

This FAQ provides information about **“Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction.”** This FAQ is a live document that might be updated in response to feedback from the scholarly community, journalists, and members of the public.

## Whom do I contact if I have more questions?

This paper was written by the [Externalizing Consortium](#), an international team of scientists that is led by Dr. Danielle Dick, Dr. Paige Harden, Dr. Philipp Koellinger, and Dr. Abraham Palmer. The members of the Externalizing Consortium have different areas of expertise, so questions and comments about the paper or this FAQ should be directed [as follows](#):

Dr. [Danielle Dick](#): General inquiries about the consortium, Substance Use and Disorder/Addiction, Childhood behavior problems, Impulsivity

Dr. [Philipp Koellinger](#): Methodological inquiries/data requests, Economic Behaviors and Outcomes

Dr. [Paige Harden](#): Sexual and reproductive behavior, antisocial behavior and crime, ethical issues in social science genetics

Dr. [Abraham A. Palmer](#): Biological basis of behavior, medical outcomes, substance abuse, impulsivity

For the particularly interested reader, this FAQ contains links to articles from the scholarly literature, which can be inaccessible to people without a university library subscription. Please contact [edgelab@vcu.edu](mailto:edgelab@vcu.edu) if you need assistance accessing one of the cited papers.

## What is externalizing?

“Externalizing” refers to a spectrum of behaviors and disorders related to impulse control. It is a word that psychologists have been using for decades. [In the 1960s](#), a psychologist named Thomas Achenbach conducted a study of different types of emotional and behavioral problems observed in children. On the basis of this study, he coined the terms “internalizing” and “externalizing.” Internalizing referred to problems that children experienced internally, such as feeling sad, worried, or anxious. Externalizing, on the other hand, referred to problems that children manifested externally, such as getting into fights or breaking rules at school or at home.

One of Achenbach’s insights was that children who showed one type of externalizing problem were more likely to show others. His approach to understanding child psychology was different than the typical psychiatry approach – still predominant today -- which focuses on diagnosing individuals with discrete disorders. In the typical psychiatric approach, people are classified as either having a disorder or not. Take, for instance, a disorder such as Attention Deficit Hyperactivity Disorder (ADHD). If you take your child to a doctor to be assessed for ADHD, the doctor will compare your child’s behaviors to a checklist of symptoms and determine whether your child qualifies for an ADHD diagnosis. Achenbach’s insight was

that children’s problems didn’t fit into tidy diagnostic boxes. Instead, children’s behavior varied along a continuum.

It has now been widely demonstrated that this distinction between externalizing and internalizing behaviors doesn’t just apply to children – it also applies to [adult psychiatric diagnoses](#). Just like in children, internalizing disorders like depression and anxiety tend to co-occur in adults. Similarly, externalizing disorders co-occur. Externalizing disorders include childhood behavior disorders like ADHD and Conduct Disorder, as well as different forms of [Substance Use Disorders](#) and [Antisocial Behavior](#) in adolescents and adults. In addition, [personality features related to impulsivity and behavioral undercontrol](#) are also related to the externalizing spectrum. These traits vary continuously in the population. The word “externalizing” thus refers to a *constellation of behaviors and psychiatric symptoms that co-occur across the lifespan and that involve difficulties with self-regulation and impulse control*.

## What did this study do?

This study is based on method called a “[genome-wide association study](#),” or **GWAS** (pronounced JEE-wahs). Your *genome* is the complete sequence of DNA that you have in all of the cells in your body. A DNA strand is made up of a unique sequence of DNA “letters” — abbreviated as G, C, T, and A. People can differ in single DNA letters. For example, at a particular spot in the genome, you might have an A, whereas someone else has a C. These differences in single DNA letters are called “single nucleotide polymorphisms,” or *SNPs* (pronounced “snips.”). SNPs are one type of *genetic variant*, or DNA difference among people. A [GWAS](#) measures millions of SNPs and correlates each one with the characteristic that is being studied. Typically, each SNP in a GWAS shows only a very small association with the characteristic being studied. If, for example, one is studying diabetes, a GWAS tests which SNPs are more common in people who have diabetes compared to people who don’t have diabetes. A GWAS is a correlational study, which means that it does not necessarily identify SNPs that cause the outcome that is being studied. It only identifies SNPs that are more common in one group of people versus another.

GWAS is a method that has been applied to study thousands of *phenotypes*, which are measured characteristics of a person. In this study, we chose seven phenotypes that previous research indicates are related to externalizing. One is a disorder of childhood – (1) Attention Deficit Hyperactivity Disorder. Three are substance use behaviors – (2) lifetime history of smoking, (3) lifetime history of cannabis use, and (4) problematic alcohol use. One is a personality characteristic called (5) “risk tolerance,” which is simply whether or not someone describes themselves as someone who likes to take risks. And the last two phenotypes are aspects of people’s sexual behavior – (6) how old they were when they first had sexual intercourse and (7) how many sexual partners they’ve had. We included these sexual behaviors because [previous research](#) has found that sexual behavior is genetically correlated with other aspects of externalizing, such as delinquent behavior and substance use.

We then used a method called [Genomic Structural Equation Modeling](#) to pool information about these seven phenotypes into what psychologists call a *latent factor*. (“Latent” means unobserved, whereas “Factor” is a technical term for how the co-occurrence between different measured variables is represented statistically.) The result is an estimate of how correlated each SNP is with a general tendency toward externalizing.

Pooling information about many phenotypes using the method Genomic Structural Equational Modeling increases the *statistical power* of the analysis, thereby allowing us to estimate even tiny associations quite precisely and enabling us to identify at least some of the many SNPs that are related to externalizing. Because the different externalizing phenotypes are influenced by many of the same genetic variants, samples collected based on one phenotype (e.g., ADHD in children) also carry information about genes that influence other externalizing phenotypes (e.g., Alcohol Use Disorder in adolescents or adults). Genomic Structural Equational Modeling aggregates this information across different phenotypes. By using this approach, our analysis is able to detect much smaller effects and many more SNPs than would be possible if we were studying just one phenotype. One way of expressing statistical power is in terms of *sample size* – how many people are we effectively studying? Our effective sample size is 1.5 million people, which makes it one of the largest GWAS ever conducted.

## Why study the genetics of externalizing?

The ultimate goal of all scientific research is to increase understanding of the world around us and to improve human lives. Externalizing disorders and behaviors are associated with [profound consequences](#) for affected individuals, their families, communities, and society at large. Accordingly, psychologists, educators, policymakers, and parents are interested in finding ways to reduce externalizing problems in both children and adults. Unfortunately, most existing psychological interventions on externalizing problems are not very effective. For example, [one review of psychological interventions](#) designed to reduce teenagers' externalizing behavior concluded that, "Even the best programs are successful at changing adolescents' knowledge but not at altering their behavior." The review went on to note that failure isn't free: "Most taxpayers would be surprised — and rightly angry — to learn that vast expenditures of their dollars are invested in ... programs that either do not work ... or are, at best, of unproven or unstudied effectiveness."

In order to design more effective interventions and policies, we must have a better scientific understanding of what actually *causes* externalizing problems. We know from previous research that externalizing problems are influenced by both genetic and environmental factors. The underlying liability to externalizing has a heritability of up to [80%](#), meaning that a large part of the reason that people differ from one another in their predisposition to develop ADHD, alcohol or other drug use disorders, or other risky behaviors related to externalizing is due to differences in their DNA. Heritability does not mean that, for example, 80% of the reason any one individual develops an externalizing disorder is because of their genes; it means that a large part of the *differences between people* in how likely they are to show externalizing problems is due to *differences between their DNA*. In this way, heritability is about variation across a population – how much does variation in people's genotypes correlate with variation in outcomes. A particular outcome can be more or less heritable in different environments. For example, imagine a society where alcohol is completely unavailable. It wouldn't matter what your genetic liability to alcohol use disorder is, the likelihood of developing alcohol problems would be determined entirely by the environment (in this case, eliminated, so the heritability would be 0 since it would not matter what genotype people had). This is why it is so important to study both genes and the environment.

But, most interventions for externalizing disorders focus entirely on environmental factors, while ignoring genetic influences. These genetic differences are treated as "noise" in research designed to detect

environmental “signal.” Studies like the one that we conducted give future researchers a new tool to measure and take into account genetic differences among individuals, which – perhaps unintuitively – may make it easier to detect environmental causes that can be targeted to design more effective prevention and intervention strategies.

Additionally, the tool that we used here – GWAS – is widely applied in the field of medicine to discover genes that make people more vulnerable to diseases like cancer or cardiovascular disease. The same strategies are now being applied to [find genes involved in why some people are more at risk for psychiatric and substance use disorders](#). This can help us understand the underlying biology related to the development of devastating disorders like opioid addiction, and this knowledge can aid in designing better treatments. As we show in this paper, many of the genes discovered as associated with biomedical conditions are also associated with behavior, and vice versa. Scientists cannot hope to develop a comprehensive understanding of how people’s genetics influence their risk for disease without taking into account the genetics of behavior.

The valid scientific reasons for studying genetics are sometimes overshadowed by what some people falsely assume are the reasons for studying genetics. We are emphatically NOT studying the genetics of externalizing to make statements about people being “innately” antisocial or prone to addiction, or to forecast the life outcomes of particular individuals (see the section below labeled “**What is a polygenic score?**”). Our study also tells nothing about any race or ethnic group differences in rates of externalizing problems or contact with the criminal justice system (see “**Who are the people in this study and why does it matter?**”).

## **Who are the people in this study, and why does it matter?**

This study uses genetic data from people who all were of “European” genetic ancestry. Genetic ancestry is measured by analyzing statistical patterns of genetic similarity and dissimilarity in a given sample of people, often in comparison to a reference panel of global genetic variation. The label “European,” in this context, means that the people in this study have patterns of genetic variation most similar to people who currently live in Europe and whose recent ancestors in the last few hundred years all lived in Europe. People in the U.S. who have only “European” genetic ancestry will most often self-identify their race as “White.”

There are [important reasons](#) for conducting a GWAS within a group of people who are all somewhat genetically similar by virtue of their relatively recent ancestors coming from the same part of the world (in this case, people with relatively recent European ancestors). The focus on only European-ancestry people, however, has several important implications and limitations.

First, the results of this study might not generalize to people with different genetic ancestry. If, as hoped, genetic research is ultimately useful for improving people’s lives, it is [problematic if those benefits are not available to minority populations](#) that are already marginalized. Second, attempts to compare racial or ethnic groups on their genetic scores are [scientifically meaningless](#). Consequently, the results of this study **do not** and **cannot** tell us anything about the sources of disparities between different racial or ethnic groups.

In addition, boys and men have higher rates of externalizing problems on average than girls and women do. In our study, we also pooled together data from both biological sexes. (In this study, we did not assess people's gender identity. We only have information about their sex chromosomes.) And, we only analyzed data on the autosomes, *i.e.*, on genetic variation in chromosomes that are not sex chromosomes. As a result, this study identified genetic variants that are associated with higher externalizing on average across both sexes. Our study thus is not informative about any cultural or biological factors that might contribute to sex-differences or gender-differences in externalizing. Previous genetic research has emphasized that [“genetic influence must be understood through the lens of historical change, the life course, and social structures like gender.”](#) How the genetic variants we identify here play out differently, depending on social structures like gender, remains an important topic for future research.

## What did this study find?

There were three main results to this study.

- (1) Our GWAS identified 579 SNPs associated with a general tendency toward externalizing. Using tools that map SNPs to genes and tissues in the body, we found that genes associated with externalizing were expressed in the brain and were involved in biological pathways related to neurodevelopment. Our GWAS found that each individual SNP shows only a very small association with externalizing problems. This is in contrast to previous studies and popular accounts which claimed to find very strong causal effects of particular genes on aggression, antisocial behavior, and alcoholism and other forms of substance abuse. For example, [there is no “warrior gene”](#), despite claims to the contrary. Similar to the GWAS of other phenotypes, we find that any individual gene exerts only a small predisposing influence on externalizing, not a large or deterministic effect.
- (2) We used two independent follow-up data sets (see, **“Who are the people in this study and why does it matter?”**) and showed that a *polygenic score* created from our GWAS of externalizing was correlated with a wide variety of important life outcomes. These life outcomes included substance use behaviors (e.g., ever using opioids), contact with the criminal justice system (e.g., ever being arrested, convicted, or incarcerated), and employment outcomes (e.g., ever being fired from work).
- (3) We matched genetic data with electronic health records and showed that people who have many genetic variants associated with externalizing (see, **“What is a polygenic score?”**) are more likely to experience a variety of diseases, including cirrhosis of the liver and HIV infection, and are more likely to attempt suicide.

## What is a polygenic score?

A polygenic score is a single number that adds up information from a person's entire genome, in order to create a single score that represents how many of the genetic variants that have been connected to particular outcome that person carries, weighted by how big of an effect each of the genetic variants they carry has. These scores reflect scientists' best estimate of a person's genetic liability to that outcome based on their DNA. [Polygenic scores are only as good as the GWAS used to create them.](#) As GWAS sample

sizes for externalizing keep increasing, our polygenic scores will become more accurate indicators of an individual's liability to develop that outcome.

It is also important to realize what polygenic scores are *NOT*. Here we discuss three misinterpretations of polygenic scores.

## 1. Polygenic scores are NOT “fortune-tellers.”

Polygenic scores typically account for only a small proportion of the total amount of variation in an outcome. In our case, our polygenic score captures up to 10% of the variance of the general tendency towards externalizing behaviors, but less than that for any specific behavior or outcome (*e.g.*, ADHD or aggression). [This is not a strong enough effect](#) to be certain about what is going to happen in the life of any individual person. Nevertheless, these variables are [meaningful for research purposes](#).

Instead of being a “fortune teller”, a polygenic score is a risk factor. By way of analogy, having high cholesterol makes it more likely that you'll have a heart attack, but it doesn't determine that outcome — lots of people have high cholesterol but don't have a heart attack, and you can take steps to prevent a heart attack if you are at high risk. Similarly, a high polygenic score “for” externalizing means you have a (slightly) higher probability of experiencing outcomes like substance use problems *if* you are also experiencing similar environmental conditions as people in the original study. But that higher probability does not mean destiny!

## 2. Polygenic scores are NOT free of environmental or social processes.

There are two components to creating a polygenic score. The first component is information about a person's SNPs, which are fixed at conception and do not change over the course of their life. The second component is a set of weights that reflect the correlation between each SNP and a target phenotype, like externalizing. These weights carry information not just about biological processes but also about the social environment.

For example, let's imagine a society in which children who go through puberty earlier are more likely to be perceived by teachers and parents as aggressive, simply because they are bigger and taller. As a result, these children are punished more harshly, and this exposure to harsh punishment makes them more likely to engage in aggressive behaviors in the future. In this example, a genetically-influenced characteristic (pubertal timing) is acted on by the social environment (via teacher and parent perceptions/punishment), producing a correlation between puberty-influencing genes and aggressive behavior. If, then, someone conducted a GWAS in this society, they would detect these puberty-accelerating SNPs as being associated with aggressive behavior. Accordingly, the polygenic score “for” aggressive behavior would include these puberty-accelerating SNPs, which are only associated with aggression through a social process (punishment). The polygenic score, then, would capture some of the *social* biases that create disparities in children's behavior problems.

### 3. Polygenic scores are NOT a measure of a child’s “innate” or “inborn” potential.

We know that parenting styles, cultural practices, medical systems, criminal justice systems, and educational systems differ around the world and have differed throughout human history — and could be modified in the future. The word “innate” implies that there is something about a certain child’s biology that pre-determines them to be high in externalizing no matter what. But we know that genes are not destiny. Supportive and nurturing environments provided by parents, peers, romantic partners, and communities can reduce the likelihood that an individual who is at risk will develop problems. Conversely, environmental adversities can increase the likelihood that at an at-risk individual will develop problems. Nothing about a child’s polygenic score tells you about what could be true about that child if their social context were different. This is true of all genetic studies, not just ours. Genetic research cannot tell you what *could be*, if social systems were different. And genetic research cannot tell you how social systems *should be* structured.

### How can the externalizing polygenic score be used?

The creation of DNA-based predictors of complex human outcomes like ADHD and aggression has sparked both excitement and dismay. Myriad potential applications of polygenic scores have been proposed, including their use in:

- Insurance (*e.g.*, setting rates for car insurance or life insurance)
- Education (*e.g.*, assigning students to curricular tracks or forms of personalized education)
- Criminal justice proceedings (*e.g.*, making decisions about sentencing or parole)
- Reproductive medicine (*e.g.*, selecting IVF embryos for implantation)
- Medicine (*e.g.*, prescribing a personalized/precision medicine or intervention)

**Currently, we urge extreme caution about using an externalizing polygenic score.** Generally, there are on-going ethical debates about the fairness of using someone’s genetic information – which they cannot control or alter – to make decisions about them. More specifically, there are three reasons to be extremely cautious about using an externalizing polygenic score.

- (1) The externalizing polygenic score does not predict outcomes for an *individual* with a sufficient level of certainty to be useful for high-stakes decisions about them, [particularly when you consider that polygenic scores in general currently add little additional information above and beyond other sources of information](#) that might be available.
- (2) We do not know *how* the polygenic score is associated with externalizing. We don’t know the mechanisms by which the identified genetic variants increase risk for externalizing. As described above (see, “What is a polygenic score?”), differences in DNA might come to be associated with externalizing because of social biases that are generally considered to be unfair. Using the polygenic score for decision-making, then, has the potential to create a feedback loop, where those social biases are perpetuated under the guise of an apparently “objective” score.

- (3) The externalizing polygenic score is currently valid only for people of “European” genetic ancestry. (Again, by “valid,” we mean that the polygenic score predicts average trends, not individual outcomes.) Deploying a tool that applies to only part of the U.S. population is, in our view, not consistent with the goals of diversity, equity, and inclusion.

At the same time, we do think there are useful applications of the externalizing polygenic score in research settings. For example, a researcher might be interested in whether a school anti-bullying intervention preferentially reduces aggressive behavior among adolescents who are at highest genetic risk for aggression. As another example, a researcher might be interested in studying whether the association between an externalizing polygenic score and likelihood of being arrested is reduced, increased, or unchanged by legislation to decriminalize cannabis use. In these types of applications, the polygenic score is being used to evaluate a program or a policy, not to evaluate an individual person.

## Could studying genetic influences on externalizing lead to discrimination?

We acknowledge that genetic research has a [long history](#) of being misinterpreted and misused to argue that social inequality is inevitable, that social programs designed to improve people’s lives are bound to fail, and that some people are “naturally” inferior to other people. **We wholeheartedly reject these claims on both scientific and moral grounds.**

First, the existence of genetic associations within a group of people (see “**Who are the people in this study and why is that important?**”, above) does *not* tell you anything about whether there are average differences between racial or ethnic groups, or why such differences, if they are observed, occur. This is an important point because [racist](#) and [classist](#) ideas about the allegedly “inferior” character of people of color and the poor have been used to justify violence and oppressive policies against them. Nothing about this study gives any sort of empirical support to these ideas.

Furthermore, we emphasize that the environment can and does make a difference. New [research](#) has shown that many polygenic scores capture what most people would think of as environmental effects. For example, parents with particular genetic variants may be more likely to use harsh and coercive discipline with their children, and these parenting practices might lead to escalations of rule-breaking and aggressive behavior in their children. In this scenario, because children receive both their genes and their parenting from their parents, the children’s polygenic scores would be correlated with their aggression for environmental reasons – a phenomenon known as “indirect genetic effects” or “genetic nurture.” The role of these indirect genetic effects in externalizing should be investigated in future studies.

Finally, we emphasize that interventions or policy reforms designed to change people’s behavior can be successful. In fact, as we explained above (see “Why study the genetics of externalizing?”), improving interventions for externalizing is a major goal of this type of research. [To use a classic example](#), just because you might be genetically predisposed to poor eyesight doesn’t mean that your eyeglasses won’t work. ***Nothing about the current study undermines support for investments in bettering human lives.*** We argue that it is by understanding the full range of factors – both genetic and environmental – that contribute to human behavioral outcomes, that we will be able to best support individuals and improve human lives.