	.32 (.02)	**<2x10^{-16}**	1877	14.97%
Self-Rated Health	.01 (.03)	.74	.20 (.03)	**2.61x10^{-13}**	722	6.36%	.03 (.02)	.04	.21 (.02)	**<2x10^{-16}**	1879	8.52%
Loneliness	.06 (.04)	.14	-.79 (.04)	**<2x10^{-16}**	719	36.66%	.03 (.03)	.36	-.86 (.03)	**<2x10^{-16}**	1876	29.74%
Depressive problems	.10 (.12)	.43	-1.79 (.12)	**<2x10^{-16}**	652	24.09%	-.02 (.09)	.80	-2.14 (.09)	**<2x10^{-16}**	1795	21.15%

4. DISCUSSION

This study examined the potential of combining multiple environmental correlates of well-being into well-being poly-environmental scores (PESs). To this end, we constructed two different PESs: one reflecting self-reported socio-environmental factors (the PES-S), and one reflecting objective (postal-code level) physical environmental factors (the PES-O). Moreover, we examined potential gene-environment correlation by associating well-being PESs with well-being PGSs. Lastly, we performed replication efforts in a UKB sample.

With respect to the predictive power of the PESs, we found a large difference between the two scores. While the score based on self-reported socioenvironmental factors explained over 35% of the variation in well-being scores, the score based on objective, physical environmental factors explained less than 1% of the variation. It is not entirely surprising that the PES-S explained such a large part of the variation in well-being: this score contained variables that are consistently associated with well-being in previous research, such as social support[202,230], feelings of stress at home/work[231], and negative and positive life events[232,233]. Since environmental factors have been found to account for 60-70% of individual differences in well-being[24,25], the included socioenvironmental exposures were able to explain approximately half of the environmental variation in well-being. In our own work on the relation between the social environment and well-being and adolescents, we found that genetic factors were able to explain a significant part of these associations[234]. An interpretation of this finding is that these associations are partly explained by a genetic predisposition for appraising one's life positively or negatively.

As a follow-up, we repeated the same analysis in a sample of UKB participants, where we included the available socioenvironmental factors that most closely resembled the ones we included in the NTR PES-S. The socioenvironmental UKB PES explained approximately 12% of the variance in well-being, which is a substantial amount but considerably less than in NTR. This difference can be traced back to differences in the amount and content of the included variables in the two PESs, where variables with large contributions to the elastic net models in NTR, such as having a partner and stress at home/work, were not available in the UKB dataset (see Supplementary Tables 4-5).

It is also not entirely surprising that the PES-O explains only a small part of the variation in well-being when used individually, and fails to predict well-being when combined with the PES-S in one model. In our previous work, similar variables on the postal code level explained only 1.45% of the variance in well-being[205]. Existing literature examining associations between well-being and spatial measures offers somewhat mixed results. In a British study, Ballas & Tranmer examined the extent to which variation in happiness and well-being was explained by four different levels in a multilevel design: region, district, household, and individual[235]. They found that almost all of the variation in well-being and happiness was attributable to the individual level, and some to the household level. A very small part of the variation in well-being, and none of the variation in happiness, was attributable to district/region. They conclude that, in the British context, well-being varies between people but not places. Similarly, in a Dutch study comparing the effect of subjective and objective spatial characteristics on well-being, the effect of subjective spatial characteristics on well-being was much larger than the effect of objective spatial characteristics[236]. In contrast, in an Irish sample, the inclusion of objective spatial indicators (such as mean annual precipitation and proximity to coast) in a model where life satisfaction was predicted using socioenvironmental indicators, while controlling for socio-economic and demographic characteristics of the individuals, led to a large increase in explained variance[237]. Moreover, researchers have found evidence for associations between well-being and different objective environmental indicators, such as air pollution[238], urban green space[239], and noise levels[240]. Nevertheless, the general consensus seems to be that subjective environmental indicators are better suited for explaining individual differences in well-being than objective ones. While these perceptions of the environment might be stronger predictors of physical environmental ones, the latter might also include more measurement error and thus be less reliable. For example, it has been shown that reliability of neighborhood condition measures is lower in rural than urban samples[241]. Nevertheless, an interesting future endeavor would be to create a PES based on subjective, instead of objective, physical environmental indicators (i.e. the perceived safety instead of the actual crime rate). In addition, it might be worth studying if big increases in sample size would improve the prediction of a PES based on objective indicators.

We first examined potential gene-environment correlation in the NTR sample. The well-being PGS did not significantly predict either well-being or the PESs. It

is likely that our limited sample size we did not have sufficient power for such endeavors. However, when performing similar analyses in the larger UKB sample (using happiness as a well-being measure instead of satisfaction with life), the well-being PGS (based on the somewhat less powered summary statistics due to the exclusion of UKB participants) was associated with both well-being itself and the well-being socioenvironmental PES. This finding supports the notion of gene-environment correlation for well-being, where a person's exposure to the environment depends on their genetic predisposition for well-being. Given that our sample consists of an adult sample, it is unlikely that we would identify passive rGE effects. The correlation between the well-being PGS and well-being PES is thus most likely to reflect either active or evocative rGE. In case of active rGE, this would mean that people's genetic disposition for well-being results in them seeking out certain types of (social) environments. For example, those with a higher genetic predisposition for well-being might seek out environments that stimulate their well-being, such as supporting relationships. In the case of evocative rGE, people's genetic predisposition for well-being would elicit certain types of environmental reactions, which in turn influences their well-being. For example, people with a high genetic predisposition for well-being might elicit positive social relations with others, which in turn would be beneficial for well-being. Our analyses indicate there potentially is gene-environment correlation, but does not allow us to give any kind of conclusive statement on the nature of this correlation and which rGE scenario is most likely.

Besides examining rGE, we also used the PESs for several other purposes. One of the research questions we were interested in was how well the well-being exposome in one context predicts well-being in other contexts. For the current project, we examined if the PESs would predict well-being to a similar extent during the COVID-19 pandemic and found that this was indeed the case. Other potential interesting applications could be to compare predictions across different ages, personality types, or other personal characteristics. We additionally used the PESs to examine overlap between different well-being constructs. By comparing if the PES for one well-being phenotype is as predictive for another well-being phenotype, we are provided with new information on the overlap/distinction between these phenotypes. Our results showed that the satisfaction with life PES predicted other well-being related phenotypes to varying degrees. For example,

psychological well-being was predicted to a much lesser extent than satisfaction with life, indicating that the environmental exposures associated with satisfaction with life only partly overlap with those for psychological well-being. This is in line with previous research that found only partly overlapping unique environmental effects between subjective and psychological measures of well-being[242]. An interesting future application would be to use PESs for follow-up analyses for answering research questions about phenotypes such as resilience, e.g., why do some people still thrive despite low environmental opportunities for well-being, and why do others score relatively low on well-being in "high well-being" contexts?

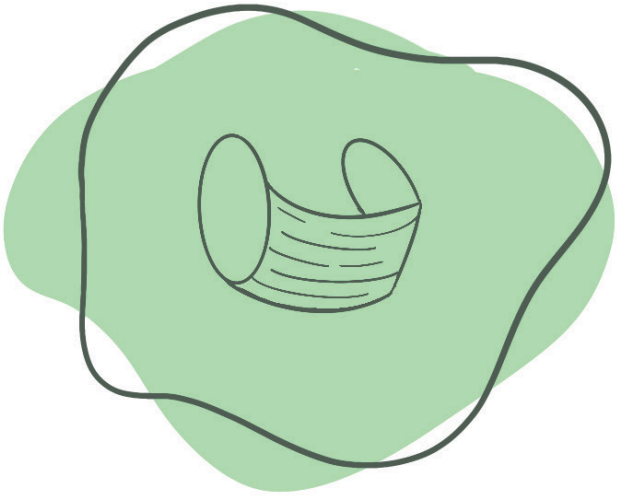
Our results should be interpreted in the light of multiple important limitations. First, the Netherlands and the UK are Western, Educated, Industrialized, Rich, and Democratic (WEIRD) countries, and results likely do not translate well to non-WEIRD contexts. One of the ways in which the findings might not translate well is with respect to the physical environmental exposures. The Netherlands is both a WEIRD and small country, meaning that the average distance to most amenities is relatively short. For example, the *maximum* distance between any participant 4-numeric postal code and a primary school is 7 km (Supplementary Table 1). For larger and less developed countries, distance to most amenities might be longer and less homogeneous across the country. In that case, there would be more individual differences in physical environmental measures possibly indicating a more important role in explaining individual differences in well-being. It would be interesting to construct well-being PESs in different cultures/contexts and compare results across these contexts. With respect to the PES-S, another limitation is that both the dependent and independent variables were obtained from the same self-report survey, meaning that the analyses might suffer from common method bias. When there is common method bias, correlations between variables can be inflated because of different types of response bias (e.g. question order bias). In our case, this would result in the PESs explaining more variance than actually is the case. With respect to the PES-O, our postal code linkage suffers from two limitations. First, for the pre-pandemic sample, we used last known postal codes for linkage. It is possible that postal code was not up-to-date for all participants, in which case the linkage would have been incorrect. Second, the GECCO data was not always available for all the years in which we collected phenotype data. For example, the pandemic dataset was collected between 2020 and 2022, but

the GECCO data was only available until 2020. In this case, we had to link the phenotype data to earlier years. While it is unlikely that there were large changes in the physical environmental data in such brief periods of time, it is possible that some error was introduced there. Lastly, while we speak of prediction models, associations between the environmental exposures and well-being should not be interpreted in a causal manner. The associations between the included environmental exposures and well-being could be causal in one or the other direction or bi-directional, and are not necessarily direct.

In summary, this study provides the first attempt to combine different environmental exposures into well-being poly-environmental scores. We find that a subjectively assessed socioenvironmental PES explains around half of the environmental variation in well-being in a Dutch sample, but that a PES based on objective physical environmental indicators does not predict well-being (when combined in one model with the PES-S). The socioenvironmental PES predicted well-being during the pandemic to a similar extent, and also predicted other well-being related phenotypes, albeit to varying extents. Additionally, we find that a PGS and PES for well-being are correlated in a UKB sample, suggesting the presence of gene-environment correlation. While our WEIRD sample has limited representability, this work shows the usefulness of using PESs for studying the well-being exposome. Future research could be conducted to examine the potential of subjective physical environmental indicators, and to study how these environmental scores vary across different cultures, contexts, and ages.

PART III

Well-Being in Light of the COVID-19 Pandemic



CHAPTER 6

Self-rated health when population health is challenged by the COVID-19 pandemic; A longitudinal study

Published as: Van De Weijer, M. P., de Vries, L. P., Pelt, D. H. M., Ligthart, L., Willemsen, G., Boomsma, D. I., de Geus, E.J.C. & Bartels, M. (2022). Self-rated health when population health is challenged by the COVID-19 pandemic; a longitudinal study. *Social Science & Medicine,* 306, 115156.

*supplementary materials accessible at:
https://doi.org/10.1016/j.socscimed.2022.115156

ABSTRACT

The coronavirus disease 2019 (COVID-19) pandemic and consequent lockdown measures have had a large impact on people's lives. Recent evidence suggests that self-rated health (SRH) scores remained relatively stable or increased during the pandemic. For the current project, we examine potential changes in the variance decomposition of SRH before and during the COVID-19 pandemic in the Netherlands. We analyse data from the Netherlands Twin Register to examine pre-pandemic SRH scores (*N*=16,127), pandemic SRH scores (*N*=17,451), and SRH difference scores (*N*= 7,464). Additionally, we perform bivariate genetic analyses to estimate genetic and environmental variance components in pre-pandemic and pandemic SRH, and estimate the genetic correlation to assess potential gene-environment interaction. The majority of the sample (66.7%) reported the same SRH before and during the pandemic, while 10.8% reported a decrease, and 22.5% an increase. Individuals who reported good/excellent SRH before the pandemic were most likely to report unchanged SRH during the pandemic, and individuals with bad/mediocre/reasonable SRH more often reported increased SRH. The bivariate longitudinal genetic model reveals no significant change in variance decomposition of SRH from before to during the pandemic, with a heritability estimate of 45% (CI 36% to 52%). We found that the genetic correlation could be constrained to 1, and a moderate unique environmental correlation (r_E = .49, CI = .37 to .60). We theorize that the increases in SRH are explained by uninfected individuals evaluating their health more positively than under normal circumstances (partly through social comparison with infected individuals), rather than actual improvements. As the same genes are expressed under different environmental exposures, these results imply no evidence for gene-environment interaction. While different environmental factors might influence SRH at the two time-points, the influence of environmental factors does not become relatively more important during the pandemic.

1. INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic has had, and continues to have, an enormous impact worldwide. Even when not infected, people suffer from the consequences the preventive measures have on their daily life. Many efforts have been taken to slow down infection rates, such as social distancing policies, and the shutting down of schools, restaurants, and other public facilities. While these regulations are necessary to ensure sufficient capacity in intensive care (IC) units, they also have large economic and public health consequences. An important question, in this regard, is what the effect of the COVID-19 pandemic is on population health and well-being[243]. Answering this question requires world-wide research in various populations and settings due to differences in policies and (lockdown) measures for different populations. The consequences for the populations' health and well-being are expected to vary from country to country not only due to differences in COVID-19 prevalence and regulations, but also because differences in well-being and health care already existed before the commencement of the pandemic[244].

In the Netherlands, COVID-19 spread rapidly in March 2020, leading to a so-called *intelligent lockdown* with stay-at-home and social distancing measures starting on March 12, 2020. During this intelligent lockdown, people were allowed to leave their homes and go outside for walks or work-outs, but public spaces such as shops, schools, bars, and restaurants were closed and people were asked to work from home. In this way, people were much more restricted than in their usual pre-pandemic lives, but the government also appealed to a self-discipline principle that allowed them to retain some authority over their lives. Studies that investigated the health consequences of the COVID-19 pandemic in the Netherlands have mainly focussed on the consequences for mental health. For example, in a study in Dutch older adults, it was found that measures for social distancing led to higher levels of loneliness, but that mental health remained relatively stable[245]. In another study, mental health status during the pandemic was compared to retrospective reports of pre-pandemic mental health in a population-representative sample, with 80% of the participants reporting no change in mental health since the beginning of the pandemic. Moreover, being male and having high pre-pandemic levels of positive well-being seemed to act as protective factors for well-being[246].

While mental health is a very important aspect of people's health, it does not cover the concept of health in its entirety. While it is obvious that many people with a current

or past COVID-19 infection suffer from health consequences, people who have not been infected might also suffer from indirect health consequences, for example due to changes in diet, exercise, and sleeping patterns, stress and loneliness[247–251]. An interesting question in the context of the COVID-19 pandemic and consequent lockdown measures is to what extent it has affected people's Self-Rated Health (SRH). Typically measured using a single Likert scale question, SRH is a reliable and valid measure of subjective health as measured by other indicators in many population groups[252]: it is a good predictor of mortality and chronic or severe diseases[253,254], and higher SRH is associated with better mental health and well-being[32,255]. Thus, SRH predominantly measures people's subjective perceptions of health, but is also associated with objective health status. In addition, SRH is widely used to study trends and socio-economic inequalities in population health[256]. Existing literature suggests that the concept of SRH is useful both as a spontaneous assessment of health and for more enduring evaluations of one's health. Measures of SRH are responsive to changes in health status such as changes in mental well-being, but it also seems to be a relatively stable measure over time, supporting the role for an enduring self-concept of health[257]. Changes in SRH have been linked to several factors, such as changes in income[258], different physical and psychosocial work factors[259], and lifestyle characteristics such as changed physical activity or dietary habits[260].

With respect to the COVID-19 pandemic, a study in large German sample examined changes in SRH from before to during the pandemic[261]. More than half of the participants (56%) reported no changes in their SRH, while 32% reported improved SRH, and 12% reported a decrease. Most participants who reported worsened SRH had been tested for COVID-19. Similarly, in a study with French respondents, more people reported to be in very good health during lockdown compared to between 2017 and 2019[262]. The authors refer to this finding as an "eye of the hurricane" paradox, where individuals who are not infected by the COVID-19 virus might evaluate their health more positively than they normally would. A number of studies have assessed factors that potentially predict changes in SRH in (the beginning of) the pandemic. Bierman and colleagues find that baseline SRH and baseline psychological distress are associated with SRH during the pandemic, with individuals reporting greater distress and lower SRH before the pandemic also reporting lower SRH during the pandemic[263]. A similar result was found in a study by Szwarcwald and colleagues, where the proportion of individuals reporting decreased SRH during pandemic was larger for individuals reporting bad baseline SRH compared to those reporting good baseline SRH[264].

Individual differences in SRH are accounted for by both genetic and environmental factors unique to the individual, with heritability estimates ranging from 25% to 64%[265,266]. The variation in heritability estimates may reflect population differences or changes in the relative role of genetic and environmental factors across the lifespan. In a large longitudinal study of Finnish twins, the heritability of SRH peaked at 63% at age 16, but declined to 33% at age 25[265]. The study also found that genetic factors were primarily responsible for moderate correlations between health ratings at different life stages. In contrast, Mosing et al[267] observed a heritability of 46% in a sample of elderly Australian twins, and observed increasing heritability and genetic variance of SRH in older age groups among Swedish twins. It is important to keep in mind that heritability reflects the *relative* influence of genetic factors. This means that if environmental variance increases, the relative influence of genetic factors will decrease. In case of a large environmental change, such as a pandemic, heritability estimates may thus change. Additionally, new genetic variation might emerge in different environmental situations, e.g. the presence of stressors in an environment might lead to stress-specific genetic variation[268], a phenomenon that is known as gene by environment (GxE) interaction.

In this paper, we examine the effect of the COVID-19 pandemic on SRH during the first intelligent lockdown in the Netherlands in persons that did not have noticeable COVID-19 symptoms. SRH scores from before the pandemic are compared to scores during the first months of the intelligent lockdown. Moreover, a genetically informative design is applied that allows us to decompose variance in, and covariance between, SRH at the two time-points. More specifically, we assess whether the total genetic and environmental variance changes (quantitative gene-environment interaction). If the pandemic leads to an increase in unique environmental variance, and genetic effects remain stable, then the relative influence of environmental variance will increase while the relative influence of genetic factors on individual differences (the heritability) will decrease. An increase in environmental variance is expected if the environmental changes brought by the pandemic do not impact everybody in the same way (e.g. people with different professions and different household compositions may be differently impacted by work-from-home policies, effectively amplifying existing differences between individuals). Additionally, the genes and environmental factors that influence SRH under "normal conditions" may be at least partially different from those influencing SRH under a different, perhaps more stressful, environment during the pandemic. We therefore also

examine whether different genes influence SRH during the pandemic by assessing the genetic correlation (qualitative gene-environment interaction).

2. MATERIALS AND METHODS

2.1. Sample

All study participants were registered with the Netherlands Twin Register (NTR)[63]. Every couple of years, NTR participants are asked to fill out a survey including questions about their health, lifestyle, personality, well-being, and other life domains. For the present study, we compared pre-pandemic SRH data collected between 2014 and 2020 to pandemic SRH data collected in April and May 2020 (the first lockdown in the Netherlands). Data were collected in twins and multiples and family members who were 16 years or older. For the pre-pandemic sample, we used data collected in two questionnaires: one collected in 2014-2015, and one collected in 2019-2020. The means and variances of SRH were very similar for the observations from 2014-2015 (M = 3.98, SD = .72, variance = .52) and the observations from 2019-2020 (M = 3.95, SD = .72, variance = .52). In predicting pandemic SRH from pre-pandemic SRH, adding the number of years between the pre-pandemic and pandemic data-points as a predictor significantly improved the prediction model (p = .005), but the change in R^2 (.001) was negligible. When participants filled out both pre-pandemic questionnaires, we used data from the last questionnaire. For twin analyses (see below), we only included twin pairs where both twins had data available from the same survey.

Sample characteristics can be found in Table 1. Since the focus of our study was to examine the effect of the pandemic in general, and not the disease itself, we excluded individuals who tested positive for COVID-19 or had an expected COVID-19 diagnosis based on the Menni model[269] (more information in measures section). In total, 517 participants were excluded due to (expected) COVID-19 infection, of whom 217 had data on both time-points. As seen in table 1, these individuals were on average younger than the general sample and scored lower on SRH. In total, pre-pandemic SRH data were available for 16,127 participants (5,602 males and 10,525 females). After excluding cases, pandemic SRH data were available for 17,451 participants (5,065 males and 12,386 females). Of these people, 7,464 had data available at both time-points (2,214 males and 5,250 females).

With respect to missingness, 8,623 individuals responded to a pre-pandemic survey but did not respond to the pandemic survey, and 8,798 individuals failed to respond to the pre-pandemic survey but did respond to the pandemic survey (see Table 1). Logistic regression indicates that missingness for the pandemic survey was not completely at random, with pre-pandemic SRH (β = -.16, *SE* = .02, p < .001), age (β = -.02, *SD* = .001, p < .001), and gender (β = -.51, *SE* = .04, p < .001) predicting missingness for the pandemic survey. However, individuals who had pre-pandemic data available but did not respond to the pandemic questionnaire scored very similar on SRH (*M* = 3.96, *SD* = .72) as individuals who filled out both questionnaires (*M* = 3.99, *SD* = .71). The group that filled out both questionnaires had a lower percentages of males (29.7%) and were slightly older (*M* = 44.63, *SD* = 16.44) compared to the group that only responded to the pre-pandemic surveys (38.9% males, M_{age} = 38.91, SD_{age} = 15.85). With respect to missingness for the pre-pandemic survey, we could predict this missingness with age (β = -.04, SD = .001, p < .001), but not with pandemic SRH (β = -.03, *SD* = .02, p = .22) or sex (β = -.06, *SD* = .04, p = .08). Compared to respondents with both pandemic and pre-pandemic data, individuals that did not have pre-pandemic data were younger (*M* = 40.62, *SD* = 13.16) than individuals who had data at both time-points (*M* = 47.46, *SD* = 15.13).

Table 1.

Descriptive statistics for SRH

	***N* (males/females)**	***M*(*SD*) age**	***M*(*SD*) SRH**	**Var (Range) SRH**
Excluded COVID cases	517 (241/330)	36.02 (12.90)	4.01 (.81)	.66 (1 to 5)
Pre-pandemic questionnaire	16127 (5602/10525)	41.47 (16.37)	3.97 (.72)	.52 (1 to 5)
Pandemic questionnaire (excl. cases)	17451 (5064/12387)	44.63 (14.80)	4.12 (.68)	.46 (1 to 5)
Overlap (excl. cases)*	7464 (2214/5250)	44.63 (16.44)	3.99 (.71)	.51 (1 to 5)
Overlap (excl. cases)**	7464 (2214/5250)	47.46 (15.13)	4.11 (.70)	.49 (1 to 5)
Non-overlap (only pre-pandemic)	8623 (3353/5270)	38.91 (15.85)	3.96 (.72)	.52 (1 to 5)
Non-overlap (only pandemic)	8798 (2433/6365)	40.62 (13.16)	4.14 (.66)	.43 (1 to 5)
Difference scores			.12 (.61)	.38 (-3 to 3)

Note. *N*= Sample Size, *M*=Mean, *SD*=Standard Deviation, Var=Variance, SRH= Self-Rated Health

* pre-pandemic descriptives

** pandemic descriptives

2.2.Measures

SRH was measured with the single item 'In general, how would you rate your health?'. In both questionnaires, there were five answer options which are scored on a five point scale with 1="bad", 2="mediocre", 3="reasonable", 4="good", and 5="excellent". This single item assessment of SRH is recommended by the World Health Organization (WHO) and validated across many studies and contexts[252,270,271].

COVID-19 infection status was assessed by two methods: First, by asking participants if they had been tested for COVID-19, and if so, whether an infection was confirmed (0=No, 1=Yes). Since there was limited testing in the Netherlands at the time of data collection, it is likely that many people remained undiagnosed at that time. Therefore, we also enquired the extent to which participants experienced a range of symptoms since February 20 (on a 5-point scale) and used the Menni self-reported symptom-based prediction model[269] to predict whether a person likely had COVID-19 (see original paper for more details). We excluded individuals if they reported having been tested positive, or were predicted to have been infected based on the Menni model.

2.3. Statistical Analyses

2.3.1. Pre-pandemic to pandemic comparison

For the pre-pandemic and pandemic SRH, we computed the means, variances, and min-max range. For the subset of participants with data available for both surveys, we calculated within-person difference scores by subtracting the pre-pandemic questionnaire scores from the pandemic questionnaire scores for SRH. Means were compared in a genetically unrelated subsample using a paired-samples t-test. Statistical tests were performed in R[272].

2.3.2 Bivariate genetic models

In a bivariate genetic model for twin data, we quantified the contribution of genetic and environmental factors to pre-pandemic and pandemic SRH (excluding COVID-19 cases) and the stability of SRH over time. These models rely on the fact that monozygotic (MZ) twins share (nearly) 100% of their genes, while dizygotic (DZ) twins share on average 50% of their segregating genes. This makes it possible to decompose (co)variation in a set of traits into four potential sources: additive

genetic (A) factors (shared 100% by MZ twins and 50% by DZ) twins, dominant genetic (D) factors (shared 100% by MZ twins and 25% by DZ twins), common environmental factors (C) (shared completely by both types of twins) and unique environmental (E) factors (unshared environmental factors and measurement error. When the MZ correlation is less than twice the DZ correlation, an ACE model is used. When the MZ correlation is twice or more the DZ correlation, and ADE model is used. Based on earlier research on SRH in the Netherlands, we expect the twin correlations to reflect an ADE model [273]. Twin correlations and cross-twin cross-trait correlations were estimated in saturated models in which all parameters (means, covariates, variances, and covariances) were freely estimated.

We performed the bivariate genetic analyses using the variance component approach[274] in OpenMx[275] (see Figure 1). Since SRH was measured on an ordinal scale, we fitted threshold models to the data, with gender and age as covariates. These models assume that categorical variables have an underlying liability with a continuous and standard normal distribution. We used 1 threshold to divide the liability distribution into two discrete categories, one representing less than good health and one representing good/excellent health. The contribution of the A and D variance components was estimated using full-information maximum-likelihood estimation and tested for significance by dropping these components one by one. By fitting the model with and without the constraints of interest, a log-likelihood ratio test (LRT) can be used to compare the nested sub-models models. The more parsimonious model is rejected if the log-likelihood statistic exceeds the chosen p-value threshold. We chose a p-value threshold of p=.005, in line with the reasoning described in Benjamin et al.[276].

We tested for potential gene-environment interaction in two steps. First, we constrained the genetic correlation to 1 and compared the fit of the model where the genetic correlation could be freely estimated to the fit of the model where the genetic correlation was constrained to 1 with a log-likelihood ratio test. If the fit of the constrained model is significantly worse than the fit of the unconstrained model, it indicates the genetic correlation cannot be constrained to 1 and thus that different genes influence SRH at the two time-points, pointing at qualitative gene-environment interaction[277]. That is, given a change in environmental conditions, we can test in the longitudinal data if the environmental change triggers a change in the genes that are expressed. In the same model, we also tested for quantitative

gene-environment interaction by comparing the contribution of the variance components during the pandemic compared to pre-pandemic. If the amount of genetic/environmental variance changes significantly from pre-pandemic to pandemic, it indicates that genes interact with environmental change in the form of quantitative gene-environment interaction. We tested this using a log-likelihood ratio test where a model where the genetic variance components were constrained to be equal were compared to the unconstrained model. This constraint was applied by setting the variance explained by genetic factors pre-pandemic (A11) equal to the variance explained by genetic factors during the pandemic (A22).

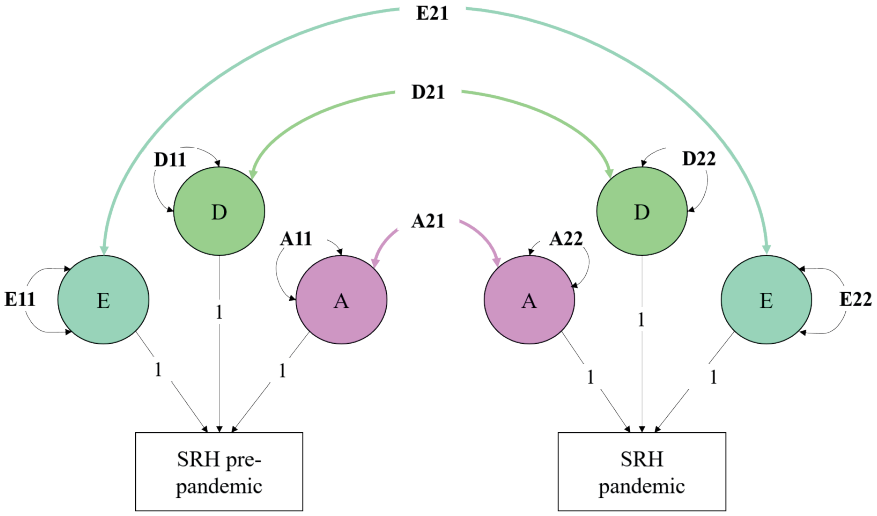


Figure 1. Variance decomposition of SRH into additive genetic (A), dominant genetic (D), and unique environmental (E) variance components.

3. RESULTS

3.1. Pre-COVID to COVID-19 Comparison

Descriptive statistics for the full sample for SRH at both time-points and the SRH difference scores can be found in Table 1. Excluding (suspected) COVID cases, 807 participants (10.8%) scored lower, 4,975 participants (66.7%) scored the same, and 1,682 participants (22.5%) scored higher on the pandemic SRH measure vs. the pre-pandemic SRH measure. Figure 2 and Table 2 depict percentages of respondents per pre-pandemic SRH category (the different colours) categorized

by their pandemic SRH score (the different columns). To illustrate, the red bar in the column "mediocre" visualizes the percentage of individuals that indicated feeling bad before the pandemic, but indicated feeling mediocre during the pandemic (50%). Figure 3 shows the percentage of individuals from each pre-pandemic SRH category with decreased, increased, and stable SRH during the pandemic. Respondents who indicated having good or excellent SRH before the pandemic were relatively most stable, with 72.8% (*N* = 3,353) and 72% (*N* = 1,098) of participants scoring in these respective categories scoring in the same category during the pandemic. About half of the respondents indicating bad (50%, *N* = 18), mediocre (45.7%, *N* = 101) or reasonable SRH (52.1%, *N* = 559) before the pandemic scored one category higher on SRH during the pandemic. For those with decreased SRH levels during the pandemic, the most common decrease was from excellent to good (*N* = 406, 26.6% of individuals with excellent pre-pandemic SRH). In a genetically unrelated sample of participants who provided data at both time-points, mean SRH scores were significantly lower in the pre-pandemic questionnaire (*M* = 3.96, *SD* = .72) compared to the pandemic questionnaire (*M* = 4.09, *SD* = .68) (M_{diff} **= -**.12, *t*(4025) = -12.67, $p < 2.2x10^{-16}$). Supplementary Figure 1 provides histograms of the distribution of pre-pandemic SRH, pandemic SRH, and SRH difference scores. These figures reveal that SRH is not normally distributed at both time-points, but that the difference scores are approximately normally distributed (as assumed by a paired-samples t-test).

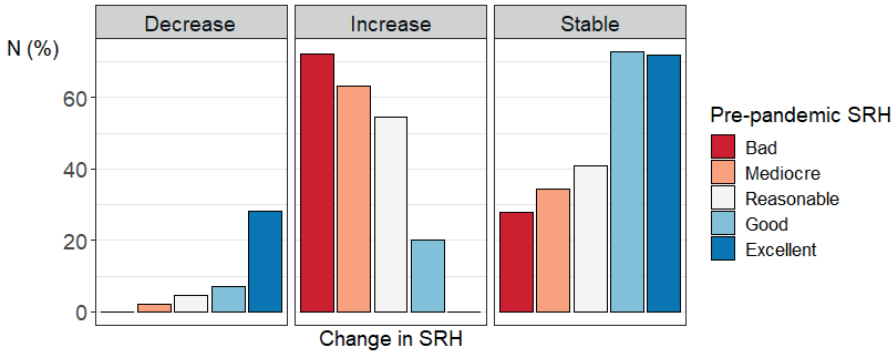


Figure 2. The percentage of individuals from each pre-pandemic self-rated health (SRH) category with decreased, increased, and stable SRH during the pandemic.

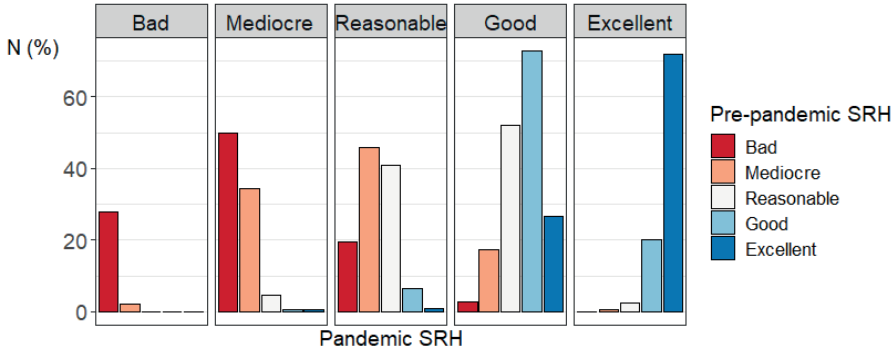


Figure 3. Percentage of individuals per pre-pandemic self-rated health (SRH) category categorized by pandemic SRH score.

Table 2

Cross table of pre-pandemic and pandemic self-rated health (SRH) scores.

Pandemic SRH		Bad	Mediocre	Reasonable	Good	Excellent	Total
Pre-pandemic SRH	**Bad**	10 (27.8%)	18 (50%)	7 (19.4%)	1 (2.8%)	0 (0%)	36
	Mediocre	5 (2.3%)	76 (34.4%)	101 (45.7%)	38 (17.2%)	1 (0.5%)	221
	Reasonable	0 (0%)	49 (4.6%)	438 (40.8%)	559 (52.1%)	27 (2.5%)	1073
	Good	2 (0%)	21 (0.5%)	302 (6.6%)	3353 (72.8%)	930 (20.2%)	4608
	Excellent	0 (0%)	8 (0.5%)	14 (0.9%)	406 (26.6%)	1098 (72%)	1526
	Total	17	172	862	4357	2056	7464

3.2 Bivariate genetic models

The overall phenotypic correlation between SRH at the two time-points in our sample is .72 (CI .66 to .77). The twin correlations and cross-twin cross-trait correlations from the saturated model are displayed in Table 3. The pre-pandemic MZ correlation (r_{MZ} = .54, CI = .42 to .64) and pandemic MZ correlation (r_{MZ}= .44, CI = .31 to.56) were larger than twice the pre-pandemic DZ correlation (r_{DZ} = .12, CI = -.11 to .33) and pandemic DZ correlation (r_{DZ}= .15, CI = -.09 to .38), indicating both the presence of additive (A) and dominant genetic influences (D).

Table 3

Twin correlations

	MZ	
	SRH pre-pan	**SRH pan**
SRH pre-pan	.54 (.42 to .64)	
SRH pan	.37 (.23 to .50)	.44 (.31 to .56)
	DZ	
	SRH pre-pan	**SRH pan**
SRH pre-pan	.12 (-.11 to .33)	
SRH pan	.22 (-.06 to .47)	.15 (-.09 to .38)

SRH pre-pan = Self-rated health pre-pandemic*, SRH pan*= Self-rated during the pandemic, *MZ*= monozygotic, *DZ*=dizygotic.

The full model fitting results can be found in Table 4. Dropping the D component, resulting in an AE model, did not lead to significantly worse model fit compared to the full ADE model (∆-*2LL*(∆*df*) = 2.47(3), *p* = 0.48). Additionally, constraining the genetic correlation to 1 also did not result in a significantly worse model fit (∆-*2LL*(∆*df*) = 1.27(1), *p* = 0.26), indicating an absence of qualitative gene-environment interaction. Lastly, constraining the variance components to be equal also did not result in a worse model fit (∆-*2LL*(∆*df*) = 2.05(1), *p* = 0.15), indicating the absence of quantitative gene-environment interaction. In this final model, the heritability for both traits was A = .45 (CI .36 to .52), indicating that 45% of individual differences in SRH could be explained by genetic factors, both before and during the pandemic. The other 55% could be explained by unique environmental differences (E = .55, CI = .48 to .64).We found a moderate unique environmental correlation (r_E =.49, CI =.37 to .60) indicating that partly different environmental factors influence SRH at the two time-points.

Table 4

Bivariate model fitting results and parameter estimates

Model fitting results							**Standardized parameter estimates**					
							SRH pre-pandemic			**SRH pandemic**		
Model	vs.	*-2LL*	*df*	χ^2	Δ *df*	*p*	A	D	E	A	D	E
1. Saturated	-	4057.41	5617	-	-	-	-	-	-	-	-	-
2. ADE	1	4069.94	5628	12.53	11	0.33	.01	.53	.46	.16	.27	.57
3. AE	2	4072.41	5631	2.47	3	0.48	.52	-	.48	.42	-	.58
4. AE, rA=1	3	4073.68	5632	1.27	1	0.26	.51	-	.49	.38	-	.62
5. AE, rA=1, A11=A21	**4**	**4075.73**	**5633**	**2.05**	**1**	**0.15**	**.45**	**-**	**.55**	**.45**	**-**	**.55**

Note. Best fitting model is presented in bold. Vs.= versus, -2LL = -2 Log Likelihood, df = degrees of freedom, χ^2 = chi-square statistic, Δ df = difference in degrees of freedom, p = p-value, SRH = self-rated health, A = proportion of variance due to additive genetic factors, D = proportion of variance due to dominant genetic effects E = proportion of variance due to unique environmental effects.

4. DISCUSSION

In this study, we examined changes in SRH from before the COVID-19 pandemic to during the beginning of the pandemic in a Dutch sample. When we compared the average SRH before the pandemic with the average SRH during the pandemic, we find that (on average) SRH has increased. We observed individual differences in how people's reports of SRH changed. While the majority of the sample (66.7%) did not report a change in their SRH, about one in ten (10.8%) reported a decrease, and about two in ten (22.5%) reported an increase.

The finding that most people's SRH did not change suggests that individuals were quite resilient during the beginning of the COVID-19 pandemic. Importantly, these results only pertain to the beginning of the pandemic, and it is possible that changes in SRH might only reveal themselves over a longer period of time. Having said this, we did find that more participants report an increase rather than a decrease in SRH, which is consistent with previous studies on SRH in the beginning of the pandemic[261,262]. More specifically, we found that individuals who already reported good or excellent SRH before the pandemic were most likely to report unchanged SRH during the pandemic, and that individuals with bad/mediocre/reasonable SRH were most likely to report increases in SRH. These effects might partially reflect floor and ceiling effects, where we would not be able to detect increases in health

for individuals indicating excellent pre-pandemic health, and where we would not be able to detect decreases in health for individuals indicating bad pre-pandemic health. However, since we found a similar result for those with bad pre-pandemic scores as those with mediocre and reasonable scores, and a similar result for those with excellent and good scores, it is unlikely that floor or ceiling effects are the primary explanation for these results. When comparing pre-pandemic SRH in the Netherlands to SRH of other European countries based on Eurostat data, the Netherlands scores higher than most other European Union countries with 77.2% of males and 72.6% of females indicating good or very good self-perceived health in 2019[278]. Similarly, the 2019 OECD report indicates the Netherlands (together with Japan, Spain, and Switzerland) to have the best overall health outcome globally based on life expectancy, avoidable mortality, chronic disease morbidity, and SRH[279]. In the context of earlier research identifying a positive association between baseline SRH and pandemic SRH[263], the relatively high baseline SRH in the Netherlands might have served as a protective mechanism for maintaining good health during the pandemic. However, since our current dataset does not allow for such cross-country evaluations, we can only speculate on this point.

There are different explanations for why individuals might evaluate their health more positively during the pandemic. First, it is possible that people adapted different health habits (e.g. an altered diet, changed physical activity patterns) that improved their health, thus leading to an increase in SRH. The current literature on this topic is mixed. For example, while there are studies reporting increases in physical activity during the lockdown[280,281], the majority of studies report decreased physical activity and increased sedentary behaviour during the COVID-19 lockdowns[282]. Additionally, it is possible that health conditions that were present in the pre-pandemic measure (i.e. disease or illness), were no longer present of improved at the time of the pandemic measure. Since we did not include objective disease indicators, we cannot rule out the possibility that the observed average increase in SRH reflects objective health increases. However, given that many diseases or illnesses have longer lasting effect, it is more to be expected that health deteriorates over time than that is improves. If, for example, we compare the number of individuals with one or more chronic illnesses (associated with COVD-19 related death) before the pandemic to during the pandemic, this number increases from 1103 to 1247 during the pandemic (see supplementary analyses). While this does not tell us anything

about symptom severity, it is at least an indication that the number of individuals with a chronic condition did not decrease.

Another explanation is that people's perception of their health might have changed, even if their objective health did not change. The item we used to measure health was designed to measure subjective, rather than objective, health. While this is an approximation of one's objective health, there are also other factors that contribute to subjective health, such as the context in which one finds themselves. As mentioned in the introduction, a previous study explained the apparent increase in SRH during the pandemic as an "eye of the hurricane" paradox, where individuals who are not infected by the virus evaluate themselves more positively than under normal circumstances[262]. A mechanism that might contribute to this paradox is social comparison: people partly rate their health based on how healthy they perceive their peers[283]. With respect to the pandemic, individuals who remain uninfected by the virus might rate their health more positively than before, as they can now compare themselves to those who have been infected. This is in line with our finding that it was especially those with bad/mediocre/reasonable pre-pandemic SRH that indicated higher pandemic SRH, while respondents indicating good/excellent pre-pandemic SRH more often remained stable. While we did not collect data on changes in health patterns or comparative SRH ratings (i.e. where people explicitly rate themselves as compared to those around them), it would be an interesting direction for future research to elucidate which mechanisms might be at play.

Second, we examined the genetic and environmental sources of individual differences in SRH across the two time-points. Our results indicate that the genetic architecture of SRH does not change from before to during the first lockdown. We report heritability estimates of 45% (CI 35 to 52%), which is well within the range of findings from previous research[265–267]. It seems that the early stages of the pandemic did not moderate the strength of the relative influence of genes and the environment on SRH. As mentioned in our introduction, a change in variance decomposition was to be expected if people were impacted dissimilarly by the pandemic, leading to an increase of environmental variance. However, the fact that we did not find an increase in total environmental variance does not necessarily mean that the pandemic impacted all respondents in the same way. The unchanged variance may be explained by the high baseline levels of SRH,

which potentially served as a protective mechanism for environmental change, even if environmental circumstances did not change similarly for different respondents. Additionally, the genetic correlation indicates that it were still the same genes that influenced differences in SRH at the two time-points. Lastly, environmental correlations indicate that it is (partly) different environmental factors that influence differences in SRH during the pandemic compared to before the pandemic.

While our bivariate genetic model indicates that partially different environmental factors influence SRH during the pandemic compared to before the pandemic, it does not provide information about which particular factors might be different. Previous research suggests that older cohorts are more likely to report changes in behaviour during the COVID-19 pandemic in terms of stress, sleep, physical activity, diet and alcohol intake compared to younger cohorts[284,285]. Moreover, differences between males and females have been found to be larger during the lockdown compared to before lockdown, with females reporting more atypical sleep levels and higher stress levels[285,286]. In this way, the pandemic might have caused existing differences between age and gender groups to become enlarged. With respect to potential environmental factors uniquely influencing SRH during the pandemic, existing research has pointed out several COVID-related stressors that might impact people's health. Examples include worry and psychological distress about risk for COVID-19 and the consequences of the pandemic[287,288], working in a high-risk profession, e.g. healthcare[289], and social distancing with consequent impaired social connectedness[290]. While population-level environmental variance did not change significantly during the pandemic compared to before the pandemic, different environmental factors became important in explaining individual differences in SRH during the pandemic. Research into identifying these specific environmental factors is important since it can be used to inform policy makers on SRH variation during crisis-situations like the COVID-19 pandemic.

Our findings should be interpreted in light of some limitations. First, as we mentioned earlier in the discussion, SRH ratings in general are partly due to comparison to other people. During the pandemic part of the 'other people' suddenly became ill of COVID-19. This resulted in an overall increase in SRH for those not affected at the moment of measurement. Of course, this does not have to reflect an absolute increase in health but probably reflects the relative

change of self-rated health in comparison to others. Following this logic, the issue is not that we are measuring something different at both time-points, but that different mechanisms influence the construct at the two time-points (a reasoning consistent with our finding that different environmental factors influence SRH at the two time-points). In addition, these findings may be somewhat limited by the representativeness of our sample. The sample used for this study was a subset of NTR participants that had both pre-pandemic and pandemic SRH data available. This particular subset unfortunately included more women (±70%) than men (±30%). Moreover, almost 60% of the sample indicated they attended higher vocational school or university, while in the average Dutch population, only about 30% of the population attends higher vocational school/university[291]. Since both education attainment and gender are associated with SRH, caution must be applied in interpreting our findings. Additionally, pre-pandemic SRH, gender, and age were associated with missing SRH pandemic data, and age was also associated with missing pre-pandemic data. However, since the differences between the overlapping and non-overlapping sample on these variables were very minor, we do not expect this had a large influence on our results. With respect to the potential influence of these confounders on our results, we ran supplementary analyses where we regressed gender, age, the presence of chronic illnesses, and educational attainment on pre-pandemic SRH, pandemic SRH, and SRH difference scores (see Supplementary Analyses). While all these factors were significant predictors of SRH at both time-points, none of the variables predicted SRH difference scores. Furthermore, although it has been observed that people from disadvantaged sociodemographic groups are more likely to change their SRH score over time[292], we do expect less of an effect of such inequalities in our analyses in a Dutch population based sample, because of the health care system in the Netherlands which provides basic health insurance to all citizens at affordable costs. Lastly, it is well possible that the influence of the pandemic and accompanying lockdowns on SRH changes over time. The results of this study pertain to the first lockdown in the Netherlands and thus reflect the immediate impact of environmental change in the form of a lockdown. Both the immediate impact and the longer term impact are interesting topics for the study of SRH, and we encourage researchers with multiple time point data during the pandemic to further explore individual differences in SRH during the pandemic.

These findings re-confirm that in the study of complex human traits, such as SRH, it is important to not only examine mean changes, but also examine individual differences. The finding that many people's SRH remained unchanged shows that there was quite a resilient response to the first stages of the COVID-19 pandemic in the Netherlands, likely driven by more positive perceptions of health during the pandemic, instead of actual health improvements. Moreover, the finding that the variance decomposition in terms of the relative influence of genetic and environmental factors does not change significantly between these two time-points indicates that, at least during the first lockdown, environmental influences did not become relatively more important. It would be interesting to see if this remains stable during longer time-frames, or whether as more time passes, the pandemic does start to moderate the strength of the relative influence of genes and the environment. Either way, our results indicate that while some people may be affected by the challenges posed by COVID-19 to the perception of their health, others are not.

CHAPTER 7

Genetic and environmental influences on quality of life: The COVID-19 pandemic as a natural experiment

Published as: van de Weijer, M. P., Pelt, D. H. M., de Vries, L. P., Huider, F., van der Zee, M. D., Helmer, Q., Ligthart, L., Willemsen, G., Boomsma, D.I., de Geus, E.J.C. & Bartels, M. (2022). Genetic and environmental influences on quality of life: The COVID-19 pandemic as a natural experiment. *Genes, Brain and Behavior,* e12796.

*supplementary materials accessible at: https://doi.org/10.1111/gbb.12796

ABSTRACT

By treating the coronavirus disease 2019 (COVID-19) pandemic as a natural experiment, we examine the influence of substantial environmental change (i.e., lockdown measures) on individual differences in Quality of Life (QoL) in the Netherlands. We compare QoL scores before the pandemic (*N*=25,772) to QoL scores during the pandemic (*N*=17,222) in a sample of twins and their family members. On a 10-point scale, we find a significant decrease in mean QoL from 7.73 (SD =1.06) before the pandemic to 7.02 (SD = 1.36) during the pandemic (Cohen's *d*= .49). Additionally, variance decomposition reveals an increase in unique environmental variance during the pandemic (0.30 to 1.08), and a decrease in the heritability estimate from 30.9% to 15.5%. We hypothesize that the increased environmental variance is the result of lockdown measures not impacting everybody equally. Whether these effects persist over longer periods and how they impact health inequalities remain topics for future investigation.

1. INTRODUCTION

Natural experiments pose a particularly interesting set of circumstances where an intervention is implemented that is not under the control of researchers[293]. With respect to research in the domain of public health and human behaviour, a great advantage of research on natural experiments is that it corresponds to "real world" conditions, in contrast to many controlled experiments. Additionally, natural experiment studies are essential for evaluating population-scale (health) interventions and changes where experimental manipulation or random allocation is not feasible. As a result, natural experiments can provide unique ecologically valid insights into health processes as they are naturally occurring.

A well-known example of a population-level natural experiment is the compulsory schooling age reform in the United Kingdom, where the minimum age at which students were allowed to leave school increased from 15 to 16 for everyone born on or after September 1st, 1957. An interesting finding in the context of this reform is that the additional year of education reduced the gap in unhealthy body size between those in the top and bottom terciles of genetic risk for obesity from 20 to 6 percentage points, thus benefitting those with a higher genetic risk for obesity[294]. Another interesting set of natural experiments is the introduction of national tobacco control policies in different countries. For example, a workplace smoke-free legislation was introduced in Ireland in March 2004. One of the results of this legislation was an immediate significant reduction in small-for-gestational birth rates, which was sustained over the post-ban period[295]. In the Netherlands, smoking prevalence decreased from 40-51% to 22-23% between 1993-1995 and 2009-2010, but no effect was seen on the heritability estimates of smoking[296].

These examples involve national-level policy changes aimed at improving population health. Another set of natural experiments is (natural) disasters with population-level consequences. For example, on March 11, 2011, Japan was struck by an earthquake and consequent tsunami, leading to the loss of ± 18,500 lives and ± 345,000 people suffering damages to (or loss of) their house[297] and many people suffered from posttraumatic stress disorder (PTSD) after this disaster. Hikichi and colleagues studied these events from a natural experiment perspective in order to gain knowledge on the association between social cohesion and the risk for PTSD[297]. They found that individual- and community-level social cohesion before the disaster were associated with a lower risk of showing PTSD symptoms following the disaster. Another disastrous event with population-level consequences was World War 2 (WW2). During the horrific events of WW2, many children were separated from their

parents. Pesonen and colleagues[298] studied the effects of being separated from both parents or only one's father (due to military service) on depressive symptoms later in life (around 60 years of age) in a Finnish cohort. They found that being separated from both parents (but not from only the father) led to higher levels of depressive symptoms later in life, illustrating the prolonged effects of early life stress on later-in-life outcomes. These examples illustrate how natural experiments can provide novel insights that would have been difficult to study under "normal" circumstances.

The difficulty in studying population-level changes is that rapid, large-scale policy or environmental changes are relatively rare. In the past year, large environmental changes occurred on a global scale due to the coronavirus disease 2019 (COVID-19) pandemic. In March 2020, the World Health Organization (WHO) officially declared a pandemic, as the virus spread quickly across many countries in the world. As a result, many countries enforced a lockdown with varying levels of regulations. In the Netherlands, a so-called "*intelligent lockdown*" was installed, meaning that public spaces, schools, restaurants, etc. were closed and that people were encouraged to work from home, but could still leave their house for walks and other outdoor activities. As a result, many people's lives changed profoundly from an economic, social, and physical perspective.

What these different aspects (economic, social, physical) have in common is that they are all related to mental health and well-being. In a meta-analysis by Prati and Mancini[299], the psychological impact of the COVID-19 pandemic lockdowns across 25 studies was evaluated in terms of both positive and negative psychological functioning. They found that lockdowns had a small but detrimental effect on mental health, as expressed in negative psychological functioning (i.e. anxiety, depression, substance use, sleep disturbances, suicide risk, negative affect, and general distress), but surprisingly the effects on positive psychological functioning were not significant. In a Dutch sample, specifically people without severe or chronic mental health disorders showed a slight increase in depression, anxiety, worry, and loneliness symptoms, whereas people with depressive, anxiety, or obsessive-compulsive disorders did not seem to have increased symptom severity during the pandemic compared to before[300]. Besides the effects of this large natural experiment on mean population levels of mental health, such an impactful natural experiment enables a unique study into causes of individual differences in mental health.

From a behaviour genetic perspective, the focus goes beyond mean levels changes to explain the causes of individual differences. It is well established that individual

differences in well-being are influenced by both environmental factors and genetic factors: research indicates that about 40% of individual differences in well-being is explained by genetic factors (the heritability), with the other 60% being explained by non-shared/unique environmental factors[301]. Research combining behaviour genetics and experiments is relatively scarce, and typically focuses on short-term interventions. For example, one might use the "method of co-twin control", where only one member of an identical twin pair receives an intervention[302]. This is an interesting way of studying the possible effect of the intervention while controlling for genetic confounding. Alternatively, we can study individual differences in the effect of an intervention by applying an intervention in a classical twin design. This design also provides information on stability and change of the sources of individual differences pre- and post-intervention. For example, Haworth and colleagues examined the influence of a 10-week positive psychology intervention on well-being in a sample of 750 twins, and found that the relative influence of genetic and environmental influences remained stable, but that (partly) different non-shared environmental factors influenced well-being post-intervention[303]. In a more recent study, a brief online mindset intervention increased the relative influence of additive genetic factors to individual differences in mindset[304]. The COVID-19 pandemic can serve as a natural experiment for the investigation of absolute and relative changes in the genetic and environmental causes of variation in well-being since we can compare the variance decomposition during the pandemic to before the pandemic. For two well-being related constructs, optimism and meaning in life, it was already found that the heritability during the pandemic was slightly lower compared to before the pandemic[305]. In addition to estimates of quantitative change such as lower heritability estimates, a study focusing on the qualitative aspects of the psychological responses to the COVID-19 crisis in young adults found a genetic correlation of 1 between pre-pandemic and pandemic purpose in life, indicating that the same genes affect this trait before and during the pandemic[306]. Optimism and meaning in life can be viewed as facets of well-being[307], but whether these effects are similar for other well-being measures, such as Quality of Life (QoL), remains unexplored.

In the present study, we explore the impact of the COVID-19 pandemic on individual differences in well-being, quantified as QoL, in the Netherlands. We use a unique dataset that is comprised of data from twin families (e.g. twins, siblings, parents, aunts, uncles, nephews, nieces: pedigree data) both before and during the pandemic to provide a useful account of how genetic and environmental influences may be impacted by substantial environmental change.

2. MATERIALS AND METHODS

2.1 Participants

Participants were voluntary registrants of the Netherlands Twin Register (NTR)[63]. NTR participants are recruited through birth felicitation services, city councils, and online platforms. Every couple of years, biological and non-biological family members are invited to partake in survey research on development, health, behaviour, and lifestyle. Relations among participants, i.e., pedigree structure information, is stored in the "Person Administration of the Netherlands Twin Register" (PANTER) database[308]. Within this database, family roles and relations (e.g. mother-offspring, sibling-sibling) among participants are stored, with unlimited one-to-one relation possibilities for each individual. Participants can have multiple roles and relations in the database. For example, a person can be a mother and a twin.

For the current project, we selected a sample with pre-pandemic QoL data, and a (partly overlapping) sample with pandemic QoL data. All participants were 16 years or older. For the pre-pandemic sample, QoL data were available for multiple waves of data collection. If multiple observations were available for an individual, we selected the most recent pre-pandemic observation (assessment data between January 2014 and February 2020). Within each family, if data for multiple siblings were available, we only selected data collected in the same data collection wave, in order to reduce potential time-dependent confounders. Additionally, if data from both parent or spouses were available, we selected their data such that the data from both parents/spouses were included from the same wave of data collection.

During the pandemic, we made use of a single wave of data collection, which took place in April and May 2020, during the first lockdown in the Netherlands. Because we were interested in the effects of the lockdown on genetic and environmental influences on QoL, and not the effect of being infected itself, we excluded individuals with an (expected) COVID-19 infection (see below for details). We included twins and higher-order multiples (e.g. triplets), parents, siblings, and spouses (of multiples). Nuclear family information and age per type of family member is presented in Table 1. In total, pre-pandemic QoL data were available for 25,772 individuals, and pandemic QoL data were available for 17,222 individuals, of whom 11,232 had data available at both time points. Across the whole sample, age ranged between 16 and 102. Supplementary Figures 1 and 2 visualize the pre-pandemic and pandemic age distributions, respectively.

Table 1

Pedigree composition

	Full Data				Sample overlap 2 time-points		
	pre-pandemic		pandemic		pre-pandemic	pandemic	
	N	M(SD) age	N	M(SD) age	N	M(SD) age	M(SD) age
Families*	16297	-	11960	-	8317	-	
Individuals	25772	41.73 (14.57)	17222**	44.79 (14.70)	11232	42.75 (15.30)	45.74 (14.66)
MZ males	1259	35.76 (17.56)	800	41.66 (17.71)	575	41.37 (18.69)	44.00 (17.41)
MZ females	3258	35.60 (16.28)	2522	38.93 (15.83)	1950	36.87 (16.84)	39.79 (15.67)
DZ males from DZ male pairs	896	30.09 (14.69)	481	35.94 (15.11)	329	33.82 (16.50)	37.04 (15.44)
DZ females from DZ female pairs	1740	30.99 (14.31)	1161	34.53 (13.82)	851	31.98 (14.90)	35.17 (13.66)
DZ males from DZ opposite-sex twin pairs	855	29.82 (13.41)	480	35.53 (14.95)	323	33.50 (16.30)	36.76 (15.08)
DZ females from DZ opposite-sex twin pairs	1488	27.95 (12.15)	940	32.01 (12.79)	689	29.17 (13.25)	32.58 (12.32)
Fathers	5709	49.87 (10.69)	2742	54.76 (10.97)	1762	53.36 (10.91)	56.35 (10.57)
Mothers	10087	45.88 (10.07)	6677	48.58 (10.43)	4493	46.78 (10.29)	49.74 (10.21)
Brothers	148	37.99 (17.28)	291	42.99 (16.84)	71	42.56 (18.04)	45.13 (16.92)
Sisters	254	36.48 (17.06)	711	39.47 (15.41)	156	39.29 (17.91)	42.37 (16.68)
Spouses of Twins	78	52.88 (13.10)	250	55.69 (12.40)	33	56.61 (12.77)	58.09 (12.47)

*Families could exist of only 1 person

** There are some individuals that fall outside of the pre-specified categories (i.e. children of twins), but these groups are very small.

2.2 Measures

Quality of Life

Well-being was assessed as QoL using a Dutch version of Cantril's Self-Anchoring Striving Scale[65]. Participants were asked the question: 'Where on the scale would you place your life in general? A score of 10 means the best life you can imagine, 1 means the worst life you can imagine.'. In one of the pre-pandemic questionnaires, the question was scored on a scale from 0 to 10 instead of 1 to 10. Since almost no participant scored a 0 (*N*=6) or 1 (*N*=3) on this question, these two answers were pooled together as one so that the question was scored similarly from 1 to 10 across the different questionnaires.

COVID-19 infection status

COVID-19 infection status was assessed by asking participants if they had been tested positive for COVID-19 based on a PCR-test. Additionally, since there was little testing in the Netherlands at the time of our pandemic data collection, we also enquired about the extent (on a 5-point scale) to which participants had experienced a range of symptoms since February 20 and used the Menni self-reported symptom-based prediction model[269,309] to predict whether a person likely had COVID at the time of assessment. Detailed information on the development and application of this variable can be found in the original study paper[269]. We excluded individuals from the pandemic sample if they reported having been tested positive (*N*=85), or were predicted to have been infected based on the Menni model (*N*=436).

2.3 Statistical analyses

2.3.1 Pre-pandemic to pandemic comparison

Means and standard deviations for pre-pandemic and pandemic QoL for individuals with different roles within families were calculated using R[74]. We selected a subsample of genetically unrelated individuals (n=8,529) and performed a paired-samples t-test to examine if QoL significantly changed from before to during the pandemic. Effect sizes were calculated using Cohen's d for paired samples. Additionally, we calculated within-individual difference scores that reflect individual change from before to during the pandemic.

2.3.2 Kinship Correlations

Kinship correlations were obtained to get a first indication of familial resemblance for QoL during and before the pandemic. We calculated the kinship correlations using the Kinship Correlation Generator Tool (https://github.com/matthijsz/KinshipCorrelationGenerator). This tool uses a pedigree file with parent-offspring relations and an individual level phenotype file as input to estimate correlations for different familial pairs, e.g. mono- and dizygotic twins, parent-offspring or cousin pairs. Weights are assigned to each pair of observations based on the number of times each individual is included in relation to different people. Per kinship relation we obtained: 1) a correlation between relatives for pre-pandemic QoL, 2) a correlation for pandemic QoL, 3) a correlation between pre-pandemic QoL in individual 1 and pandemic QoL in individual 2 for each pair of relatives, and 4) a correlation between pandemic QoL in individual 1 and pre-pandemic QoL in individual 2 for each pair of relatives. Thus, correlation 1) and 2) were correlations within time-points, while 3) and 4) were cross-time correlations. These last two correlations between pre-pandemic and pandemic QoL were pooled with fixed effect meta-analysis in the *meta* package in R so that one cross-phenotype correlation is computed to be used for further interpretation. Since the tool does not provide standard errors or confidence intervals (CIs), these were calculated manually ($s_r = \sqrt{\frac{1-r^2}{n-2}}$).

2.3.3 Genetic analyses

We used the Mendel 16.0 software package "Variance Components" analysis option[310] to decompose (co)variation in QoL into additive genetic (A), dominant genetic (D), common/household environmental (C), and unique environmental (E, also includes measurement error) sources of (co)variation. Effects of age and sex were regressed out prior to the Mendel analyses, and subsequent analyses were conducted on the residual QoL scores[26]. Shared environmental influences were defined as influences that are shared by members of the same household. Since we are examining adults only, most (adult-aged) children within a nuclear family will not live in the same household. Therefore, a household effect was specified for spouses.

To perform variance decomposition in Mendel, three input files are required: 1. an input pedigree file, 2. a control file, and 3. a definition file.

1. The input pedigree file contains all the familial and phenotype data, grouped by family ID. Within the pedigree file, the following variables are required: family ID, person ID, and a Father and Mother ID, sex, and Twincode (an identifier for MZ twin pairs indicating which individuals are part of the same MZ twin pair). Genetic relationships between individuals within a pedigree with the same family ID are traced based on parental IDs (e.g. individuals within the same family with the same 2 parents are inferred to be full siblings). Our input pedigree file further specifies the household indicator field (in our case, spouse ID), and two fields for the (residualized) phenotype values for QoL before and during the pandemic.

2. The control file indicates all the analysis parameters. In our case, this includes the relevant variance components (A/C/D/E), the column names for the two quantitative traits present in the input pedigree file, the group factor specification (spousal household) and the way missing values are defined.

3. Lastly, the definition file provides information on non-mandatory variables in the pedigree file: the variable types (factor/variable), and the associated levels and bounds. Mendel uses the control and definition files to read in the data from the input pedigree file, and estimates variance components based on variance-covariance matrices for relatives with different degrees of genetic relatedness based on classical biometrical genetics[311].

We analysed four different models: 1) an ACDE model where C indicates the common household for spouses, 2) an ACE model, 3) an ADE model, and 4) an AE model. We compared the different nested models by comparing the log likelihood (LL) of the full ACDE model to the LL of the nested sub-models using a -2 log likelihood (-2LL) test that approximately follows a χ^2 distribution. Genetic and environmental correlations between the variance components were calculated by dividing the covariance of between pre-pandemic and pandemic QoL variables by the square root of its underlying variances[312]. These genetic and environmental correlations reflect the extent to which similar genes and environmental factors influence QoL at the two time-points.

3. RESULTS

3.1 Pandemic to pre-pandemic comparison

Mean QoL scores for the different groups can be found in Supplementary Table 1. Across the full sample, mean QoL decreased from 7.73 (SD =1.06) before the pandemic to 7.02 (SD = 1.36) during the pandemic. A paired samples t-test in an unrelated subsample (n=8,529) indicates this difference to be significant ($t(8528)=45.57$, $p<2.2\times10^{-16}$), indicating that QoL scores significantly decreased during the pandemic. The Cohen's *d* statistic (.49) indicated a medium effect size. Individuals with pre-pandemic data but without pandemic data (non-responders) did not score differently on the pre-pandemic QoL measure than individuals who provided data for both time-points (responders) ($M=7.71$, $SD=1.09$).

Within-individual change scores for the whole sample are visualized in Figure 1. A negative score indicates QoL decreased from before to during the pandemic, while a positive score indicates an increase in QoL. In total, QoL scores decreased for 6,183 (55.05%) individuals, remained stable for 3,239 (28.84%) individuals, and increased for 1,810 (16.11%) individuals. From the group of individuals that indicated decreased QoL, 1,158 (18.73%) individuals went from "sufficient" QoL before the pandemic (indicated by a 6 or higher), to "insufficient" QoL during the pandemic (indicated by a 5 or lower).

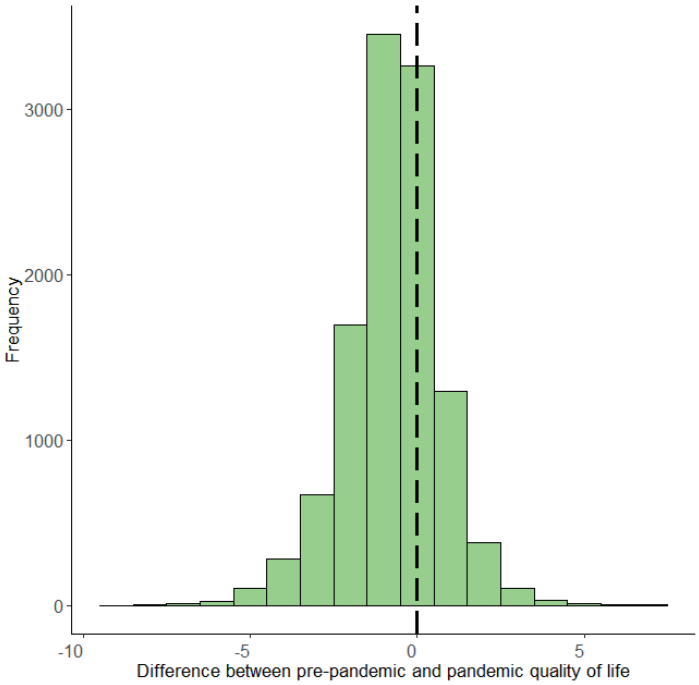


Figure 1. Histogram of Quality of Life (QoL) difference scores. The black dashed line indicates a change score of 0 or no change.

Figures 2 and 3 depict the number of individuals and percentage of individuals, respectively, that increased, decreased, and remained stable per QOL pre-pandemic score. As can be seen in Figure 2, pre-pandemic QOL scores are relatively skewed with most people indicating good pre-pandemic QOL. In general, the most common change was a decrease in QOL. Examining the group of respondents with decreased QOL during the pandemic in more detail (Figure 3), we see that individuals with high pre-pandemic QOL scores more often decreased during the pandemic compared to individuals with lower pre-pandemic QOL scores. With respect to the group of respondents that indicated increased QOL during the pandemic, it was especially individuals with lower pre-pandemic QOL scores that indicated higher scores during the pandemic. We also plotted the percentage of individuals that decreased, increased, or remained stable for QoL for different age groups separately in Supplementary Figure 3. Visual inspection of the plot does not reveal large differences between the age groups, with only a very slight trend of younger individuals being more negatively impacted in terms of QoL.

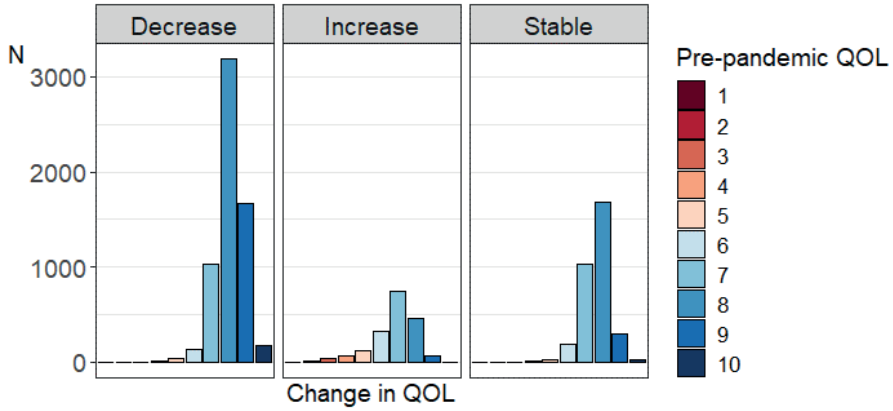


Figure 2. Number of individuals for whom Quality of Life (QOL) decreased, increased, and remained stable per pre-pandemic QOL score.

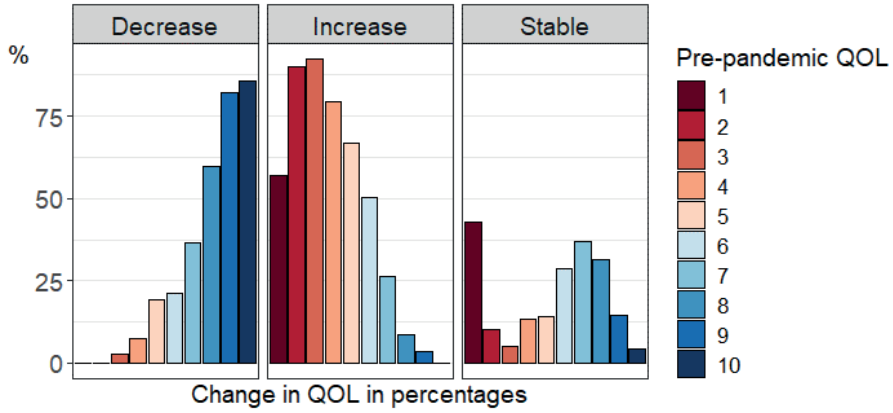


Figure 3. Percentage of individuals for whom Quality of Life (QOL) decreased, increased, and remained stable per pre-pandemic QOL score. Each colour presents a pre-pandemic QOL score, and is divided in percentages over the three change categories.

3.2 Longitudinal and Kinship correlations

Across the whole sample, the correlation (*r*) between pre-pandemic and pandemic QoL was .28 (*CI*= .26-.30). Number of pairs for pre-pandemic and pandemic QoL and correlations for the different relationship types are presented in Table 2. The pre-pandemic MZ correlations for males (*r*=.46, *CI*=.37-.55) were more than twice as high as the correlations for DZ male (DZM) pairs (*r*=.10, *CI*=-.05-.25), suggesting

a role for additive and dominant genetic influences. Correlations for MZ females (*r*=.29, *CI*=.23-.35) were slightly less than twice the female (DZF) pair correlations (*r*=.15, *CI*=.05-.25), suggesting a role for additive genetic and shared environmental influences. The DZ opposite sex (DOS) pair correlation (*r*=.20, *CI*=.11-.29) was slightly higher than the DZM and DZF correlations, albeit with overlapping CIs. The spousal correlation (*r*=.35, *CI*=.32-.38) was relatively high, and was modelled as a common household variable in later analyses. Parent-offspring and sibling-sibling correlations were in the same range as DZ correlations.

Pandemic QoL correlations were similar to or lower than pre-pandemic QoL correlations. As seen in Table 2, twin- and spousal correlation estimates decreased, indicating a larger role for the non-shared environment during the pandemic. An exception is the DZF correlation (*r*=.29, *CI*=.16-.42), which seemed to increase. The overlapping CIs for most twin correlations do suggest that this might not be a significant decrease. A larger role for E was also suggested by the parent-offspring correlations, with the correlations with daughters no longer being significantly different from zero. Sibling correlations were similar during and before the pandemic.

The separate cross-time correlations (and sample sizes), and the meta-analysed cross-correlation estimates can be found in Supplementary Table 2 and Table 2. Correlations between pre-pandemic and pandemic QoL were lower than the correlations for pre-pandemic QoL, and comparable to pandemic QoL correlations. The relatively low correlations between pre-pandemic and pandemic QoL suggest a large role for unique environmental influences, as these are not shared between different family members and thus introduce differences between family members.

Table 2

Kinship correlations for Quality of Life

	pre-pandemic			pandemic			cross-correlation*	
	N pairs (weighted)	r (SE)	95% CI	N pairs (weighted)	r (SE)	CI	r(SE)	95% CI
Spouses (incl. parents of twins)	3951	.35 (.01)	.32-.38	1428	.24 (.03)	.19-.29	.11 (.02)	.08-.14
MZM	341	.46 (.05)	.37-.55	168	.15 (.08)	0-.30	.14 (.05)	.04-.23
MZF	1063	.29 (.03)	.23-.35	719	.21 (.04)	.14-.28	.11 (.03)	.06-.16
DZM	178	.10 (.08)	-.05-.25	63	-.03 (.13)	-.28-.22	.15 (.08)	0-.29
DZF	416	.15 (.05)	.05-.25	197	.29 (.07)	.16-.42	.09 (.05)	0-.18
DOS	454	.20 (.05)	.11-.29	200	.17 (.07)	.03-.31	.10 (.04)	.02-.19
Mother - Daughter	1849	.14 (.02)	.09-.19	1209	.03 (.03)	-.03-.09	.08 (.02)	.04-.12
Mother - Son	1023	.21 (.03)	.15-.27	492	.10 (.04)	.01-.19	.11 (.03)	.06-.16
Father - Daughter	1174	.15 (.03)	.09-.21	736	.04 (.04)	-.03-.11	.04 (.02)	0-.09
Father - Son	689	.12 (.04)	.05-.19	354	.12 (.05)	.02-.22	.09 (.03)	.02-.15
Brother - Brother	156	.04 (.08)	-.12-.20	71	.05 (.12)	-.19-.29	.09 (.08)	-.07-.24
Brother - Sister	650	.18 (.04)	.10-26	444	.08 (.05)	-.01-.17	.08 (.03)	.02-.14
Sister - Sister	460	.20 (.05)	.11-29	319	.22 (.05)	.11-.33	.12 (.04)	.04-.19

Note. r = correlation, SE=standard error, CI= Confidence Interval, MZM= monozygotic males, MZF= monozygotic females, DZM= dizygotic males, DZF= dizygotic females, DOS= dizygotic opposite sex.

* see methods for an explanation of how the cross-correlation was computed.

3.3 Genetic analyses

The total phenotypic variance in QoL increased from 1.13 before the pandemic to 1.83 during the pandemic. The full model comparison results and variance decomposition for all the different models can be found in Supplementary Tables 3-4. The best fitting model was the ACE model including additive genetic, common household for spouses, and unique environmental variance components. The variance component estimates for the ACE model can be found in Table 3. The increase in total variance is attributable to a large increase in unique environmental variance (from 0.30 to 1.08). While unique environmental variance increased, the common environmental variance remained stable, and the genetic variance decreased from 0.35 to 0.28.

Table 3

Unstandardized (incl. SE) and standardized variance components

	pre-pan	**pan**	**pre-pan**	**pan**	**pre-pan**	**pan**	**pre-pan**	**pan**
	A	**A**	**C**	**C**	**E**	**E**	**P**	**P**
				Unstandardized				
pre-pan	0.3503 (.0205)		0.4837 (.0172)		0.2983 (.0250)		1.1323	
pan	0.1828 (.0222)	0.2835 (.0430)	0.1936 (.0234)	0.4731 (.0479)	0.0292 (.0310)	1.0754 (.0629)	0.4056	1.8320
	Standardized (unstandardized estimate/total phenotypic variance)							
pre-pan	0.3094		0.4272		0.2634			
pandemic	0.4507	0.1547	0.4774	0.2582	0.0720	0.5871		

Note. pre-pan= pre-pandemic, pan=pandemic, A= additive genetic variance, C= common environment variance, E= unique environment variance, P=phenotypic variance

Consequently, the standardized variance decomposition of QoL also changed from before to during the pandemic. Relatively, the magnitude of the total variance that was explained by genetic differences, or the heritability, decreased from 30.9% to 15.5%, and the magnitude of common environmental influences decreased from 42.7% to 25.8%. Unique environmental factors became relatively more important, with 58.7% of the variance in QoL being explained by unique environmental factors during the pandemic, compared to 26.3% before the pandemic.

Most of the covariance between QoL before and during the pandemic is explained by genetic (45.1%) and shared environmental (47.7%) factors. Only 7.2% of the

covariance was explained by unique environmental factors. Lastly, we found moderate genetic (rA=.58) and common environmental correlations (rC=.40), and a small unique environmental correlation (rE=.05) between pre-pandemic and pandemic QoL scores.

4. DISCUSSION

The present study set out to examine the impact of an impactful natural experiment (the COVID-19 pandemic) on the genetic architecture of well-being, measured as QoL. We find that on average, QoL decreased from 7.73 to 7.02 in the first months after the onset of the pandemic, reflecting a medium decrease (*d* = .49). QoL scores decreased for more than half the sample (55.05%), remained stable for 28.84% of the sample, and increased for 16.11% of the sample. Additionally, bivariate variance decomposition in Mendel showed a large increase in unique environmental variance during the pandemic. As a result, the relative proportion of individual differences explained by genetic factors (i.e. the heritability) decreased during the pandemic.

So far, the existing literature comparing pre-pandemic and pandemic well-being has produced mixed results. A meta-analysis by Prati & Mancini[299] did not find a significant effect on positive psychological functioning across 6 studies. However, positive psychological functioning was assessed in different countries using diverging well-being definitions, for example mental well-being measured in the United Kingdom (UK) using the Short Warwick-Edinburgh Mental Wellbeing Scale[313], subjective well-being was measured in France based on the frequency participants reported feeling 'nervous', 'low', 'relaxed', 'sad', 'happy', and 'lonely'[262], and positive affect measured using the Positive Affect Negative Affect Scale in China[314]. In a similar fashion, Aknin and colleagues[315] performed a review on mental health during the pandemic and concluded that life satisfaction was largely unchanged in many countries during the first year of the pandemic, but that people did experience more unpleasant emotions during the pandemic.

In the present study, we find less optimistic results for QoL than expected based on these reviews, with the majority of individuals reporting lower QoL during the pandemic compared to before the pandemic. While the reason for this discrepancy is not clear, it might have something to do with the time period in which we collected the pandemic data in the Netherlands. Since the data were collected

during the first lockdown accompanying the first wave of COVID-19 infections, these results represent the first response of participants to the pandemic and consequent lockdown measures. In the earlier stages of the pandemic, there were a lot of uncertainties and fears over the virus, infection rates were high, and the lockdown measures were highly disruptive. As such, changes in QoL may have been especially pronounced in these beginnings of the pandemic, and may have returned to more normal levels later on. It should be mentioned that some of the studies included in the reviews above also examined effects during the first lockdown. Therefore, it is likely that there are different effects across different countries, even during similar lockdown periods.

Important in the context of these results is that our sample, and the Netherlands in general, scores relatively high on QoL and other well-being measures compared to other countries. The Netherlands scores among the top happiest countries according to the 2019 World Happiness Report[316], which was unchanged in the 2021 World Happiness Report that reported on the data collected in 2020 (during the pandemic)[317]. Importantly, we found a pandemic average QoL of 7.02 which is significantly lower than the pre-pandemic average, but still a good score indicating that people were still quite satisfied with their QoL. We found that especially those with higher QoL scores were prone to decreases in QoL during the pandemic. Given the skewed distribution of QoL in our sample, this was the majority of our sample. Increases in QoL, however, were found mostly for individuals with lower QoL scores. While this is a relatively small part of our sample, it was surprising that it was especially individuals with lower baseline QoL that showed improvements in QoL during the pandemic. A potential explanation is that we only examined individuals that did not have a COVID-19 infection around the time of assessment. Individuals with low levels of baseline QoL might have evaluated their QoL differently during the pandemic as they started comparing themselves with others that did become ill. In this way, they might have altered their perception, causing them to provide a different judgment during the pandemic[262]. Individuals with higher levels of baseline QoL, on the other hand, might not have focused on these kinds of comparative mechanisms since they did not think their QoL was worse than average to begin with. Importantly, another possibility is that these findings might (partly) result from regression to the mean (RTM), the phenomenon whereby the second assessment of a trait results in values closer to the mean

than at initial assessment purely by chance. However, pre-pandemic QoL was measured on multiple occasions for some individuals, in which case we chose the latest available time-point. By selecting participants from different measurement occasions, we attempt to get a better estimate of the participants' true baseline mean, which in turn decreases RTM[318]. In a way, our pre-pandemic QoL measure is formed by taking a random sample around an individual's baseline mean levels. This makes it less likely that high/low scorers will inevitably go down/up at the next measurement occasion (i.e., during the pandemic).Therefore, while we cannot rule out regression to the mean completely, this does make it less likely that our results are (fully) attributable to this phenomenon.

We used Mendel, instead of the classical twin design (CTD), to decompose the variance into genetic and environmental sources of variation. In the CTD, the variance components are estimated based only on the MZ and DZ twin covariances. As a result, only three parameters can be estimated simultaneously, so that an a priori choice needs to be made between an ACE or ADE model. The advantage of the Mendel software is that it allows for efficient analysis of whole pedigree data, allowing us to examine a large sample and estimating A, C, D, and E simultaneously. There are extensions of twin designs where other family members can be included, such as the Cascade model [319] that are more flexible in terms of model specification (e.g. constraining paths and sex-specific heritability). However, the advantage of Mendel is that it easily allows for the inclusion of complex family relations and irregular pedigrees, as are present in large twin-family registers like the Netherlands Twin Register. Yet, we did not find any evidence for dominant genetic effects (D), i.e. alleles acting in a multiplicative fashion (dominance or epistasis). Based on the correlations between the different types of family members (Table 2), there was some suggestive evidence for D in the male twin correlations, but not the female twin correlations (which were based on a much larger sample). Results from existing twin- and family studies on the contribution of non-additive genetic effects to well-being have been very mixed, but the largest twin-family study to date did find evidence for non-additive genetic effects (Nes et al. 2010). Importantly, the study by Nes and colleagues focused on a happiness measure, while in the current study we examined individual differences in quality of life, potentially explaining this discrepancy. This is in line with our earlier work where we also report stronger evidence for dominant genetic effects on happiness versus quality of life[82].

A striking finding was the large increase in unique environmental variance – estimates of E more than doubled – during the pandemic, which resulted in a decreased heritability, indicating an increase of the relative importance of unique environmental factors. Similar results were found in a large etiological study by Carroll and colleagues showing that unique environmental influences were amplified for emotional symptoms and conduct problems in youth (but not for attention-deficit hyperactivity problems) as a result of pandemic-relation disruption among multiple life domains[320]. This phenomenon, where the total genetic/environmental variance is dependent on the environment, in this case pandemic-related environmental change, is reflective of quantitative gene-environment interaction (GxE). Our quantitative GxE finding is in line with the bioecological model that postulates that genetic influences are maximized in stable and adaptive environments, and non-shared environmental influences are greatest in more "risky" environments[321]. Clearly, pandemics such as the COVID-19 pandemic, can be viewed as more risky environments characterized by high levels of disruption and uncertainty about the future. Alternatively, the findings can also be framed in a social control model, where genetic influences are relatively dampened as the result of social constraints imposed by the environment[321]. While it was clear at the beginning of this study that the lockdown introduced many social constraints and consequently large environmental change in people's everyday lives, we did not yet know whether this would lead to an increase or a decrease in environmental variance. Theoretically, the restrictions imposed by the lockdown measures could have reduced the environmental variance by making everyone's lives more similar to each other. However, the finding that these measures led to a large environmental increase suggests that the pandemic and consequent lockdown measures did not impact everybody in a similar way. It is important to identify such factors, since they potentially enlarge health inequalities during the pandemic. For example, a study by Ravens-Sieberer found that children with low socioeconomic status, migration background, and limited living space were affected significantly more by the pandemic in terms of health-related QoL, mental health problems, anxiety, and depression[322].

Several potential explanations for individuals' different reactions to the environmental change imposed by the pandemic can be proposed. For example, people were encouraged to work from home, but only if possible. Before the pandemic, 1 in 3

people in the Netherlands (occasionally) worked from home. This increased to 1 in 2 people in the beginning of the pandemic, with the strongest increase in people with higher educational attainment and people using public transport to commute between home and work [323]. Thus, the "working from home" policy did not affect everyone in the population equally, potentially leading to increased differences in reported QoL across individuals. Additionally, with schools and day-care centres closed, people with children were likely impacted in a different way than people without children. Parents working from home (with children also staying at home) presumably had more trouble concentrating and were less productive, but the extent to which depended on different factors, like the age of the child and the age of parents[324]. While the present study cannot pinpoint what exactly caused the increased environmental variance, these factors might serve as suggested causes of increased environmental variance in QoL in follow-up research.

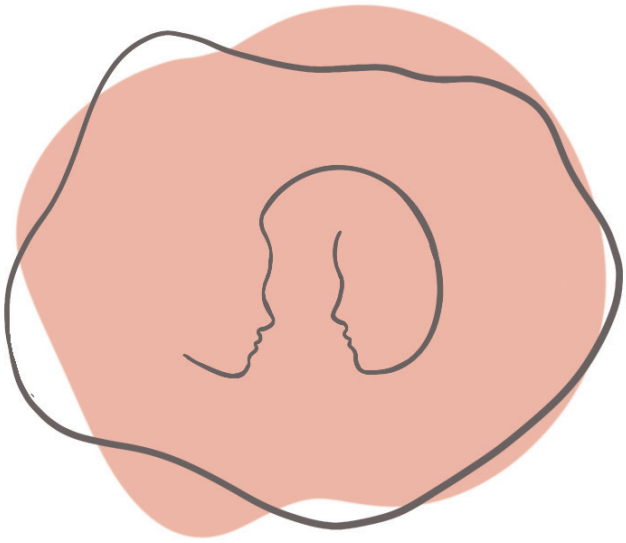
Based on the variance and covariance estimates provided by Mendel, we were able to calculate genetic and environmental correlations, which tells us something about the extent to which the same genetic and environmental factors influence QoL at the two time-points. We found a moderate genetic correlation (r_A=.58) and common environmental correlation (r_C=.40), and a small unique environmental correlation (r_E=.05). To test if a correlation is significantly different from zero (or one), one would normally fit a model where the relevant correlation is constrained to zero (or one) and compare the fit of this model to the fit of the unconstrained model. Unfortunately, since Mendel does not allow for the inclusion of such constraints, we were not able to perform such model comparisons. However, based on the point estimate (E_{cov}=.029) and standard error (*SE*=.031) of the unique environmental covariance, we can conclude that the unique environmental correlation is not significantly different from zero. In other words, the unique environmental factors influencing individual differences in QoL during the pandemic are completely different from the unique environmental factors influencing individual differences in QoL before the pandemic. This is important to consider when thinking about potential (positive) psychological prevention and intervention strategies to harness people from the negative effects of extreme environmental change, such as the COVID-19 pandemic lockdowns. These strategies rely on existing research by focusing their strategies on existing evidence on correlates of well-being. However, as indicated by this study, the environmental factors that determine individual differences in well-being under "normal circumstances" are likely not the same as during crisis situations like these.

It is important to interpret these results within the context of our sample and the time-frame in which we collected the data. Since it was the first lockdown, when the WHO had just announced a pandemic, individuals were likely still psychologically adjusting to the new situation. Whether the effects found in this study would be similar in later stages of the pandemic is a question that remains to be answered. Additionally, different countries employed different strategies to contain the virus, with the Netherlands installing the intelligent lockdown where people were encouraged to stay at home, but were still allowed to freely move around outside at all times of day. In this light, the finding of the large increase in environmental variance is even more remarkable, since the regulations in the Netherlands were less strict than those in many other countries. As such, the environmental effects of more stringent lockdown measures may be even larger. In any case, it is reasonable to expect that countries with different regulations will find different results than presented here, since these regulations impact the extent to which people had to alter their lives. Finally, a limitation of our sample was that we had more female respondents than male respondents in both the pre-pandemic sample (65% female, 35% male), and the pandemic sample (71% female, 29% male). The representativeness was further limited by there being roughly twice as many highly educated individuals in the sample than expected based on the Dutch population.

In conclusion, in this study we used data from before and during the COVID-19 pandemic as a natural experiment to add to our understanding of genetic and environmental influences on QoL. By treating the COVID-19 pandemic as a natural experiment, we were able to demonstrate the dynamics of environmental change on individual differences and heritability. The most prominent finding to emerge is that unique environmental factors became relatively more important in explaining differences in QoL during the pandemic, with genetic factors becoming less important. Additionally, it seems that different unique environmental factors become relevant to QoL during compared to before the pandemic. Further research is required to determine if these effects are similar in the long term. Additionally, future research should explore what environmental factors are important for QoL during the pandemic, as these factors likely increase health inequalities in the population.

PART IV

Moving Towards Causal Approaches



CHAPTER 8

Genetic factors explain a significant part of associations between adolescent well-being and the social environment

Published as: Van De Weijer, M. P., Pelt, D. H.M., Van Beijsterveldt, C. E., Willemsen, G., & Bartels, M. (2021). Genetic factors explain a significant part of associations between adolescent well-being and the social environment. *European child & adolescent psychiatry*, 1-12.

*supplementary materials accessible at: https://doi.org/10.1007/s00787-021-01798-3

ABSTRACT

Socio-environmental factors play an important role in adolescent well-being, but potential genetic contributions to these associations are rarely assessed. To address this gap in the literature, associations between well-being and family conflict and functioning, number of friends, friendship importance and satisfaction, and leisure time variables were studied in N=~4700 twin-pairs from the Netherlands Twin Register, using generalized estimating equations and twin-difference scores. When twin-difference scores indicated a role for genetic factors, we used bivariate genetic models to quantify genetic and environmental contributions to these associations. We identify significant associations between well-being and family functioning, family conflict, different leisure time activities, number of friends, and satisfaction with friendships. Additionally, we find evidence for large (73-91%) genetic influence on the associations between well-being and family conflict and functioning, leisure time sport/scouting clubs, and satisfaction with friendships. Lastly, findings support the hypothesis of a causal association between well-being and family conflict and functioning. These findings have important implications for research into the social correlates of well-being in adolescence, as not taking genetic factors into account leads to overestimations of the influence of identified correlates and consequently to recommendations of these correlates as intervention targets.

1. INTRODUCTION

Adolescence, defined by the WHO as the period between age 10 and 19, marks a period in life where a person transitions from childhood to adulthood. During this transition period the body rapidly develops, and there is accumulating evidence that adolescence is a critical period for later health and disease[325]. For example, half of the cases of lifetime DSM-IV anxiety, mood, impulse-control, and substance use disorders, have had their onset by age 14. While this period of pubertal mental and bodily maturation thus represents a period full of risk, it can also be interpreted as a period that holds great potential for interventions.

The focus of adolescent mental health research so far has mainly been on mental illness. For example, there is abundant research into how depression in adolescence might lead to adult depression, comorbid disorders, and suicide[326–329]. With this emphasis on mental illness, it is easy forgotten that most adolescents develop relatively well, with only a small proportion of adolescents reporting low levels of well-being[330,331]. In addition, large genetically informed studies find genetic correlations of ~.7 between well-being and depression, suggesting that, although they are substantially related, the genetic predisposition for well-being is partly independent from the genetic predisposition for depression[32,332] and that well-being is more than just the absence of depression. Therefore, in addition to studying mental illness and its risk factors, it is valuable to study the determinants of mental health and well-being.

Creating adolescent interventions to improve adult outcomes requires in-depth understanding of the determinants of adolescent well-being. This is supported by findings that adolescent well-being predicts adult well-being and general health[333,334]. Given the importance of adolescent well-being for later-in-life outcomes, it is essential to identify its correlates and determinants. One of the most studied factors in relation to well-being is one's social environment. For example, a meta-analysis on the associations between well-being and social support measures in children and adolescents across 246 studies found that social support from parents, peers, and teachers is positively associated with well-being[335]. Moreover, a review focusing on the connection between well-being and friendships concludes that children's friendships are associated with their happiness, and that negative social relationships have an adverse effect on their well-being[336].

The nature of these associations, however, often remains unexplored. Are these relations causal or is there an unmeasured third factor that is related to both, resulting in the observed association? For example, resilience and well-being are often observed to be strongly associated, accompanied by firm conclusions about the direction of causation. However, 51% of the phenotypic association between resilience and well-being is accounted for by a third factor: genetic influences[337]. For socio-environmental factors, it was traditionally assumed that epidemiological associations between individuals and their environment could only be explained in an unidirectional manner, with the environment affecting the individual[338]. We have since learned that these associations are bidirectional, with environments also being subject to heritable influences, through individuals' behavior[40,338,339]. Research showed that the heritability of well-being is 40%[24,25], meaning that about 40% of the individual differences in well-being can be explained by genetic differences between people. Thus, if we study well-being in relation to another heritable trait (e.g., family conflict[40]) it may occur that the observed phenotypic association is (partly) due to overlapping genetic factors. If these genetic factors are not taken into account, one might overestimate the (causal) influence of identified social factors and consequently recommend these correlates as targets for interventions, even though they carry small or no direct (causal) effect on well-being.

In the current study, we investigate the underlying sources of associations between adolescent well-being and various socio-environmental factors. Using twin data from the Netherlands Twin Register (NTR), we examine monozygotic (MZ) and dizygotic (DZ) difference scores to explore the possibility that these associations are (partly) attributable to genetic factors. When evidence for genetic influences is seen, we use bivariate genetic models to quantify the genetic and environmental influences on the covariance between well-being and socio-environmental factors.

2. MATERIALS AND METHODS

2.1 Sample

Study participants are voluntarily registrants at the NTR[63]. We selected a subset of NTR participants who filled out the Dutch Health and Behavior Questionnaire, administered to adolescent participants aged 13 to 17 (for more data collection information, see[340]). In total, well-being data were available for 11406 adolescents

from 4739 complete twin pairs and 1928 incomplete twin pairs (*M* age=15.66, *SD* age=1.31, *N* males/females= 4855/6551). Sample size per analysis varied depending on sample size available per social variable (see Online Resources, eTable 1). For each social variable, the sample size reflects the number of complete twin pairs that also have well-being data. If an individual had data available at more than one time-point, we used data from the last time-point. We made sure that within twin-pairs, data were selected from the same time-point to reduce bias due to differences in timing of the survey. Zygosity in same-sex twin pairs was determined based on DNA genotyping (34.4%) or, when DNA samples were not available, by previously collected questionnaires containing parental-reports about same-sex twin similarity in physical characteristics and frequency of mistaking one twin for another by parents, relatives and strangers. Based on these self-report questions the accuracy of classification is 95.9%[63].

2.2 Variables

Well-being was assessed using the Dutch version of the satisfaction with life (SWL) scale[12]. This scale contains 5 items that assess SWL on a 7-point Likert scale. The scale has good internal consistency in the sample (α=.87). Scores on the individual items are summed to create SWL scores for each respondent. An example of an item is: 'I am satisfied with my life'. Well-being scores were standardized to z-scores in all analyses.

In the Dutch Health and Behavior Questionnaire (DHBQ)*, social variables* are available for the following categories: leisure time activities, family functioning, family conflict, and friendships. Scores for all variables were standardized to z-scores in all analyses.

Leisure time activities (LT) are assessed by self-report on how much time participants spend on the following activities: a) watching TV – videos – DVDs, b) computer games, c) computer/ internet d) making music/choir, e) reading, f) drawing/ painting, g) handicrafts, h) being at home with friends, i) visiting friends, j) on the street with friends, k) sports club or scouting, l) chess, board games and m) going out (disco, cafe, bar). For each activity, participants can choose from the following answer categories: 1) never, 2) only once until now, 3) less than once a week, 4) once a week, 5) a few days per week, 6) almost every day, 7) every day. Since

some activities can be categorized under broader categories, we summed some of the categories together into 1) computer games and computer/the internet, 2) reading and chess, board games (hereafter referred to as indoor games), 3) drawing/painting and handicrafts, 4) being at home with friends, visiting friends and on the street with friends.

Family functioning is assessed using a Dutch translation of the subscale General Functioning of the Family Assessment Device (FAD)[341]. This 12-item scale measures overall (un)healthy family functioning, with items assessing problem solving, communication, roles, affective responsiveness, affective involvement, and behavior control. The subscale holds high reliability in our sample (α=.88). The items are answered on a scale from 1-4, 1 representing strong agreement with the item, and 4 representing strong disagreement. Since 6 items measure healthy functioning, and 6 measure unhealthy functioning, we recoded half of the items as 5 – [item score] so that all questions were scored in the same direction. After this transformation, items are summed to create a total family functioning score, with higher scores indicating higher levels of dysfunction. An example of an item is 'planning family activities is difficult because we misunderstand each other'.

Family conflict is assessed using a Dutch translation of the subscale Conflict of the Family Environment Scale (FES)[342]. The subscale contains 11 items with a 2-point scale, with 1=No and 2=Yes. For each item, the participant indicates whether the presented statement is true for their family. The subscale shows acceptable reliability in our sample (α=.0.73). An example an of item is: "In our family we argue a lot". One item ("We seldom get openly angry at each other at home") is reverse-coded so that answering yes implies low family conflict, whereas yes on the other items indicates high family conflict. The Dutch translation of this item was misinterpreted by a lot of participants, leading to inconsistent data patterns or missing data[343]. Therefore, during data collection, this item was changed to "we often get openly angry at each other at home", so that all items were collected in the same direction. For this study, we used this reworded version of the item. Scores on the 11 items were summed to get a total score for family conflict, with higher scores indicating higher levels of conflict.

Friendship is assessed in three ways: 1) "How many good male/female friends do you have?"; 2) "In general, how satisfied are you with your female/male friends?",

and 3) "In general, how important are your female/male friends to you?". To minimize the number of statistical tests, we summed the responses for male and female friends for (1), and took the mean for (2) and (3). For question 1, the participant could answer within the following categories: 0) I don't have any good friends; 1) 1 or 2; 2) 3 or 4; 3) 5 or 6; 4) 7 or 8; 5) 9 or 10; 6) 11 to 15; 7) more than 15. For questions (2) and (3) the answering categories were as follows: 0) very dissatisfied/unimportant; 1) dissatisfied/unimportant, 2) somewhat satisfied/important; 3) satisfied/important; 5) very satisfied/important.

2.3 Statistical analyses

2.3.1 Phenotypic associations

Using the full sample (including incomplete twin pairs, see Table 1), we applied linear regression analysis to identify associations between well-being and the social variables. To correct for familial dependency in the observations, we used the generalized estimating equation (GEE) function in R[344]. In GEE, an exchangeable conditional covariance matrix is used to account for relatedness, and tests are based on sandwich-corrected, robust standard errors [185]. Sex was included as a covariate in the regression analyses.

2.3.2 Intra-pair difference scores

Intra-pair difference scores were used to get a first indication of the nature of the association between well-being and different social variables. Since MZ twins share both their genetic makeup (additive genetic effects A) and their common environment (C), intra-pair difference scores between the twins must be the result of unique environmental experiences (E). On the other hand, intra-pair difference scores in DZ twins can be a result of differences in unique environmental influences (E), but also a result of differences in their genetic makeup (A), since they only share 50% of their genetic material on average. In a twin difference design, intra-pair difference scores for one trait are regressed on the intra-pair difference scores of another trait. Based on these analyses, we expect the following given there is an observed phenotypic association between the traits (see Figure 1):

I. If there is a significant association between the intra-pair difference score of well-being and the intra-pair difference score of a social variable in both MZ and

DZ twins, this supports the hypothesis there is a causal relation between the two traits, or a large role for E. Genetic factors are additionally likely to contribute when the absolute DZ regression coefficient is larger than the absolute MZ regression coefficient.

II. If there is a significant association between the intra-pair difference score of well-being and the intra-pair difference score of a social variable in DZ twins solely, it suggests A plays a large role in the association, supporting the hypothesis that genetic factors act as a third unobserved variable underlying the association.

III. If in both MZ and DZ twins the intra-pair difference score regression returns non-significant results, the covariance between two traits is likely caused by C as these are 100% shared in both types of twin pairs.

We calculated difference scores for all variables in all twin-pairs by subtracting the score of one twin from the score of the other twin. Next, for each social variable, we regressed the social variable difference score on the well-being difference score using linear regression while correcting for a variable that reflects whether the twin pair is same-sex or different-sex. Since MZ twins are always same sex, we did not control for sex in MZ difference score analyses. We used a significance threshold of α = .05/13 = 0.0038 to correct for multiple testing (13 tests in total).

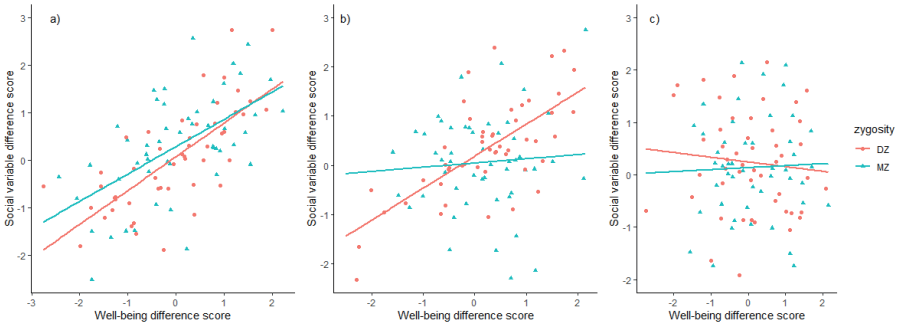


Figure 1. Expectations based on different scenarios: a) significant MZ and DZ difference score associations suggest a causal effect between well-being and a social variable; b) a significant DZ difference score association but no MZ difference score association suggests a large role for genetic factors; c) lack of association of both MZ and DZ difference scores suggests a large role for common environmental influences.

2.3.3. Bivariate Twin Models

For associations where we find significant intra-pair difference regression results in both MZ and DZ twins or in DZ twin solely, we use bivariate genetic models to quantify genetic and environmental influences on the covariance between well-being and social variable. The difference in genetic relatedness between MZ and DZ twins enables decomposing the (co)variance of the traits under investigation into additive genetic (A), dominant genetic (D), common environmental (C), and unique environmental factors (E, including measurement error). Since C and D cannot simultaneously be estimated based on MZ and DZ covariance alone, either an ACE or ADE model is fit. When the MZ correlation is more than twice the DZ correlation, an ADE model is fit. When the MZ correlation is less than twice the DZ correlation, an ACE model is fit. Based on the literature, it was likely that the twin correlations for well-being might suggest an influence of D[82], while the twin correlations for some social traits might suggest an influence of C[343]. As we cannot model both C and D in the bivariate model, and since we are most interested in potential common environmental influences, we a priori chose to model bivariate ACE models (see Online Resources, eFigure 1).

Twin correlations and cross-twin cross trait (CT-CT) correlations were estimated in saturated models in which all parameters (means, variances, and covariances) are freely estimated. We modeled variance components separately for males and females. For satisfaction with friendships, scores were highly skewed. In order to prevent bias[345] we transformed this variable into an ordinal variable with three categories (low, middle, high), and applied a liability threshold model with 2 thresholds. Under this model, it is assumed that there is an underlying continuous liability distribution for this trait, with two thresholds that that define three categories. The thresholds divided the data into three groups of equal sizes (33%).

Since the variance component approach does not yet allow for the inclusion of opposite-sex twins in estimating the variance components, DZ opposite-sex twins were excluded when estimating the variance components. We therefore could not test for qualitative sex differences. To test for quantitative sex differences (i.e. the same genetic and environmental factors exert influence of different magnitudes in males and females), we constrained variance components to be equal across males and females and compared the fit of the constrained model to that of the less restrictive model. Next, we tested whether C significantly contributed to the

(co)variance by dropping common environmental components in three steps (see Online Resources, eFigure 1): 1) we dropped C for well-being (*c22*), 2) we dropped the covariance explained by C (*c21*), and 3) we dropped C for the social variable (*c11)*. If a C component could not be dropped for both sexes, we tested whether it could be dropped for males or females only. Additionally, we computed genetic and environmental correlations, reflecting the extent to which there is overlap in the latent genetic and environmental factors influencing the traits. Parameters were estimated using maximum likelihood estimation in OpenMx[275] using the variance component approach[274]. By fitting the model with and without the constraints of interest, a log-likelihood ratio test can be used to compare models. The more parsimonious model is rejected if the log-likelihood statistic exceeds the chosen threshold. In line with the reasoning by Benjamin and colleagues[276] that the traditional p-value threshold of .05 leads to a high false-positive rate, we used a *p*-value threshold of α=.005.

3. RESULTS

3.1. Phenotypic associations

In the GEE analyses (Table 1), higher well-being was significantly associated with less family dysfunction (FAD, β=-.35, *SE*=.01, $p=2.23\times10^{-181}$), less family conflict (FES, β=-.26, *SE*=.01, $p= 2.25\times10^{-98}$), more leisure time indoor games (β=.03, *SE*=.01, $p=7.43\times10^{-4}$), more leisure time contact with friends (β=.05, *SE*=.01, $p=4.60\times10^{-6}$), more leisure time sports club/scouting (β=.13, *SE*=.01, $p= 3.70\times10^{-33}$), a higher number of friends (β=.12, *SE*=.01, $p=1.25\times10^{-21}$), and higher satisfaction with friendships (SWF) (β=.16, *SE*=.01, $p=9.33\times10^{-27}$). Leisure time crafts, leisure time making music, leisure time computer, leisure time going out, leisure time TV, and importance of friendships were not associated with well-being.

Table 1

Associations between well-being and all variables

	GEE (whole sample)			DZ difference			MZ difference		
	β(SE)	p	N	β(SE)	p	N (pairs)	β(SE)	p	N (pairs)
FAD (family functioning)	-0.35 (.01)	$2.23x10^{-181*}$	10478	-0.23 (.02)	$<2x10^{-16*}$	2493	-0.15 (.02)	$3.18x10^{-10*}$	1596
FES (family environment)	-0.26 (.01)	$2.25x10^{-98*}$	7479	-0.20 (.02)	$<2x10^{-16*}$	1707	-0.13 (.03)	$2.14x10^{-6*}$	1088
Leisure time – indoor games	0.03 (.01)	$7.43x10^{-4*}$	11044	0.02 (.02)	0.341	2707	-0.01 (.02)	0.758	1768
leisure time - contact with friends	0.05 (.01)	$4.60x10^{-6*}$	10965	0.05 (.02)	.008	2689	0.003 (.02)	0.886	1716
leisure time - crafts	-0.01 (.01)	0.145	11100	-0.01 (.02)	0.457	2744	0.01 (.02)	0.622	1775
leisure time - making music/choir	.004 (.01)	0.715	11178	-0.06 (.02)	.004	2769	0.01 (.02)	0.558	1798
leisure time - computer	-0.02 (.02)	0.278	3553	0.04 (.03)	0.285	1006	0.02 (.04)	0.559	679
leisure time - going out (dancing)	0.03 (.01)	0.005	11197	0.01 (.02)	0.467	2784	0.01 (.02)	0.735	1800
leisure time - sport/scouting club	0.13 (.01)	$3.70x10^{-33*}$	11182	0.08 (.02)	$8.79x10^{-5*}$	2776	0.03 (.02)	0.224	1806
leisure time - TV	0.03 (.02)	0.072	3608	0.03 (.03)	0.391	1041	-0.06 (.04)	0.155	701
Number of friends	0.12 (.01)	$1.25x10^{-21*}$	7690	0.06 (.02)	.005	1771	0.07 (.03)	0.012	1143
Importance of friendships	0.06 (.03)	0.062	1322	-0.05 (.06)	0.406	293	0.09 (.06)	0.143	209
Satisfaction with friendships	0.16 (.01)	$9.33x10^{-27*}$	6519	0.12 (.03)	$6.65x10^{-6*}$	1331	0.02 (.03)	0.468	863

Note. GEE= generalized estimating equation, DZ=dizygotic, MZ=monozygotic, β=beta, *SE*=standard error, p=p-value, N=sample size

* significant after correction for multiple testing (α=.0038)

3.2 Intra-pair difference scores

Differences scores for two social variables were significantly associated with well-being difference scores in both MZ and DZ twin pairs: the FAD difference score (MZ: β=-.15, p=3.18x10^{-10}, DZ: β=-.23, p= <2x10^{-16}) and the FES difference score (MZ: β=-.13, p=2.14x10^{-6}, DZ: β=-.20, p=<2x10^{-16}). This supports the hypothesis there is either a causal relation between the two traits, or a large role for E. Given that the DZ difference scores are more strongly associated than the MZ difference scores, there is also a potential role for A.

Difference scores for two variables were significantly associated with well-being difference scores in DZ twins, but not MZ twins: leisure time sport/scouting club (β=.08, p= 8.79x10^{-5}), and the SWF difference score (β=.12, p=6.65x10^{-6}). This suggests that A plays a large role in the association between well-being and these social variables.

Lastly, three of the variables for which we observed a phenotypic association with well-being were not significant in the intra-pair difference score analyses for both MZ and DZ twins (Table 1): leisure indoor games (MZ β=-.01, p=.758, DZ: β=.02, p=.341), leisure time contact with friends (MZ: β=.003, p=.886, DZ: β=.05, p=.008), and number of friends (MZ: β=.07, p=.012, DZ: β=.06, p=.005). The lack of MZ and DZ difference score association indicates that shared environmental factors are likely the underlying source of the observed association between well-being and these social variables.

3.3 Bivariate Twin Models.

Based on the difference score analyses, four traits were followed-up with bivariate genetic model fitting: FAD, FES, leisure time spend at sports/scouting club, and satisfaction with friendships. From the saturated models (model fitting results in Online Resources, eTable 2), we estimated the cross-twin and CT-CT correlations for each trait (Online Resources, eTable 3). For all traits, MZ correlations were higher than DZ correlations, indicating a role for A. Twin correlations for well-being and satisfaction with friendships in males indicated a potential role for D in the variance decomposition. However, as explained in the methods section, we only fit bivariate ACE models.

All MZ CT-CT correlations were significantly different from zero. DZ correlations were either non-significant or smaller than MZ correlations for all traits except family conflict in males, suggesting that A has a substantial influence on the covariance between well-being and the social variables (see Figure 2). For family conflict in males, MZ and DZ CT-CT correlations were of similar magnitude, suggesting a potential role for C on the covariance. The bivariate model fit comparisons can be found in eTable 4 in the Online Resources. For all traits, constraining the variance components to be equal across sex resulted in a significantly worse model fit. The full model variance decompositions can be found in eTable 5 in the Online Resources. In the full models, the C component for well-being and SWF in males is negative, likely due to genetic dominance[274].

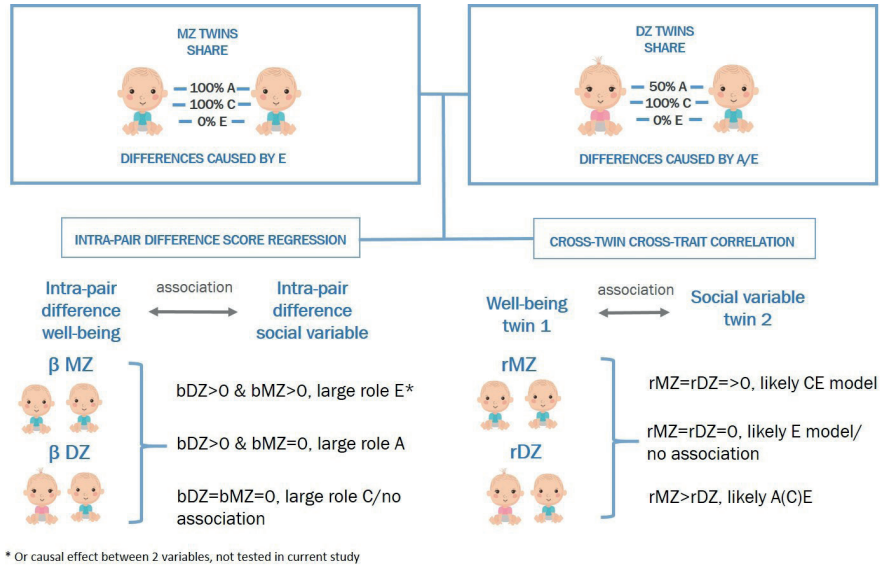


Figure 2. Overview of expectations based on DZ and MZ twin similarity and differences.

The final model (co)variance decomposition results for all traits can be found in Figure 3 and Table 2. Across all bivariate models, C could be dropped for well-being and for the covariance between well-being and the social trait in question. Well-being heritability estimates (A) were in line with previous studies[24], and we found slightly lower estimates for males (*A*=35%, CI 29-41%) than females (*A*=43%, CI 39-47%). In the bivariate model with well-being and FAD scores, all C components

could be dropped for males, while the C component for FAD could not be dropped in females (C=20%, CI=8-32%). The heritability (A) of FAD was 47% (CI 41-52%) for males, and 25% (CI 11-39%) for females. The covariance with well-being was mainly explained by genetic factors (males: 76%, CI=63-89%, females: 73%, CI 49-95%). For FES scores, the best fitting model in males was an AE model (*A*=55%, CI=49-61%), while the best fitting model for females was an ACE model (*A*=24% [CI 11-37%], *C*=41% [CI=29-52%]). The phenotypic correlation between well-being and FES was explained mostly by genetic factors (males: 73%, CI 49-95%, females: 81%, CI 69-93%).

For SWF, all C components could be dropped. The heritability (A) of SWF was estimated at 25% (CI 12-38%) in males and 35% (CI 29-41%) in females. Again, genetic factors explained the largest part of the covariation with well-being (males: 70%, CI 34-108%, females: 82%, CI 58-107%). Lastly, for leisure time sport/scouting clubs, C could be dropped for males, but not females (*C*= 33%, CI 23-43%). The heritability (A) was estimated at 60% (CI 56-64%) in males, and 33% (CI 22-44%) in females. Genetic factors contributed to 91% of the covariance between well-being and leisure time sports/scouting clubs in males (CI 60-126%) and to 89% of the covariance in females (CI 69-108%). The estimates of our variance components were unbounded, which led to confidence intervals outside the usual range of 0-1 for these last two traits. This, together with the twin and CT-CT correlations, indicates a potential role for D.

eTable 6 (Online Resources) contains the genetic and environmental correlations between well-being and the other traits for males (r_{Am} and r_{Em}, respectively) and females (r_{Af} and r_{Ef}, respectively) separately. All genetic correlations were significant, with negative genetic correlations between well-being and FAD (r_{Am}= -.60, r_{Af}= -.94) and FES (r_{Am}= -.52, r_{Af}= -.71), and positive genetic correlations between well-being and SWF (r_{Am}= .54, r_{Af}= .54) and leisure time sports/scouting club (r_{Am}= .22, r_{Af}= .30). Unique environmental correlations were significant for FAD (r_{Em}= -.13, r_{Ef}= -.12) and FES (r_{Em}= -.12, r_{Ef}= -.12) only.

Table 2

Standardized covariation decomposition of SWL with the different traits

		A		C		E	
		Social trait	**WB**	**Social trait**	**WB**	**Social trait**	**WB**
males	**FAD**	.47 [.41-.52]		-		.53 [.48-.59]	
	WB	.76 [.63-.89]	.35 [.29-.41]	-	-	.24 [.11-.37]	.65 [.59-.71]
females	**FAD**	.25 [.11-.39]		.20 [.08-.32]		.55 [50-.60]	
	WB	.82 [.74-.90]	.43 [.39-.47]	-	-	.18 [.11-.26]	.57 [.53-.61]
males	**FES**	.55 [.49-.61]		-		.45 [.39-.51]	
	WB	.73 [.49-.95]	.35 [.29-.41]	-	-	.27 [.05-.51]	.65 [.59-.71]
females	**FES**	.24 [.11-.37]		.41 [.29-.52]		.35 [.31-.40]	
	WB	.81 [.69-.93]	.43 [.39-.47]	-	-	.19 [.07-.31]	.57 [.53-.61]
males	**SWF**	.25 [.12-.38]		-		.75 [.62-.88]	
	WB	.70 [.34-1.08]	.35 [.29-.41]	-	-	.30 [-.08 -.66]	.65 [.59-.71]
females	**SWF**	.34 [.24-.43]		-		.66 [.57-.76]	
	WB	.82 [.58-.1.07]	.44 [.39-.48]	-	-	.18 [-.07-.42]	.56 [.52-.61]
males	**LT-SP**	.60 [.56-.64]		-		.40 [.36-.44]	
	WB	.91 [.60-1.26]	.35 [.29-.41]	-	-	.09 [-.26-.40]	.65 [.59-.71]
females	**LT-SP**	.33 [.22-.44]		.33 [.23-.43]		.34 [.31-.37]	
	WB	.89 [.69-1.08]	.43 [.39-.48]	-	-	.11 [-.08 -.31]	.57 [.52-.61]

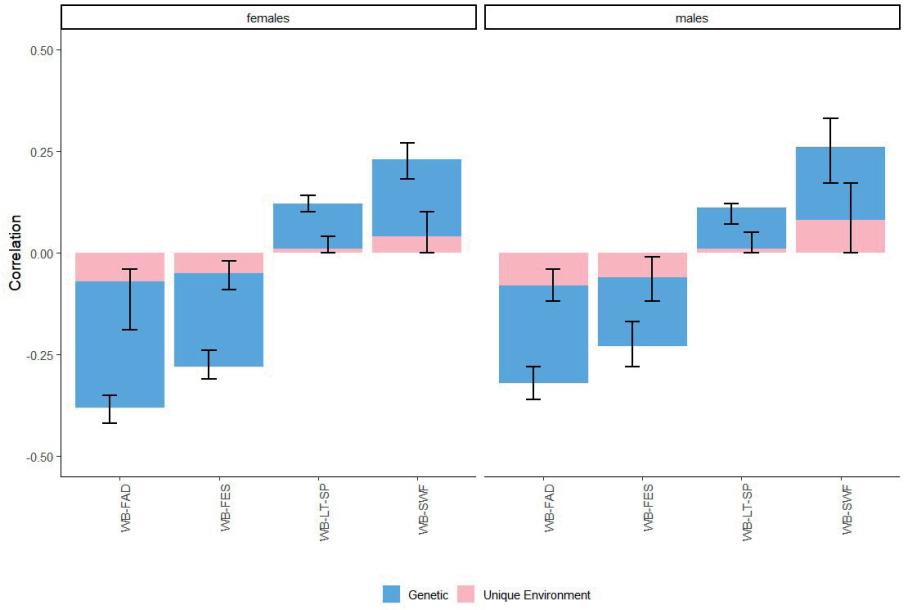


Figure 3. The contribution of genetic and environmental factors to correlations between well-being (WB) and family functioning (FAD), family conflict (FES), leisure time sport/scouting club (LT-SP) and satisfaction with friendships (SWF).

4. DISCUSSION

In the present study, we examined the relation between adolescent well-being and various social variables. We identified significant associations between well-being and family functioning, family conflict, leisure time indoor games, leisure time contact with friends, leisure time sports club/scouting, number of friends, and satisfaction with friendships. Well-being was not associated with leisure time crafts, leisure time making music/choir, leisure time spend on the computer, leisure time going out, leisure time watching TV, and importance of friendships.

Adolescent leisure time physical activity[346,347], different aspects of adolescent friendships[348], and going out[349] have all previously been associated with well-being, just as familial and friendship variables[350–352]. These studies did not, however, examine potential genetic influences on these associations. Based on earlier research indicating that well-being and socio-environmental factors are subject to heritable influences[24,40], we hypothesized that the observed associations might be partially explained by genetic factors. Intra-pair difference score analyses indicated genetic influences on the association between well-being and leisure time spend at sport/scouting clubs, and satisfaction with friendships. Intra-pair difference score associations for family functioning and family conflict suggested a role for both genetic and unique environmental influences. Moreover, these analyses reveal that genes do not seem to play a substantial role in the association of well-being with leisure time indoor games, leisure time contact with friends, and number of friends. Based on the difference score analyses, the relation between well-being and those three variables is most likely explained by common environmental influences. With respect to those friendship variables, a potential explanation is that it is siblings close in age spend time with the same peers. Additionally, parents might stimulate contact with peers through stimulating them to participate in outdoor activities, or alternatively limit time spent with peers based on how strict they are. With respect to leisure time indoor games, which includes chess, board games and reading: the extent to which these things are present in a household is highly influenced by parents, which explains a large role for common environmental influences.

For traits where there was an indication that genetic factors played a role in the association with well-being (i.e., family functioning, family conflict, satisfaction with friendships, leisure time spend at sport/scouting clubs), we performed

bivariate genetic analyses. Common environmental influences did not contribute to associations with well-being, with genetic and unique environmental factors explaining the associations fully. For all traits, the largest part of the association was explained by genetic factors (between 73% and 91%). For females, a higher proportion of the association between well-being and the social traits was explained by genetic factors. For males, twin correlations indicated that D might contribute to variation in well-being and satisfaction with friendships. Additionally, CT-CT correlations for satisfaction with friendships and leisure time sport/scouting club also indicate a potential role for D. This is in line with previous studies on well-being in adolescence, where a role for D was also indicated[82,353].

While we did not directly test for causality in this study, we can draw some inferences based on the genetic and environmental correlations. If there is a causal relation between two traits, it is expected that genetic and environmental factors influencing one trait also influence the other trait (i.e. the genetic and environmental correlation should be significant[354]). If the genetic correlation is significant but the environmental correlation is not, this falsifies the hypothesis of a causal effect. In line with the difference score analyses, we found significant genetic correlations but non-significant unique environmental correlations between well-being and satisfaction with friendships and leisure time sports/scouting club, indicating that genetic factors play a dominant role in these associations. Additionally, we found significant genetic and unique environmental correlations between well-being and family conflict and family functioning, supporting a role for causality in these associations. Yet, it is important to mention that the significant unique environmental correlations with both FAD and FES were small (r_{Em}= -.13, r_{Ef}= -.12 and r_{Em}= -.12, r_{Ef}= -.12, respectively). While this does not falsify the claim that there might be a role for causality, this does indicate that a potential causal association will likely also be of small magnitude. Interestingly, we did not find a significant unique environmental correlation between well-being and satisfaction with friendships, suggesting that the association between those two traits is non-causal, at least in adolescence. While multiple studies identify an association between well-being and friendship quality/satisfaction[351,355], these studies did not yet take into account the potential role of genetic factors. Based on what we find here, the most likely explanation for this association is that those who consider themselves to be satisfied with their lives are more likely to also

consider themselves satisfied with their friendships due to them having a general (genetic) predisposition for positive ratings of life domains. This does not have to come as a surprise since it has been shown that several well-being domains, such as satisfaction with life, satisfaction with friendship, and happiness are significantly associated both phenotypically as well as genetically[82,157]. To check if our averaging the friendship variables over gender did not impact our conclusions, we performed supplementary GEE analyses where we examined same-sex and opposite-sex associations in males and females separately. Our results did not change when we examined these associations separately, even though the association between well-being and satisfaction with friendships was somewhat stronger for same-sex friendships than opposite-sex friendships (see Online Resource eTable 7).

These findings show that there is an important third factor in the association between well-being and several social variables in adolescence that is often unmeasured in psychological research: heritable influences. Phenotypic associations between well-being and different social variables are often found, but it appears that large parts of these associations are attributable to genetic factors. An interpretation of this genetic overlap is that the association between well-being and these variables is likely largely due to a genetic predisposition for appraising one's life positively or negatively. For example, one might evaluate his or her well-being and friendship environment more positively in general because of their genetic predisposition for doing so. While this only pertains to the traits we now studied in more detail (i.e. family environment, friendship satisfaction, and leisure time sport/scouting club), an interesting question for future research would be to study if this genetic influence is also be present for associations with other social variables (e.g. perceived social support in adolescence). Additionally, genetic and environmental correlations indicated causality might be at play in the associations between well-being and family conflict and family functioning, with similar genetic and environmental factors influencing both traits, potentially through a causal chain.

An interesting follow-up for these findings is longitudinal studies, preferably using genetically sensitive designs. For example, if twin data are available, direction of causation models (if there are different modes of inheritance for the traits under study)[356] and genetic cross-lagged models[357] provide genetically sensitive methods for studying causality. In the absence of family data, one can still try to separate genetic from environmental effects if DNA data are available, for example by

incorporating the effect of polygenic scores (scores that reflect individuals' genetic predisposition for a trait based on results from genome-wide association studies) in mediation models[358] or Mendelian randomization models[359]. From a research perspective, it is important that investigations into adolescent mental health correlates take into consideration that these associations might reflect a shared genetic liability. In this study, we aimed to provide more information on these genetic influences, and confirmed that these cannot be ignored whilst studying these traits. This is also important from a clinical perspective, as the aim is to identify modifiable environmental factors in adolescence that improve well-being. What is important to keep in mind is that the mechanism behind (adolescent) well-being is very complex and multifaceted, with every relevant part only inducing a small, if any, change. Based on our results, the family environment seems a valuable part of the "well-being mechanism" that potentially has a small causal influence. This is interesting from an intervention perspective. However, as with any complex mechanism, the influence of a single aspect cannot be interpreted separate from all other effects. This means that its influence is different for different types of people, with strong causal effects being unlikely. Moreover, in this system, we cannot yet say anything about the potential direction of causality: while the family environment might influence well-being, this might also be the other way around or bidirectional. Moreover, the Netherlands is a country with relatively high levels of individualism according to Hofstede's individualism index[360], and it is important to interpret our findings within this a Western context, where relationships and group prosperity have a lower priority [361] than an Eastern context. Satisfaction with life is also known to be more suitable to measure well-being in the context of Western compared to Eastern cultures[362]. An interesting endeavor for future research would thus be to see how these associations vary across cultures and measures of well-being on both a phenotypic and genetic level.

In conclusion, we examined associations between well-being and a set of socio-environmental variables and find that genetic factors play a large role in several of these associations, confirming the importance of taking genetic differences into account. Additionally, we find a potential role for causality in the association between family conflict/functioning and well-being, with overlap in the genetic and environmental factors that influence these traits. From a clinical perspective, the family environment thus forms an interesting target for improving adolescent well-being.

CHAPTER 9

Disentangling potential causal effects of educational duration on well-being, and mental and physical health outcomes

Submitted as: Van de Weijer, M.P., Demange, P.A.D., Pelt, D.H.M., Bartels, M., & Nivard, M. G. (*under revision*). Disentangling potential causal effects of educational duration on well-being, and mental and physical health outcomes.

*supplementary materials accessible at: https://drive.google.com/drive/folders/1Kr7yboBztrnB9MQefOL3TWcPGmD4oPBq?usp=sharing

ABSTRACT

Extensive research has focused on the potential benefits of education on various mental and physical health outcomes. However, whether the associations reflect a causal effect is harder to establish. To examine associations between educational duration and specific aspects of well-being, anxiety and mood disorders, and cardiovascular health in UK Biobank data, we apply four different causal inference methods (a natural policy experiment leveraging the minimum school leaving age, a sibling-control design, mendelian randomization (MR), and within-family MR), and assess if the methods converge on the same conclusion. A comparison of results across the four methods reveals that associations between educational duration and these outcomes appears predominantly to be the result of confounding or bias rather than a true causal effect of education on well-being and health outcomes. Whereas we do consistently find no associations between educational duration and happiness, family satisfaction, work satisfaction, meaning in life, depression, anxiety, and bipolar disorder, we do not find consistent significant associations across all methods for the other phenotypes (health satisfaction, financial satisfaction, friendship satisfaction, neuroticism, and cardiovascular outcomes). We discuss inconsistencies in results across methods considering their respective limitations and biases, and additionally discuss the generalizability of our findings in light of the sample and phenotype limitations. Overall, this study strengthens the idea that triangulation across different methods is necessary to enhance our understanding of the causal consequences of educational duration.

1. INTRODUCTION

There is an extensive body of research examining associations between educational attainment (EA) and mental and physical health outcomes. Existing studies have pointed to EA (measured as years of education, age at leaving education, or diploma obtained) as a correlate of well-being[363], depression[364], quality adjusted life years[365], different cardiovascular outcomes[366], and a wide range of other diseases and disorders[367–369]. Often, EA is interpreted as a modifiable risk factor that might improve outcomes in these different domains, but confounding and reverse causation are difficult to rule out.

Correlational evidence provides us with a first indication of associations between education and (mental) health outcomes. For example, a meta-analysis by Bücker and colleagues suggests a small to medium positive correlation between academic achievement and subjective well-being (SWB) that was stable across different measures of academic achievement and SWB[363]. Similarly, a small but significant correlation has been found between academic achievement and subsequent depression through meta-analysis[370]. In addition, lower education has been associated with a higher risk of different cardiovascular outcomes[366], and lower self-reported health[365].

Such meta-analytic studies offer the opportunity to evaluate and summarize the existing literature, which allows us to identify correlations worth exploring in more detail. However, it is difficult to establish whether these associations reflect causal associations or whether they might be caused by residual confounding (e.g., genetics, socioeconomic status)[371,372]. While confounders can be considered in meta-analysis, it is rarely the case that a large number of studies include the same confounders. Moreover, even if confounding factors could be ruled out, correlational studies would not offer clarity on the direction of causation. For example, while higher levels of education might lead to better access to healthcare, less health problems, and higher health[373], the reverse could also be true: for example, people in good health might have better possibilities to focus on education and reach higher levels of education than those in poor health[374].

A quasi-experimental design that has been applied widely in educational research is to consider compulsory schooling laws where the legal minimum school leaving age is increased[375] as an exposure over which individuals can be reasonably

assumed to have no control. The implementation of these laws serves as a natural experiment where people are quasi-randomly separated in two groups (before and after, or subject to or not subject to the policy change). Assuming that this policy change only directly impacts the number of years someone stays in education, and assuming that is unrelated to confounding factors, this policy change can be used to estimate the direct effect of educational duration on diverse outcomes. Using this design, researchers have found positive effects of educational duration on mental health[376,377], cognitive abilities[378], mortality[379], income[379,380], and cardiovascular health[381]. Nevertheless, there is still considerable disagreement across different studies employing this design due to heterogeneity in study features such as the included instrument, the examined number of years around the reform, or the populations included (see[382]. Additionally, the policy shift only affects those that would otherwise have left school earlier. This is an important caveat that should be kept in mind when interpreting results, since this limits the generalizability of findings to those not affected by the reform[383].

Another quasi-experimental design controlling for several forms of confounding using observational data is the sibling-control design. Comparing outcomes of biological siblings brought up in the same family allows to control for shared environmental confounding (e.g., socioeconomic conditions during childhood), and for shared genetic predispositions. However, factors unique to one of the siblings but not the other and measurement error can still bias the results of sibling-control studies[384]. Additionally, even if we could control for all unshared confounders, the method would not help us determine the direction of causation. If we find that siblings who score higher on well-being also stay in school longer, this could be because well-being causally increases school-leaving age, but the reverse is as likely: school-leaving age might causally increase well-being.

In Mendelian Randomization (MR), one or more genetic variant(s) robustly associated with a predictor variable are used as instrumental variables to examine a potentially causal association between a predictor and outcome. The approach relies on Mendel's laws of segregation and independent assortment, which assume that genetic variants are inherited randomly from one's parents and independent from other genetic variants. Assuming that the genetic variants are 1) robustly associated with the exposure, 2) not associated with potential confounders, and 3) not associated with the outcome of interest other than via

the exposure (no pleiotropy), the genetic variants for an exposure can be used as instruments to examine potential causality between the exposure and an outcome. For example, a genetic variant associated with educational duration that is also indirectly associated with higher well-being (through its association with educational duration) provides supportive evidence of a causal association from education on well-being. Multiple studies have used MR to examine causal links between EA and health-related traits, with suggestive evidence for causal influences on traits like alcohol consumption, physical activity, and cardiovascular outcomes[385,386]. Importantly, these associations are only valid if the three key assumptions mentioned above are met. Unfortunately, it is often difficult to evaluate if the assumption of no pleiotropy is met, as many, or even most, genetic variants exert pleiotropic effects. In addition, unmodeled assortative mating, dynastic effects, and population stratification can spuriously induce associations between the genetic variant(s) and outcomes[387].

A further development of MR is the application of this method in the context of within-family analysis[387]. By performing genetic instrumental variable within sibling pairs, we directly control for the influences of assortative mating, population stratification (siblings share the same population background) and dynastic effects. First, since genetic variants inherited by siblings are random within a family, genotype differences between siblings will be independent of assortative mating. Second, since the effects of parental wealth and status on their offspring is likely similar across siblings, genetic differences between siblings will be independent of dynastic effects. Lastly, genetic differences between siblings are independent of population stratification. Using within sibling MR, Brumpton and colleagues demonstrate that conventional non-family MR estimates for the association between taller height/lower BMI and increased EA were almost entirely attenuated in the context of within-family MR[387]. While within-family MR has important advantages over conventional MR, it is nevertheless still fallible to unmet assumptions (e.g., the presence of pleiotropy) and is also less powerful as it is applied only in siblings within a larger sample.

There are various methods for examining causality in observational data, but all rely on strict assumptions that often are difficult to meet or evaluate. A way in which we can reduce our reliance on these individual assumptions is by applying multiple methods and evaluate the consistency of results and potential discrepancies

therein, in light of the biases that accompany each of these methods. In a study where the effect of body mass index (BMI) on different outcomes was assessed, the authors used both MR (subject to family-level confounding) and non-genetic and genetic within-family analyses (subject to reverse causation)[388]. By verifying that these methods converge upon the same conclusion, the authors increase the certainty that the results were not a by-product of their respective biases. In a similar fashion, Davies and colleagues examined potential causal effects of education on health, mortality, and income using both a design where they leverage the raising of school leaving age and MR, with both methods suggesting similar effects for almost all outcomes[389].

For the current project, we are interested in causal influences on specific aspects of well-being, anxiety and mood disorders, and cardiovascular health. As educational effects on well-being are of primary interest to us, we depart from treating "well-being" as a single unified outcome and separately consider effects on satisfaction with family relations, work, friendships, health, and finances[390]. We rely on four widely accepted techniques for causal inference: we make use of a random natural policy shift in England and Wales in September 1972 that raised school leaving age from 15 to 16 but is unlikely to be related to confounding factors. We perform analyses within sibships to control for shared environmental confounders, and partly control for shared genetics. We make use of an index of genetic variation related to educational attainment as an instrumental variable in MR. Finally, we combine the genetic instrumental variable with within family analysis in sibling pairs. We apply those techniques in a single homogenously measured sample (the UK Biobank), minimizing variation in results due to differences in measurement. By assessing if these different methods converge on the same conclusion, we can be more confident in our conclusions on the potential causal relation between education and the different outcomes.

2. MATERIALS AND METHODS

This project was pre-registered at the Open Science Framework (https://osf.io/s6gha). Deviations from the pre-registration are indicated throughout the manuscript.

2.1 Sample

We used data from the UK Biobank, a large UK cohort study which collected genetic and phenotypic data on ± 500,000 participants between 40 and 69 years old at recruitment[112]. For the current project, we selected individuals of European ancestry (a decision taken to minimize ancestral confounding in genetic analyses) that were born in England and Wales (to ensure participants were likely affected by the school leaving age reform). Specific further sample selection procedures for the four different analyses are described below per analysis.

Education variable

We used UKB data-field 845 "age completed full time education" as our education exposure variable. Participants were asked to answer the question "at what age did you complete your continuous full-time education?". If someone provided an answer below 5, or an answer higher than their age, the answer was rejected. If someone answered with an age higher than 40, the participant was asked to confirm their answer. Since the question was not collected in participants who indicated having a college or university degree, we, in line with the literature[379,391] imputed their age at completed full time education as 21. In case someone provided an answer on more than 1 instance, we used the last available answer as the age at which one completed their full-time education. If the answer at the later time-point indicated a lower age than a previous answer (*N*=72), we coded the answer as missing.

2.2 Outcome variables

General information on item construction and cleaning procedures for these variables can be found in the Supplementary Methods. The following self-report items were included as well-being outcome variables: **general happiness** based on happiness (*UKB ID 4526*) and general happiness (*UKB ID 20459*), **family relationship satisfaction** (*UKB ID 4559*), **financial situation satisfaction** (*UKB ID 4581*), **friendship satisfaction** (*UKB ID 4570*), **work/job satisfaction** (*UKB ID 4537*), **health satisfaction** based on health satisfaction (*UKB ID 4548*) and general happiness with own health (*UKB ID 20459*), and **belief that own life is meaningful** (*UKB ID 20460*). All items were coded so that a higher score indicated a higher level of well-being. For **neuroticism**, we included a summary score (*UKB ID 20127*) that

was based on 12 neurotic domain self-report items. We used a combination of medical record data (*UKB ID 41270*) and self-report data (*UKB ID 20002*) to create binary variables reflecting if someone was ever diagnosed with **depression**, **anxiety**, or **manic or bipolar disorder.** Lastly, a binary variable indicating **cardiovascular problems** was constructed based on vascular/heart problems diagnosed by a doctor (*UKB ID 6150*) or self-reported (*UKB ID 20002*).

2.3 Control variables

We selected four negative control variables: **height** (*UKB ID 50*), **birthweight** (*UKB ID 20022*), **comparative body size at age 10** (*UKB ID 1687*), and **comparative height size at age 10** (*UKB ID 1697*). It is unlikely these variables are causally influenced by additional years of schooling, but the presence of confounding parental variables (e.g. parental SES) might lead to observable but false positive associations. As a positive control variable, we included **average total household income before tax** (*UKB ID 738*), which was split into the four yes/no dichotomous variables: income over 18k, income over 31k, income over 52k, and income over 100k. General information on item construction and cleaning procedures for these variables can also be found in the Supplementary Methods.

2.4 Covariates

As phenotypic covariates, we included sex (*UKB ID 31*), assessment centre (*UKB ID 54*), family size (based on number of (adopted) siblings, *UKB IDs 1873, 3972, 1883 & 3982*) season of birth (based on month of birth, *UKB ID 52*), and year of birth (UKB ID *34*). Genetic covariates included the first 10 genomic PCs and batch (*UKB ID 22000*).

2.5 Genotype data

SNPs from HapMap3 CEU (1,345,801 SNPs) were filtered out of the imputed dataset. A pre-PCA QC was done on unrelated individuals, filtering out SNPs with MAF < .01 and missingness > .05, leaving 1,252,123 SNPs. After filtering out individuals with non-European ancestry, the SNP QC was repeated on unrelated Europeans (N = 312,927). SNPs with MAF < .01, missingness >.05, and HWE p < 10^{-10} were filtered, leaving 1,246,531 SNPs. The HWE p-value threshold of 10^{-10} was

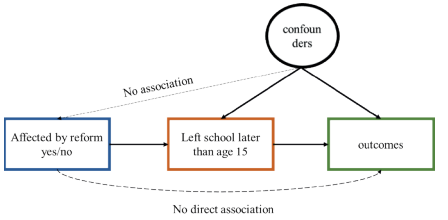
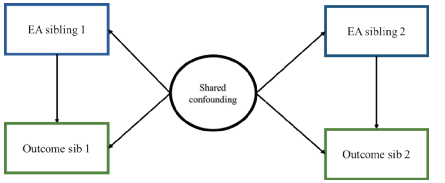
based on: http://www.nealelab.is/blog/2019/9/17/genotyped-snps-in-uk-biobank-failing-hardy-weinberg-equilibrium-test. A final dataset of 1,246,531 QC-ed SNPs was created for 456,028 UKB subjects of European ancestry.

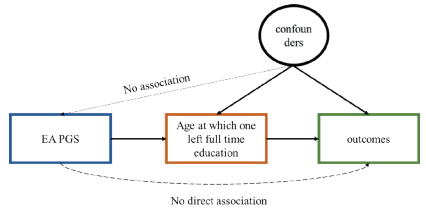
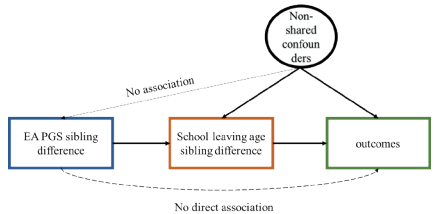
2.6 Analyses

We use four different methods to examine potential causal effects between educational duration and our outcomes. Table 1 provides an overview of these four methods, including their respective advantages and limitations. Sample descriptives per method can be found in Supplementary Table 1. Below, we describe each of the four methods in more detail. All analysis code is available at https://github.com/margotvandeweijer/EA_causality. All continuous outcomes were standardized so that the resulting effect sizes reflect the SD increase in the outcomes for each additional year of education (see Supplementary Table 1 for an overview of the SDs of the included variables).

Table 1

Overview of different methods used in the present study

Method	Short summary	Core assumptions	Core limitations	Visual description
ROSLA reform IV analysis	On 1 September 1972, the raising of school leaving age in England and Wales was raised from 15 to 16. As a result, the compulsory school stay for individuals born in September 1957 and later was a year longer than for those born before September 1957. We used this policy shift as an instrument in instrumental variable analysis, where the reform directly affects school leaving age, but does not directly affect any of our outcomes.	1. The ROSLA reform is not directly associated with any of the outcomes. 2. The ROSLA reform is not associated with potential confounders. 3. The ROSLA reform directly affects school leaving age.	The reform only directly impacts those who would have otherwise left school at age 15. Thus, the instrument only affects a small part of the population and results are not generalizable to the entire population.	
Sibling control design	The chances and support provided by the (early) childhood (shared) environment is considered one of the primary causes of confounding in educational research, these can be controlled for by relating outcomes to differences in the educational measure within sibling pairs.	Familial influences (e.g., parental SES) affect siblings within the same family similarly	Confounders unshared by family members and measurement error can still bias the results	

Method	Short summary	Core assumptions	Core limitations	Visual description
Mendelian Randomization (MR)	MR is a special form of instrumental variable analysis, where the instrument is based on genetic variants associated with the exposure. By using genetic variants as instrumental variants, MR is unlikely to be subject to reverse causality. We used a PGS for EA based on Lee et al. (2018) as the instrument in our analyses.	1. The PGS is not directly associated with any of the outcomes other than through the exposure (no pleiotropy). 2. The PGS is not associated with potential confounders. 3. The PGS directly affects school leaving age.	Key assumptions of the method, like no pleiotropy, do not always hold. Additionally, residual confounders such as dynastic effects can influence the results.	
Mendelian Randomization in sibships	This method is a combination of MR and the sibling control method; it has the same advantages as MR but additionally controls for assortative mating, dynastic effects, and population stratification. We perform MR in a sample of sibships where we take the difference between the sibships on the PGS and school leaving age to remove the effect of family-level confounders.	The same assumptions as MR and the sibling control design	Can still be confounded by unmet assumptions, and is less powerful than conventional MR as the sample is reduced to only sibships.	

2.6.1 Instrumental variable analysis leveraging the raising of school leaving age (ROSLA)

We used the raising of school leaving age (ROSLA) policy reform where the minimum school leaving age was increased from 15 to 16 in England and Wales to examine the effects of longer schooling on our different outcomes. We selected a sample of UKB participants born in a 5 year window (1 February 1955 to 1 February 1960) around the reform (1 September 1972), and excluded related individuals (KING kinship coefficient > .0884) using the *ukbtools* package in R[392]. A binary ROSLA indicator was created for this subset of participants, that indicates if a participant was born before (affected=0) or after (affected=1) 1 September 1957 and was thus affected by the reform or not. Additionally, we transformed the age at which one left fulltime education variable into a binary variable that indicates if an individual stayed in school after age 15 or not[393]. Next, we used two-stage least squares (2SLS) instrumental variable analyses using the *fixest* R package[394], where in the first stage the binary education variable was included as the dependent variable and the binary ROSLA indicator was included as the instrument. In the second stage, we regressed all our standardized outcome variables on the fitted education values from the first stage regression. Both stages included the phenotypic covariates. For comparative purposes, we also run regular (non pre-registred) OLS regression in the same sample where age at which one left full-time education was used to predict the different outcomes (including the same covariates as the ROSLA analyses). To examine the robustness of the ROSLA results, we repeated the analyses using samples born in a 2 and 10 years window around the reform.

2.6.2 Sibling control design

We perform analyses within sibships to control for shared familial background characteristics, and partly control for genetic effects. Biological sibships in the UKB dataset are defined as participants with a kinship coefficient between $\frac{1}{2^{5/2}}$ and $\frac{1}{2^{3/2}}$ and a probability of zero identical-by-state (IBS) sharing > 0.0012[112,395]. Individuals indicating they were adopted were removed from this sample. For each sibship j with i siblings, we start by calculating the average age at which sibships left full-time education $\overline{edu_{oj}} = \sum_1^m edu_{ij}/m$. Next, we calculate each sibling's deviation from the sibship average: $edu_{\Delta ij} = edu_{ij} - \overline{edu_{0j}}$. We use

these estimates in a linear model where each outcome Y_{ij} for sibling i in sibship j is predicted as follows:

$$Y_{ij} = \beta_{00} + \beta_B \overline{edu_{oj}} + \beta_W edu_{\Delta ij} + covariates + e$$

, where β_B is the between-sibship effect estimating if the average school leaving age within sibships is associated with our outcomes, and β_W is the within-sibship effect estimating if a sibling deviating from the sibship school-leaving age average is associated with our outcome measures. Since we examine the effect of these within- and between-sibship estimates on the outcomes of individual siblings, we excluded sibships where only one sibling reported on educational duration, but we did not exclude sibships where not all siblings reported on one or more outcome measures. All phenotypic covariates were included in the analyses.

2.6.3 Mendelian randomization

We used polygenic scores (PGS) for EA in 2SLS instrumental variable analysis as genetic instruments for testing a directed causal association between educational duration and the outcomes. Polygenic scores are aggregate measures of genetic susceptibility for a trait of interest weighted by effect size estimates from genome-wide association studies[396]. To calculate the PGS for EA, we used the summary statistics from the Genome Wide Association Study (GWAS) of years of education by Lee et al.[397], excluding 23andme and British cohorts (*N*= ~245k). Polygenic scores were constructed from the set of genome-wide significant HapMap3 SNPs ($p < 5x10^{-8}$), pruned to be independent (using the package *TwoSampleMR*[398]) using a clumping window of 1000kb and an LD cut-off of R^2= .1. The PGS prediction accuracy for EA was assessed based on the incremental R^2 when including the PGS in a regression with all covariates.

Next, the PGS was used as a genetic instrument in 2SLS instrumental variable analysis in a sample of unrelated UKB participants (KING kinship coefficient > .0884). In the first stage, we predicted age at which one left full-time education (standardized) from the PGSs. In the second stage, the outcome and control variables were predicted from the fitted education values. All phenotypic and genetic covariates were included as covariates in both stages. The MR analyses were conducted using the *fixest* package in R[394]. For comparison, we also

perform regular (non pre-registered) OLS regression in the same sample, where standardized age at which one left full time education is used to predict the outcomes, whilst correcting for the phenotypic covariates.

2.6.4 Mendelian randomization in sibships

Since one of the limitations of MR is its susceptibility to residual confounding stemming from dynastic effects, population stratification, and assortative mating, we additionally perform MR within sibships. We identify siblings in UKB and calculate each sibling's deviation from the sibship average using the same methodology as used for the sibling control design (see "2.6.2 Sibling control design"). Additionally, we use the PGSs calculated for the MR analyses (see "2.6.3 Mendelian randomization") to calculate a PGS average within sibships: $\overline{PGS_{oj}} = \sum_1^m PGS_{ij}/m$, and each sibling's deviation from the sibship average: $PGI_{\Delta ij} = PGI_{ij} - \overline{PGI_{0j}}$. We use these deviation estimates in instrumental variable regression (using the *fixest* package), where in the first stage we predict the sibling education deviation from the sibling PGS deviation. Next, the outcome and control variables were predicted from the first stage fitted education values. Similar to the within-sibling analyses, we excluded sibships where only one sibling reported on EA, but we did not exclude sibships where not all siblings reported on one or more outcome measures. Both the phenotypic and genetic covariates were included.

2.7 Pre-registered interpretation of results

We define an unambiguous causal association as one where the policy shift, the sibling control design, and the mendelian randomization analyses all imply a significant result in the same direction. The absence of significance across these methods would imply the absence of such a result. Due to the lower power associated with our within-sibship MR analyses, we are satisfied if the magnitude and direction of the mendelian randomization within siblings is consistent with the other methods. With respect to statistical significance and multiple testing, we use two significance thresholds: 1) a suggestive threshold where we correct for the number of outcomes (15), so that α=.05/15=.003, and 2) a conservative threshold where we correct for the number of outcomes (15) and analysis types (4), so that α=.05/60=.0008. Inconsistencies across results will be interpreted along the potential biases and assumptions that accompany the different methods.

3. RESULTS

3.1 Instrumental variable analysis leveraging the raising of school leaving age (ROSLA)

Table 2 depicts the results of the ROSLA instrumental variable analyses. Based on the 2SLS models, none of the outcomes are significantly predicted by age at which one left full-time education. This contrasts our comparative OLS analyses, which do not control for unmeasured confounders, where most associations were significant. The F-statistic of the 2SLS analyses ranged from 104.3 to 694.1 depending on the outcome of interest, indicating that our instrument is unlikely to suffer from weak instrument bias. Since the standard errors are relatively large and the Wu-Hausman statistics, which test for the absence of endogeneity, were almost always non-significant at α=.05, it is suggested that the 2SLS and OLS models do not statistically differ. However, the methods do lead to different estimates, suggesting the OLS results are nonetheless subject to considerable bias. Examining these associations in a 2- or 10-year window around the reform did not change our conclusions (see Supplementary Table 2).

These findings contrast earlier findings by Davies and colleagues[379]. Using instrumental variable regression in UK Biobank, they did observe an effect of remaining in school after age 15 on different cardiovascular outcomes and income. The main difference between the current study and the Davies et al. study is the method of correcting for year of birth, where they used a difference-in-difference approach instead of including this variable as a covariate. Therefore, we performed supplementary (non-preregistered) analyses where we in a step-wise fashion added season of birth and year of birth. The results are shown in Supplementary Table 3 and Figure 1. While adding year of birth as covariates might increase the chance that we are overcorrecting, it is evident from these results that the use of a policy experiment as an instrumental variable is very sensitive to the model specification: inclusion year of birth renders previously significant associations with happiness, familial, financial, and work satisfaction, cardiovascular problems, income, birthweight, and height non-significant.

Table 2

Results ROSLA instrumental variable analyses

	Main outcomes								
	Education (fitted)			F-test (1st stage)		Wu-Hausman[a]		Regular OLS education	
	β(SE)	*p*	*N*	*F*	*p*	*wh*	*p*	*β(SE)*	*p*
Happiness	.03 (.21)	.876	28336	352.2	$<2.2x10^{-16}$	.0003	.985	-.005 (.002)	.067
Health satisfaction	.14 (.21)	.508	28420	355.3	$<2.2x10^{-16}$	.116	.733	.03 (.002)	**$<2.2x10^{-16}$**
Family satisfaction	-.11 (.23)	.633	20123	282.6	$<2.2x10^{-16}$	.114	.736	-.01 (.003)	**.0007**
Financial satisfaction	-.33 (.24)	.157	20207	276.8	$<2.2x10^{-16}$	7.18	.007	.05 (.003)	**$<2.2x10^{-16}$**
Friendship satisfaction	.18 (.23)	.424	20038	280.6	$<2.2x10^{-16}$	.978	.323	-.02 (.002)	**$3.82x10^{-12}$**
Work satisfaction	-.35 (.26)	.182	17462	227.6	$<2.2x10^{-16}$	2.25	.134	.01 (.003)	**.0005**
Meaning in life	.33 (.42)	.428	16864	126.1	$<2.2x10^{-16}$	.466	.495	.005 (.003)	.179
Neuroticism	.13 (.15)	.392	39438	568.5	$<2.2x10^{-16}$	5.52	.019	-.03 (.002)	**$<2.2x10^{-16}$**
Depression	-.02 (.04)	.566	47586	739.1	$<2.2x10^{-16}$	.274	.600	-.004 (.0005)	**$3.51x10^{-11}$**
Anxiety	-.008 (.02)	.742	47586	739.1	$<2.2x10^{-16}$	.136	.713	-.002 (.0003)	**$2.68x10^{-9}$**
Bipolar or manic disorder	.01 (.009)	.141	47586	739.1	$<2.2x10^{-16}$	3.35	.067	-.0001 (.0001)	.418
Cardiovascular problems	-.05 (.05)	.301	47586	739.1	$<2.2x10^{-16}$	.056	.813	-.01 (.0007)	**$<2.2x10^{-16}$**
	Control outcomes								
Income over 18k	.03 (.05)	.572	43272	637.2	$<2.2x10^{-16}$	9.41	.002	.17 (.006)	**$<2.2x10^{-16}$**
Income over 31k	.14 (.06)	.026	43272	637.2	$<2.2x10^{-16}$	3.25	.071	.26 (.008)	**$<2.2x10^{-16}$**
Income over 52k	.08 (.07)	.263	43272	637.2	$<2.2x10^{-16}$	4.76	.029	.22 (.008)	**$<2.2x10^{-16}$**
Income over 100k	-.001 (.04)	.973	43272	637.2	$<2.2x10^{-16}$	2.69	.101	.06 (.005)	**$<2.2x10^{-16}$**
Birthweight	.17 (.18)	.337	27852	411.9	$<2.2x10^{-16}$	.139	.709	.10 (.02)	**$1.83x10^{-6}$**
Height	-.04 (.09)	.675	47483	737.5	$<2.2x10^{-16}$	5.48	.019	.17 (.011)	**$<2.2x10^{-16}$**
Comparative body size at age 10	.16 (.09)	.788	46317	729.6	$<2.2x10^{-16}$	2.94	.087	-.01 (.01)	.537
Comparative height size at age 10	-.04 (.09)	.672	46471	717.9	$<2.2x10^{-16}$	.827	.363	.05 (.01)	**$6.98x10^{-7}$**

Note. All **continuous** outcomes were standardized. Assessment center, sex, season of birth, and year of birth were included as covariates.

a. H0 is the absence of endogeneity of the instrumented variables

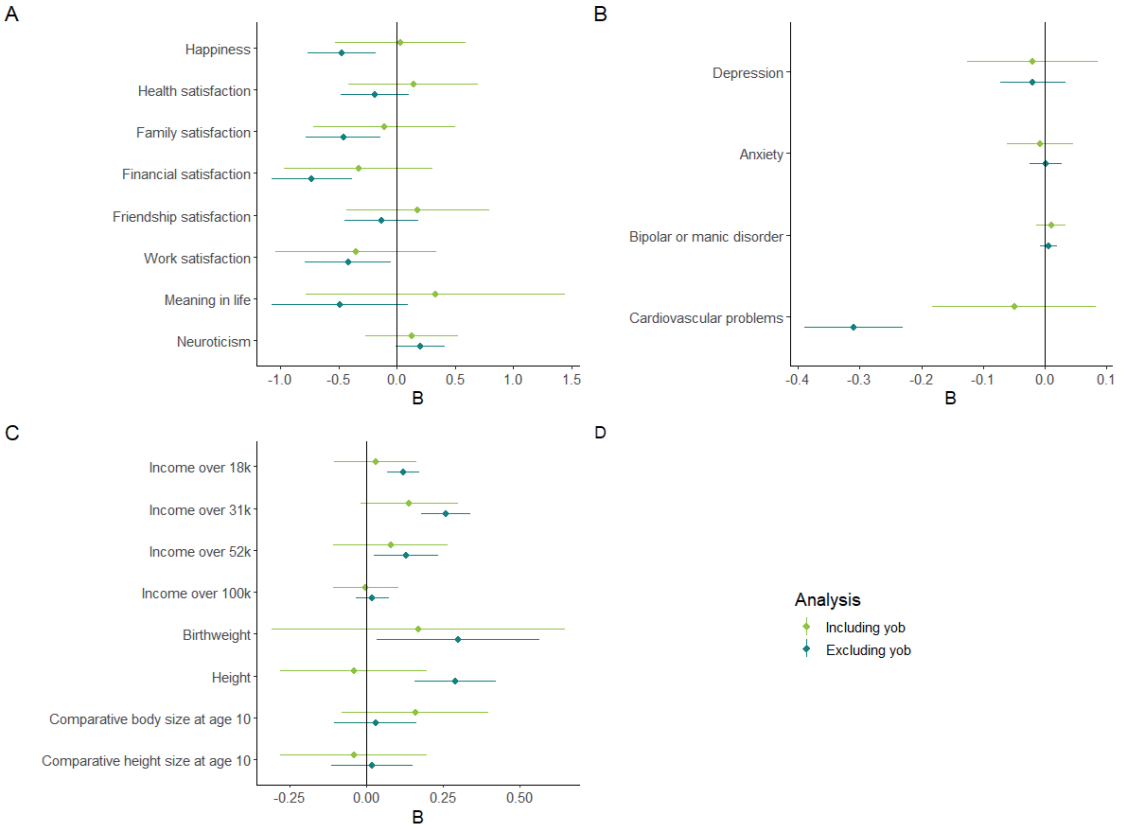


Figure 1. Comparison of ROSLA results including and excluding year of birth (yob) as a covariate for a) continuous outcome measures, b) binary outcome measures, and c) control measures.

3.2 Sibling control design

In total, there were 15,237 families with sibships of at least 2 siblings. The number of included individuals per outcome varied (see Table 3 for the sample size per outcome). The intra-class correlation (ICC) for education, reflecting the amount of total variation in education explained by the family-level, was .40. Table 3 presents the within- and between-sibship estimates from the sibling control analyses. For the main outcomes, the between-sibling estimates for school leaving age were significantly associated (based on the conservative α=.0008 threshold) with happiness, health satisfaction, family satisfaction, financial satisfaction, work satisfaction, neuroticism, anxiety, and cardiovascular problems. However, the within-sibling estimates (indicating a potential causal effect) were only significant for financial satisfaction (β=.024, *SE*=.006, p=1.58x10^{-5}), and neuroticism (β=-.017, *SE*=.004, p=2.4x10^{-5}). With respect to the positive control outcomes, all between-sibship and within-sibship estimates were significant. Negative control outcomes also showed significant positive associations with age at leaving school, except for comparative body size at age 10. As expected, the within-sibling estimates were however not significant for birthweight and comparative height/body size at age 10, but surprisingly still significant for height.

Table 3

Results sibling control analyses

	Main outcomes						
		Within-sibling estimate			Between-sibling estimate		
	N	*β(SE)*	*t*	*p*	*β(SE)*	*t*	*p*
Happiness	17595	-.005 (.005)	-.951	.341	-.014 (.003)	-4.026	**5.69x10^{-5}**
Health satisfaction	17620	.002 (.005)	.426	.670	.015 (.003)	4.391	**1.14x10^{-5}**
Family satisfaction	12531	-.002 (.006)	-.415	.678	-.022 (.004)	-5.573	**2.55x10^{-8}**
Financial satisfaction	12518	.024 (.006)	4.319	**1.58x10^{-5}**	.034 (.004)	8.669	**4.89x10^{-18}**
Friendship satisfaction	12458	-.007 (.006)	-1.176	.240	-.034 (.004)	-8.266	**1.53x10^{-16}**
Work satisfaction	8222	-.0004 (.007)	-.051	.960	.00004(.005)	.008	.993
Meaning in life	10088	.010 (.007)	1.449	.147	-.014 (.005)	-2.915	.004
Neuroticism	25478	-.017 (.004)	-4.224	**2.4x10^{-5}**	-.033 (.003)	-11.182	**5.86x10^{-29}**
Depression	31337	.0004 (.001)	.430	.667	-.001 (.0007)	-1.941	.052
Anxiety	31337	.0009 (.0007)	1.270	.204	-.002 (.0005)	-3.888	**.0001**
Bipolar or manic disorder	31337	.00007 (.0002)	-286	.775	.0006 (.0001)	3.694	.0002
Cardiovascular problems	31337	-.003 (.001)	-1.694	.090	-.012 (.001)	-10.37	**3.62x10^{-25}**
	Control outcomes						
Income over 18k	27348	.019 (.002)	12.397	**3.38x10^{-35}**	.040 (.001)	35.706	**6.11x10^{-273}**
Income over 31k	27348	.024 (.002)	13.463	**3.49x10^{-41}**	.052 (.001)	40.882	**0**
Income over 52k	27348	.021 (.002)	13.433	**5.25x10^{-41}**	.040 (.001)	36.577	**5.22x10^{-286}**
Income over 100k	27348	.005 (.0008)	6.976	**3.11x10^{-12}**	.011 (.0006)	19.185	**1.71x10^{-81}**
Height	31249	.012 (.003)	4.598	**4.28x10^{-6}**	.041 (.002)	22.910	**3.31x10^{-115}**
Birthweight	15543	.008 (.005)	1.411	.158	.020 (.004)	5.320	**1.05x10^{-7}**
Comparative body size at age 10	30470	.003 (.003)	1.247	.213	.004 (.002)	2.227	0.26
Comparative height size at age 10	30432	.007 (.003)	2.564	.010	.02 (.002)	8.848	**9.40x10^{-19}**

3.3 Mendelian Randomization

The EA PGS predicted 0.28% of the variance in school leaving age, which is similar to the predictive power of the EA PGS from Lee et al.[399] that was based on the genome-wide significant SNPs based on a similarly-sized (*N*=293,723) discovery GWAS. Despite the relatively low predictive power, the F-values from our MR analyses ranged from 258.1 to 1032.1 for the different outcomes, indicating that the PGS did not suffer from weak instrument bias.

The full results from the MR analyses are shown in Table 4. Five outcomes were significantly associated with school leaving age based on our conservative significance threshold of α=.0008: health satisfaction (β=.18, *SE*=.04, $p=3.21 \times 10^{-5}$), financial satisfaction (β=.26, *SE*=.05, $p=3.25 \times 10^{-7}$), friendship satisfaction (β=-.23, *SE*=.05, $p=1.08 \times 10^{-5}$), neuroticism (β= -.20, *SE*=.04, $p=1.15 \times 10^{-8}$), and cardiovascular problems (β=-.06, *SE*=.01, $p=2.13 \times 10^{-5}$). All positive control variables, and all negative control variables except comparative body size at age 10 were significantly associated with school leaving age. For comparison, we also associated the outcomes with age at which one left fulltime education in regular OLS regression. We did so to examine if analyses that do not take into account causality through a genetic instrument would indicate an association. When doing so, all outcomes except meaning in life were significantly associated.

Table 4

Results mendelian randomization analyses

	Main outcomes								
	Education (fitted)			F-test (1st stage)		Wu-Hausman		Regular OLS education	
	β(*SE*)	*p*	*N*	*F*	*p*	*wh*	*p*	*β*(*SE*)	*p*
Happiness	-.10 (.04)	.02	192465	586.3	$<2.2x10^{-16}$	4.24	.039	-.02 (.002)	**$8.14x10^{-11}$**
Health satisfaction	.18 (.04)	**$3.21x10^{-5}$**	192643	584.1	$<2.2x10^{-16}$	8.70	.003	.06 (.002)	**$<2.2x10^{-16}$**
Family satisfaction	-.08 (.05)	.13	140133	394.7	$<2.2x10^{-16}$	.881	.348	-.03 (.003)	**$<2.2x10^{-16}$**
Financial satisfaction	.26 (.05)	**$3.25x10^{-7}$**	140843	403.5	$<2.2x10^{-16}$	10.3	.001	.11 (.003)	**$<2.2x10^{-16}$**
Friendship satisfaction	-.23 (.05)	**1.08x10-5**	139869	402.2	$<2.2x10^{-16}$	11.8	6x10-4	-.05 (.003)	**$<2.2x10^{-16}$**
Work satisfaction	.05 (.06)	.459	92871	258.1	$<2.2x10^{-16}$	.12	.734	.03 (.003)	**$2.22x10^{-14}$**
Meaning in life	-.06 (.07)	.348	106741	269.7	$<2.2x10^{-16}$	.86	.354	-.0003 (.003)	.920
Neuroticism	-.20 (.04)	**$1.15x10^{-8}$**	272531	836.3	$<2.2x10^{-16}$	11.9	$5.69x10^{-4}$	-.082 (.002)	**$<2.2x10^{-16}$**
Depression	-.03 (.01)	.004	335076	1032.1	$<2.2x10^{-16}$	4.98	.026	-.006 (.0005)	**$<2.2x10^{-16}$**
Anxiety	-.002 (.006)	.768	335076	1032.1	$<2.2x10^{-16}$	.15	.697	-.004 (.0003)	**$<2.2x10^{-16}$**
Bipolar or manic disorder	.002 (.002)	.290	335076	1032.1	$<2.2x10^{-16}$	.56	.452	-.0006 (.0001)	**$3.7x10^{-8}$**
Cardiovascular problems	-.06 (.01)	**$2.13x10^{-5}$**	335076	1032.1	$<2.2x10^{-16}$	5.16	.023	-.003 (.0008)	**$<2.2x10^{-16}$**
	Control outcomes								
Income over 18k	.18 (.01)	**$<2.2x10^{-16}$**	290541	862.0	$<2.2x10^{-16}$	42.1	$8.47x10^{-11}$	.09 (.0007)	**$<2.2x10^{-16}$**
Income over 31k	.23 (.02)	**$<2.2x10^{-16}$**	290541	862.0	$<2.2x10^{-16}$	47.8	$4.64x10^{-12}$	.12 (.0008)	**$<2.2x10^{-16}$**
Income over 52k	.19 (.01)	**$<2.2x10^{-16}$**	290541	862.0	$<2.2x10^{-16}$	48.4	$3.46x10^{-12}$	.10 (.0008)	**$<2.2x10^{-16}$**
Income over 100k	.07 (.01)	**$<2.2x10^{-16}$**	290541	862.0	$<2.2x10^{-16}$	35	$3.22x10^{-9}$	.003 (.0004)	**$<2.2x10^{-16}$**
Birthweight	.18 (.05)	**$6.68x10^{-5}$**	165453	528.8	$<2.2x10^{-16}$	11.5	$6.95x10^{-4}$	.03 (.003)	**$<2.2x10^{-16}$**
Height	.16 (.02)	**$1.34x10^{-13}$**	334307	1033.3	$<2.2x10^{-16}$	14.4	$1.50x10^{-4}$	.08 (.001)	**$<2.2x10^{-16}$**
Comparative body size at age 10	.04 (.02)	.061	325062	1022.2	$<2.2x10^{-16}$	2.88	.089	.004 (.001)	.002
Comparative height size at age 10	.10 (.02)	**$7.39x10^{-6}$**	326456	1003.9	$<2.2x10^{-16}$	978	.002	.03 (.001)	**$<2.2x10^{-16}$**

Note. sex, family size, season of birth, year of birth, assessment center, batch, and the first 10 genomic PCs were included as covariates for the MR analyses. The OLS regression is the prediction of the outcomes with education including the same covariates, with the exception of batch and the genomic PCs. All **continuous** outcomes and age at which one left fulltime education were standardized

Table 5

Results mendelian randomization analyses within sibships

	Main outcomes						
	Education deviation (fitted)			F-test (1st stage)		Wu-Hausman	
	β(SE)	p	N	F	p	wh	p
Happiness	.06 (.13)	.652	17595	27.2	$1.82x10^{-7}$	.239	.625
Health satisfaction	.02 (.13)	.887	17620	27.2	$1.85x10^{-7}$	.017	.897
Family satisfaction	-.06 (.14)	.672	12562	19.5	$9.89x10^{-6}$	.170	.680
Financial satisfaction	.24 (.15)	.112	12549	19.4	$1.04x10^{-5}$	2.29	.130
Friendship satisfaction	-.32 (.16)	.040	12458	21.1	$4.48x10^{-6}$	5.06	.025
Work satisfaction	.05 (.19)	.768	8253	11.9	$5.72x10^{-4}$	.088	.767
Meaning in life	.36 (.19)	.060	10119	15.9	$6.70x10^{-5}$	4.23	.040
Neuroticism	-.11 (.12)	.322	25478	32.7	$1.09x10^{-8}$	.736	.391
Depression	.00007 (.03)	.998	31337	35.6	$2.42x10^{-9}$	.0002	.989
Anxiety	-.01 (.02)	.510	31337	35.6	$2.42x10^{-9}$	.501	.479
Bipolar or manic disorder	.01 (.01)	.162	31337	35.6	$2.42x10^{-9}$	2.10	.147
Cardiovascular problems	-.10 (.05)	.067	31337	35.6	$2.42x10^{-9}$	3.49	.062
	Control outcomes						
Income over 18k	.09 (.05)	.080	27348	31.5	$2.03x10^{-8}$	1.97	.160
Income over 31k	.14 (.06)	.020	27348	31.5	$2.03x10^{-8}$	4.20	.041
Income over 52k	.11 (.05)	.028	27348	31.5	$2.03x10^{-8}$	3.56	.059
Income over 100k	.002 (.02)	.936	27348	31.5	$2.03x10^{-8}$	.022	.882
Birthweight	.13 (.15)	.391	15543	19.7	$9.01x10^{-6}$	.681	.409
Height	.03 (.07)	.701	31280	36.2	$1.84x10^{-9}$	.057	.811
Comparative body size at age 10	.12 (.08)	.137	30432	34.5	$4.28x10^{-9}$	2.24	.135
Comparative height size at age 10	.09 (.08)	.224	30470	34.1	$5.29x10^{-9}$	1.34	.246

3.4 Mendelian randomization in sibships

The results from the MR analyses within sibships can be found in Table 5. While our instrument was much less powerful than in regular MR, all F-statistics were higher than 10 (which is commonly used as a rule of thumb to avoid bias[400]). None of the associations (for both control and outcome variables) were significant after correcting for multiple testing.

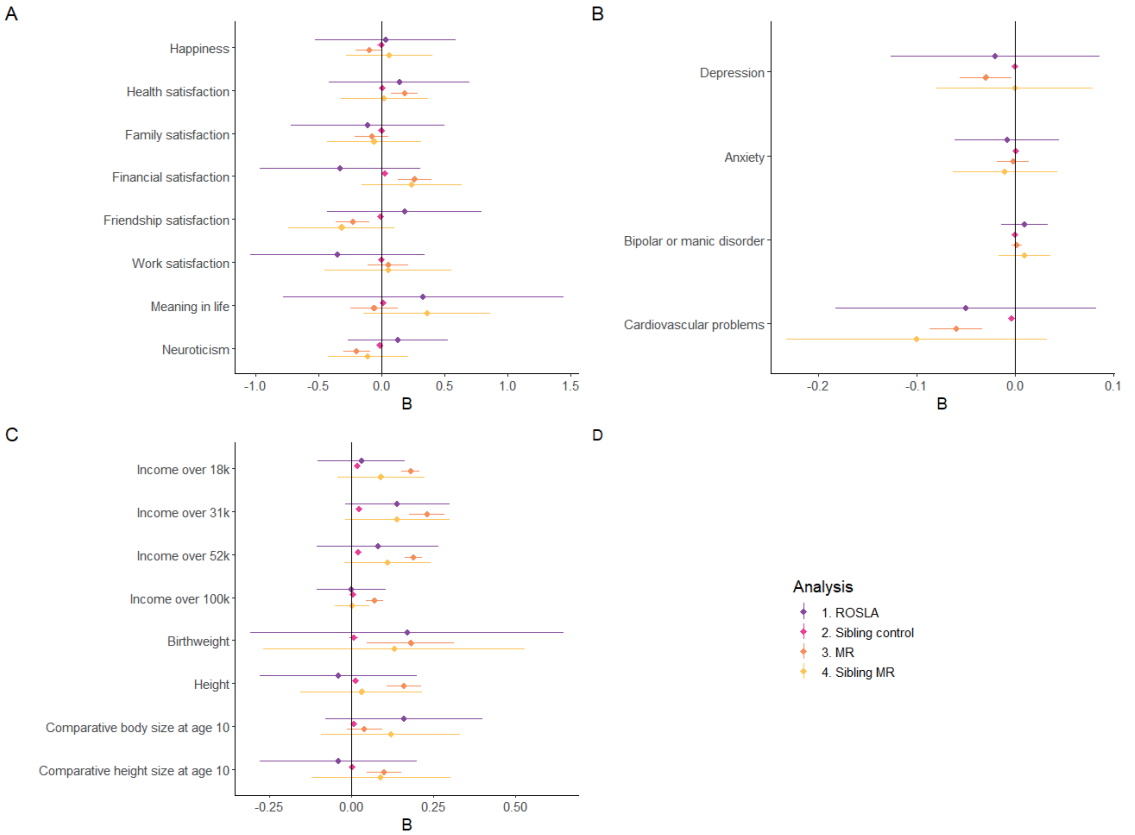


Figure 2. Overview of the results from the different analyses for a) continuous outcome measures, b) binary outcome measures, and c) control measures

3.5 Interpretation of results

An overview of all results from the four different methods can be found in Figure 2. As mentioned in the methods section, we define an unambiguous result as one which is consistent across all methods. Additionally, due to the lower power associated with our within-sibship MR analyses, we are satisfied if the magnitude and direction of the mendelian randomization within siblings is consistent with the other methods.

The ROSLA estimates displayed in Figure 2 reflect the associations where year of birth was included as a covariate (as pre-registered). With respect to our main outcomes, we found non-significant associations across all four methods for: happiness, family satisfaction, work satisfaction, meaning in life, depression, anxiety, and bipolar disorder. Educational duration was positively associated with financial satisfaction, and negatively associated with neuroticism in the sibling-control and MR analyses, but these associations were non-significant in both the ROSLA and within-sibling MR. The within-sibling MR estimates for these two variables were however in the same direction, and of comparable magnitude as the conventional MR results. Lastly, educational duration was significantly positively associated with health satisfaction and significantly negatively associated with cardiovascular outcomes in the conventional MR analyses only. Thus, overall, the different analyses do not seem to converge on a consistent conclusion.

We included different income classes as positive control variables, as we expected these to be causally and positively associated with educational duration. Only in the sibling control and MR analyses were the different income variables significantly associated with educational duration. The non-significant within-family MR estimates for income were in the same direction but of slightly smaller magnitude as the conventional MR. When not including year of birth as a covariate in the ROSLA analyses, the first three income classes were significantly associated with educational duration, suggesting a potential overcorrection in our ROSLA analyses (Supplementary Table 3). With respect to our negative controls, height was significantly associated with education in both the sibling-control and MR analyses. Additionally, birthweight and comparative body height at age 10 were significantly associated with education in the MR analyses. Associations with these negative control phenotypes suggests the possible presence of residual bias.

4. DISCUSSION

Our study was designed to disentangle causal effects from confounding in the association between educational duration and different well-being, and mental and physical health indicators. To this end, we applied four established techniques for causal inference to a homogeneous sample, the UK Biobank. We find consistent non-significant associations for happiness, family satisfaction, work satisfaction, meaning in life, depression, anxiety, and bipolar disorder. However, we do not find robust significant associations across all four methods for health satisfaction, friendship satisfaction, financial satisfaction, neuroticism, and cardiovascular outcomes. The absence of significant consistent results suggests that associations between educational duration and well-being, mental and physical health are largely confounded or biased by reverse causation. Alternatively, a small causal effect may exist but power in one or some of our techniques may have been insufficient to detect it.

Overall, in our first set of analyses (based on the ROSLA), we do not find significant associations with any of the outcomes, including our positive controls. This contradicts an earlier study similarly examining the causal effect of education in the UK in light of the ROSLA reform, where a causal effect was found for different cardiovascular outcomes, income, and height[17]. The main difference between the current study and the Davies et al. study is the method of correcting for year of birth. Whereas Davies and colleagues employ a difference-in-difference approach where the data is stratified by year of birth, we directly include year of birth as a covariate. To examine the effect of year of birth on the associations, we ran supplementary analyses where we compare the associations with and without year of birth as a covariate. When including year of birth, educational duration no longer significantly influenced our positive control variables (the four income classes), suggesting that by including year of birth in this set of analyses we might be overcorrecting. This degree of sensitivity of the model to inclusion of covariates, and to what way covariates are taken into account, does complicate the interpretation of our findings.

Although the ROSLA analyses did not result in significant associations, we found significant associations between educational duration and health satisfaction, friendship satisfaction, and cardiovascular outcomes, but only in the conventional

(non-family) MR analyses. Moreover, we found significant associations with financial satisfaction and neuroticism in both the conventional MR and within-sibship analyses. For the latter, we found that people who stayed in school longer had higher financial satisfaction and lower neuroticism. While these associations were not significant in the within-sibling MR analyses, the direction of effect was consistent. Besides our potential overcorrection issue, it could be argued that discrepancies with the ROSLA analyses are caused by the caveat that the ROSLA results only apply to those who would have left school at age 15 in the absence of the reform. In this sense, the ROSLA results are less generalizable to the population than the other methods as the reform did not affect those who would have stayed in school until age 16 or later irrespective of the reform. Additionally, one of the possibilities that comes to mind when interpreting the results for financial satisfaction is the potential mediation of income. We therefore re-ran the sibling control analyses including the sibling deviation in income from the sibship income average as a covariate[401] (see Supplementary Table 4). We found that the standardized estimates for both financial satisfaction and neuroticism decreased substantially (with similar standard errors), resulting in non-significant associations controlling for income, suggesting that the association between educational duration and these two outcomes may be mediated through income. Since financial satisfaction is partly the result of one's income, this finding is not surprising. With respect to neuroticism, a previous MR study found evidence for bidirectional causality between education and neuroticism, but did not consider potential mediation of income[402].

The most consistent finding to emerge from the data is the lack of evidence for a causal effect of educational duration on happiness, family satisfaction, work satisfaction, meaning in life, depression, anxiety, and bipolar disorder. The association between educational duration and these outcomes was non-significant, irrespective of which method was applied, while OLS did confirm there was an association. In these OLS analyses, we observe that almost all outcomes were significantly predicted by school leaving age. It is therefore likely that these associations are subject to confounding and/or reverse causation, and are unlikely to reflect direct, causal effects. Alternatively, our analysis may have lacked the statistical power to detect a small true causal effect. One important note in the context of our findings is that we examine variation in educational duration from a

minimal schooling leaving age onward, and do not examine the effect of attending education in general. It is therefore important to note that schooling in general has important pecuniary and non-pecuniary consequences[403], but that our study suggests a lack of evidence for causal effects of variation in educational duration on variation in (mental) health outcomes beyond a minimum school leaving age.

When applying the methods used here in isolation, it is often difficult or impossible to evaluate all the respective limitations and assumptions. A strength of this study is that we try to minimize our reliance on any one set of assumptions by applying various existing approaches for causal inference that rely on different assumptions to account for possible confounding and bias, and triangulate results. In doing so, we found that the different causal inference approaches led to heterogeneous results. Since we investigate the same measures in (largely) the same population, differences in results across methods are most likely attributable to the methods themselves. Importantly, if we decided to focus on only one of these methods for the current paper, we would have drawn very different conclusions than we do now. With respect to health satisfaction, friendship satisfaction, and cardiovascular outcomes, these were only significantly predicted by educational duration in the conventional MR analyses, but not in any of the other analyses. Evaluating this discrepancy in light of the characteristics of the different methods, it is possible that these associations are caused by a familial or population effect that is uncontrolled for in conventional MR but is controlled for in the other analyses. Additionally, the MR sample was the largest sample we examined, and it might be the case that an increase in sample size for the ROSLA and within-sibship analyses would allow us to detect smaller effects that remain undetected using the current sample size. Regardless of power, the negative control traits suggest a reliance on MR alone risks false positive results for obvious reasons: the significant causal effects of educational duration on birthweight and height at age 10 cannot be true effects.

While we tried to account for the limitations of the separate methods by means of triangulation, our results are still sensitive to our sample and measurement characteristics. We used a relatively homogeneous sample that allowed for a straightforward comparison between methods, but this also limits the generalizability of our findings. More specifically, the UK Biobank sample is known to suffer from a "healthy volunteer" bias, where participants are more healthy

than the general population[404,405]. Additionally, participants are more likely to be older, female, and live in more socioeconomically advantaged areas than non-participants[404]. Moreover, we used relatively broad, imprecise phenotype and disease definitions. For example, we included all depression diagnoses present in UK Biobank under the umbrella "depression", and all cardiovascular-related diagnoses under the umbrella "cardiovascular outcomes". For our continuous phenotypes (except neuroticism), we used single items to measure the phenotypes. It is possible that more precise phenotype and disease definitions could reduce measurement error and influence power. Additionally, it has been argued that quantitative education measures such as years of education is not an optimal measure of education, especially in the context of nonpecuniary returns of education[406]. More qualitative measures of education, such as teaching methods or curricula differences might be better suited in this context, but these data are difficult to acquire and analyse on a large scale. We did use single items for our well-being phenotypes, but we did not treat well-being as a unidimensional construct. Alternative to looking at a general well-being item or sum-score, we assessed if educational duration influenced specific well-being aspects, such as work satisfaction and meaning in life.

We applied instrumental variable analysis in the context of a natural experiment, sibling-control analysis, mendelian randomization, and within-sibship mendelian randomization to a large UK sample to disentangle potential causal effects of education duration on several mental and physical health outcomes. A comparison of results across these four methods illustrates that 1) associations between education and these several outcomes are largely confounded, and 2) triangulation of evidence across different methods is necessary to examine the results in light of their respective limitations. Notwithstanding the relatively limited generalizability of our findings across different cultures, time frames, and educational systems, this work provides valuable insight into the complexities of establishing the causal effects of educational attainment on important life outcomes.

CHAPTER 10

Summary and General Discussion

SUMMARY

When thinking about biological psychology, or behavior genetics, well-being might not be the first phenotype that comes to mind for most people. If the goal of well-being research is to increase population (or individual level) well-being, why would we want to know anything about the genetics of well-being? Since we are certainly not going to use this knowledge to change people's DNA or biology, what is the added value of this research?

Fortunately, my predecessors in this field had the answers to this question years before I even started thinking about well-being. As discussed in the introduction (**Chapter 1**) of this thesis, we know that a substantial part of individual differences in well-being can be attributed to genetic differences between people[24,25]. While we could choose to focus our efforts on the part of well-being variation that is concerned with environmental factors and ignore the part associated with genetic factors, this would never provide us with a complete picture of well-being. What is more, genetic- and environmental factors do not act in solitude: genetic predispositions influence what type of environments we are exposed to, and our environment can also influence the relative extent to which genetic differences are important for well-being, and whether we get the opportunity to live up to our "genetic potential". Thus, we can study and manipulate the environment in relation to well-being, but if our research designs neglect genetics, we risk drawing faulty conclusions.

In 2019, Røysamb and Nes published an important correspondence on the role of genetics in subjective well-being[407]. They discuss three topics that need to be addressed: 1) the specific genetic variants and pathways important for well-being, 2) gene-environment correlations and interactions, and 3) how we can use gene-environment knowledge to develop tailored interventions that can increase and sustain well-being. In that same year, my (co-)promotors dr. Bart Baselmans and prof. dr. Meike Bartels (and colleagues, including myself) published a genome-wide association meta-analysis that vastly increases our knowledge on the first topic[108]. The study identifies over 300 genetic variants involved in the well-being spectrum, and additionally find that genes differentially expressed in the subiculum and GABAergic interneurons are enriched in their effect on well-being. Since my own work on this thesis started around the same time this paper was published, I

was in the fortunate position to build on this knowledge and focus my efforts on the second topic: how genes and the environment dynamically interact in their influence on well-being. By summarizing and discussing the findings presented in this thesis, I hope to also shine some light on the third topic in this discussion by theorizing how these findings might be translated to improve individual and population well-being.

Section I: Re-evaluating well-being phenotypes and genetics

In **Chapters 2 and 3**, I start with an assessment of where we are in the field on both a genetic and phenotypic level. In the introduction of this thesis, I devoted considerable attention to how philosophical ideas about well-being transformed into modern-day psychological constructs. While the ancient separation between hedonism and eudaimonism is theoretically explained by data-free philosophical perspectives, the question remains whether this structure can also be captured by modern-day questionnaires on constructs such as subjective- and psychological well-being. This is a question that has traditionally been tackled using factor analytic designs, where commonalities between different well-being constructs is explained by one or multiple overarching factors. While this research has led to important insights, consensus with respect to how we should define the well-being framework has not been reached. In **Chapter 3**, we aimed to get more insight in the well-being spectrum by taking a network approach. First, in a sample of *N*=1343 Netherlands Twin Register (NTR)[63] participants, we examined potential item redundancy based on associations between satisfaction with life, subjective happiness, quality of life, flourishing, self-rated health, depressive symptoms, neuroticism, and loneliness items. After excluding redundant items, we fitted a network in an independent sample of *N*=759 participants. We found a final network consisting of a positive cluster including satisfaction with life, subjective happiness, and flourishing items, and a negative cluster including depressive symptoms, loneliness, and neuroticism items. Nevertheless, the two clusters were also densely connected through multiple connections. For comparative purposes, we also run factor analyses where we find that eight independent but moderately to strongly correlated factors (corresponding to the included constructs) were a better fit to the data than a model with one or two overarching well-being factors.

As mentioned above, there was already substantial genome-wide evidence for many well-being related genetic variants at the start of this thesis. While the data-driven genome-wide approach was already common practice in the field of behavior genetics at that point, this had not yet translated into all other fields, such as positive psychology. Within this field, common practice was (and still sometimes is) the hypothesis-driven candidate gene approach. In a candidate gene study, a single gene is examined in relation to a phenotype based on a biological hypothesis. While this sounds theoretically intuitive, genome-wide evidence soon revealed that it is extremely difficult to hypothesize which variants might influence a phenotype[100]. Most psychological phenotypes are influenced by hundreds to thousands genetic variants, many with unknown or unclear biological functions. Since many candidate gene studies had been performed for well-being and since the results of these studies had never been reviewed in light of genome-wide evidence, we first performed a systematic literature review of candidate gene studies for well-being in **Chapter 2.** We identified 41 well-being candidate gene studies examining different genetic variants, variable number tandem repeats (VNTRs), and gene-environment interactions. For the candidate genetic variants, we performed a look up in the summary statistics from the Baselmans and colleagues genome-wide meta-analysis[108]. Moreover, we re-examined the VNTRs and several gene-environment interactions in UK Biobank data[112], a dataset much larger than typically used in candidate gene studies. We did not find support for any of the candidate genes or interactions, suggesting that earlier candidate gene findings are false positives. Based on these results, we strongly advice researchers in the well-being field to abandon the candidate gene approach and focus their efforts on genome-wide approaches.

Section II: The well-being exposome

In **Chapters 4 and 5**, I evaluate the impact of the environment on well-being in order to contribute to a more complete picture of the well-being exposome. Before I started working on this thesis, associations between well-being and environmental factors had been studied across various contexts and timeframes. However, clear consensus on which specific environmental factors are important (and in which context) was still lacking. This lack of consensus has multiple reasons, for instance variations in study design, well-being measures, or environmental measures and

the pick-and-choose approach, with studies focusing on single environmental factors in isolation. In an attempt to more systematically study environmental associations with well-being, we conducted a data-driven environment-wide association study (EnWAS) in **Chapter 4**. In this EnWAS, we examined satisfaction with life data from the NTR (*N*=11,975) in relation to 139 neighborhood factors extracted from the Geoscience and Health Cohort Consortium (GECCO)[175]. The neighborhood indicators, which were linked to NTR data based on 4-numeric postal code, were variables in various domains, such as the physical environment, socioeconomic factors, accessibility, education, livability, care, and sports. We identified 21 environmental factors significantly associated with well-being. After taking into account multicollinearity between the variables, socioeconomic status and safety were indicated as the most important factors in explaining individual differences in well-being. Additionally, we used a well-being spectrum polygenic score (PGS)[108] to examine if we could find evidence for gene-environment correlation, which would reflect a role of the genetic predisposition for well-being in the exposure to certain environmental factors. We did not find any evidence for gene-environment correlation. This could be due to an absence of these effects, but more likely, this is due to a lack of power. By performing EnWAS studies across multiple contexts and populations, we can slowly start piecing together the well-being exposome, which can in turn can be used to inform public policy.

One of the things we noticed when evaluating the results of our EnWAS was that all the associations we identified between neighborhood factors and well-being were based on very small effect sizes. In genome-wide research, it is common to find small genetic effects that by themselves do not contribute a lot to variation in an outcome. Therefore, it is common practice to aggregate individual genetic effects into polygenic scores and use these scores for (clinical) prediction and for follow-up analyses. Inspired by this approach, we aggregate individual social and physical environmental effects into poly-environmental scores (PESs) in **Chapter 5**. We calculated two PESs for satisfaction with life in an NTR sample: one based on subjective socioenvironmental indicators (PES-S), and one based on objective physical environmental indicators (PES-O). We found that the PES-S explained approximately 36% of the variance in satisfaction with life, which is about half of the environmental variance. The PES-O, on the other hand, did not explain any variance when combined in one model with the PES-S. The PESs predicted pre-

pandemic and pandemic well-being to a similar extent, and predicted different well-being (-related) phenotypes to varying degrees. As a follow-up, we calculated a similar PES-S in UKB data, which explained approximately 12% of the variation in happiness. The smaller amount of variance explained in UKB compared to NTR can be traced back to the unavailability of important socioenvironmental predictors such as having a partner and stress at work/home. Lastly, we examined potential gene-environment correlation in both the NTR and UKB dataset by associating PGSs with the PES-Ss. The PGS predicted the PES in the UKB, but not in the smaller NTR, dataset, suggesting the presence of gene-environment correlation. Overall, the study demonstrates the usefulness of (socioenvironmental) PESs for studying the well-being exposome, with multiple potential future research applications such as how these environmental scores vary across different cultures, contexts, and ages.

Section III: Well-being in light of the COVID-19 pandemic

In the middle of my PhD trajectory, the COVID-19 pandemic struck. While this unfortunately put a damper on my well-being data collection, it also led to a novel "NTR COVID-19 data collection". As part of this data collection, participants were asked to report on their well-being and health during the first lockdown of the pandemic. In **Chapters 6 and 7**, we treat the first lockdown of the COVID-19 pandemic as a natural experiment where people's environments radically and abruptly changed. First, in **Chapter 6**, we use a bivariate twin design to compare Self-Rated Health (SRH) (*N*=16,127) before and during the pandemic (*N*=17,451) in individuals without a suspected COVID-19 infection. Surprisingly, we found that the majority of the sample (66.7%) reports increased SRH during the first lockdown. We theorize that these increases do not necessarily reflect objective increases in health, but that uninfected individuals might evaluate their health more positively than under usual circumstances, a phenomenon that Recchi and colleagues refer to as "the eye of the hurricane paradox"[262]. Another question we were interested in answering was if we could detect any changes in the genetic architecture of SRH before compared to during the pandemic. To this end, we fitted a bivariate twin model where we decomposed the (co-)variance of SRH into genetic and environmental sources of (co-)variation. The model revealed no significant changes in the variance decomposition, with a heritability estimate of

45% at both time-points. In other words, the relative extent to which genetic and environmental factors explain individual differences in well-being did not change during the first lockdown.

Another phenotype we measured in the COVID-19 survey was quality of life (QoL), based on the Cantril ladder[65], a single-item indicator of subjective well-being that is scored on a scale from 1 (lowest QoL) to 10 (highest QoL). In **Chapter 7**, we took a closer look at what happened to QoL scores during the first lockdown. When comparing pre-pandemic QoL (*N*=25,772) with pandemic QoL (*N*=17,222) in the sample of multiples and their family members, we found that average QoL decreased from 7.73 to 7.02. Comparing this to our results for SRH, it is interesting to see that during the first lockdown, it was especially mental (and not physical) health that seemed to decrease. Similar to our SRH project, we again examined potential changes in the genetic architecture of QoL. For this project, we made use of full family pedigree data and the "Mendel" program, taking into account correlations between all family members. We observed an increase in total variance in QoL, mainly driven by an increase of unique environmental variance during the pandemic. Relatively, the unique environmental influence increased from 26.3% to 58.7%, whereas the heritability decreased from 30.9% to 15.5%. We hypothesize that this increase in unique environmental variance is the result of the lockdown measures not impacting everybody equally. For example, while some people had to work from home, others were still expected at their usual workplace. In addition to informing us about the impact of the COVID-19 pandemic on well-being, this study is also an interesting example of the dynamic and context-dependent nature of genetic influences on complex traits.

Section IV: Moving towards causal approaches

In the previous chapters, we examine different types of associations: associations between family members, associations between well-being items, and associations between well-being and different environmental and genetic factors. However, since this is all correlational evidence, we refrained from making statements about causality. In **Chapters 8 and 9**, we move toward more causality-oriented approaches. First, in **Chapter 8**, we were interested in the role of socio-environmental factors in adolescent well-being. In *N*=4,700 twin pairs from the NTR, we examine associations between well-being and family conflict

and functioning, number of friends, friendship importance and satisfaction, and leisure time variables. Next, when twin-difference scores indicated a role for genetic factors, we used bivariate genetic models to quantify genetic and environmental contributions to these associations. Using these methods, we find evidence for large (73-91%) genetic influence on the associations between well-being and family conflict and functioning, leisure time sport/scouting clubs, and satisfaction with friendships. While these methods are not indicative of a causal relation between well-being and these variables, it does bring us a step closer to understanding where the observed associations stem from.

Lastly, in **Chapter 9**, we apply four different causal inference methods to examine potential causal associations between educational duration and several mental and physical health outcomes in UK Biobank data. We make use of 1) a natural policy experiment leveraging the minimum school leaving age ($N \approx 30{,}000$), 2) a sibling-control design ($N \approx 18{,}000$), 3) mendelian randomization (MR) ($N \approx 200{,}000$), and 4) within-family MR ($N \approx 18{,}000$). By comparing results across the four methods, we aimed to examine if observed associations are confounded or biased by reverse causation. While we do find consistent non-significant associations between educational duration and happiness, family satisfaction, work satisfaction, meaning in life, depression, anxiety, and bipolar disorder, we do not find consistent significant associations across all the different methods for the other outcomes (health satisfaction, financial satisfaction, friendship satisfaction, neuroticism, and cardiovascular outcomes). In this way, triangulating across different causal inference designs allowed us to conclude that most of the observed associations between educational duration and mental and physical health outcomes are confounded.

GENERAL DISCUSSION

My goal was to increase our knowledge of the well-being exposome and gene-environment interplay, and in turn use that knowledge to inform policy and practice. In what follows, I elaborate on how I believe the work in this thesis has advanced the field, and I provide some suggestions for directions for future research based on the knowledge gaps that still exist.

Genetics and positive psychology

One of the ways this thesis has aided progress is by bringing more clarity on well-being genetics, especially to the field of positive psychology. By reviewing candidate gene literature in the context of better-powered genome-wide evidence, we were able to debunk some myths about popular hypotheses such as the serotonin-transporter gene hypothesis. In this way, we hope to inspire well-being researchers to replace candidate gene approaches with genome-wide approaches. Performing analyses like GWAS might sound daunting to those novel to this approach: it requires one to not only invest in genotyping, but also to become an expert in complex methods unfamiliar to them. However, this is not necessarily true. One of the great developments, in my opinion, of the past years is that we are entering the era of *open science*. Amongst other things, this means that researchers everywhere are freely making their code and data available. In the context of genetics, this means that one could download summary statistics from large GWAS efforts online for free and can use these data without ever genotyping an individual. For instance, innovative methods such as genomic structural equation modelling[408] allow you to use GWAS summary statistics to model genetic associations between traits in a structural equation modelling framework. What is more, as I demonstrate especially in Chapter 9, genetic data can also be used as a tool for examining causality. Pingault and colleagues have written a useful paper on the use of genetic data for causal inference[409]. In this paper, they describe how family-based designs (sibling and twin designs, adoption at birth and in vitro fertilization designs, and the direction of causation model), mendelian randomization, extensions of mendelian randomization, and other emerging approaches can be used for causal inference. In this way, genetic research and data is not only interesting for those who want to know more about the genetics/biology underlying a trait, but also for those interested in studying causal associations.

Mapping a data-driven well-being exposome

Both theory-driven and data-driven research have their own advantages, and most well-being research on environmental exposures has traditionally taken a theory-driven approach. While this is definitely valuable research, there still seem to be many inconsistencies and we are far from having a complete and comprehensive overview of the well-being exposome. Inspired by the theory- to data-driven shift in genetics, we explored the value of taking a data-driven approach, in the form of an EnWAS, for studying environmental factors. In light of open science, we found it important to pre-register our design and report all results, regardless of their statistical significance. Multiple advantages accompany this data-driven approach: first, by examining a wide range of environmental factors instead of zooming in on a few, we might find unexpected associations that would otherwise have been overlooked. Relatedly, we can use the outcomes to generate new hypotheses that can be explored in more depth in follow-up analyses. Lastly, it offers a replicable means of mapping the exposome where we can compare results of EnWAS studies across different contexts and populations. It would for instance be interesting to see if our results on the importance of safety and socioeconomic status also translate to different countries or even geographical levels (e.g., by zooming in on streets, or zooming out by comparing different municipalities). Just as interesting as the significant associations are the non-significant associations: finding out that an environmental factor does not contribute to individual differences at a certain (geographical) level tells us that this factor might not be as important to look into for future research and policy.

Gene-environment interplay

Examining genetic and environmental influences individually is already a complex task, but examining their interplay is even more challenging. We examined two different types of gene-environment interplay: gene-environment interaction and gene-environment correlation. First, we examined how (relative) contributions of genes and the environment to phenotypic variance changed in the presence of big environmental change (i.e., COVID-19 lockdown). Doing so allowed us to study (quantitative) gene-environment interaction, in our case if the genetic effect is dependent on the environment. We find evidence for such an effect for quality of life, but not self-rated health. Besides gene-environment interaction, we also

studied gene-environment correlation: the phenomenon whereby genes influence exposure to the environment. We aimed to study gene-environment correlation in our EnWAS by examining if a polygenic score for the well-being spectrum predicted the well-being-associated environmental variables. Such an approach, where polygenic scores are correlated with environmental factors, had been used to study gene-environment correlation earlier for traits such as schizophrenia and major depression[410,411]. However, we did not find any evidence for gene-environment correlation in our own study, but this could also reflect power limitations. Similarly, we aimed to examine potential correlations between a well-being polygenic score and a well-being environmental score but were again limited by power. However, we did find an indication for gene-environment correlation in chapter 5 when we correlated poly-environmental scores with polygenic scores in a better-powered UKB sample. Taken together, while we started unravelling gene-environment interplay for well-being in this thesis, it is obvious that we are still far from having a comprehensive understanding on this topic.

Improving well-being

The previous sections describe how the work in this thesis has impacted research in the scientific field. While this is an important consequence of my work, it is also important to reflect upon the implications of this work in terms of individual-level and population well-being. I purposefully separate the two, as the findings presented here do not translate similarly to the individual-level as they do on a population level. Starting with the latter: all the work in this thesis was performed in population-based data with the aim of explaining individual differences in well-being. It logically follows that our conclusions are mostly relevant at the population level. This comes at a fortunate time, as many governments are gradually starting to direct policy based on the idea of a "well-being economy". A great example of this shift is the Well-Being Economy Governments Partnership (WEGo), currently comprised of Finland, Scotland, New Zealand, Iceland, and Wales. The movement was instigated by the well-being economy alliance, a global collaboration of 200 individuals, organizations, governments, academics, communities and businesses[412]. The goal of this movement and alliance is to shift economies away from a focus on marketed goods and services, and towards a broad well-being economy that creates well-being for both people and the planet. Initiatives like

this demonstrate that in the new global economy, well-being has become a central issue that requires informative, replicable research. The movement toward well-being-based economies is not the only indicator for the need of well-being research. The past three years, lives have globally been impacted by the COVID-19 pandemic. In our own work, we demonstrate that the accompanying measures especially impacted people's well-being (at least in the first lockdown). Now that we are starting to adapt to a "new normal" where governments acknowledge we might have to continue dealing with the virus in the future, we are at an opportune moment where we need to re-evaluate our priorities under both normal and crisis circumstances. As stated by the well-being economy alliance in a briefing paper, societies can choose two possible paths: they either go "back to worse", or try to "build back better"[413]. In short, "worse", in this case, would be to prioritize economic growth, while "better" would be to prioritize well-being. Central to the theme of this thesis, this means that we need a better understanding of the well-being exposome. Efforts like our EnWAS can help map this exposome, which could in turn be used to inform policy. Importantly, before we use results like these to inform policy, we also need to invest more effort into examining causality. Our own work on causality demonstrated that there is little reason to assume a causal association between educational duration and well-being, and different mental and physical health outcomes. This does not mean educational outcomes should be disregarded in the context of well-being, but rather that we should perhaps focus our efforts on different aspects of education, such as the quality of education.

Using these and other research outcomes to inform policy is an obvious implication on the population-level. Whether the goal is to improve well-being for targeted population groups or to increase average population well-being, this does not directly translate into individual-level well-being or clinical outcomes. From a more clinical perspective, there is a substantial amount of research focused on so-called positive psychological interventions. Meta-analyses have shown that these interventions can be effective both in terms of increasing well-being as well as in helping reduce depressive symptoms[414,415]. On the other hand, the effect sizes found are typically quite small, meaning there is still room for improvement for these interventions[416]. In my own work, I have not focused on these interventions or clinical populations. However, this does not mean that the work presented in this thesis cannot contribute to progress in this area. One of

the ways in which our work might contribute to this area is by highlighting the need to consider the sources of individual differences in well-being. For example, we demonstrate that the association between adolescent well-being and different socio-environmental factors is partly genetic in nature. A potential interpretation of this finding is that some individuals might be predisposed to evaluate both their well-being and social interactions more positively than others. It is important to interpret this finding in light of two sidenotes. First of all, genetic predispositions do not equal genetic determinism. What it does mean is that people are impacted differently by the similar environments, and that we need to take this into account when evaluating the environment. This also means that we can try to identify environmental factors that can make a difference for people's outcomes. Second, we examined these associations in light of individual differences in the general population. This is not the same as comparing well-being in adolescents with wildly differing environmental circumstances (e.g. extremely supportive family and social contacts vs. extreme neglect). Keeping in mind these two sidenotes, I believe that using designs such as the within-family design might enable us to better evaluate positive psychological interventions. For example, by comparing the effectiveness of a positive psychological intervention between siblings, we can be sure that an identified effect cannot be attributed to confounding factors such as parental socioeconomic status.

FUTURE RESEARCH

The picture I paint in this discussion has two sides. On the one hand, we find ourselves in a period of unprecedented scientific progress with respect to well-being research. On the other hand, there is still much we still have to discover and study. In what follows, I reflect upon the ways in which I believe we can continue this progress in terms of future research.

Diversity

First, the majority of research to date (including my own) has focused on WEIRD (White, Educated, Industrialized, Rich, and Democratic) samples. This is problematic because findings in this very specific group of individuals likely do not translate well to other groups. Notably, this is not a problem specific to well-being research, but a problem that is prominent in science in general and

in genetic research specifically. Fortunately, the first steps toward more inclusive and diverse research are already being taken. For example, the most recent World Happiness Report (WHR) does not only report on satisfaction with life (a well-being measure known to be a good indicator for Western cultures but less suitable for Eastern cultures), but also reports on balance and harmony (well-being indicators believed to be more relevant to Eastern cultures)[417]. Moreover, the first cross-population GWAS for well-being was recently performed by comparing the genetic architecture of subjective well-being across European and east-Asian contexts[418]. They found a significant cross-ancestry genetic correlation (r_g=0.78) between the samples. While efforts like these are a great first step, we are only beginning to explore well-being across different cultural contexts. However, we can still learn from the progress we have made in WEIRD contexts. For example, one of the efforts that has greatly sped up progress in the field of genetics is the introduction of large research consortia where research groups across the globe standardize their research methods and (to a certain extent) share data. It would be extremely useful to introduce this practice to the well-being field as this would improve both replicability and allow for more useful comparisons across contexts.

Combining different data types

Second, an exciting development in the field is that researchers are starting to use and combine different data types to explore individual differences in well-being. In this thesis, I combine genetic, environmental, and psychological data to study well-being. By combining these different types of data, we were able to explore novel research questions that touch upon different scientific domains. Besides these types of data, there are many other different directions worth exploring. In what follows, I explore an option I believe to be an exciting prospect for future research: ecological momentary assessments. When we examine well-being, we are often limited by survey data that reflect a person's well-being at a given moment in time. However, well-being is known to fluctuate over seasons, months, days, or even times of day[419]. Therefore, we might reach a better understanding of well-being if we evaluate it in the context of fluctuations over time. To accomplish this, we can make use of ecological momentary assessments (EMA). EMA is a broad term that encompasses different methods, but one of the ways in which EMA can be applied is in the context of smartphone-based assessment. By asking people

to report on their well-being through smartphone-based assessments multiple times a day, we can create a dataset that reflects people's well-being fluctuations on an hourly basis. At the same time, our smartphones can measure aspects of our real-world environment such as where we are and how much we move (i.e., GPS or accelerometer data), how much time we spend on our phones (screen time), and who in our social network we are with (e.g., based on Bluetooth data). In the context of the well-being exposome, one of the big limitations of my current work is that we link a single measure of well-being with different aspects of their environment. EMA initiatives open up a whole new set of possibilities since this allows us to track real-time well-being variation across different environmental contexts.

Interdisciplinary research

Relatedly, a last topic I would like to broach is the added value of interdisciplinary research. Whereas traditionally, scientists were inclined to work within their own scientific domain, we see more and more that the research questions we are trying to tackle require an interdisciplinary approach. While I study well-being from a behavior genetic perspective, it is a phenotype that is relevant across many other disciplines, such as economy, philosophy, biology, and sociology. In my own work, it is easy to see how interdisciplinary effort helps improve work. For example, it is difficult to understand why we distinguish certain types of well-being without any background knowledge on the philosophical origin of the construct. There are different ways in which interdisciplinary efforts can expand our knowledge of well-being. First, as I already mentioned, different disciplines tend to work with different types of data and different methods. All these data types and methods tell a different side of the story and linking them together might lead to new and surprising conclusions. Second, researchers in different fields have different types of background knowledge. Whereas a psychological researcher might have intricate knowledge on psychological processes that contribute to well-being on a more individual level, economists have detailed knowledge on policy changes or macroeconomic processes that influence population-level well-being. By collaborating on a project, we can benefit from the strengths of these individual researchers, meaning that one does not need to make themselves an expert on all thinkable aspects of a phenotype. In this way, it is easy to see how interdisciplinary research aids progress and prevents researchers from "reinventing the wheel".

CONCLUSION

Four years ago, when I started working on this thesis, the field of behavior genetics/biological psychology seemed to have just entered a paradigm shift from theory- to data-driven approaches. We had moved from the candidate gene approach to the genome-wide approach and there was an explosion of new analytic methods and novel genetic variants being identified for traits all across the psychological spectrum. Moreover, interest in well-being was rapidly increasing in both science and policy. Taking these two things together, I was in the fortunate position to continue progress by applying genetically informative designs to large well-being datasets. Fast forward four years, and I am able to present some interesting novel research findings in this thesis. First, by reviewing well-being candidate gene literature and the well-being network/factor structure, we reached a better understanding of the well-being phenotype and genetics. Second, by applying a data-driven approach to studying environmental correlates of well-being, we took a step towards mapping the complexities of the well-being exposome. Third, by treating the COVID-19 pandemic as a natural experiment, we found out more about how (the variance decomposition of) well-being changes in response to environmental change. Lastly, by shifting our focus from correlational evidence to more causality-oriented work where we combine different methods, we demonstrate how correlational evidence, or even individual causal inference methods, can lead to bias or confounding. While these novel insights certainly help us understand well-being better, it also re-affirms that well-being is a very complicated phenotype that requires interdisciplinary efforts powered by large sample sizes. In the future, it will be exciting to see how research across diverse cultural contexts, multiple data types and different scientific disciplines will allow us further see the pieces of the puzzle coming together.

NEDERLANDSE SAMENVATTING

In dit hoofdstuk vat ik de bevindingen van mijn proefschrift kort in het Nederlands samen. In dit vakgebied maken we graag gebruik van termen die niet heel vaak voorkomen in het dagelijks leven. Om de onderstaande samenvatting iets leesbaarder te maken voor iedereen die niet dagelijks bezig is met genetisch of psychologisch onderzoek heb ik aan het eind van dit hoofdstuk een toelichting van alle *schuingedrukte* termen en begrippen bijgevoegd.

Deel 1: Een her-evaluatie van welbevinden fenotypen en genetica.

In **hoofdstuk 2 en 3** begin ik met een evaluatie van de stand van zaken in het veld op zowel genetisch als *fenotypisch* niveau. In de introductie van dit proefschrift wijd ik veel aandacht aan hoe moderne psychologische *constructen* zijn beïnvloed door filosofische ideeën over welbevinden. Het is echter maar de vraag of het filosofische, oude onderscheid tussen *hedonisme* en *eudaimonisme* ook terug te vinden is in moderne vragenlijsten over constructen zoals subjectief- en psychologisch welbevinden. Traditioneel gezien werd deze vraag vaak onderzocht door middel van zogeheten factor analytische studies. Een factor analyse is een methode die je kunt gebruiken als je in een groep mensen meerdere dingen hebt gemeten, bijvoorbeeld hoe goed leerlingen scoren op verschillende vakken op de basisschool, en je je afvraagt of er iets overkoepelend is wat dit kan verklaren, bijvoorbeeld algemene cognitieve vaardigheden of intelligentie. Met betrekking tot welbevinden werd hier dus gekeken of overeenkomsten tussen verschillende welbevinden meetinstrumenten verklaard wordt door één of meerdere overkoepelende factoren. Hoewel dit onderzoek tot veel belangrijke inzichten heeft geleid is er echter nog steeds geen consensus over hoe we het welbevinden kader het beste kunnen definiëren. Een is een alternatief voor factor analyse is netwerkanalyse, waar er juist niet van uit wordt gegaan dat er een overkoepelende factor is die de samenhang tussen allerlei maten kan verklaren. Een eigenschap of aandoening wordt hierbij als netwerk gezien met allerlei verbindingen tussen de losse elementen. Depressie is een voorbeeld waar vaak netwerkanalyse is toegepast: hierbij wordt ervan uit gegaan dat verschillende symptomen (bijvoorbeeld somberheid en verminderde concentratie) elkaar constant beïnvloeden, in plaats van dat zij allemaal beïnvloed worden door één overkoepelde “depressie” oorzaak.

In **hoofdstuk 3** hebben we netwerkanalyse toegepast om meer inzicht in het welbevinden spectrum te verkrijgen. Allereerst onderzochten we in een steekproef van *N*=1343 Nederlandse Tweelingen Register (NTR) deelnemers associaties tussen *items* van vragenlijsten over tevredenheid met het leven, subjectief geluk, kwaliteit van leven, floreren, zelf beoordeelde gezondheid, depressieve symptomen, neuroticisme, en eenzaamheid items. We begonnen hierbij met kijken of sommige van deze items redundant waren (niets toevoegen aan het netwerk). Na het uitsluiten van deze overbodige items pasten we een netwerk toe in een onafhankelijke steekproef van *N*=759 NTR deelnemers. We vonden een netwerk bestaande uit een positief cluster welke tevredenheid met het leven, subjectief geluk en floreren items bevat en een negatief cluster met depressieve symptomen, eenzaamheid en neuroticisme items. De twee clusters waren nauw met elkaar verbonden via meerdere verbindingen. Voor vergelijkingsdoeleinden voeren we ook factoranalyses uit waarbij acht onafhankelijke maar matig tot sterk ge*correleerde* factoren (overeenkomend met de opgenomen constructen) beter bij de gegevens pasten dan een model met een of twee overkoepelende welbevinden factoren.

Voordat ik aan mijn proefschrift begon was er al veel bewijs voor *genetische varianten* die gerelateerd zijn aan welzijn. Hoewel de data-gedreven *genoom-brede* aanpak in het veld van gedragsgenetica al veel werd toegepast, was dit nog niet het geval in andere velden, zoals positieve psychologie. In dit veld was de hypothese-gedreven "*kandidaat-gen*" benadering gebruikelijk (en soms nog steeds), waarbij op basis van biologische hypothesen slechts één gen of een handje vol genen wordt onderzocht in relatie tot een fenotype. Hoewel dit theoretisch intuïtief klinkt, bleek uit genoom-breed bewijs al snel dat het extreem moeilijk is om een hypothese op te stellen over welke genetische varianten een fenotype mogelijk beïnvloeden. De meeste psychologische fenotypen worden beïnvloed door honderden tot duizenden genetische varianten, waarvan velen een onbekende of onduidelijke biologische functie hebben. Aangezien er veel kandidaat-gen studies waren uitgevoerd voor welbevinden en de resultaten van deze studies nog nooit in het licht van genoom-breed bewijs waren beoordeeld, hebben we eerst een systematische literatuurstudie uitgevoerd van kandidaat-gen studies voor welbevinden in **Hoofdstuk 3**. We hebben 41 kandidaat-gen studies geïdentificeerd die verschillende genetische varianten, *variable number tandem repeats* (VNTR's) en *gen-omgevings interacties* onderzochten in relatie tot welbevinden. Voor de

kandidaat-genetische varianten hebben we gegevens opgezocht in de uitkomsten (summary statistics) van de genoom-brede associatie (genome-wide association, GWA) meta-analyse van Baselmans en collega's. In zo'n genoom-brede associatie studie kijken we naar de stukjes DNA (over het gehele genoom) die tussen mensen verschillen en proberen we te achterhalen of deze DNA verschillen leiden tot verschillen in een uitkomst, bijvoorbeeld welbevinden. Daarnaast hebben we de VNTR's en verschillende gen-omgevingsinteracties opnieuw onderzocht in de UK Biobank (UKB), een dataset die veel groter is dan doorgaans gebruikt in kandidaat-gen studies. We vonden geen bewijs voor alle kandidaat-gen of interactie-effecten, wat suggereert dat eerdere kandidaat-gen bevindingen *fout-positieven* zijn. Op basis van deze resultaten adviseren we onderzoekers in het welbevinden om de kandidaat-gen benadering niet meer te gebruiken en zich te concentreren op genoom-brede benaderingen.

Deel II: Het welbevinden exposoom

In **hoofdstuk 4 en 5** evalueer ik de impact van de omgeving op welbevinden om een completer beeld van het welbevinden *exposoom* te krijgen. Voordat ik aan dit proefschrift begon was er al veel onderzoek gedaan naar verbanden tussen welbevinden en omgevingsfactoren in verschillende contexten en tijdsperiodes. Echter, er was nog steeds geen duidelijke consensus over welke specifieke omgevingsfactoren belangrijk zijn (en in welke context). Dit gebrek aan consensus heeft meerdere oorzaken, zoals variaties in onderzoeksopzet, hoe welbevinden wordt gemeten, hoe de omgeving wordt gemeten, en "cherry-picking", waar selectief naar bepaalde omgevingsfactoren wordt gekeken.

In een poging om omgevingsfactoren meer systematisch te bestuderen in relatie tot welbevinden hebben we in **hoofdstuk 4** een data-gedreven omgevings-brede associatie studie, ofwel een Environment-Wide Association Study (EnWAS), uitgevoerd. Een EnWAS is een methode waarbij we kijken of we een verband kunnen vinden tussen een uitkomst (bijvoorbeeld welbevinden) en een grote set omgevingsfactoren, zonder van tevoren hypothesen op te stellen over welke van deze omgevingsfactoren van belang zijn. In onze EnWAS onderzochten we tevredenheid met het leven scores van NTR-deelnemers (N=11,975) in relatie tot 139 omgevingsfactoren verkregen via het Geoscience and Health Cohort Consortium (GECCO). De omgevingsfactoren, welke werden gekoppeld

aan NTR-gegevens op basis van de 4-cijferige postcode, waren variabelen in verschillende domeinen, zoals de fysieke omgeving, sociaaleconomische factoren, toegankelijkheid, onderwijs, leefbaarheid, zorg en sport. We hebben 21 omgevingsfactoren geïdentificeerd die significant geassocieerd waren met welbevinden. Na het evalueren van de *multicollineariteit* tussen deze variabelen werden sociaaleconomische status en veiligheid aangewezen als de belangrijkste factoren voor het verklaren van individuele verschillen in welbevinden. Daarnaast hebben we een poly-genetische score (PGS) voor welbevinden gebruikt om te onderzoeken of we bewijs konden vinden voor *gen-omgevingscorrelatie*, wat zou wijzen op een rol van de genetische predispositie voor welbevinden in de blootstelling aan bepaalde omgevingen. Een PGS is een score welke genetische gevoeligheid voor een uitkomst weerspiegelt: een welbevinden PGS geeft dus iemands genetische predispositie voor welbevinden aan. Een belangrijke voetnoot hierbij is dat de voorspelling van zulke scores maar zo correct en verklarend is als de onderliggende (GWA) studie waarop deze gebaseerd is, en dat we tot op heden nog maar rond 1% van de verschillen tussen mensen kunnen verklaren op basis van deze scores. We hebben op basis van deze analyse geen bewijs gevonden voor dergelijke gen-omgevingscorrelatie. Dit kan betekenen dat er geen gen-omgevingscorrelatie is, maar het is ook waarschijnlijk dat we een gebrek aan statistische *power* hadden. Door in de toekomst EnWAS-studies over meerdere contexten en populaties uit te voeren, kunnen we langzaam beginnen met het samenstellen van het welbevinden exposoom, dat op zijn beurt kan worden gebruikt om het beleid te informeren.

Een van de dingen die we opmerkten bij het evalueren van de resultaten van onze EnWAS was dat alle verbanden die we identificeerden tussen omgevingsfactoren en welbevinden gebaseerd waren op zeer kleine *effectgroottes*. Bij genoom-brede onderzoeken is het gebruikelijk om kleine genetische effecten te vinden die op zichzelf niet veel bijdragen aan verschillen tussen mensen. Daarom is het gebruikelijk om individuele genetische effecten samen te voegen tot poly-genetische scores en deze scores te gebruiken voor (klinische) voorspelling en voor vervolganalyses. Geïnspireerd door deze aanpak voegen we individuele sociale en fysieke omgevingseffecten samen tot poly-omgeving scores (PES) in **hoofdstuk 5**. We berekenden twee PES's voor tevredenheid met het leven in een NTR-steekproef: één op basis van subjectieve sociale omgevingsfactoren (PES-S)

en één op basis van objectieve fysieke omgevingsfactoren (PES-O). We ontdekten dat de PES-S ongeveer 36% van de individuele verschillen in tevredenheid met het leven verklaarde, wat ongeveer de helft van de omgevingsvariantie is. De PES-O verklaarde daarentegen geen enkele variantie wanneer gecombineerd in één model met de PES-S. De PES's voorspelden pre-pandemisch en pandemisch welbevinden in gelijke mate en voorspelden verschillende welbevinden(-gerelateerde) fenotypen in verschillende mate. Als vervolg berekenden we een vergelijkbare PES-S in de UKB-dataset, welke ongeveer 12% van de variatie in geluk verklaarde. Het kleinere percentage *verklaarde variantie* in UKB in vergelijking met NTR kan worden teruggeleid op het ontbreken van belangrijke sociale omgevingsfactoren in de UKB-dataset, zoals het hebben van een partner en stress op het werk/thuis. Ten slotte onderzochten we potentiële *gen-omgevingscorrelatie* in zowel de NTR- als UKB-dataset door PGS's te associëren met de PES-Ss. De PGS voorspelde de PES-S in de UKB, maar niet in de kleinere NTR-dataset, wat wijst op de aanwezigheid van gen-omgevingscorrelatie. Over het algemeen toont de studie de bruikbaarheid aan van (sociale) omgevings-PESs voor het bestuderen van het welbevinden exposoom, met meerdere mogelijke toekomstige onderzoeks-toepassingen, zoals hoe deze omgevingsscores verschillen tussen verschillende culturen, contexten en leeftijden.

Sectie III: Welbevinden en de COVID-19-pandemie

Midden in mijn promotietraject begon de COVID-19 pandemie. Hoewel dit helaas voor vertraging in mijn welbevinden dataverzameling zorgde, leidde het ook tot een nieuwe "NTR COVID-19-dataverzameling". Als onderdeel van deze dataverzameling werd de deelnemers gevraagd om te rapporteren over hun welbevinden en gezondheid tijdens de eerste lockdown van de pandemie. In de **hoofdstukken 6 en 7** behandelen we de eerste lockdown van de COVID-19-pandemie als een zogeheten natuurlijk experiment, een soort interventie die ons allen beïnvloedde doordat de omgeving radicaal en abrupt veranderde. Allereerst gebruiken we in **hoofdstuk 6** een bivariaat tweeling ontwerp om zelf-beoordeelde gezondheid, ofwel Self-Rated Health (SRH), voor (*N*=16,127) en tijdens de pandemie (*N*=17,451) te vergelijken bij personen zonder een vermoedelijke COVID-19-infectie.

Verrassend genoeg ontdekten we dat de meerderheid van de steekproef (66,7%) een verhoogde SRH meldt tijdens de eerste lockdown. We theoretiseren dat deze

toenames niet noodzakelijkerwijs een objectieve toename van de gezondheid weerspiegelen, maar dat niet-geïnfecteerde personen hun gezondheid positiever beoordelen dan onder normale omstandigheden, een fenomeen waarnaar Recchi en collega's verwijzen als "*het oog van de orkaan paradox*". Een andere vraag die we graag wilden beantwoorden was of we veranderingen in de genetische architectuur van SRH konden detecteren als we vóór de pandemie vergeleken met tijdens de pandemie. In een tweelingmodel maken we gebruik van het feit dat eeneiige tweelingen nagenoeg 100% genetisch identiek zijn, terwijl tweeeiige tweelingen ongeveer 50% genetisch identiek zijn. Door deze twee soorten tweelingparen met elkaar te vergelijken kunnen we uitspraken doen over in hoeverre genen, de gedeelde (familie) omgeving en de unieke (niet gedeeld door broers/zussen uit hetzelfde gezin) omgeving individuele verschillen in een uitkomst als welbevinden verklaren. In dit geval gebruiken we een bivariaat tweelingmodel, wat in deze context betekent dat we niet alleen individuele verschillen in op 1 tijdspunt kijken, maar op 2 tijdspunten (voor en tijdens de pandemie). Zo kunnen we ook uitspraken doen over in hoeverre dezelfde genen en omgevingsinvloeden op beide tijdspunten invloed hebben. Het model onthulde geen significante veranderingen in de variantie decompositie, met een erfelijkheidsschatting van 45% op beide tijdstippen. Met andere woorden, de relatieve mate waarin genetische en omgevingsfactoren individuele verschillen in welbevinden verklaren veranderde niet tijdens de eerste lockdown.

Een ander fenotype dat we in de COVID-19 vragenlijst hebben gemeten, was kwaliteit van leven, ofwel Quality of Life (QoL). Kwaliteit van leven werd gemeten met de zogeheten Cantril ladder, waarbij naar subjectief welbevinden wordt gevraagd op een schaal van 1 (laagste kwaliteit van leven) tot 10 (hoogste kwaliteit van leven). In **hoofdstuk 7** hebben we nader bekeken wat er met de QoL scores gebeurde tijdens de eerste lockdown. Bij het vergelijken van pre-pandemie QoL (*N*=25,772) met pandemie QoL (*N*=17,222) in de steekproef van meerlingen en hun gezinsleden, ontdekten we dat het gemiddelde QoL daalde van 7.73 naar 7.02. Als we dit vergelijken met onze resultaten voor SRH is het interessant om te zien dat tijdens de eerste lockdown vooral de mentale (en niet de fysieke) gezondheid leek af te nemen. Net als bij ons SRH-project hebben we ook mogelijke veranderingen in de genetische architectuur van QoL onderzocht. Voor dit project hebben we gebruik gemaakt van volledige stamboomgegevens

van de families door middel van het programma “Mendel”, welke rekening houdt met correlaties tussen alle gezinsleden (in plaats van alleen de tweelingen). We zagen een toename in de totale variantie in QoL, voornamelijk als gevolg van een toename van de unieke omgevingsvariantie tijdens de pandemie. Relatief namen de unieke omgevingsinvloeden toe van 26.3% naar 58.7%, terwijl de erfelijkheid afnam van 30.9% naar 15.5%. We veronderstellen dat deze toename in unieke omgevingsvariantie het gevolg is van het feit dat de lockdownmaatregelen niet iedereen in gelijke mate troffen. Terwijl sommige mensen bijvoorbeeld thuis moesten werken werden anderen op hun gebruikelijke werkplek verwacht. Deze studie informeert ons niet alleen over de impact van de COVID-19-pandemie op het welbevinden, maar is ook een interessant voorbeeld van de dynamische en contextuele aard van genetische invloeden op complexe eigenschappen.

Sectie IV: Richting Causaliteit

In de voorgaande hoofdstukken hebben we verschillende soorten verbanden onderzocht: verbanden tussen gezinsleden, verbanden tussen welbevinden items en verbanden tussen welbevinden en verschillende omgevings- en genetische factoren. Aangezien dit echter allemaal correlationeel bewijs betrof, hebben we hierbij geen uitspraken gedaan over causaliteit. In de **hoofdstuk 8 en 9** werken we naar meer causale benaderingen toe. Allereerst waren we in **hoofdstuk 8** geïnteresseerd in de rol van sociale omgevingsfactoren in het welbevinden van adolescenten. In *N*=4,700 tweelingparen van de NTR database onderzochten we verbanden tussen welbevinden en gezinsconflict en -functioneren, het aantal vrienden, vriendschapsbelang en -tevredenheid, en vrijetijds variabelen. Hierbij hebben we eerst gekeken naar tweeling verschilscores: we kijken hierbij binnen zowel eeneiige als twee-eiige tweelingparen of een verschil in welbevinden samenhangt met een verschil in de andere variabelen. Als deze verschilscores alleen samenhangen in twee-eiige tweelingen en niet (of in mindere mate) in eeneiige tweelingen, dan suggereert dit een invloed van genetische factoren. Wanneer tweeling-verschilscores een rol voor genetische factoren suggereerde gebruikten we bivariate tweeling modellen om genetische en omgevings bijdragen aan deze associaties te kwantificeren. Met behulp van deze methoden vinden we bewijs voor een grote (73-91%) genetische invloed op de associaties tussen welzijn en gezinsconflict en -functioneren, vrijetijds sport/scoutingclubs en tevredenheid

met vriendschappen. Hoewel deze methoden geen indicatie zijn voor een causaal verband tussen welbevinden en deze variabelen, brengt het ons wel een stap dichter bij het begrijpen waar de waargenomen associaties vandaan komen.

Ten slotte passen we in **hoofdstuk 9** vier verschillende causale inferentie methoden toe om mogelijke causale verbanden tussen opleidingsduur en verschillende mentale en fysieke gezondheidsuitkomsten in UK Biobank data te onderzoeken. We maken gebruik van 1) een natuurlijk beleidsexperiment dat gebruikmaakt van een verhoging in de minimumleeftijd voor het verlaten van school ($N \approx 30{,}000$), 2) een studie waar we biologische broers/zussen opgegroeid in hetzelfde gezin met elkaar vergelijken ($N \approx 18{,}000$), 3) Mendeliaanse Randomisatie (MR) ($N \approx 200{,}000$), en 4) MR binnen families ($N \approx 18{,}000$). Een uitgebreide uitleg van deze vier methoden valt buiten de scope van deze samenvatting, maar van belang is dat dit vier verschillende methoden zijn die naar causaliteit kijken waarbij elke methode zwakke en sterke punten heeft. Door de resultaten van de vier methoden met elkaar te vergelijken wilden we onderzoeken of waargenomen associaties worden beïnvloed of zogeheten *confounders* of door omgekeerde causaliteit. Hoewel we consistente niet-significante verbanden vinden tussen opleidingsduur en geluk (happiness), gezinstevredenheid, werktevredenheid, zinvol leven, depressie, angst en bipolaire stoornis, vinden we geen consistente significante associaties bij alle verschillende methoden voor de andere uitkomsten (tevredenheid over gezondheid, financiële tevredenheid, vriendschapstevredenheid, neuroticisme en cardiovasculaire uitkomsten). Op deze manier konden we door middel van triangulatie over verschillende causale inferentie methoden concluderen dat de meeste van de waargenomen verbanden tussen onderwijsduur en mentale en fysieke gezondheid *confounded* zijn.

Conclusie

Vier jaar geleden, toen ik aan dit proefschrift begon, leek het gebied van gedragsgenetica/ biologische psychologie net een paradigmaverschuiving te hebben ondergaan van theorie- naar data gedreven onderzoek. We waren overgestapt van de kandidaat-gen benadering naar de genoom-brede benadering en er was een explosieve toename van nieuwe analytische methoden en nieuwe genetische varianten die werden geïdentificeerd voor eigenschappen over het hele psychologische spectrum. Bovendien nam de belangstelling voor

welbevinden snel toe, zowel in de wetenschap als in beleid. Met deze twee ontwikkelingen in het achterhoofd, bevond ik me in de gunstige positie waar ik genetisch informatieve en innovatieve onderzoeks-designs kon toe passen op grote datasets over welbevinden. Vier jaar later kan ik in dit proefschrift een aantal interessante nieuwe onderzoeksresultaten presenteren. Ten eerste hebben we, door de literatuur over kandidaat-genen op het gebied van welbevinden en de netwerk en factor structuur van welbevinden te bekijken, een beter begrip gekregen van het welbevinden fenotype en de genetica die hier een rol in speelt. Ten tweede hebben we, door een data-gedreven benadering toe te passen bij het bestuderen van omgevingsfactoren op welbevinden, een stap gezet in de richting van het in kaart brengen van de complexiteit van het welbevinden exposoom. Ten derde kwamen we, door de COVID-19-pandemie als een natuurlijk experiment te behandelen, meer te weten over hoe (de variantie decompositie van) welbevinden verandert als reactie op veranderingen in de omgeving. Ten slotte hebben we, door onze focus te verleggen van correlationele methoden naar meer causaliteitsgerichte methoden waarbij we verschillende methoden combineren, laten we zien hoe correlationeel bewijs, of zelfs individuele causale inferentiemethoden, kunnen leiden tot vertekende resultaten. Hoewel deze nieuwe inzichten ons zeker helpen welbevinden beter te begrijpen, bevestigt het ook opnieuw dat welbevinden een zeer gecompliceerd fenotype is dat interdisciplinaire inspanningen grote steekproeven vereist. In de toekomst zal het interessant zijn om te zien hoe onderzoek in meer diverse culturele contexten, meerdere datatypes en verschillende wetenschappelijke disciplines ons in staat zal stellen om de stukjes van de puzzel samen te zien komen.

Term	Uitleg
Confounder	Een variabele die zowel gerelateerd is aan de uitkomst als een voorspeller die het verband tussen twee variabelen verstoort of veroorzaakt. Bijvoorbeeld: stel er is een verband tussen hoeveel ijsjes er verkocht worden op een dag en hoeveel mensen er verdrinken. Dit verband is echter niet oorzakelijk, maar wordt verklaard door een derde, confounding, variabele: stijgende temperatuur.
Construct	Een psychologisch construct is een theoretisch concept of idee welke niet rechtstreeks observeerbaar is, zoals in dit geval welbevinden.
Correlaties, gecorreleerd	Een correlatie is een maat van samenhang tussen twee variabelen. Een correlatie tussen twee variabelen (bijvoorbeeld depressie en welbevinden) betekent dat deze variabelen dus met elkaar in verband staan (hoe lager iemand scoort op een depressie vragenlijst, hoe hoger deze persoon waarschijnlijk scoort op welbevinden vragenlijst).
Hedonisme	Een filosofisch gedachtegoed over welbevinden waarbij streven naar geluk het grootste levensdoel is. Geluk wordt hierbij gedefinieerd als een maximalisatie van genot en minimalisatie van pijn.
Effectgrootte	Wanneer we een verband tussen twee variabelen vinden zijn we geïnteresseerd in de grootte/sterkte van dit verband. Dit noemen we de effectgrootte.
Eudaimonisme	Een filosofisch gedachtegoed over welbevinden ontwikkelt als alternatief voor het hedonisme. Geluk wordt hierbij gedefinieerd als breder dan alleen de maximalisatie van genot, waarbij moraliteit en zelfontplooiing centraal staan.
Exposoom	Het geheel van omgevingsinvloeden die samenhangen met een uitkomst. Het welbevinden exposoom heeft dus betrekking op alle omgevingsinvloeden die in relatie staan tot welbevinden.
Fenotypisch, fenotype	De term fenotype wordt gebruikt om kenmerken of eigenschappen van mensen te beschrijven. Dit fenotype is het resultaat van iemands genen en alle ondervonden omgevingsinvloeden. In dit geval kijk ik in dit proefschrift naar het fenotype “welbevinden”.
Fout-positief	Wanneer we in een analyse een effect vinden terwijl deze er in de werkelijkheid niet is. In dit geval zou een foutpositieve associatie tussen een gen en welbevinden betekenen dat dit gen eigenlijk geen verband heeft met welbevinden.
Genetische varianten	Als we het DNA van verschillende personen vergelijken dan zijn er veel plekken waarop dat DNA niet hetzelfde is. Deze genetische verschillen tussen mensen verklaren een deel van de uiterlijke en persoonlijke verschillen tussen ons, onder andere in welbevinden. Als we inzoomen op een enkel stukje DNA dat verschilt tussen mensen kan je daar dus een andere genetische variant hebben dan iemand anders.
Genoom breed	Methoden die naar het gehele DNA (het hele genoom) kijken (om verschillen tussen mensen te verklaren) noemen we genoom-breed.
Gen-omgevingscorrelatie	Het fenomeen waarbij blootstelling aan bepaalde omgevingen beïnvloed wordt door iemands genetische profiel. In het geval van welbevinden zou een voorbeeld van gen-omgevingscorrelatie zijn dat mensen met een hoge genetische predispositie voor welbevinden eerder positieve, welbevinden-stimulerende omgevingen uitkiezen dan mensen met een lage genetische predispositie voor welbevinden.

Term	Uitleg
Gen-omgevingsinteractie	Het fenomeen waarbij dezelfde omgeving een ander effect heeft op mensen met een verschillend genetisch profiel, of andersom: het verschijnsel waarbij dezelfde genen een ander effect hebben in verschillende omgevingen.
Items, item	In de samenvatting zal ik het soms hebben over "items" van vragenlijsten. Hierbij bedoel ik afzonderlijke vragen binnen die vragenlijsten.
Multicollineariteit	We spreken van multicollineariteit tussen 2 variabelen als deze erg sterk met elkaar samenhangen. Als deze variabelen samen gebruikt worden om een uitkomst te voorspellen in 1 model kan dit voor problemen zorgen doordat het de betrouwbaarheid van het model vermindert.
Power	De kans dat we een effect vinden als deze ook in de werkelijkheid bestaat.
Variable number tandem repeats	Stukken DNA (bestaande uit meerdere genetische varianten) die zich herhalen in ons DNA. Hoe vaak deze stukken DNA herhaald wordt verschilt tussen personen en kan invloed hebben op verschillen tussen mensen.
Verklaarde variantie	De verschillen tussen mensen (in bijvoorbeeld welbevinden) die verklaar kunnen worden door een verschil in een bepaalde andere variabele.

REFERENCES

1. Zencey, E. Measuring what matters. in *Consumption and Well-Being in the Material World* vol. 9789400773684 153–174 (Springer Netherlands, 2014).
2. Stiglitz, J. E., Sen, A. K. & Fitoussi, J.-P. *Report by the commission on the measurement of economic performance and social progress*. (2009).
3. Boelhouwer, J. Wellbeing in the Netherlands : the SCP life situation index since 1974. (Utrecht University, 2010).
4. Lambert, L., Passmore, H.-A. & Holder, M. D. Foundational frameworks of positive psychology: Mapping well-being orientations. *Canadian Psychology/Psychologie canadienne* **56**, 311–321 (2015).
5. Biswas-Diener, R., Kashdan, T. B. & King, L. A. Two traditions of happiness research, not two distinct types of happiness. *Journal of Positive Psychology* **4**, 208–211 (2009).
6. Berryman, S. Democritus. in *The Stanford Encyclopedia of Philosophy* (ed. Zalta, E. N.) (Metaphysics Research Lab, Stanford University, 2016).
7. Tännsjö, T. Narrow Hedonism. *Journal of Happiness Studies* **8**, 79–98 (2007).
8. Crimmins, J. E. Jeremy Bentham. in *The Stanford Encyclopedia of Philosophy* (ed. Zalta, E. N.) (Metaphysics Research Lab, Stanford University, 2017).
9. West, H. R. *An introduction to Mill's utilitarian ethics*. (Cambridge University Press, 2004).
10. Linton, M.-J., Dieppe, P. & Medina-Lara, A. Review of 99 self-report measures for assessing well-being in adults: exploring dimensions of well-being and developments over time. *BMJ open* **6**, e010641 (2016).
11. Diener, E. Subjective well-being. *Psychological bulletin* **95**, 542–75 (1984).
12. Diener, E., Emmons, R. A., Larsen, R. J. & Griffin, S. The Satisfaction With Life Scale. *Journal of Personality Assessment* **49**, 71–75 (1985).
13. Feldman, F. *What Is This Thing Called Happiness?* (OUP Oxford, 2010).
14. Veenhoven, R. *Conditions of happiness*. (D. Reidel, 1984).
15. *Handbook of Eudaimonic Well-Being*. (Springer International Publishing, 2016). doi:10.1007/978-3-319-42445-3.
16. Waterman, A. S. & S., A. The relevance of Aristotle's conception of eudaimonia for the psychological study of happiness. *Theoretical & Philosophical Psychology* **10**, 39–44 (1990).
17. McMahon, D. M. Living well in light of science. *Annals of the New York Academy of Sciences* **1384**, 36–38 (2016).
18. Huta, V. & Ryan, R. M. Pursuing Pleasure or Virtue: The Differential and Overlapping Well-Being Benefits of Hedonic and Eudaimonic Motives. *Journal of Happiness Studies* **11**, 735–762 (2010).
19. Maslow, A. H. & H., A. A theory of human motivation. *Psychological Review* **50**, 370–396 (1943).
20. Ryff, C. D. & Singer, B. H. Know Thyself and Become What You Are: A Eudaimonic Approach to Psychological Well-Being. *Journal of Happiness Studies* **9**, 13–39 (2008).
21. Ryan, R. M. & Deci, E. L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* **55**, 68–78 (2000).
22. Ryff, C. D. Psychological well-being revisited: advances in the science and practice of eudaimonia. *Psychotherapy and psychosomatics* **83**, 10–28 (2014).

23. Fowers, B. J. & Lefevor, G. T. The inescapability of intersubjectivity in meaning. *American Psychologist* **70**, 573–574 (2015).
24. Bartels, M. Genetics of wellbeing and its components satisfaction with life, happiness, and quality of life: a review and meta-analysis of heritability studies. *Behavior genetics* **45**, 137–156 (2015).
25. Nes, R. B. & Røysamb, E. The heritability of subjective well-being: review and meta-analysis. in *Genetics of Psychological Well-Being* 75–96 (Oxford University Press, 2015). doi:10.1093/acprof:oso/9780199686674.003.0005.
26. Baselmans, B. M. L. *et al.* Unraveling the Genetic and Environmental Relationship Between Well-Being and Depressive Symptoms Throughout the Lifespan. *Frontiers in Psychiatry* **9**, (2018).
27. Røysamb, E., Nes, R. B. & Vittersø, J. Chapter 2 - Well-Being: Heritable and Changeable. in *Stability of Happiness* (eds. Sheldon, K. M. & Lucas, R. E.) 9–36 (Academic Press, 2014). doi:10.1016/B978-0-12-411478-4.00002-3.
28. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**, 76–82 (2011).
29. Rietveld, C. A. *et al.* Molecular genetics and subjective well-being. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9692–7 (2013).
30. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses. *Nature Genetics* **48**, 624–633 (2016).
31. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine* **12**, 44 (2020).
32. Baselmans, B. M. L. *et al.* A Genetic Investigation of the Well-Being Spectrum. *Behav Genet* **49**, 286–297 (2019).
33. Sanderson, E. *et al.* Mendelian randomization. *Nat Rev Methods Primers* **2**, 1–21 (2022).
34. Pinquart, M. & Sörensen, S. Influences of socioeconomic status, social network, and competence on subjective well-being in later life: a meta-analysis. *Psychology and aging* **15**, 187–224 (2000).
35. Wang, H. H., Wu, S. Z. & Liu, Y. Y. Association between social support and health outcomes: A meta-analysis. *Kaohsiung Journal of Medical Sciences* **19**, 345–350 (2003).
36. Lovell, R., Wheeler, B. W., Higgins, S. L., Irvine, K. N. & Depledge, M. H. A Systematic Review of the Health and Well-Being Benefits of Biodiverse Environments. *JOURNAL OF TOXICOLOGY AND ENVIRONMENTAL HEALTH-PART B-CRITICAL REVIEWS* **17**, 1–20 (2014).
37. Vanaken, G.-J. & Danckaerts, M. Impact of Green Space Exposure on Children's and Adolescents' Mental Health: A Systematic Review. *International journal of environmental research and public health* **15**, (2018).
38. Houlden, V., Weich, S., Porto de Albuquerque, J., Jarvis, S. & Rees, K. The relationship between greenspace and the mental wellbeing of adults: A systematic review. *PloS one* **13**, e0203000 (2018).
39. Ni, M. Y. *et al.* Determinants of Physical, Mental and Social Well-Being: A Longitudinal Environment-Wide Association Study. *International Journal of Epidemiology* **49**, 380–389 (2020).

40. Vinkhuyzen, A. A. E., van der Sluis, S., de Geus, E. J. C., Boomsma, D. I. & Posthuma, D. Genetic influences on `environmental' factors. *Genes Brain and Behavior* **9**, 276–287 (2010).
41. Deci, E. L. & Ryan, R. M. Hedonia, eudaimonia, and well-being: An introduction. *Journal of Happiness Studies* **9**, 1–11 (2008).
42. Diener, E. & Ryan, K. Subjective well-being: A general overview. *South African Journal of Psychology* **39**, 391–406 (2009).
43. Seligman, M. *Flourish : a visionary new understanding of happiness and well-being*. (Free Press, 2011).
44. Goodman, F. R., Disabato, D. J., Kashdan, T. B. & Kauffman, S. B. Measuring well-being: A comparison of subjective well-being and PERMA. *https://doi-org.vu-nl.idm.oclc.org/10.1080/17439760.2017.1388434* **13**, 321–332 (2017).
45. Seligman, M. PERMA and the building blocks of well-being. *https://doi-org.vu-nl.idm.oclc.org/10.1080/17439760.2018.1437466* **13**, 333–335 (2018).
46. Ryff, C. D. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology* **57**, 1069–1081 (1989).
47. Keyes, C. Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of consulting and clinical psychology* **73**, 539–548 (2005).
48. Keyes, C. Social well-being. *Social Psychology Quarterly* **61**, 121–137 (1998).
49. Kim, K., Lehning, A. J. & Sacco, P. Assessing the factor structure of well-being in older adults: findings from the National Health and Aging Trends Study. *Aging & mental health* **20**, 814–822 (2016).
50. Vanhoutte, B. & Nazroo, J. Cognitive, affective and eudemonic well-being in later life: Measurement equivalence over gender and life stage. *Sociological Research Online* **19**, (2014).
51. McMahan, E. A. & Estes, D. Hedonic Versus Eudaimonic Conceptions of Well-being: Evidence of Differential Associations With Self-reported Well-being. *Social Indicators Research* **103**, 93–108 (2011).
52. Joshanloo, M. Revisiting the Empirical Distinction Between Hedonic and Eudaimonic Aspects of Well-Being Using Exploratory Structural Equation Modeling. *Journal of Happiness Studies* **17**, 2023–2036 (2016).
53. Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R. & Borsboom, D. The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine* **50**, 353–366 (2020).
54. Borgatti, S. P. & Halgin, D. S. On Network Theory. *Organization Science* **22**, 1168–1181 (2011).
55. Epskamp, S., Borsboom, D. & Fried, E. I. Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods* **50**, 195–212 (2018).
56. Fried, E. I. *et al.* Mental disorders as networks of problems: a review of recent insights. *Social Psychiatry and Psychiatric Epidemiology* **52**, 1–10 (2017).
57. Zeng, G., Peng, K. & Hu, C.-P. The Network Structure of Adolescent Well-Being Traits: Results From a Large-Scale Chinese Sample. *Frontiers in Psychology* **10**, 2783 (2019).
58. Stochl, J. *et al.* Identifying key targets for interventions to improve psychological wellbeing: replicable results from four UK cohorts. *Psychological Medicine* **49**, 2389–2396 (2019).

59. Giuntoli, L. & Vidotto, G. Exploring Diener's Multidimensional Conceptualization of Well-Being Through Network Psychometrics. *Psychological reports* **124**, 896–919 (2021).
60. Van Woerkom, M. *et al.* Networks of happiness: applying a Network Approach to Well-Being in the General Population. *J Happiness Stud* (2022) doi:10.1007/s10902-022-00546-x.
61. Bringmann, L. F. *et al.* What Do Centrality Measures Measure in Psychological Networks? *J Abnorm Psychol* **128**, 892–903 (2019).
62. Neal, Z. P. & Neal, J. W. Out of bounds? The boundary specification problem for centrality in psychological networks. *Psychological Methods* (2021) doi:10.1037/met0000426.
63. Ligthart, L. *et al.* The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-Family Designs. *Twin Research and Human Genetics* 1–14 (2019) doi:10.1017/thg.2019.93.
64. Lyubomirsky, S. & Lepper, H. S. A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research* **46**, 137–155 (1999).
65. Cantril, H. *The pattern of human concerns*. (Rutgers University Press, 1965).
66. Diener, E. *et al.* New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research* (2010) doi:10.1007/s11205-009-9493-y.
67. Rescorla, L. A. & Achenbach, T. M. The Achenbach System of Empirically Based Assessment (ASEBA) for Ages 18 to 90 Years. . in *The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults* (ed. Maruish, M. E.) 115–152 (Lawrence Erlbaum Associates Publishers, 2004).
68. Hughes, M. E., Waite, L. J., Hawkley, L. C. & Cacioppo, J. T. A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies. *Research on aging* **26**, 655–672 (2004).
69. Costa, P. T. & McCrae, R. R. *Revised NEO Personality Inventory (NEO-PIR) and NEO Five Factor Inventory (NEO-FFI) professional manual*. (Psycological Assessment Resources, 1992).
70. Eriksson, I., Undén, A. L. & Elofsson, S. Self-rated health. Comparisons between three different measures. Results from a population study. *International Journal of Epidemiology* (2001) doi:10.1093/ije/30.2.326.
71. Epskamp, S., Fried, E. & Epskamp, M. Package 'bootnet'. Preprint at (2015).
72. Jones, P. & Jones, M. Package 'networktools'. Preprint at (2017).
73. Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D. & Borsboom, D. Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software* **48**, 1–18 (2012).
74. RStudio Team. RStudio: Integrated Development for R. Preprint at (2020).
75. Haslbeck, J. M. B. & Waldorp, L. J. Structure estimation for mixed graphical models in high-dimensional data. (2015).
76. Robinaugh, D. J., Millner, A. J. & McNally, R. J. Identifying Highly Influential Nodes in the Complicated Grief Network. *J Abnorm Psychol* **125**, 747–757 (2016).
77. Costantini, G. *et al.* Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences* **136**, 68–78 (2019).

78. Zhang, B. & Horvath, S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology* **4**, (2005).

79. Revelle, W. & Revellle, M. W. Package 'psych'. (2022).

80. Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* **48**, 1–36 (2012).

81. Jones, P. J., Mair, P. & McNally, R. J. Visualizing Psychological Networks: A Tutorial in R. *Frontiers in Psychology* **9**, 1742 (2018).

82. Bartels, M. & Boomsma, D. I. Born to be Happy? The Etiology of Subjective Well-Being. *Behav Genet* **39**, 605 (2009).

83. Kokko, K., Korkalainen, A., Lyyra, A. L. & Feldt, T. Structure and Continuity of Well-Being in Mid-Adulthood: A Longitudinal Study. *Journal of Happiness Studies* **14**, 99–114 (2013).

84. Weinberger, A. H., Darkes, J., Boca, F. K. Del, Greenbaum, P. E. & Goldman, M. S. Items as Context: Effects of Item Order and Ambiguity on Factor Structure. *http://dx.doi.org/10.1207/s15324834basp2801_2* **28**, 17–26 (2010).

85. Hamer, D. H. The heritability of happiness. *Nature Genetics* vol. 14 125–126 Preprint at https://doi.org/10.1038/ng1096-125 (1996).

86. Lykken, D. & Tellegen, A. Happiness Is a Stochastic Phenomenon: *https://doi.org/10.1111/j.1467-9280.1996.tb00355.x* **7**, 186–189 (2016).

87. Lin, S.-H., Lee, L.-T. & Yang, Y. K. Serotonin and Mental Disorders: A Concise Review on Molecular Neuroimaging Evidence. *Clinical Psychopharmacology and Neuroscience* **12**, 196 (2014).

88. Cowen, P. J. & Browning, M. What has serotonin to do with depression? *World Psychiatry* **14**, 158 (2015).

89. De Neve, J.-E. Functional polymorphism (5-HTTLPR) in the serotonin transporter gene is associated with subjective well-being: evidence from a US nationally representative sample. *Journal of human genetics* **56**, 456–459 (2011).

90. De Neve, J. E., Christakis, N. A., Fowler, J. H. & Frey, B. S. Genes, economics, and happiness. *Journal of Neuroscience, Psychology, and Economics* **5**, 193–211 (2012).

91. Gartner, M. *et al.* The interplay of genetic and environmental factors in shaping well-being across the lifespan: Evidence from the serotonin transporter gene. *Aging & mental health* **22**, 1216–1222 (2018).

92. Matsunaga, M., Isowa, T., Yamakawa, K. & Ohira, H. Association between the serotonin transporter polymorphism (5HTTLPR) and subjective happiness level in Japanese adults. *Psychology of Well-Being: Theory, Research and Practice* **3**, 5 (2013).

93. Gohier, B. *et al.* Genetic modulation of the response bias towards facial displays of anger and happiness. *European Psychiatry* **29**, 197–202 (2014).

94. Hankin, B. L. *et al.* Differential susceptibility in youth: evidence that 5-HTTLPR x positive parenting is associated with positive affect 'for better and worse'. *Translational psychiatry* **1**, e44 (2011).

95. Hartmann, J. A. *et al.* The serotonin transporter 5-HTTLPR polymorphism in the association between sleep quality and affect. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* **24**, 1086–1090 (2014).

96. Roekel, E., MariekeWichers, Hartmann, J. A. & Verhagen, M. The 5-HTTLPR genotype moderates the association between sleep quality and positive affect: A replication study. *European Neuropsychopharmacology* **26**, 1350–1351 (2016).

97. Border, R. *et al.* No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry* **176**, 376–387 (2019).

98. Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nature Genetics 2018 50:9* **50**, 1318–1326 (2018).

99. Hewitt, J. K. Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behavior Genetics* vol. 42 1–2 Preprint at https://doi.org/10.1007/s10519-011-9504-z (2012).

100. Duncan, L. E., Ostacher, M. & Ballon, J. How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. *Neuropsychopharmacology* **44**, 1518–1523 (2019).

101. Dick, D. M. *et al.* Candidate Gene–Environment Interaction Research: Reflections and Recommendations. *Perspectives on Psychological Science* **10**, 37–59 (2015).

102. Belsky, J. The Differential Susceptibility Hypothesis: Sensitivity to the Environment for Better and for Worse. *JAMA Pediatrics* **170**, 321–322 (2016).

103. Sheffer-Matan, L., Gilad-Lysy, R., Guy, A., Guy, U. & Kohn, Y. The Role of 5-HTTLPR and MAOA in Moderating the Association Between Environment and Well-being. *Israel Journal of Psychiatry and Related Sciences* **56**, 11–18 (2019).

104. Ryan, R. M. & Deci, E. L. On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. *Annual review of psychology* **52**, 141–166 (2001).

105. Magyar, J. L. & Keyes, C. L. M. Defining, measuring, and applying subjective well-being. *Positive psychological assessment: A handbook of models and measures (2nd ed.).* 389–415 (2019) doi:10.1037/0000138-025.

106. Visscher, P. M. & Montgomery, G. W. Genome-wide association studies and human disease: From trickle to flood. *JAMA - Journal of the American Medical Association* vol. 302 2028–2029 Preprint at https://doi.org/10.1001/jama.2009.1643 (2009).

107. Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J. & Visscher, P. M. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell* **173**, 1573–1580 (2018).

108. Baselmans, B. M. L. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nature genetics* **51**, 445–451 (2019).

109. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics 2018 50:2* **50**, 229–237 (2018).

110. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nature Genetics 2018 50:11* **50**, 1593–1599 (2018).

111. Farrell, M. S. *et al.* Evaluating historical candidate genes for schizophrenia. *Molecular Psychiatry* **20**, 555–562 (2015).

112. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

113. Border, R. *et al.* Imputation of behavioral candidate gene repeat variants in 486,551 publicly-available UK Biobank individuals. *European Journal of Human Genetics* **27**, 963–969 (2019).

114. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).

115. Bradley, B., Davis, T. A., Wingo, A. P., Mercer, K. B. & Ressler, K. J. Family environment and adult resilience: contributions of positive parenting and the oxytocin receptor gene. *European journal of psychotraumatology* **4**, (2013).

116. Kuepper, Y. *et al.* 5-HTTLPR S-allele: a genetic plasticity factor regarding the effects of life events on personality? *Genes, brain, and behavior* **11**, 643–650 (2012).

117. Sicorello, M. *et al.* Highs and lows: Genetic susceptibility to daily events. *PLOS ONE* **15**, (2020).

118. Martin, P. *et al.* APOE ϵ4, rated life experiences, and affect among centenarians. *Aging & mental health* **18**, 240–247 (2014).

119. Keller, M. C. Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. *Biological Psychiatry* **75**, 18–24 (2014).

120. Page, M. J. *et al.* The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery* **88**, 105906 (2021).

121. Alexander, K. E. *et al.* Association between single-nucleotide polymorphisms in growth factor genes and quality of life in men with prostate cancer and the general population. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* **24**, 2183–2193 (2015).

122. Conner, T. S. *et al.* The Oxytocin Receptor Gene (OXTR) Variant rs53576 Is Not Related to Emotional Traits or States in Young Adults. *Fronteers in Psychology* **9**, (2018).

123. Lucht, M. J. *et al.* Associations between the oxytocin receptor gene (OXTR) and affect, loneliness and intelligence in normal subjects. *Progress in neuro-psychopharmacology & biological psychiatry* **33**, 860–866 (2009).

124. Whillans, A. V, Aknin, L. B., Ross, C. J., Chen, L. H. & Chen, F. S. Common Variants of the Oxytocin Receptor Gene Do Not Predict the Positive Mood Benefits of Prosocial Spending. *Emotion* **20**, 734–749 (2020).

125. Love, T. M. *et al.* Oxytocin gene polymorphisms influence human dopaminergic function in a sex-dependent manner. *Biological psychiatry* **72**, 198–206 (2012).

126. Hill, L. D. *et al.* Catechol-O-methyltransferase Val158Met polymorphism associates with affect and cortisol levels in women. *Brain and Behavior* **8**, (2018).

127. Jimenez, K. M., Pereira-Morales, A. J. & Forero, D. A. Val158Met polymorphism in the COMT gene is associated with hypersomnia and mental health-related quality of life in a Colombian sample. *Neuroscience letters* **644**, 43–47 (2017).

128. Liu, J., Gong, P., Gao, X. & Zhou, X. The association between well-being and the COMT gene: Dispositional gratitude and forgiveness as mediators. *Journal of affective disorders* **214**, 115–121 (2017).

129. Turan, B., Sims, T., Best, S. E. & Carstensen, L. L. Older age may offset genetic influence on affect: The COMT polymorphism and affective well-being across the life span. *Psychology and aging* **31**, 287–294 (2016).

130. Lachmann, B. *et al.* The Molecular Genetics of Life Satisfaction: Extending Findings from a Recent Genome-Wide Association Study and Examining the Role of the Serotonin Transporter. *Journal of Happiness Studies* (2020) doi:10.1007/s10902-020-00231-x.

131. Wang, Y. *et al.* The KSR2-rs7973260 Polymorphism is Associated with Metabolic Phenotypes, but Not Psychological Phenotypes, in Chinese Elders. *Genetic testing and molecular biomarkers* **21**, 416–421 (2017).

132. Lacker, T. J., Walther, A., Fiacco, S. & Ehlert, U. The Relation Between Steroid Secretion Patterns and the Androgen Receptor Gene Polymorphism on Physical Health and Psychological Well-Being-Longitudinal Findings From the Men's Health 40+Study. *Frontiers in Human Neuroscience* **14**, (2020).

133. Matsunaga, M. *et al.* Genetic variations in the human cannabinoid receptor gene are associated with happiness. *PloS one* **9**, e93771 (2014).

134. Matsunaga, M. *et al.* Culture and cannabinoid receptor gene polymorphism interact to influence the perception of happiness. *PloS one* **13**, e0209552 (2018).

135. Matsunaga, M. *et al.* Serotonin Receptor (HTR2A) Gene Polymorphism Modulates Social Sharing of Happiness in Both American and Japanese Adults. *Japanese psychological research* (2021) doi:10.1111/jpr.12389.

136. Matsunaga, M. *et al.* A Genetic Variation in the Y Chromosome Among Modern Japanese Males Related to Several Physiological and Psychological Characteristics. *Frontiers in behavioral neuroscience* **15**, 774879 (2021).

137. Sleijpen, M., Heitland, I., Mooren, T. & Kleber, R. J. Resilience in refugee and Dutch adolescents: Genetic variability in the corticotropin releasing hormone receptor 1. *Personality and Individual Differences* **111**, 211–214 (2017).

138. Wouters, H. J. C. M. *et al.* No Effect of the Thr92Ala Polymorphism of Deiodinase-2 on Thyroid Hormone Parameters, Health-Related Quality of Life, and Cognitive Functioning in a Large Population-Based Cohort Study. *Thyroid : official journal of the American Thyroid Association* **27**, 147–155 (2017).

139. Yang, X. *et al.* Independent self-construal mediates the association between CYP19A1 gene variant and subjective well-being. *Consciousness and cognition* **55**, 205–213 (2017).

140. Lazar, A. S. *et al.* Sleep, diurnal preference, health, and psychological well-being: a prospective single-allelic-variation study. *Chronobiology international* **29**, 131–146 (2012).

141. Chen, H. *et al.* The MAOA gene predicts happiness in women. *Progress in neuro-psychopharmacology & biological psychiatry* **40**, 122–125 (2013).

142. Gureev, A. S. *et al.* Association of MAOA-uVNTR Polymorphism with Subjective Well-Being in Men. *Russian Journal of Genetics* **54**, 562–567 (2018).

143. Hu, Y. *et al.* The effect of childhood maltreatment on college students' depression symptoms: The mediating role of subjective well-being and the moderating role of MAOA gene rs6323 polymorphism. *European Journal of Developmental Psychology* (2021) doi:10.1080/17405629.2021.1928491.

144. Kimbrel, N. A. *et al.* Effect of the 5-HTTLPR polymorphism on posttraumatic stress disorder, depression, anxiety, and quality of life among Iraq and Afghanistan veterans. *Anxiety, stress, and coping* **28**, 456–466 (2015).

145. O'Hara, R. *et al.* 5-HTTLPR short allele, resilience, and successful aging in older adults. *American Journal of Geriatric Psychiatry* **20**, 452–456 (2012).

146. Perez, M. *et al.* Associations of serotonin transporter genotypes to components of the tripartite model of depression and anxiety. *Personality and Individual Differences* **43**, 107–118 (2007).

147. Nestor, P. G. *et al.* In search of positive mental health: Personality profiles and genetic polymorphisms. *Stress and Health* **37**, 310–319 (2021).

148. Ohtsubo, Y. *et al.* Test of the Serotonin Transporter Gene x Early Life Stress Interaction Effect on Subjective Well-Being and Loneliness Among Japanese Young Adults. *Japanese Psychological Research* (2021) doi:10.1111/jpr.12376.

149. Blazer, D. G., Fillenbaum, G. G., Gold, D. T., Burchett, B. M. & Hays, J. C. APOE epsilon 4 as a predictor of subjective quality of life in a biracial older person community sample. *Journal of Aging and Health* **15**, 645–660 (2003).

150. Kuo, C.-L., Pilling, L. C., Atkins, J. L., Kuchel, G. A. & Melzer, D. ApoE e2 and aging-related outcomes in 379,000 UK Biobank participants. *Aging (Albany NY)* **12**, 12222 (2020).

151. Aschard, H. *et al.* Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. *Human genetics* **131**, 1591–1613 (2012).

152. Okbay, A. & Rietveld, C. A. On improving the credibility of candidate gene studies: A review of candidate gene studies published in Emotion. *Emotion (Washington, D.C.)* **15**, 531–537 (2015).

153. Jamshidi, J. *et al.* Diverse phenotypic measurements of wellbeing: Heritability, temporal stability, and the variance explained by polygenic scores. *Genes, Brain and Behavior* (2020) doi:10.1111/gbb.12694.

154. Patel, P. C., Rietveld, C. A., Wolfe, M. T. & Wiklund, J. The Polygenic Risk Score of Subjective Well-Being, Self-Employment, and Earnings Among Older Individuals*: *https://doi.org/10.1177/1042258720936984* **45**, 440–466 (2020).

155. de Vries, L. P. *et al.* Genetic evidence for a large overlap and potential bidirectional causal effects between resilience and well-being. *Neurobiology of Stress* **14**, 100315 (2021).

156. Zhou, F., Guo, Y., Wang, Z., Liu, S. & Xu, H. Assessing the causal associations of insomnia with depressive symptoms and subjective well-being: a bidirectional Mendelian randomization study. *Sleep Medicine* **87**, 85–91 (2021).

157. Baselmans, B. M. L. & Bartels, M. A genetic perspective on the relationship between eudaimonic –and hedonic well-being. *Scientific Reports* **8**, (2018).

158. Duncan, L. E. & Keller, M. C. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry* **168**, 1041–1049 (2011).

159. Guo, J. *et al.* Quantifying genetic heterogeneity between continental populations for human height and body mass index. *Scientific Reports 2021 11:1* **11**, 1–9 (2021).

160. Fan, P. *et al.* Urbanization, economic development, environmental and social changes in transitional economies: Vietnam after Doimoi. *Landscape and Urban Planning* **187**, 145–155 (2019).

161. Liang, W. & Yang, M. Urbanization, economic growth and environmental pollution: Evidence from China. *Sustainable Computing: Informatics and Systems* **21**, 1–9 (2019).

162. Gamm, L., Stone, S. & Pittman, S. Mental health and mental disorders—A rural challenge: A literature review. *Rural healthy people,* **2**, 97–113 (2010).

163. Srivastava, K. Urbanization and mental health. *Ind Psychiatry J* **18**, 75–76 (2009).

164. Peen, J. *et al.* Is the prevalence of psychiatric disorders associated with urbanization? *Social Psychiatry and Psychiatric Epidemiology* **42**, 984–989 (2007).

165. Alcock, I., White, M. P., Wheeler, B. W., Fleming, L. E. & Depledge, M. H. Longitudinal Effects on Mental Health of Moving to Greener and Less Green Urban Areas. *Environmental Science and Technology* **48**, 1247–1255 (2014).

166. Lederbogen, F. *et al.* City living and urban upbringing affect neural social stress processing in humans. *Nature* **474**, 498–501 (2011).

167. Chen, J. & Chen, S. Mental health effects of perceived living environment and neighborhood safety in urbanizing China. *Habitat International* **46**, 101–110 (2015).

168. Whitley, R. & Prince, M. Fear of crime, mobility and mental health in inner-city London, UK. *Social Science and Medicine* **61**, 1678–1688 (2005).

169. Colodro-Conde, L. *et al.* Association between population density and genetic risk for schizophrenia. in *JAMA Psychiatry* vol. 75 901–910 (American Medical Association, 2018).

170. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**, e1001779 (2015).

171. Howard, D. M. *et al.* Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nature Communications* **9**, 1470 (2018).

172. Shen, X. *et al.* Subcortical volume and white matter integrity abnormalities in major depressive disorder: Findings from UK Biobank imaging data. *Scientific Reports* **7**, 1–10 (2017).

173. Bakker, B. F. M., Van Rooijen, J. & Van Toor, L. The System of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics. *Statistical Journal of the IAOS* **30**, 411–424 (2014).

174. De Bruin, A. *et al.* Record linkage of hospital discharge register with population register: Experiences at Statistics Netherlands. *Statistical Journal of the United Nations Economic Commission for Europe* **21**, 23–32 (2004).

175. Timmermans, E. J. *et al.* Cohort profile: the Geoscience and Health Cohort Consortium (GECCO) in the Netherlands. *BMJ open* **8**, e021597 (2018).

176. Penninx, B. W. J. H. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *International Journal of Methods in Psychiatric Research* **17**, 121–140 (2008).

177. Dolan, P., Peasgood, T. & White, M. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology* **29**, 94–122 (2008).

178. Helliwell, J. F. & Putnam, R. D. The social context of well-being. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359**, 1435–46 (2004).

179. Patel, C. J. & Ioannidis, J. P. A. Studying the elusive environment in large scale. *JAMA - Journal of the American Medical Association* vol. 311 2173–2174 Preprint at https://doi.org/10.1001/jama.2014.4129 (2014).

180. Abdellaoui, A. *et al.* Genetic correlates of social stratification in Great Britain. *Nature Human Behaviour* **3**, 1332–1342 (2019).

181. Abdellaoui, A. *et al.* Population structure, migration, and diversifying selection in the Netherlands. *European journal of human genetics : EJHG* **21**, 1277–85 (2013).

182. Ioannidis, J. P. A., Loy, E. Y., Poulton, R. & Chia, K. S. Researching genetic versus nongenetic determinants of disease: A comparison and proposed unification. *Science Translational Medicine* vol. 1 Preprint at https://doi.org/10.1126/scitranslmed.3000247 (2009).

183. Willemsen, G. *et al.* The Adult Netherlands Twin Register: Twenty-Five Years of Survey and Biological Data Collection. *Twin Research and Human Genetics* **16**, 271–281 (2013).

184. Consortium, T. G. of the N. *et al.* Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature Genetics* **46**, 818–825 (2014).

185. Minicə, C. C., Boomsma, D. I., Vink, J. M. & Dolan, C. V. MZ twin pairs or MZ singletons in population family-based GWAS? More power in pairs. *Molecular Psychiatry* vol. 19 1154–1155 Preprint at https://doi.org/10.1038/mp.2014.121 (2014).

186. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American journal of human genetics* **97**, 576–92 (2015).

187. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize implements and enhances circular visualization in R | Bioinformatics | Oxford Academic. *Bioinformatics* **30**, 2811–2812 (2014).

188. Choi, Y. J. & Matz-Costa, C. Perceived Neighborhood Safety, Social Cohesion, and Psychological Health of Older Adults. *The Gerontologist* **58**, 196–206 (2018).

189. Wilson-Genderson, M. & Pruchno, R. Effects of neighborhood violence and perceptions of neighborhood safety on depressive symptoms of older adults. *Social Science and Medicine* **85**, 43–49 (2013).

190. Cheng, Z. & Smyth, R. Crime victimization, neighborhood safety and happiness in China. *Economic Modelling* **51**, 424–435 (2015).

191. Ludwig, J. *et al.* Neighborhood effects on the long-term well-being of low-income adults. *Science (New York, N.Y.)* **337**, 1505–10 (2012).

192. Dittmann, J. & Goebel, J. Your House, Your Car, Your Education: The Socioeconomic Situation of the Neighborhood and its Impact on Life Satisfaction in Germany. *Social Indicators Research* **96**, 497–513 (2010).

193. Generaal, E., Timmermans, E. J., Dekkers, J. E. C., Smit, J. H. & Penninx, B. W. J. H. Not urbanization level but socioeconomic, physical and social neighbourhood characteristics are associated with presence and severity of depressive and anxiety disorders. *Psychological medicine* **49**, 149–161 (2019).

194. van Kamp, I., Leidelmeijer, K., Marsman, G. & de Hollander, A. Urban environmental quality and human well-being: Towards a conceptual framework and demarcation of concepts; a literature study. *Landscape and Urban Planning* **65**, 5–18 (2003).

195. Liao, P. shan, Shaw, D. & Lin, Y. ming. Environmental Quality and Life Satisfaction: Subjective Versus Objective Measures of Air Quality. *Social Indicators Research* **124**, 599–616 (2015).

196. Zeidner, M. & Shechter, M. Psychological responses to air pollution: Some personality and demographic correlates. *Journal of Environmental Psychology* **8**, 191–208 (1988).

197. Barger, S. D., Donoho, C. J. & Wayment, H. A. The relative contributions of race/ethnicity, socioeconomic status, health, and social relationships to life satisfaction in the United States. *Quality of Life Research* **18**, 179–189 (2009).

198. Navarro-Carrillo, G., Alonso-Ferres, M., Moya, M. & Valor-Segura, I. Socioeconomic Status and Psychological Well-Being: Revisiting the Role of Subjective Socioeconomic Status. *Frontiers in Psychology* **11**, 1303 (2020).

199. McMahan, E. A. & Estes, D. The effect of contact with natural environments on positive and negative affect: A meta-analysis. *The Journal of Positive Psychology* **10**, 507–519 (2015).

200. Welsch, H. Environment and happiness: Valuation of air pollution using life satisfaction data. *Ecological Economics* **58**, 801–813 (2006).

201. Dennis, K. K. *et al.* The importance of the biological impact of exposure to the concept of the exposome. *Environmental Health Perspectives* vol. 124 1504–1510 Preprint at https://doi.org/10.1289/EHP140 (2016).

202. Jiameng, S. & Huiyong, F. a. N. A Meta-analysis of the Relationship between Social Support and Subjective Well-being. *Advances in Psychological Science* **21**, 1357 (2013).

203. Tan, J. J. X., Kraus, M. W., Carpenter, N. C. & Adler, N. E. The association between objective and subjective socioeconomic status and subjective well-being: A meta-analytic review. *Psychological Bulletin* **146**, 970–1020 (2020).

204. Wild, C. P. The exposome: from concept to utility. *International Journal of Epidemiology* **41**, 24–32 (2012).

205. van de Weijer, M. P. *et al.* Expanding the environmental scope: an environment-wide association study for mental well-being. *Journal of Exposure Science and Environmental Epidemiology* **32**, 195–204 (2022).

206. Abdellaoui, A. *et al.* Predicting loneliness with polygenic scores of social, psychological and psychiatric traits. *Genes, Brain and Behavior* **17**, (2018).

207. Akingbuwa, W. A. *et al.* Genetic Associations between Childhood Psychopathology and Adult Depression and Associated Traits in 42998 Individuals: A Meta-analysis. *JAMA Psychiatry* **77**, 715–728 (2020).

208. Song, L. *et al.* Polygenic Score of Subjective Well-Being Is Associated with the Brain Morphology in Superior Temporal Gyrus and Insula. *Neuroscience* **414**, 210–218 (2019).

209. Padmanabhan, J. L., Shah, J. L., Tandon, N. & Keshavan, M. S. The "polyenviromic risk score": Aggregating environmental risk factors predicts conversion to psychosis in familial high-risk subjects. *Schizophrenia Research* **181**, 17–22 (2017).

210. Pries, L.-K. *et al.* Estimating Exposome Score for Schizophrenia Using Predictive Modeling Approach in Two Independent Samples: The Results From the EUGEI Study. *Schizophrenia Bulletin* **45**, 960–965 (2019).

211. Pries, L.-K., Erzin, G., Rutte, B. P. F., van Os, J. & Guloksuz, S. Estimating Aggregate Environmental Risk Score in Psychiatry: The Exposome Score for Schizophrenia. *Frontiers in Psychiatry* **12**, (2021).

212. Vassos, E. *et al.* The Maudsley environmental risk score for psychosis. *Psychological Medicine* **50**, 2213–2220 (2020).

213. Jeon, E.-J. *et al.* Development of the Korea-Polyenvironmental Risk Score for Psychosis. *Psychiatry Investig* **19**, 197–206 (2022).

214. Deckers, K. *et al.* Gender and Educational Differences in the Association between Lifestyle and Cognitive Decline over 10 Years: The Doetinchem Cohort Study. *Journal of Alzheimer's Disease* **70**, S31–S41 (2019).

215. Oliver, D., Radua, J., Reichenberg, A., Uher, R. & Fusar-Poli, P. Psychosis Polyrisk Score (PPS) for the Detection of Individuals At-Risk and the Prediction of Their Outcomes. *Frontiers in Psychiatry* **10**, (2019).

216. He, Y., Lakhani, C. M., Manrai, A. K. & Patel, C. J. Poly-Exposure and Poly-Genomic Scores Implicate Prominent Roles of Non-Genetic and Demographic Factors in Four Common Diseases in the UK. 833632 Preprint at https://doi.org/10.1101/833632 (2019).

217. Dick, D. M. Gene-Environment Interaction in Psychological Traits and Disorders. *Annu Rev Clin Psychol* **7**, 383–409 (2011).

218. Jaffee, S. R. & Price, T. S. Gene–environment correlations: a review of the evidence and implications for prevention of mental illness. *Mol Psychiatry* **12**, 432–442 (2007).

219. van de Weijer, M. P. *et al.* Genetic and environmental influences on quality of life: The COVID-19 pandemic as a natural experiment. *Genes, Brain and Behavior* **21**, e12796 (2022).

220. Keyes, C. L. M. The mental health continuum: from languishing to flourishing in life. *J Health Soc Behav* **43**, 207–222 (2002).

221. Achenbach, T. M., Bernstein, A. & Dumenci, L. DSM-Oriented Scales and Statistically Based Syndromes for Ages 18 to 59: Linking Taxonomic Paradigms to Facilitate Multitaxonomic Approaches. *Journal of Personality Assessment* **84**, 49–63 (2005).

222. Via, M., Gignoux, C. & Burchard, E. G. The 1000 Genomes Project: new opportunities for research and social challenges. *Genome Medicine* **2**, 3 (2010).

223. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443–1448 (2016).

224. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).

225. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284–1287 (2016).

226. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).

227. Kuhn, M. caret: Classification and Regression Training. (2022).

228. Olkin, I. & Finn, J. D. Correlations redux. *Psychological Bulletin* **118**, 155–164 (1995).

229. Robinson, M. R. *et al.* Genetic evidence of assortative mating in humans. *Nat Hum Behav* **1**, 1–13 (2017).

230. Harandi, T. F., Taghinasab, M. M. & Nayeri, T. D. The correlation of social support with mental health: A meta-analysis. *Electron Physician* **9**, 5212–5222 (2017).

231. Gerhardt, C. *et al.* How are social stressors at work related to well-being and health? A systematic review and meta-analysis. *BMC Public Health* **21**, 890 (2021).

232. Clark, A. E. & Oswald, A. J. A simple statistical method for measuring how life events affect happiness. *International Journal of Epidemiology* **31**, 1139–1144 (2002).

233. Ballas, D. & Dorling, D. Measuring the impact of major life events upon happiness. *International Journal of Epidemiology* **36**, 1244–1252 (2007).

234. van de Weijer, M. P., Pelt, D. H. M., van Beijsterveldt, C. E. M., Willemsen, G. & Bartels, M. Genetic factors explain a significant part of associations between adolescent well-being and the social environment. *Eur Child Adolesc Psychiatry* **31**, 1611–1622 (2022).

235. Ballas, D. & Tranmer, M. Happy People or Happy Places? A Multilevel Modeling Approach to the Analysis of Happiness and Well-Being. *International Regional Science Review* **35**, 70–102 (2012).

236. Ettema, D. & Schekkerman, M. How do spatial characteristics influence well-being and mental health? Comparing the effect of objective and subjective characteristics at different spatial scales. *Travel Behaviour and Society* **5**, 56–67 (2016).

237. Brereton, F., Clinch, J. P. & Ferreira, S. Happiness, geography and the environment. *Ecological Economics* **65**, 386–396 (2008).

238. Orru, K., Orru, H., Maasikmets, M., Hendrikson, R. & Ainsaar, M. Well-being and environmental quality: Does pollution affect life satisfaction? *Qual Life Res* **25**, 699–705 (2016).

239. Bertram, C. & Rehdanz, K. The role of urban green space for human well-being. *Ecological Economics* **120**, 139–152 (2015).

240. Benita, F., Bansal, G. & Tunçer, B. Public spaces and happiness: Evidence from a large-scale field experiment. *Health & Place* **56**, 9–18 (2019).

241. Pruitt, S. L., Jeffe, D. B., Yan, Y. & Schootman, M. Reliability of perceived neighbourhood conditions and the effects of measurement error on self-rated health across urban and rural neighbourhoods. *J Epidemiol Community Health* **66**, 342–351 (2012).

242. Keyes, C. L. M., Myers, J. M. & Kendler, K. S. The Structure of the Genetic and Environmental Influences on Mental Well-Being. *Am J Public Health* **100**, 2379–2384 (2010).

243. Holmes, E. A. *et al.* Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The Lancet Psychiatry* **7**, 547–560 (2020).

244. Helliwell, J. F., Huang, H., Wang, S. & Norton, M. Social Environments for World Happiness. in *World Happiness Report 2020* (2020).

245. van Tilburg, T. G., Steinmetz, S., Stolte, E., van der Roest, H. & de Vries, D. H. Loneliness and mental health during the COVID-19 pandemic: A study among Dutch older adults. *The Journals of Gerontology: Series B* **XX**, 1–7 (2020).

246. Gijzen, M. *et al.* The bittersweet effects of COVID-19 on mental health: Results of an online survey among a sample of the Dutch population. *OSF preprint* (2020) doi:10.31219/OSF.IO/7JSNG.

247. Cancello, R., Soranna, D., Zambra, G., Zambon, A. & Invitti, C. Determinants of the Lifestyle Changes during COVID-19 Pandemic in the Residents of Northern Italy. *International Journal of Environmental Research and Public Health* **17**, 6287 (2020).

248. Martínez-de-Quel, Ó., Suárez-Iglesias, D., López-Flores, M. & Pérez, C. A. Physical activity, dietary habits and sleep quality before and during COVID-19 lockdown: A longitudinal study. *Appetite* **158**, 105019 (2021).

249. Poelman, M. P. *et al.* Eating behavior and food purchases during the COVID-19 lockdown: A cross-sectional study among adults in the Netherlands. *Appetite* **157**, 105002 (2021).

250. Kowal, M. *et al.* Who is the Most Stressed During the COVID-19 Pandemic? Data From 26 Countries and Areas. *Applied Psychology: Health and Well-Being* **12**, 946–966 (2020).

251. Bu, F., Steptoe, A. & Fancourt, D. Who is lonely in lockdown? Cross-cohort analyses of predictors of loneliness before and during the COVID-19 pandemic. *Public Health* **186**, 31–34 (2020).

252. Ahmad, F., Jhajj, A. K., Stewart, D. E., Burghardt, M. & Bierman, A. S. Single item measures of self-rated mental health: A scoping review. *BMC Health Services Research* vol. 14 398 Preprint at https://doi.org/10.1186/1472-6963-14-398 (2014).

253. Jylhä, M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science and Medicine* **69**, 307–316 (2009).

254. Idler, E. L. & Benyamini, Y. Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. *Journal of Health and Social Behavior* vol. 38 21–37 Preprint at https://doi.org/10.2307/2955359 (1997).

255. Van Lente, E. *et al.* Measuring population mental health and social well-being. *International Journal of Public Health* **57**, 421–430 (2012).

256. Assari, S. & Lankarani, M. M. Does multi-morbidity mediate the effect of socioeconomics on self-rated health? Cross-country differences. *International Journal of Preventive Medicine* **2015-September**, (2015).

257. Perruccio, A. V., Badley, E. M., Hogg-Johnson, S. & Davis, A. M. Characterizing self-rated health during a period of changing health status. *Social Science & Medicine* **71**, 1636–1643 (2010).

258. Gunasekara, F. I., Carter, K. & Blakely, T. Change in income and change in self-rated health: Systematic review of studies using repeated measures to control for confounding bias. *Social Science & Medicine* **72**, 193–201 (2011).

259. Dieker, A. C. M. *et al.* The contribution of work and lifestyle factors to socioeconomic inequalities in self-rated health - a systematic review. *Scandinavian Journal of Work, Environment and Health* **45**, 114–125 (2019).

260. Pisinger, C., Toft, U., Aadahl, M., Glümer, C. & Jørgensen, T. The relationship between lifestyle and self-reported health in a general population: The Inter99 study. *Preventive Medicine* **49**, 418–423 (2009).

261. Peters, A., Rospleszcz, S., Greiser, K. H., Dallavalle, M. & Berger, K. The Impact of the COVID-19 Pandemic on Self-Reported Health. *Deutsches Arzteblatt international* **117**, 861–867 (2020).

262. Recchi, E. *et al.* The "Eye of the Hurricane" Paradox: An Unexpected and Unequal Rise of Well-Being During the Covid-19 Lockdown in France. *Research in Social Stratification and Mobility* **68**, 100508 (2020).

263. Bierman, A., Upenieks, L., Glavin, P. & Schieman, S. Accumulation of economic hardship and health during the COVID-19 pandemic: Social causation or selection? *Social Science & Medicine* **275**, 113774 (2021).

264. Szwarcwald, C. L. *et al.* Factors affecting Brazilians' self-rated health during the COVID-19 pandemic. *Cadernos de Saúde Pública* **37**, (2021).

265. Silventoinen, K., Posthuma, D., Lahelma, E., Rose, R. J. & Kaprio, J. Genetic and environmental factors affecting self-rated health from age 16-25: A longitudinal study of Finnish twins. *Behavior Genetics* **37**, 326–333 (2007).

266. Romeis, J. C. *et al.* Heritability of self-reported health. *Health Services Research* **5**, 995–1010 (2000).

267. Mosing, M. A., Zietsch, B. P., Shekar, S. N., Wright, M. J. & Martin, N. G. Genetic and environmental influences on optimism and its relationship to mental and self-rated health: A study of aging twins. *Behavior Genetics* **39**, 597–604 (2009).

268. De Geus, E. J. C., Kupper, N., Boomsma, D. I. & Snieder, H. Bivariate Genetic Modeling of Cardiovascular Stress Reactivity: Does Stress Uncover Genetic Variance? *Psychosomatic Medicine* **69**, 356–364 (2007).

269. Menni, C. *et al.* Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine* **26**, 1037–1040 (2020).

270. Chandola, T. & Jenkinson, C. Validating self-rated health in different ethnic groups. *Ethnicity and Health* **5**, 151–159 (2000).

271. Bardage, C. *et al.* Self-rated health among older adults: A cross-national comparison. *European Journal of Ageing* **2**, 149–158 (2005).

272. R Core Team. R: A language and environment for statistical computing. Preprint at (2013).

273. De Moor, M. H. M., Stubbe, J. H., Boomsma, D. I. & De Geus, E. J. C. Exercise participation and self-rated health: Do common genes explain the association? *European Journal of Epidemiology* **22**, 27–32 (2007).

274. Verhulst, B., Prom-Wormley, E., Keller, M., Medland, S. & Neale, M. C. Type I Error Rates and Parameter Bias in Multivariate Behavioral Genetic Models. *Behavior Genetics* **49**, 99–111 (2019).

275. Boker, S. *et al.* OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika* **76**, 306–317 (2011).

276. Benjamin, D. J. *et al.* Redefine statistical significance. *Nature Human Behaviour* vol. 2 6–10 Preprint at https://doi.org/10.1038/s41562-017-0189-z (2018).

277. Falconer, D. S. The Problem of Environment and Selection. *The American Naturalist* (1952) doi:10.1086/281736.

278. Eurostat. Self-perceived health statistics - Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Self-perceived_health_statistics#Self-perceived_health (2020).

279. OECD Indicators. *Health at a Glance* . (2019). doi:10.1787/4DD50C09-EN.

280. Ding, D., Cruz, B. del P., Green, M. A. & Bauman, A. E. Is the COVID-19 lockdown nudging people to be more active: a big data analysis. *Br J Sports Med* **54**, 1183–1184 (2020).

281. Romero-Blanco, C. *et al.* Physical Activity and Sedentary Lifestyle in University Students: Changes during Confinement Due to the COVID-19 Pandemic. *International Journal of Environmental Research and Public Health* **17**, 6567 (2020).

282. Stockwell, S. *et al.* Changes in physical activity and sedentary behaviours from before to during the COVID-19 pandemic lockdown: a systematic review. *BMJ Open Sport & Exercise Medicine* **7**, e000960 (2021).

283. Cheng, S.-T., Fung, H. & Chan, A. Maintaining Self-Rated Health Through Social Comparison in Old Age. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* **62**, P277–P285 (2007).

284. Nwachukwu, I. *et al.* COVID-19 Pandemic: Age-Related Differences in Measures of Stress, Anxiety and Depression in Canada. *International Journal of Environmental Research and Public Health* **17**, 6366 (2020).

285. Bann, D. *et al.* Changes in the behavioural determinants of health during the coronavirus (COVID-19) pandemic: Gender, socioeconomic and ethnic inequalities in 5 British cohort studies. *medRxiv* 2020.07.29.20164244 Preprint at https://doi.org/10.1101/2020.07.29.20164244 (2020).

286. Cellini, N. *et al.* Changes in sleep timing and subjective sleep quality during the COVID-19 lockdown in Italy and Belgium: age, gender and working status as modulating factors. *Sleep Medicine* **77**, 112–119 (2021).

287. Blix, I., Birkeland, M. S. & Thoresen, S. Worry and mental health in the Covid-19 pandemic: vulnerability factors in the general Norwegian population. *BMC Public Health 2021 21:1* **21**, 1–10 (2021).

288. Cunningham, T. J., Fields, E. C., Garcia, S. M. & Kensinger, E. A. The relation between age and experienced stress, worry, affect, and depression during the spring 2020 phase of the COVID-19 pandemic in the United States. *Emotion* (2021) doi:10.1037/EMO0000982.

289. De Kock, J. H. *et al.* A rapid review of the impact of COVID-19 on the mental health of healthcare workers: implications for supporting psychological well-being. *BMC Public Health 2021 21:1* **21**, 1–18 (2021).

290. Bierman, A., Upenieks, L. & Schieman, S. Socially Distant? Social Network Confidants, Loneliness, and Health during the COVID-19 Pandemic: *https://doi.org/10.1177/23294965211011591* **8**, 299–313 (2021).

291. Maslowski, R. *Onderwijs. In: De sociale staat van Nederland 2018.* (2018).

292. Zajacova, A. & Dowd, J. B. Reliability of self-rated health in US adults. *American journal of epidemiology* **174**, 977–983 (2011).

293. Leatherdale, S. T. Natural experiment methodology for research: a review of how different methods can support real-world research. *International Journal of Social Research Methodology* **22**, 19–35 (2019).

294. Barcellos, S. H., Carvalho, L. S. & Turley, P. Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E9765–E9772 (2018).

295. Kabir, Z., Daly, S., Clarke, V., Keogan, S. & Clancy, L. Smoking Ban and Small-For-Gestational Age Births in Ireland. *PLoS ONE* **8**, e57441 (2013).

296. Vink, J. M. & Boomsma, D. I. Interplay between heritability of smoking and environmental conditions? A comparison of two birth cohorts. *BMC Public Health 2011 11:1* **11**, 1–7 (2011).

297. Hikichi, H., Aida, J., Tsuboya, T., Kondo, K. & Kawachi, I. Can Community Social Cohesion Prevent Posttraumatic Stress Disorder in the Aftermath of a Disaster? A Natural Experiment From the 2011 Tohoku Earthquake and Tsunami. *American Journal of Epidemiology* **183**, 902–910 (2016).

298. Pesonen, A.-K. *et al.* Depressive Symptoms in Adults Separated from Their Parents as Children: A Natural Experiment during World War II. *American Journal of Epidemiology* **166**, 1126–1133 (2007).

299. Prati, G. & Mancini, A. D. The psychological impact of COVID-19 pandemic lockdowns: A review and meta-analysis of longitudinal studies and natural experiments. *Psychological Medicine* vol. 51 201–211 Preprint at https://doi.org/10.1017/S0033291721000015 (2021).

300. Pan, K. Y. *et al.* The mental health impact of the COVID-19 pandemic on people with and without depressive, anxiety, or obsessive-compulsive disorders: a longitudinal study of three Dutch case-control cohorts. *The Lancet Psychiatry* **8**, 121–129 (2021).

301. van de Weijer, M., de Vries, L. & Bartels, M. Happiness and Wellbeing; the value and findings from genetic studies. *psyarxiv* (2021) doi:10.31234/OSF.IO/ZVU8J.

302. Plomin, R. & Haworth, C. M. A. Genetics and Intervention Research. *Perspectives on psychological science : a journal of the Association for Psychological Science* **5**, 557 (2010).

303. Haworth, C. M. A. *et al.* Stability and change in genetic and environmental influences on well-being in response to an intervention. *PLoS ONE* **11**, e0155538 (2016).

304. Burgoyne, A. P. *et al.* Can a brief intervention alter genetic and environmental influences on psychological traits? An experimental behavioral genetics approach. *Learning and Motivation* **72**, 101683 (2020).

305. de Vries, L. *et al.* Individual differences in the effect of the COVID-19 pandemic on optimism and meaning in life. *psyarxiv* (2021) doi:10.31234/OSF.IO/B2GE6.

306. Rimfeld, K. *et al.* Genetic Correlates of Psychological Responses to the COVID-19 Crisis in Young Adult Twins in Great Britain. *Behavior Genetics* **51**, 110–124 (2021).

307. Kafka, G. J. & Kozma, A. The construct validity of Ryff's Scales of Psychological Well-Being (SPWB) and their relationship to measures of subjective well-being. *Social Indicators Research* **57**, 171–190 (2002).

308. Boomsma, D. I. *et al.* Design and implementation of a twin-family database for behavior genetics and genomics studies. *Twin Research and Human Genetics* **11**, 342–348 (2008).

309. Blokland, I. V. van *et al.* Using symptom-based case predictions to identify host genetic factors that contribute to COVID-19 susceptibility. *PLOS ONE* **16**, e0255402 (2021).

310. Lange, K. *et al.* Mendel: the Swiss army knife of genetic analysis programs. *Bioinformatics* **29**, 1568–1570 (2013).

311. Lange, K. & Boehnke, M. Extensions to pedigree analysis. IV. Covariance components models for multivariate traits. *American Journal of Medical Genetics* **14**, 513–524 (1983).

312. Posthuma, D. *et al.* Theory and Practice in Quantitative Genetics. *Twin Research* **6**, 361–376 (2003).

313. Kwong, A. *et al.* Mental health during the COVID-19 pandemic in two longitudinal UK population cohorts. *medRxiv* 2020.06.16.20133116 (2020) doi:10.1101/2020.06.16.20133116.

314. Li, H. Y., Cao, H., Leung, D. Y. P. & Mak, Y. W. The psychological impacts of a covid-19 outbreak on college students in China: A longitudinal study. *International Journal of Environmental Research and Public Health* **17**, (2020).

315. Aknin, L. *et al.* Mental Health During the First Year of the COVID-19 Pandemic: A Review and Recommendations for Moving Forward. *PsyArXiv* (2021) doi:10.31234/OSF.IO/ZW93G.

316. Helliwell, J. F., Huang, H. & Wang, S. Changing World Happiness. in *World Happiness Report 2019* 11–46 (2019).

317. Helliwell, J. F., Huang, H., Wang, S. & Norton, M. Happiness, trust, and deaths under COVID-19. in *The World Happiness Report 2021* (2021).

318. Barnett, A. G., van der Pols, J. C. & Dobson, A. J. Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology* **34**, 215–220 (2005).

319. Keller, M. C. *et al.* Modeling extended twin family data I: Description of the cascade model. *Twin Research and Human Genetics* **12**, 8–18 (2009).

320. Carroll, S. L., Shewark, E. A., Hyde, L. W., Klump, K. L. & Burt, S. A. Understanding the Effects of the COVID-19 Pandemic on Youth Psychopathology: Genotype–Environment Interplay. *Biological Psychiatry Global Open Science* (2021) doi:10.1016/J.BPSGOS.2021.07.004.

321. South, S. C., Hamdi, N. R. & Krueger, R. F. Biometric Modeling of Gene-Environment Interplay: The Intersection of Theory and Method and Applications for Social Inequality. *Journal of Personality* **85**, 22–37 (2017).

322. Ravens-Sieberer, U. *et al.* Impact of the COVID-19 pandemic on quality of life and mental health in children and adolescents in Germany. *European Child & Adolescent Psychiatry 2021* **1**, 1–11 (2021).

323. Hamersma, M., de Haas, M. & Faber, R. Thuiswerken en de coronacrisis. Een overzicht van studies naar de omvang, beleving en toekomstverwachting van thuiswerken in coronatijd. (2020).

324. Kesselring, M. & van Spanje-Hennes, A. *Slepend en soms ook slopend. De impact van (de maatregelen tegen) COVID-19 op het gezinsleven.* (2021).

325. Viner, R. M. *et al.* Life course epidemiology: Recognising the importance of adolescence. *Journal of Epidemiology and Community Health* vol. 69 719–720 Preprint at https://doi.org/10.1136/jech-2014-205300 (2015).

326. Fergusson, D. M., Horwood, L. J., Ridder, E. M. & Beautrais, A. L. Subthreshold depression in adolescence and mental health outcomes in adulthood. *Archives of General Psychiatry* **62**, 66–72 (2005).

327. Clayborne, Z. M., Varin, M. & Colman, I. Systematic Review and Meta-Analysis: Adolescent Depression and Long-Term Psychosocial Outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry* **58**, 72–79 (2019).

328. Birmaher, B. *et al.* Childhood and adolescent depression: A review of the past 10 years. Part I. *Journal of the American Academy of Child and Adolescent Psychiatry* **35**, 1427–1439 (1996).

329. Johnson, D., Dupuis, G., Piche, J., Clayborne, Z. & Colman, I. Adult mental health outcomes of adolescent depression: A systematic review. *Depression and Anxiety* **35**, 700–716 (2018).

330. Tennant, R. *et al.* The Warwick-Dinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes* **5**, 1–13 (2007).

331. Zaborskis, A. *et al.* Social Inequality in Adolescent Life Satisfaction: Comparison of Measure Approaches and Correlation with Macro-level Indices in 41 Countries. *Social Indicators Research* **141**, 1055–1079 (2019).

332. Bartels, M., Cacioppo, J. T., Van Beijsterveldt, T. C. E. M. & Boomsma, D. I. Exploring the association between well-being and psychopathology in adolescents. *Behavior Genetics* **43**, 177–190 (2013).

333. Coffey, J. K., Warren, M. T. & Gottfried, A. W. Does Infant Happiness Forecast Adult Life Satisfaction? Examining Subjective Well-Being in the First Quarter Century of Life. *Journal of Happiness Studies* **16**, 1401–1421 (2015).

334. Hoyt, L. T., Chase-Lansdale, P. L., McDade, T. W. & Adam, E. K. Positive youth, healthy adults: Does positive well-being in adolescence predict better perceived health and fewer risky health behaviors in young adulthood? *Journal of Adolescent Health* **50**, 66–73 (2012).

335. Chu, P. Sen, Saucier, D. A. & Hafner, E. Meta-Analysis of the Relationships Between Social Support and Well-Being in Children and Adolescents. *Journal of Social and Clinical Psychology* **29**, 624–645 (2010).

336. Holder, M. D. & Coleman, B. Children's friendships and positive well-being. in *Friendship and Happiness: Across The Life-Span and Cultures* 81–94 (Springer Netherlands, 2015). doi:10.1007/978-94-017-9603-3_5.

337. Vries, D. *et al.* Genetic evidence for bidirectional causal effects between resilience and well-being. *bioRxiv* 2020.11.03.366492 (2020) doi:10.1101/2020.11.03.366492.

338. Kendler, K. S. & Baker, J. H. Genetic influences on measures of the environment: a systematic review. *Psychological Medicine* **37**, 615–626 (2007).

339. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* **47**, 702–709 (2015).

340. Bartels, M., van de Aa, N., van Beijsterveldt, C. E. M., Middeldorp, C. M. & Boomsma, D. I. Adolescent self-report of emotional and behavioral problems: interactions of genetic factors with sex and age. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Academie canadienne de psychiatrie de l'enfant et de l'adolescent* **20**, 35–52 (2011).

341. Epstein, N. B., Baldwin, L. M. & Bishop, D. S. THE McMASTER FAMILY ASSESSMENT DEVICE*. *Journal of Marital and Family Therapy* **9**, 171–180 (1983).

342. Moos, R. H. & Moos, B. S. *Family environment scale : manual*. (Consulting Psychologists Press, 1981).

343. van der Aa, N., Boomsma, D. I., Rebollo-Mesa, I., Hudziak, J. J. & Bartels, M. Moderation of Genetic Factors by Parental Divorce in Adolescents' Evaluations of Family Functioning and Subjective Wellbeing. *TWIN RESEARCH AND HUMAN GENETICS* **13**, 143–162 (2010).

344. Carey, V. J., Lumley, T. & Ripley, B. gee: Generalized estimation equation solver. *R package version* **4**, 13–20 (2019).

345. Derks, E. M., Dolan, C. V. & Boomsma, D. I. Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research* **7**, 659–669 (2004).

346. Sacker, A. & Cable, N. Do adolescent leisure-time physical activities foster health and well-being in adulthood? Evidence from two British birth cohorts. *European Journal of Public Health* **16**, 331–335 (2006).

347. Birkeland, M. S., Torsheim, T. & Wold, B. A longitudinal study of the relationship between leisure-time physical activity and depressed mood among adolescents. *Psychology of Sport and Exercise* **10**, 25–34 (2009).

348. Bagwell, C. L., Kochel, K. P. & Schmidt, M. E. Friendship and happiness in adolescence. in *Friendship and Happiness: Across The Life-Span and Cultures* 99–116 (Springer Netherlands, 2015). doi:10.1007/978-94-017-9603-3_6.

349. Engels, R. C. M. E., Knibbe, R. A. & Drop, M. J. Visiting public drinking places: An explorative study into the functions of pub-going for late adolescents. *Substance Use and Misuse* **34**, 1261–1280 (1999).

350. Shek, D. T. L. A longitudinal study of the relations between parent-adolescent conflict and adolescent psychological well-being. *Journal of Genetic Psychology* **159**, 53–67 (1998).

351. Raboteg-Saric, Z. & Sakic, M. Relations of Parenting Styles and Friendship Quality to Self-Esteem, Life Satisfaction and Happiness in Adolescents. *Applied Research in Quality of Life* **9**, 749–765 (2014).

352. Holder, M. D. & Coleman, B. The Contribution of Social Relationships to Children's Happiness. *Journal of Happiness Studies* **10**, 329–349 (2009).

353. Nes, R. B., Czajkowski, N. & Tambs, K. Family matters: Happiness in nuclear families and twins. *Behavior Genetics* **40**, 577–590 (2010).

354. De Moor, M. H. M., Boomsma, D. I., Stubbe, J. H., Willemsen, G. & De Geus, E. J. C. Testing causality in the association between regular exercise and symptoms of anxiety and depression. *Archives of General Psychiatry* **65**, 897–905 (2008).

355. Demir, M., Orthel-Clark, H., Özdemir, M. & Özdemir, S. B. Friendship and happiness among young adults. in *Friendship and Happiness: Across The Life-Span and Cultures* 117–135 (Springer Netherlands, 2015). doi:10.1007/978-94-017-9603-3_7.

356. Duffy, D. L. & Martin, N. G. Inferring the direction of causation in cross-sectional twin data: Theoretical and empirical considerations. *Genetic Epidemiology* **11**, 483–502 (1994).

357. Luo, Y. L. L., Haworth, C. M. A. & Plomin, R. A novel approach to genetic and environmental analysis of cross-lagged associations over time: The cross-lagged relationship between self-perceived abilities and school achievement is mediated by genes as well as the environment. *Twin Research and Human Genetics* **13**, 426–436 (2010).

358. Allegrini, A. G. *et al.* Multivariable G-E interplay in the prediction of educational achievement. *PLOS Genetics* **16**, e1009153 (2020).

359. Burgess, S., Timpson, N. J., Ebrahim, S. & Davey Smith, G. Mendelian randomization: where are we now and where are we going? *International Journal of Epidemiology* **44**, 379–388 (2015).

360. Hofstede, G. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. (Sage Publications, 2001).

361. Grossmann, I. & Na, J. Research in culture and psychology: past lessons and future challenges. *Wiley Interdisciplinary Reviews: Cognitive Science* **5**, 1–14 (2014).

362. Lambert, L. *et al.* Towards a greater global understanding of wellbeing: A proposal for a more inclusive measure. *International Journal of Wellbeing* **10**, 1–18 (2020).

363. Bücker, S., Nuraydin, S., Simonsmeier, B. A., Schneider, M. & Luhmann, M. Subjective well-being and academic achievement: A meta-analysis. *Journal of Research in Personality* **74**, 83–94 (2018).

364. Lorant, V. *et al.* Socioeconomic Inequalities in Depression: A Meta-Analysis. *American Journal of Epidemiology* **157**, 98–112 (2003).

365. Furnée, C. A., Groot, W. & Van Den Brink, H. M. The health effects of education: a meta-analysis. *European Journal of Public Health* **18**, 417–421 (2008).

366. Khaing, W., Vallibhakara, S. A., Attia, J., McEvoy, M. & Thakkinstian, A. Effects of education and income on cardiovascular outcomes: A systematic review and meta-analysis. *European Journal of Preventive Cardiology* **24**, 1032–1042 (2017).

367. Putrik, P. *et al.* Lower education and living in countries with lower wealth are associated with higher disease activity in rheumatoid arthritis: results from the multinational COMORA study. *Annals of the Rheumatic Diseases* **75**, 540–546 (2016).

368. Choi, A. I. *et al.* Association of Educational Attainment With Chronic Disease and Mortality: The Kidney Early Evaluation Program (KEEP). *American Journal of Kidney Diseases* **58**, 228–234 (2011).

369. Telfair, J. & Shelton, T. L. Educational Attainment as a Social Determinant of Health. *North Carolina Medical Journal* **73**, 358–365 (2012).

370. Huang, C. Academic Achievement and Subsequent Depression: A Meta-analysis of Longitudinal Studies. *Journal of Child and Family Studies* **24**, 434–442 (2015).

371. Fewell, Z., Davey Smith, G. & Sterne, J. A. C. The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study. *American Journal of Epidemiology* **166**, 646–655 (2007).

372. Sobel, M. E. Causal Inference in the Social Sciences. *Journal of the American Statistical Association* **95**, 647–651 (2000).

373. van der Heide, I. *et al.* The Relationship Between Health, Education, and Health Literacy: Results From the Dutch Adult Literacy and Life Skills Survey. *Journal of Health Communication* **18**, 172–184 (2013).

374. Kawachi, I., Adler, N. E. & Dow, W. H. Money, schooling, and health: Mechanisms and causal evidence. *Ann N Y Acad Sci* **1186**, 56–68 (2010).

375. Glymour, M. M. & Manly, J. J. Compulsory Schooling Laws as Quasi-Experiments for the Health Effects of Education: Revisiting Theory to Understand Mixed Results. *Soc Sci Med* **214**, 67–69 (2018).

376. Chevalier, A. & Feinstein, L. Sheepskin or Prozac: The Causal Effect of Education on Mental Health. *SSRN Electronic Journal* (2006) doi:10.2139/SSRN.923530.

377. Graeber, D. *Does more education protect against mental health problems?* https://www.econstor.eu/handle/10419/169442 (2017).

378. Banks, J. & Mazzonna, F. The Effect of Education on Old Age Cognitive Abilities: Evidence from a Regression Discontinuity Design. *The Economic Journal* **122**, 418–448 (2012).

379. Davies, N. M., Dickson, M., Smith, G. D., Van Den Berg, G. J. & Windmeijer, F. The causal effects of education on health outcomes in the UK Biobank. *Nature Human Behaviour 2017 2:2* **2**, 117–125 (2018).

380. Grenet, J. Is Extending Compulsory Schooling Alone Enough to Raise Earnings? Evidence from French and British Compulsory Schooling Laws*. *The Scandinavian Journal of Economics* **115**, 176–210 (2013).

381. Hamad, R., Nguyen, T. T., Bhattacharya, J., Glymour, M. M. & Rehkopf, D. H. Educational attainment and cardiovascular disease in the United States: A quasi-experimental instrumental variables analysis. *PLOS Medicine* **16**, e1002834 (2019).

382. Hamad, R., Elser, H., Tran, D. C., Rehkopf, D. H. & Goodman, S. N. How and why studies disagree about the effects of education on health: A systematic review and meta-analysis of studies of compulsory schooling laws. *Social Science & Medicine* **212**, 168–178 (2018).

383. Ichino, A. & Winter-Ebmer, R. Lower and upper bounds of returns to schooling: An exercise in IV estimation with different instruments. *European Economic Review* **43**, 889–901 (1999).

384. Frisell, T. Invited Commentary: Sibling-Comparison Designs, Are They Worth the Effort? *Am J Epidemiol* **190**, 738–741 (2021).

385. Davies, N. M. *et al.* Multivariable two-sample Mendelian randomization estimates of the effects of intelligence and education on health. *eLife* **8**, e43990 (2019).

386. Gill, D., Efstathiadou, A., Cawood, K., Tzoulaki, I. & Dehghan, A. Education protects against coronary heart disease and stroke independently of cognitive function: evidence from Mendelian randomization. *International Journal of Epidemiology* **48**, 1468–1477 (2019).

387. Brumpton, B. *et al.* Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat Commun* **11**, 3519 (2020).

388. Howe, L. D. *et al.* Effects of body mass index on relationship status, social contact and socio-economic position: Mendelian randomization and within-sibling study in UK Biobank. *International Journal of Epidemiology* **49**, 1173–1184 (2020).

389. Davies, N. M., Dickson, M., Davey Smith, G., Windmeijer, F. & van den Berg, G. J. The Causal Effects of Education on Adult Health, Mortality and Income: Evidence from Mendelian Randomization and the Raising of the School Leaving Age. *SSRN Electronic Journal* (2021) doi:10.2139/SSRN.3390179.

390. Schimmack, U. The Structure of Subjective Well-Being. in *The Science of Subjective Well-Being* 97–123 (Guilford Press, 2008).

391. Plotnikov, D. *et al.* Effect of Education on Myopia: Evidence from the United Kingdom ROSLA 1972 Reform. *Investigative Ophthalmology & Visual Science* **61**, 7–7 (2020).

392. Hanscombe, K. B., Coleman, J. R. I., Traylor, M. & Lewis, C. M. ukbtools: An R package to manage and query UK Biobank data. *PLOS ONE* **14**, e0214311 (2019).

393. Davies, N. M., Dickson, M., Davey Smith, G., Windmeijer, F. & van den Berg, G. J. The Causal Effects of Education on Adult Health, Mortality and Income: Evidence from Mendelian Randomization and the Raising of the School Leaving Age. IZA Discussion Paper No. 12192 at https://doi.org/10.2139/ssrn.3390179 (2019).

394. Bergé, L. Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm. *CREA discussion papers* **13**, (2018).

395. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

396. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* **15**, 2759–2772 (2020).

397. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics* **50**, 1112–1121 (2018).

398. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).

399. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature 2016 533:7604* **533**, 539–542 (2016).

400. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Smith, G. D. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* **27**, 1133–1163 (2008).

401. Saunders, G. R. B., McGue, M. & Malone, S. M. Sibling Comparison Designs: Addressing Confounding Bias with Inclusion of Measured Confounders. *Twin Res Hum Genet* **22**, 290–296 (2019).

402. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet* **50**, 920–927 (2018).

403. Grossman, M. Chapter 10 Education and Nonmarket Outcomes. in *Handbook of the Economics of Education* (eds. Hanushek, E. & Welch, F.) vol. 1 577–633 (Elsevier, 2006).

404. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology* **186**, 1026–1034 (2017).

405. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G. Collider scope: when selection bias can substantially influence observed associations. *International Journal of Epidemiology* **47**, 226–235 (2018).

406. Oreopoulos, P. & Salvanes, K. G. *How large are returns to schooling? Hint: Money isn't everything*. *NBER Working Papers* https://ideas.repec.org/p/nbr/nberwo/15339.html (2009).

407. Røysamb, E. & Nes, R. B. The role of genetics in subjective well-being. *Nature Human Behaviour* **3**, 3 (2019).

408. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav* **3**, 513–525 (2019).

409. Pingault, J.-B. *et al.* Using genetic data to strengthen causal inference in observational research. *Nat Rev Genet* **19**, 566–580 (2018).

410. Avinun, R. & Hariri, A. R. A polygenic score for body mass index is associated with depressive symptoms via early life stress: Evidence for gene-environment correlation. *Journal of Psychiatric Research* **118**, 9–13 (2019).

411. Machlitt-Northen, S. *et al.* Polygenic scores for schizophrenia and major depression are associated with psychosocial risk factors in children: evidence of gene–environment correlation. *Journal of Child Psychology and Psychiatry* **63**, 1140–1152 (2022).

412. Costanza, R. *et al.* Creating a Wellbeing Economy Alliance (WEAll) to motivate and facilitate the transition. *Sustainable Wellbeing Futures* 399–407 (2020).

413. Büchs, M. *et al.* Wellbeing economics for the COVID-19 recovery: Ten principles for building back better to create wellbeing economies post-covid. in *Wellbeing Economy Alliance Briefing Papers* (2020).

414. Bolier, L. *et al.* Positive psychology interventions: a meta-analysis of randomized controlled studies. *BMC Public Health* **13**, 119 (2013).

415. Carr, A. *et al.* Effectiveness of positive psychology interventions: a systematic review and meta-analysis. *The Journal of Positive Psychology* **16**, 749–769 (2021).

416. White, C. A., Uttl, B. & Holder, M. D. Meta-analyses of positive psychology interventions: The effects are much smaller than previously reported. *PLOS ONE* **14**, e0216588 (2019).

417. Lomas, T. *et al.* Insights from the First Global Survey of Balance and Harmony. in *World Happiness Report* (2022).

418. Kim, S. *et al.* Shared genetic architectures of subjective well-being in East Asian and European ancestry populations. *Nature Human Behaviour* **6**, 1014–1026 (2022).

419. de Vries, L. P., Baselmans, B. M. L. & Bartels, M. Smartphone-Based Ecological Momentary Assessment of Well-Being: A Systematic Review and Recommendations for Future Studies. *Journal of Happiness Studies* **22**, 2361–2408 (2021).

SUMMARY OF AUTHOR CONTRIBUTIONS

Chapter 1

Chapter 1 was partly based on three articles: 1) Van de Weijer, M., de Vries, L., & Bartels, M. (2022). Happiness and Wellbeing; the value and findings from genetic studies. In Tarnoki, A., Tarnoki, D. Harris, J. & Segal, S. (2022) *Twin Research for Everyone*. Academic Press, 2) Bartels, M., Nes, R. B., Armitage, J. M., van de Weijer, M. P., de Vries, L. P., & Haworth, C. (2022). Exploring the biological basis for happiness. In Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2022*). World Happiness Report 2022.* New York: Sustainable Development Solutions Network, and 3)Van de Weijer, M., Baselmans, B., van der Deijl, W., & Bartels, M. (2018). A growing sense of well-being: a literature review on the complex framework well-being. DOI: 10.31234/osf.io/3rmx9

All authors contributed to writing these three papers.

Chapter 2

Chapter 2 was based on Van de Weijer, M.P., Landvreugd, A., Pelt, D. H. M., & Bartels, M. (*submitted for publication*). Connecting the dots: Using a network approach to study the well-being spectrum.

The study was designed by all authors. The article was written by M.P van de Weijer and A. Landvreugd in collaboration with D. Pelt and M. Bartels. Analyses were performed by M.P van de Weijer and A. Landvreugd.

Chapter 3

Chapter 3 was based on Van De Weijer, M. P., Pelt, D. H. M., De Vries, L. P., Baselmans, B. M. L., & Bartels, M. (2022). A Re-evaluation of Candidate Gene Studies for Well-Being in Light of Genome-Wide Evidence. *Journal of Happiness Studies*, 1-23.

The study was designed by M.P. van de Weijer, L.P. de Vries, B.M.L. Baselmans and M. Bartels. The article was written by M.P. van de Weijer in collaboration with the other co-authors. The systematic review and analyses were performed by M.P. van de Weijer.

Chapter 4

Chapter 4 was based on Van De Weijer, M. P., Baselmans, B. M. L., Hottenga, J. J., Dolan, C. V., Willemsen, G., & Bartels, M. (2022). Expanding the environmental

scope: an environment-wide association study for mental well-being. *Journal of exposure science & environmental epidemiology*, 32(2), 195-204.

The study was designed by M.P. van de Weijer, B.M.L. Baselmans and M. Bartels. The data were prepared by G. Willemsen (GECCO data) and J.J. Hottenga (genetic data). Analyses were performed by M.P. van de Weijer, B.M.L. Baselmans, and C.V. Dolan. The article was written by M.P. van de Weijer in collaboration with all other co-authors.

Chapter 5

Chapter 5 was based on Van De Weijer, M.P., Pelt, D.H.M., Baselmans, B.M.L., Ligthart, L., Huider, F., Hottenga, J.-J., Pool, R. & Bartels, M. Capturing the well-being exposome in poly-environmental scores (*manuscript in preparation*).

The study was designed by M.P. van de Weijer, D.H.M Pelt and M. Bartels. Data were collected by M.P. van de Weijer and F. Huider. The data were cleaned by L. Ligthart. The PGSs for NTR were calculated by R. Pool. Analyses were performed by M.P. van de Weijer. The article was written by M.P. van de Weijer in collaboration with all co-authors.

Chapter 6

Chapter 6 was based on Van De Weijer, M. P., de Vries, L. P., Pelt, D. H. M., Ligthart, L., Willemsen, G., Boomsma, D. I., de Geus, E.J.C. & Bartels, M. (2022). Self-rated health when population health is challenged by the COVID-19 pandemic; a longitudinal study. *Social Science & Medicine*, 306, 115156.

The study was designed by M.P. van de Weijer, L.P. de Vries, D.H.M. Pelt and M. Bartels. Data were cleaned by L. Ligthart. Funding for NTR COVID-19 research was acquired by G. Willemsen, D.I. Boomsma, E.J.C. de Geus and M. Bartels. Analyses were performed by M.P. van de Weijer, in collaboration with L.P. de Vries. The article was written by M.P. van de Weijer in collaboration with all co-authors.

Chapter 7

Chapter 7 was based on van de Weijer, M. P., Pelt, D. H. M., de Vries, L. P., Huider, F., van der Zee, M. D., Helmer, Q., Ligthart, L., Willemsen, G., Boomsma, D.I., de Geus , E.J.C. & Bartels, M. (2022). Genetic and environmental influences on quality of life: The COVID-19 pandemic as a natural experiment. *Genes, Brain and Behavior,* e12796.

The study was designed by M.P. van de Weijer, L.P. de Vries, D.H.M. Pelt and M. Bartels. Data were cleaned by L. Ligthart. Funding for NTR COVID-19 research was acquired by G. Willemsen, D.I. Boomsma, E.J.C. de Geus and M. Bartels. Q. Helmer prepared the pedigree files, and M.D. van der Zee created the kinship correlation generation tool. The analyses were performed by M.P. van de Weijer in collaboration with F. Huider. The article was written by M.P. van de Weijer in collaboration with all co-authors.

Chapter 8

Chapter 8 was based on Van De Weijer, M. P., Pelt, D. H. M., Van Beijsterveldt, C. E., Willemsen, G., & Bartels, M. (2022). Genetic factors explain a significant part of associations between adolescent well-being and the social environment*. European child & adolescent psychiatry,* 31(10), 1611-1622.

The study was designed by M.P. van de Weijer, D.H.M. Pelt and M. Bartels. The data were cleaned by C.E. van Beijsterveldt and G. Willemsen. Analyses were performed by M.P. van de Weijer. The article was written by M.P. van de Weijer in collaboration with all co-authors.

Chapter 9

Chapter 9 was based on Van de Weijer, M.P., Demange, P.A.D., Pelt, D. H. M., Bartels, M., & Nivard, M. G. (2022). Disentangling potential causal effects of educational duration on well-being, and mental and physical health outcomes (*under revision*).

The study was designed by M.G. Nivard, in collaboration with M.P. van de Weijer. The analyses were performed by M.P. van de Weijer and P.A.D. Demange, in collaboration with M.G. Nivard. The paper was written by M.P. van de Weijer, in collaboration with all co-authors.

LIST OF PUBLICATIONS

Bartels, M., Bang Nes, R., Armitage, J.M., **van de Weijer, M.P.,** de Vries, L.P. & Haworth, C.M.A. (2022). Exploring the biological basis for happiness. World Happiness Report, 105-126.

Baselmans, B. M., Jansen, R., Ip, H. F., van Dongen, J., Abdellaoui, A., **van de Weijer, M. P**., ... & Hottenga, J. J. (2019). Multivariate genome-wide analyses of the well-being spectrum. *Nature genetics*, 51(3), 445-451.

Baselmans, B. M. L., **van de Weijer, M. P.**, Abdellaoui, A., Vink, J. M., Hottenga, J. J., Willemsen, G., ... & Bartels, M. (2019). A genetic investigation of the well-being spectrum. *Behavior genetics,* 49(3), 286-297.

Beck, J. J., Pool, R., **Van De Weijer, M.P.**, Chen, X., Krapohl, E., Gordon, S. D., ... & Hottenga, J. J. (2021). Genetic meta-analysis of twin birth weight shows high genetic correlation with singleton birth weight. *Human molecular genetics*, 30(19), 1894-1905.

Lambert, L., Lomas, T., **van de Weijer, M. P.**, Passmore, H. A., Joshanloo, M., Harter, J., ... & Kawakami, T. (2020). Towards a greater global understanding of wellbeing: A proposal for a more inclusive measure. *International Journal of Wellbeing*, 10(2).

Lomas, T., Bartels, M., **Van De Weijer, M.P.**, Pluess, M., Hanson, J., & VanderWeele, T. J. (2022). The Architecture of Happiness. *Emotion Review*, 17540739221114109.

Mbarek, H., **van de Weijer, M. P.,** van der Zee, M. D., Ip, H. F., Beck, J. J., Abdellaoui, A., ... & Bartels, M. (2019). Biological insights into multiple birth: genetic findings from UK Biobank. *European Journal of Human Genetics*, 27(6), 970-979.

de Vries, L. P., **van de Weijer, M. P.,** & Bartels, M. (2022). The human physiology of well-being: A systematic review on the association between neurotransmitters, hormones, inflammatory markers, the microbiome and well-being. *Neuroscience & Biobehavioral Reviews,* 104733.

de Vries, L. P., **van de Weijer, M. P**., & Bartels, M. (2023). A systematic review of the neural correlates of well-being reveals no consistent associations. *Neuroscience & Biobehavioral Reviews*, 105036.

de Vries, L. P., **van de Weijer, M. P**., Ligthart, L., Willemsen, G., Dolan, C. V., Boomsma, D. I., ... & Bartels, M. (2020). A Comparison of the ASEBA Adult Self Report (ASR) and the Brief Problem Monitor (BPM/18-59). *Behavior Genetics*, 1-11.

de Vries, L. P., **van de Weijer, M. P.,** Pelt, D. H. M., Ligthart, L., Willemsen, G., Boomsma, D. I., ... & Bartels, M. (2022). Gene-by-Crisis Interaction for Optimism and Meaning in Life: The Effects of the COVID-19 Pandemic. *Behavior genetics*, 52(1), 13-25.

van de Weijer, M. P., Baselmans, B. M. L., Hottenga, J. J., Dolan, C. V., Willemsen, G., & Bartels, M. (2021). Expanding the environmental scope: an environment-wide association study for mental well-being. *Journal of exposure science & environmental epidemiology*, 1-10.

van de Weijer, M.P., de Vries, L.P. & Bartels, M. (2021). Les influences génétiques et environnementales sur le bonheur et le bien-être. In Martin-Krumm, C., & Tarquinio, C. (Eds.), Grand manuel de psychologie positive: Fondements, théories et champs d'intervention (pp 65-91). Dunod.

van de Weijer, M. P., de Vries, L. P., Pelt, D. H. M., Ligthart, L., Willemsen, G., Boomsma, D. I., ... & Bartels, M. (2022). Self-rated health when population health is challenged by the COVID-19 pandemic; a longitudinal study. *Social Science & Medicine*, 306, 115156.

van de Weijer, M. P., de Vries, L. P., & Bartels, M. (2022). Happiness and well-being: The value and findings from genetic studies. In Twin Research for Everyone (pp. 295-322). Academic Press.

van de Weijer, M. P., Jansen, I. E., Verboven, A. H., Andreassen, O. A., & Posthuma, D. (2020). Genomics of Alzheimer's disease. Personalized Psychiatry, 275-283.

Van De Weijer, M. P., Pelt, D. H. M., de Vries, L. P., Baselmans, B. M., & Bartels, M. (2022). A Re-evaluation of Candidate Gene Studies for Well-Being in Light of Genome-Wide Evidence*. Journal of Happiness Studies,* 1-23.

Van de Weijer, M. P., Pelt, D. H. M., de Vries, L. P., Huider, F., van der Zee, M. D., Helmer, Q., ... & Bartels, M. (2022). Genetic and environmental influences on quality of life: The COVID-19 pandemic as a natural experiment. *Genes, Brain and Behavior*, e12796.

Van De Weijer, M. P., Pelt, D. H. M., Van Beijsterveldt, C. E., Willemsen, G., & Bartels, M. (2021). Genetic factors explain a significant part of associations between adolescent well-being and the social environment. *European child & adolescent psychiatry*, 1-12.

Van de Weijer, M.P., Demange, P.A.D., Pelt, D.H.M., Bartels, M., & Nivard, M. G. (*Under review*). Disentangling potential causal effects of educational duration on well-being, and mental and physical health outcomes.

Van de Weijer, M.P., Landvreugd, A., Pelt, D. H. M., & Bartels, M. (*under review*). Connecting the dots: Using a network approach to study the well-being spectrum.

Van De Weijer, M.P., Pelt, D. H. M., Baselmans, B. M. L., Ligthart, L., Huider, F., Hottenga, J.-J., Pool, R. & Bartels, M. (*in preparation*). Capturing the well-being exposome in poly-environmental scores.

DANKWOORD

Een proefschrift wordt zeker niet door één iemand mogelijk gemaakt, maar door samenwerkingen en steun van enorm veel mensen die ik graag wil bedanken. Dit dankwoord zal een mix van Nederlands en Engels zijn afhankelijk van wie ik bedank en dus waarschijnlijk aardig chaotisch worden – excuses!

Allereerst bedank ik graag mijn promotor en copromotoren die het 4+ jaar met me hebben volgehouden. Lieve **Meike**, zonder jou was dit proefschrift er niet geweest maar was ik ook zeker een ander persoon geweest. We werken al sinds ik aan mijn master begon aan well-being onderzoek en ik heb daar altijd enorm van genoten! Ik bedank je graag voor je lessen en hulp als promotor, maar ook vooral voor alle gezelligheid. Het is superfijn om een vrouwelijk rolmodel in de wetenschap te hebben die veel moeite steekt in het kansen creëren voor haar promovendi. Naast het wetenschappelijke werk kunnen we ook gezellig kletsen, lachen en borrelen. Een groot deel van mijn PhD was helaas online door de pandemie, maar door de extra moeite die jij stak in het organiseren van team meetings en het controleren of het wel goed met ons ging werd deze periode een stuk minder zwaar. We zullen elkaar in de toekomst ongetwijfeld nog veel tegenkomen en ik heb veel zin in alle toekomstige samenwerkingen (en op alle kerstborrels waar ik vrees dat ik in de toekomst weggeslagen zal moeten worden).

Bart, wij werken inmiddels ook al 7+ jaar samen in afwisselende frequentie! Al sinds mijn masterstage heb ik enorm veel van jou geleerd als day-to-day supervisor, van LDpred (gaat nog steeds fout) tot hoeveel barretjes je op een avond kan afgaan in Melbourne (te veel). Jouw rol als copromotor was een rollercoaster waarbij je na een jaar naar Australië verhuisde en 2 jaar later weer terugkwam maar op allerlei verschillende plekken werkte, maar gelukkig maakte je altijd tijd vrij voor mijn promotie! Naast gedeelde wetenschappelijke interesses (zowel op studiegebied als werkgebied) kunnen we het ook gewoon hartstikke goed met elkaar vinden waardoor ik zeker weet dat er in de toekomst nog veel etentjes zullen zijn waar we één wijntje gaan drinken ;).

Dirk, waar Bart 1 jaar na de aanvang van mijn promotie de biopsy verliet, begon jij juist 1,5-2 jaar na het begin van mijn promotie. Ook al was dit midden in de corona-pandemie, was ik super blij dat je gelijk kon inspringen als copromotor. Het was snel duidelijk dat je heel goed in ons team en op de afdeling past en we hebben altijd superfijn samengewerkt. Je kwam uit een ander veld, maar ik heb

nog nooit zo snel iemand zich ingewikkelde tweelingmodellen zien eigen maken en je bent altijd bezig met nieuwe leuke toevoegingen voor papers te bedenken. In het laatste jaar van mijn PhD zijn we als hoogtepunt nog naar LA BGA geweest wat enorm gezellig was. Ook wij zullen elkaar zeker nog zien in de toekomst!

In addition, I would like to thank the members of my reading committee: **prof.dr. Marit Sijbrandij, prof.dr. Gonneke Stevens, prof.dr. Christiaan Vinkers, dr. Marco Helbich, dr. Yayouk Willems & dr. Bruno Sauce** - thank you for reading and evaluating my thesis. Daarnaast bedank ik ook graag alle deelnemers van het Nederlands Tweelingen Register, zonder wiens bijdragen dit proefschrift zeker niet mogelijk was geweest.

Dan bedank ik graag twee hele belangrijke collega's en vriendinnen, mijn paranimfen Lianne en Zenab. **Lianne**, wij zijn ongeveer tegelijk begonnen als well-being AIO's en zijn in de afgelopen jaren een soort academische zussen geworden. We hebben heel veel samengewerkt aan projecten en dat ging altijd super goed en makkelijk: ik kan me niet herinneren dat we ooit ook maar een seconde discussie hebben gehad over hoe iets moest lopen! Daarnaast hebben we het ook gewoon heel erg gezellig gehad als kamergenootjes met heel veel gezamenlijke liefde voor kara-eco en kerst en kletsen. Ik ben zo dankbaar dat we onze PhDs samen kunnen afsluiten als elkaars paranimf en heb zin in het samen meemaken van alle volgende stappen die gaan meemaken! **Zenab**, direct na jouw sollicitatiegesprek bij de afdeling leerde ik je kennen bij de basket en ik wist vanaf seconde 1 dat wij meant-to-be als vriendinnen waren. Je bent niet alleen heel erg slim maar ook super sociaal en komt op voor jezelf en anderen. We hebben ontelbaar vaak tot diep in de nacht gedanst, gezongen, filosofische gesprekken gehad, gelachen en gehuild. Bedankt voor het me er altijd door heen slepen en het altijd aanbieden van een luisterend oor en advies.

Naast mijn (co)promotoren en paranimfen bedank ik graag al mijn (oud)collega's bij de afdeling biologische psychologie. **Michel**, je was niet mijn copromotor maar volgens mij fungeer jij als onofficiële copromotor bij ongeveer elk promotietraject. Iedereen weet dat je voor goed advies (over wetenschap, maar eigenlijk over alles) bij jou terecht kan en daar heb ik ook zeker goed gebruik van gemaakt. Ik denk dat zonder ons de doppio op campus waarschijnlijk failliet was gegaan aangezien we ongeveer 800 keer koffie hebben gehaald. Bedankt voor alle koffiewandelingen,

al het lachen, advies en kerstborrel escalaties. **Natascha,** als jij op de afdeling bent is het altijd gezellig. Onze relatie is die van kerstvrouw-kerstelf en moeder-dochter, en zonder jou was alles een stuk minder leuk geweest! **Michiel,** de onzichtbare kracht achter veel belangrijks op de afdeling, zoals bijvoorbeeld het hele land rondrijden zodat iedereen een kerstpakket heeft! Je neemt altijd de tijd voor iedereen en doet het ook nog met een lach. **Eco**, ik altijd het gevoel gehad dat jouw deur open stond voor advies en vond het daarnaast ook heel gezellig dat je altijd (actief) van de partij bent bij de vrijmibo, afdelingsfeesten en promotie-vieringen, dankjewel. **Dennis**, als b-these begeleider was jij de eerste waarmee ik in contact kwam op de afdeling en daarna ben ik eigenlijk nooit meer weggegaan. Eén ding is zeker: jij krijgt die lach niet van mijn gezicht! **Gonneke**, altijd de eerste op de afdeling en dus de eerste om even mee te kletsen vroeg in de ochtend. Bedankt voor de prettige samenwerkingen en gezellige momenten! **Conor,** bedankt voor het altijd de tijd nemen voor het antwoorden van ingewikkelde vragen waar niemand anders het antwoord op heeft, maar ook voor het opfleuren van de afdeling met verse bloemen en thee elke week. **Dorret,** bedankt voor de tijd als URF waar ik het genoegen had meer te leren over twinning en birthweight. **Jouke-Jan, René P, René N, Quinta, Cyrina, Lannie** en **Toos**, met jullie heb ik afwisselend samengewerkt aan projecten en data-verzameling: jullie zijn niet alleen absoluut essentieel geweest voor mijn projecten maar voor het gehele NTR! **Elsje en Eveline,** mijn eerste aanraking met de wetenschap was bij jullie als student-assistent. Bedankt voor deze tijd en de jaren daarna als collega's! **Jenny,** bedankt voor altijd op een super kalme en vriendelijke manier bereid te zijn dingen uit te leggen aan mij en aan alle andere promovendi/collega's. **Anouk**, helaas hebben we elkaar de afgelopen jaren een stuk minder gezien door het thuiswerken en doordat jij natuurlijk op 2 afdelingen werkzaam bent, maar dat maakte het echter niet minder gezellig als we er allebei wel waren! **Bruno,** not only are you in my committee, we are also biopsy colleagues. You are well known for always being up for helping people and engaging in interesting discussions but also crazy conversations, thank you for being a great colleague. **Martin,** op werkgebied hadden wij eigenlijk niet zo veel met elkaar te maken, maar we hebben wel heel veel leuke lunches en borrels meegemaakt! Aan het einde van mijn promotietraject heb je de belangrijke rol van PhD candidate advisor op je genomen, en van wat ik daar nog van heb meegekregen is dit een rol die je op het lijf geschreven is!

My TTT friends and colleagues **Wonu, Perline**, **Sofieke** and **Zenab,** thank you for the endless chatting and banter on and off work, the movie nights, "the chin

touch", the emotional support, the tiktoks and tweets, the insane (harry styles/ F1/taylor swift/etc.) fandom, the karaeco, and all the other 1000 things that made the past four years so special. We have had (and will continue to have) a lot of fun together, but I've also learned so much from you guys that have made me a different person compared to before I met all of you! My MF roommates, **Lianne, Anne, Fiona, Selim, Bodine, Hekmat** and **Susanne:** it is not easy to share a room with so many people and still do some actual work, but I think we managed quite well! It was a lot of fun to share an office together, especially around Christmas. My (ex-)PhD colleagues from the "other PhD room/ballroom", **Sofieke, Zenab, Nicole, Eshim, Wonu**, **Matthijs, Perline, Floris, Veronika, Sjors, Zoey, Camiel, Nikki, and Denise,** I am sorry for barging in and distracting all of you from time to time and want to thank you for the great times on and off work!

Anne en Yayouk, het is al een tijd geleden sinds we collega's waren, maar ooit ben ik bij jullie op de kamer in het transitorium begonnen en zijn we daarna samen verhuist naar het MF-gebouw. Onze tijd als roomies was voor mij super bijzonder en het is moeilijk om dit kort samen te vatten, maar hier wat highlights: de beroemde snoep-pot, work-out breaks van 1 minuut planken (never forget de oerkreet), koffie momentjes en wandelingen, kledingruilen, vrijdagmiddag kara-eco, de cocos en nog veel meer. Ik heb daarnaast ook heel veel van jullie geleerd waardoor mijn PhD een stuk prettiger was dan als ik jullie niet had gekend. Bedankt voor alles! Niet te vergeten als ik het over koffie wandelingen heb is natuurlijk **Lisette,** mijn "nicht". Ook jij bent helaas al een paar jaar weg maar ook wij hebben gelukkig nog contact! Bedankt voor alle gezelligheid als "OG-koffiedame". **Hill,** ook voor ons is het een tijdje geleden dat we samen in een kamer op het MF zaten, maar ook hier kijk ik naar terug als een superleuke tijd! Bedankt voor alle hulp, basket hang-outs, al het lachen en natuurlijk de 1 april pranks en al het snoep. **Matthijs,** zonder jou was mijn tijd op de afdeling een stuk minder gezellig geweest! Ik denk niet dat veel mensen een collega hebben die zowel game of thrones als temptation island poules opzetten, mario kart toernooien organiseren en daarnaast zichzelf ook (onder lichte dwang) opofferen als borrelhoofd. Daarnaast hebben wij met zijn tweeën heel veel gelachen tijdens biertjes bij de basket, over rare video's op slack en over no-look high-fives. Bedankt voor alles! Ook is er een groepje "oude/ nieuwe " collega's, **Jorien, Karin, Laura, Dirk, Abdel, Melanie** en volledig nieuwe collega's **Rada, David, Anaïs, Eva en Shu,** die ik graag bedank voor de warme verwelkoming bij het AMC. Vooral **Jorien**, waarmee ik het genoegen had samen de

GENE Amsterdam dag te organiseren en waarmee ik nu superleuke projecten aan het starten ben! Lastly, I would like to thank all my fellow PhD council colleagues, **Fiona, Rick, Perline, Nicole, and Nadia**, but also especially **Timothy and Nina**, for not only putting effort into this council, but also in setting up the whole council!

Naast mijn collega's bedank ik graag ook mijn vrienden voor hun support tijdens mijn PhD. Normaal gesproken hebben je vrienden niet heel veel te werken met je werk, maar doordat de helft van mijn promotietraject tijdens de pandemie was heb ik met een deel van jullie zelfs nog best veel samengewerkt! De soju "meiden" (waarom hebben alle groupschats zo'n vreemde naam?), **Vera**, **Ipek, Iris** en **Lisette**, die altijd klaar staan voor vakanties, etentjes, en algehele escalatie. Maar ook de rest van de Belly group, **Floris, Yanick, Mariam, Marc, Mariana, Robin, Felicien, Hinde, Francis en Benny** voor de onvergetelijke vakanties, feestjes en het altijd klaar staan op vrijdag voor de vrijmibo. **Fenna en Donna**, 16 jaar geleden leerden we elkaar kennen op het HWC, en sindsdien zijn we vriendinnen. Als we met zijn drieën zijn is het gegarandeerd gezellig en lachen, en ook al zijn we in al die jaren veel verandert, lijkt het als we samen zijn weer even net alsof we 16 zijn ;). **Tim en Wesley**, ook wij kennen elkaar al sinds de HWC-tijden. Sinds die tijd hebben we veel op het terras gezeten, op de bank gehangen, wandelingen in west gemaakt en zeker ook veel gelachen. **Lynn en Reena**, als Badhoevedorpers tegelijk aan de bachelor begonnen en daarna alle drie een andere richting op gegaan, maar nog steeds zien we elkaar geregeld (het liefst terug naar onze roots in Badhoevedorp natuurlijk). Als wij onszelf niet 4 keer per jaar een week hadden opgesloten op de VU en samen naar alle tentamens waren gegaan dan had ik waarschijnlijk de bachelor niet eens gehaald! **Soesja**, samen opgegroeid in de Reigerstraat en samen dezelfde obsessies meegemaakt door de jaren heen: Twilight, Valerio, Harry Potter, Taylor Swift, DDD etc. (sommige van deze obsessies zijn losgelaten, andere… wat minder). Ik ben heel dankbaar dat we nog steeds samen kunnen fangirl'en! **Merel,** ook oud-collega's, maar dan van de vomar! Van zondag ochtend op 16-jarige leeftijd om 6 uur s'ochtends giechelend bij het brood tot 10+ jaar later wijntjes doen op het terras. Bedankt voor alles.

Als laatst bedank ik graag mijn lieve familie en schoonfamilie. **Jos en Annette,** die mij meer dan 8 jaar geleden met open armen hebben verwelkomd in de familie en altijd klaar staan voor hulp als nodig en voor gezellig samen eten en drinken. Bedankt voor de afgelopen jaren en ik heb zin in de aankomende jaren! Hetzelfde

geldt voor **Mandy en Ruben**, die voor iedereen zorgzaam zijn en klaar staan om te helpen met van alles. **Bob en An,** ook jullie bedank graag voor de bijzondere herinneringen de afgelopen jaren. Ook al zal An dit proefschrift helaas niet meer zien, hebben jullie allebei een speciaal plekje in mijn hart.

Paul en Bernadette, die niet echt familie zijn maar wel tellen als familie: jullie zijn altijd een beetje mijn bonus ouders geweest en ik ben jullie heel dankbaar voor de speciale jeugd (en gezellige etentjes die we nog steeds hebben. Mijn lieve en gekke zusjes, Eva, Iris en Linda. **Eva,** jouw speciale talent is iedereen op zijn/haar gemak laten voelen en overal bij betrekken. Bedankt voor het er altijd voor zorgen/controleren dat het goed met me gaat. **Iris,** als jij ergens binnenloopt gaat de zon schijnen en wordt iedereen een stukje vrolijker. Bedankt voor het gezellig maken van alles en me altijd opfleuren! **Linda,** jij ziet de wereld op zo'n creatieve manier waardoor jij altijd oplossingen en ideeën ziet waar anderen dat niet zien. Bedankt voor het altijd inspireren en natuurlijk voor het ontwerpen van mijn proefschrift. Ook bedank ik natuurlijk graag jullie lieve vriendjes, **Steven**, **Thom** en **Sebas**, die soms heel wat moeten verduren in zo'n drukke en gekke familie maar hier helemaal een deel van zijn geworden. Mijn ouders, **Teus en Anna**. Een boodschap die ik altijd van jullie heb meegekregen is dat het niet uitmaakt als iets niet lukt, als je maar je best doet – een les die me altijd is bijgebleven en heeft geholpen! Jullie hebben ons altijd vooropgesteld en ervoor gezorgd dat we alle kansen hebben gekregen die we nodig hadden om alle 4 te komen waar we zijn. **Pap**, jouw discipline en doorzettingsvermogen heeft mij altijd geïnspireerd en als ik daar ook maar 10% van heb meegekregen kom ik al een heel eind. **Mam**, niemand weet zo goed als jij hoe je het gezellig moet maken, van gekke kinderverjaardagen tot thema diners tot creatieve cadeaus. Ik hoop dat ik ook zo out-of-the-box leer denken als jij. Allebei bedankt voor alles in het verleden, heden en de toekomst!

Last maar zeker not least, **Joey**. Samen zijn met iemand die bezig is met een PhD is niet altijd even gemakkelijk of gezellig, maar jij bent me altijd blijven steunen en stimuleren. Jij hebt vaak nog meer vertrouwen in mij dan ik in mezelf, en dit heeft mij niet alleen enorm geholpen maar er ook zeker voor gezorgd dat dit proefschrift bestaat. Je hebt zelf heel veel doorzettingsvermogen, creatief denkvermogen en talent voor het inspireren van niet alleen mij maar van veel anderen. Ik ben je heel dankbaar voor alles wat ik van je heb geleerd en alle bijzondere momenten die we hebben gehad de afgelopen 8 jaar, en kijk uit naar alle avonturen in de toekomst. Ik hou van je.