

# Do we measure or compute polygenic risk scores?

Citation for published version (APA):

Penders, B., & Janssens, A. C. J. W. (2022). Do we measure or compute polygenic risk scores? Why language matters. *Human Genetics*, 141(5), 1093-1097. <https://doi.org/10.1007/s00439-021-02262-7>

## Document status and date:

Published: 01/05/2022

## DOI:

[10.1007/s00439-021-02262-7](https://doi.org/10.1007/s00439-021-02262-7)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

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

[Link to publication](#)

## General rights

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

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

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

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

## Take down policy

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

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

providing details and we will investigate your claim.



# Do we measure or compute polygenic risk scores? Why language matters

Bart Penders<sup>1</sup> · A. Cecile J. W. Janssens<sup>2</sup>

Received: 23 October 2020 / Accepted: 2 February 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

## Abstract

Here, we argue that polygenic risk scores (PRSs) are different epistemic objects as compared to other biomarkers such as blood pressure or sodium level. While the latter two may be subject to variation, measured inaccurately or interpreted in various ways, blood flow has pressure and sodium is available in a concentration that can be quantified and visualised. In stark contrast, PRSs are calculated, compiled or constructed through the statistical assemblage of genetic variants. How researchers frame and name PRSs has consequences for how we interpret and value their results. We distinguish between the tangible and inferential understanding of PRS and the corresponding languages of measurement and computation, respectively. The conflation of these frames obscures important questions we need to ask: what PRS seeks to represent, whether current ways of ‘doing PRS’ are optimal and responsible, and upon what we base the credibility of PRS-based knowledge claims.

## Abbreviations

GWAS	Genome-wide association studies
PRS(s)	Polygenic Risk Score(s)
RCT	Randomized clinical trial
SNP	Single nucleotide polymorphism

## Introducing PRS

PRSs are used for investigating the genetic contribution to complex diseases and traits, the overlap in genetic contribution between multiple diseases and traits, and the contribution of genetic risk factors in their prediction. PRSs were introduced as a heuristic to handle the large number of SNPs that are implicated in complex diseases and traits. When the number of SNPs became too large to include them as separate variables in regression analysis, PRSs became a pragmatic alternative strategy for calculation. Some of the earliest studies that investigated the combined predictive performance of clinical and genetic risk factors included a handful of SNPs as separate variables in the prediction models (e.g. Maller et al. 2006; Seddon et al. 2009).

Others included unweighted risk scores where all SNPs were counted equally (Meigs et al. 2008; van Hoek et al. 2008; Lango et al. 2008), but soon the weighted scores became the standard (International Schizophrenia Consortium 2009; Talmud et al. 2010). In the weighted scores, risk alleles are first multiplied with their weight before they are summed into a score.

In recent years, the construction of the PRS has become more advanced. PRSs are still calculated as sum scores, but the estimation of the weights has become more sophisticated (Choi et al. 2020). When PRSs are constructed from genome-wide significant SNPs only, weights are generally taken from GWAS. When they are constructed from all genotyped SNPs and include millions of variants, then the estimation of the weights may take into account the weights of neighboring variants. It is known that GWAS weights may be overestimated and that the weights are pooled effects, largely dictated by the largest studies in the GWAS meta-analysis, and they may not generalize to other populations (Choi et al. 2020). The sum score is considered appropriate because it fits the distribution of genetic variation for complex diseases that was outlined by Fisher (Visscher and Goddard 2019). That this sum score is considered appropriate, does not mean that the only way to approach that distribution is by putting all SNPs into a single score, in the way it is currently done. This has generated discussion about PRS, and about the appropriateness of the particular statistical approaches, it contains in particular (e.g. Wald and Old 2019; Janssens 2019).

✉ A. Cecile J. W. Janssens  
cecile.janssens@emory.edu

<sup>1</sup> Care and Public Health Research Institute, Maastricht University, Maastricht, The Netherlands

<sup>2</sup> Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia

PRSs are also used in the context of the study of traits or characteristics that are not usually or at least not primarily considered genetic. This further stimulates continued critical examination of the aforementioned decisions and their (potential) consequences. To name but a few, PRSs are used in the study of household income (Hill et al. 2019), educational attainment (Lee et al. 2018), or even entrepreneurship (Rietveld et al. 2020). This has led to an ongoing conversation about PRS and questions about the validity, limits, character and characteristics of the method (Janssens 2019). Hence, we should ask ourselves how we should understand the method and the scientific claims it generates.

## Situating PRS

To help understand and explain different levels of credibility assigned to scientific methods, approaches and claims, the sociology of science and knowledge scholar Harry Collins proposes to differentiate between tangible and inferential experiments and observations (2019, pp. 114–122). In tangible experiments or observations, statistical claims form only a part of the support for the claims they produce. The rest of the support is not statistical, but tangible, and consists of described mechanisms, visible or understood. “When an outcome is tangible”, Collins writes, “this [understanding] can support a statistical inference” (p. 115). In inferential experiments or observations, the evidence consists solely of statistical claims. Inferential experiments or observations are not inferior to tangible ones, but the process through which they build credibility for that evidence is different. Tangible outcomes can seek alignment with existing credible word views, where inferential ones cannot.

The distinction between the tangible and the inferential is a conceptual clarification of an argument he and Trevor Pinch made over a decade ago. Collins and Pinch sketch examples of hypothetical research on a condition called ‘undifferentiated broken limb’ (Collins and Pinch 2008, pp. 32–33). First preliminary studies, they write, showed that a cast applied to the left leg led to full recovery in up to 25% of all cases. A subsequent randomised clinical trial (RCT) to consolidate the efficacy of casting the left leg confirmed the preliminary findings.

A clinical study on the efficacy of casting without checking which leg is broken is, and we hope all agree, quite absurd. We realise that it is absurd because a cast is a tangible intervention: we ‘know’ that casting a leg that is not broken will not heal the leg that is broken. In other words: one would always check which limb is broken and use this (tangible) observation to inform which limb needs a cast. Thus, the determination of what constitutes a reasonable treatment draws its credibility not from the statistical or methodological rigour an RCT can provide, but from everybody knowing

that the tangible intervention will lead to the tangible outcome: casts applied to the right limb will heal the broken limb. This formal type of evidence can be nice as an extra, a bonus, if we have some understanding of the mechanism of a problem, but the credibility of an intervention does not depend on it, whether that is a cast for a broken limb or—another famous example—the value of a parachute when jumping out of an aeroplane (Yeh et al. 2018). A broken arm or leg and the cast that accompanies them, as well as parachutes and aeroplanes for that matter, are tangible objects in our lives. We can all reasonably produce a narrative credibly connecting the intervention (a cast applied to a broken limb; or jumping from a plane without a parachute) to the tangible outcome (improved healing of that limb; death) that would be convincing to all or most audiences without any need for statistics.

Many of the research contexts in the medical, life and psychological sciences do not have the “luxury” of tangible interventions and outcomes. A treatment, a cure, or a protocol can work without tangibility, in which the meaning of ‘work’ differs per situation, but, this lack of tangibility limits ways to build credibility and trust for the intervention. For inferential interventions and outcomes, the evidence consists of statistics alone. When the credibility of a scientific claim rests only on statistical inference, Collins argues, this produces fragile credibility. It has to rest upon the assumption that everything has happened according to the rules of the game, no mistakes were made, no actors or circumstances may have had an undue influence, and a lot more. To Collins, this fragility manifests itself in ‘statistical paranoia’ in the form of explicit rules and thresholds, clear protocols, preregistrations and endless possibilities to question and discredit the claim. This contrast, Collins argues, shows that “[w]e only reach for statistical punctiliousness when we don’t know what is going on” (Collins 2019, p. 115). While these rules and thresholds are also applied to tangible experiments and observations, in that case, a little bit of “post-hocery” is less of a problem, since the credibility of the scientific claim draws from the tangible outcome, and rests less on the numbers. This stronger credibility that may withstand some “post-hocery” is not an excuse or permission to p-hack or cut corners. Rather, it helps explain why, after errors or wrongdoing have been discovered, some claims remain credible. In the case of inferential experiments, the opposite is the case and every little taint on the study quickly erodes its credibility—hence its fragility.

## Framing PRS

When does something fall in the tangible category and when in the inferential? How do we describe the differences? The answer is not straightforward since no outcome is absolutely

tangible. Even relatively direct measurements, such as height, weight or temperature are to some extent dependent on human decision-making (Chang 2004). While such decisions may influence measurements (do we weigh people with or without their underwear?), they do not seriously deny measurability. On the tangible end of the spectrum, we find height, weight, numerical counts, and on the inferential end, we find complex networked assemblages such as IQ scores or personality tests. Access to the object of study is an indicator for tangibility, as is the materiality of the object of study (arms, legs, parachutes, proteins, etc.). The use of terminology such as “measurement” rhetorically suggests such access, as does language centering on discovery (‘discovered’, ‘revealed’, ‘found’, etc.).

In the case of PRSs, the vocabulary of measurement is used by some. For instance, Frank et al. (2015) write “genetic risk was assessed by measuring the polygenic risk score” and Domingue et al. (2014) write “Genetic risk was measured using the genetic risk score for Europeans (GRS-E)”. When authors write that they measured PRS, they suggest that their analyses are tangible experiments and that PRS is “out there” to be documented. Other authors write about calculating (e.g. Derks et al. 2012), assembling (Dron and Hegele 2018), or compiling (Papiol et al. 2014) PRSs, conceptualising their method as inferential. The wording reveals how authors frame PRS (Goffmann 1974) and how they think about the credibility of their claims, intentionally or unintentionally.

In the aforementioned, we can distinguish two frames. Let us call the former “measurement” and the latter “computation”. These conceptual frames are ways to define situations, organise collective experiences about them, as well as offer avenues for action (Goffmann 1974). All our engagements with the world use frames since observing or speaking about the world in all its intricacies is impossible. Frames make understanding and communication possible and there is no science communication that is unframed. However, the same situations can be described through multiple frames. Multiple frames may coexist or one frame may dominate over others. Since frames are implicit, it is often difficult to recognise them. In the case of PRS, measurement (tangible) and computation (inferential) frames seem to co-exist. Whether such framing is intentional, is unknown.

Frames have consequences. Framing allows a certain definition of a problem, the attribution of blame, and the legitimization of some solutions while delegitimizing others (Saguy 2012). An example: discussions on research integrity distinguish between the ‘bad apples’ and the ‘bad baskets’ frames (corrupted individuals or corrupted structures). If a concrete digression is framed as a problem of a bad apple, the individual scientist is to be blamed and their moral character was the problem (Zwart and Ter Meulen 2019). Firing the scientists and educating the rest would be the obvious

solution. If the same digression is framed as a bad basket problem, the system is blamed and the rewards and incentives hardwired into science as seen as the problem—re-designing them would be the obvious solution.

PRS literature frames PRS as either “measurement” or as “computation”. These frames imply either the *presence* or the *absence* of a tangible outcome, and, therewith, a narrative explaining that outcome. In the case of PRS, the two frames position the genetic origin of complex diseases and traits either as tangible or as inferential. The “measurement” frame argues that predicted risk is such a tangible outcome and, as a consequence, provides protection against what Collins calls ‘statistical paranoia’. The “computation” frame foregoes that protection by arguing that the risk prediction is computed. That these frames exist and exist next to one another does not mean that all researchers who employ them are aware of them. Unintentional or not, framing still has consequences.

## Portraying PRS

PRSs belong, in our view, on the inferential end of the spectrum. In other words, we see PRS as a concept that is actively constructed: as “computation”. PRSs are sophisticated calculations and the product of statistical innovation. PRSs are constructed out of a variety of components, brought together as a heuristic for managing the intricacies of the genetic contribution to the origins of complex diseases and traits. In the statistical synthesis that brings them into being, we can discern various choices, all shaping a concept—the risk score—that is far from tangible. Framing PRS as measurement comes with a language of discovery and thereby, obscures those choices and obscures the many intermediaries required to produce a polygenic risk score. The computation frame offers, at the very least, a possible window on this production process and the display of the underlying choices through the language it employs: that of a process to be managed and maintained.

Language matters. All scientific language renders certain elements of the practice more visible and others less visible. The passive voice is a famous example. It renders the researcher(s) invisible, suggesting objectivity by deleting the subject. This rhetorical strategy or frame, also called the “God trick” or the “voice from nowhere”, is heavily critiqued for exactly this reason (Haraway 1991). The inferential ‘computation’ frame emphasizes that the object does not exist ‘out there’, tangibly, but that it is the result of a process of computation. Furthermore, the inferential frame allows experts and others to judge, weigh and decide how credulous or sceptical to be when approaching the knowledge claims surrounding PRSs. The computation frame

promotes continued thinking over the meek acceptance of PRS as a black box.

We have to think about *how* we speak about things, including but not limited to PRSs. Our scientific vocabulary allows us to differentiate between the tangible and the inferential and between measurement and computation frames even without explicit reference to the sociological theory we mention in this text. We invite scientists and scholars who develop and use PRSs to reflect upon the objects they are producing and discussing, and use appropriate language when they do. To compute, calculate, construct, compile, or even to build or assemble, does justice to the process and labour that results in PRSs and to the relationship they have to the biological world they aim to describe. It establishes PRS not as tangible but as inferential, the result of a distinct and specific statistical process populated by assumptions, traditions, and tools and separated from the biological world through a substantial series of intermediaries and proxies, a process that can be improved, adapted, optimised and more. This language also allows the recognition of the expertise, skill and labour it requires to compute PRSs well, as opposed to a language of discovery (Bartlett et al. 2017).

The words and numbers that we use, shape the world we aim to describe with them. When it comes to PRS, an open conversation about framing would stimulate us to continually and critically think about the constitutive ingredients of PRS, and the statistical and social translations required, as well as what this means for its legitimate and responsible use in (clinical) practice. When PRS is framed as an inferential outcome, this conversation should include questions about what PRS exactly tries to represent and whether it accomplishes this representation, and if so, how it compiles this representation optimally and responsibly.

As a final note, the way we position, interpret and describe PRS as either tangible or inferential is not necessarily forever stable. As the knowledge base underpinning genetic variance and our understanding of the mechanisms through which they contribute to phenotypic variance grows, the constitutive elements of PRS may advance slowly towards the tangible and thereby assemble defenses against statistical paranoia (Collins 2019). When this happens, vocabularies may responsibly shift along. But today is not that day.

**Acknowledgements** We thank Lotte Thissen for providing valuable and constructive feedback on earlier versions of our manuscript.

**Author contributions** BP: Conceptualization, writing – original draft, writing – review & editing; ACJWJ: Conceptualization, writing – original draft, writing – review & editing (CRediT roles).

**Funding** No specific funding was used to produce this article.

**Data availability** This conceptual analysis does not report data.

## Compliance with ethical standards

**Conflict of interest** None to report.

**Ethical approval** The article presents a conceptual, non-empirical, analysis and does not require ethics approval.

**Informed consent to participate** The article presents a conceptual analysis, the research reported has no participants.

**Informed consent for publication** The article presents a conceptual analysis and requires no such consent.

## References

- Bartlett A, Penders B, Lewis J (2017) Bioinformatics: indispensable, yet hidden in plain sight? *BMC Bioinformatics* 18:311. <https://doi.org/10.1186/s12859-017-1730-9>
- Chang H (2004) *Inventing temperature: measurement and scientific progress*. Oxford University Press, New York
- Choi SW, Mak TSH, O'Reilly PF (2020) Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* 15:2759–2772. <https://doi.org/10.1038/s41596-020-0353-1>
- Collins H (2019) *Forms of life: the method and meaning of sociology*. MIT Press, Boston
- Collins H, Pinch T (2008) *Dr. Golem: how to think about medicine*. University of Chicago Press, Chicago
- Derks EM, Vorstman JA, Ripke S, Kahn RS, Ophoff RA, Schizophrenia Psychiatric Genomic Consortium (2012) Investigation of the genetic association between quantitative measures of psychosis and schizophrenia: a polygenic risk score analysis. *PLoS ONE* 7:6. <https://doi.org/10.1371/journal.pone.0037852>
- Domingue BW, Belsky DW, Harris KM, Smolen A, McQueen MB, Boardman JD (2014) Polygenic risk predicts obesity in both white and black young adults. *PLoS ONE* 9:7. <https://doi.org/10.1371/journal.pone.0101596>
- Dron JS, Hegele RA (2018) Polygenic influences on dyslipidemias. *Curr Opin Lipidol* 29:133–143. <https://doi.org/10.1097/MOL.0000000000000482>
- Frank J, Lang M, Witt SH, Strohmaier J, Rujescu D, Cichon S, Degenhardt F, Nöthen MM, Collier DA, Ripke S, Naber D (2015) Identification of increased genetic risk scores for schizophrenia in treatment-resistant patients. *Mol Psychiatry* 20:150–151. <https://doi.org/10.1038/mp.2014.56>
- Goffmann E (1974) *Frame analysis: an essay on the organization of experience*. Harper and Row, New York
- Haraway D (1991) *Cyborgs, simians, and women: the reinvention of nature*. Free Association Books, London
- Hill WD, Davies NM, Ritchie SJ, Skene NG, Bryois J, Bell S, Di Angelantonio E, Roberts DJ, Xueyi S, Davies G, Liewald DC, Porteous DJ, Hayward C, Butterworth AS, McIntosh AM, Gale CR, Deary IJ (2019) Genetic analysis identifies molecular systems and biological pathways associated with household income. *Nat Commun* 10:5741. <https://doi.org/10.1038/s41467-019-13585-5>
- International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460:748–752. <https://doi.org/10.1038/nature08185>

- Janssens ACJW (2019) Validity of polygenic risk scores: are we measuring what we think we are? *Hum Mol Genet* 28:R143–R150. <https://doi.org/10.1093/hmg/ddz205>
- Lango H, UK Type 2 Diabetes Genetics Consortium, Palmer CN, Morris AD, Zeggini E, Hattersley AT, McCarthy MI, Frayling TM, Weedon MN (2008) Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. *Diabetes* 57:3129–3135. <https://doi.org/10.2337/db08-0504>
- Lee JJ, Wedow R, Okbay A et al (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 50:1112–1121. <https://doi.org/10.1038/s41588-018-0147-3>
- Maller J, George S, Purcell S, Fagerness J, Altshuler D, Daly MJ, Seddon JM (2006) Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* 38:1055–1059. <https://doi.org/10.1038/ng1873>
- Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB, Cupples LA (2008) Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 359:2208–2219. <https://doi.org/10.1056/NEJMoa0804742>
- Papiol S, Mitjans M, Assogna F, Piras F, Hammer C, Caltagirone C, Arias B, Ehrenreich H, Spalletta G (2014) Polygenic determinants of white matter volume derived from GWAS lack reproducibility in a replicate sample. *Transl Psychiatry* 4:e362. <https://doi.org/10.1038/tp.2013.126>
- Rietveld CA, Slob E, Thurik AR (2020) A decade of research on the genetics of entrepreneurship: a review and view ahead. *Small Bus Econ*. <https://doi.org/10.1007/s11187-020-00349-5>
- Saguy AC (2012) *What's wrong with fat?* Oxford University Press, New York
- Seddon JM, Reynolds R, Maller J, Fagerness A, Daly MJ, Rosner B (2009) Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables. *Invest Ophthalmol Vis Sci* 50:2044–2053. <https://doi.org/10.1167/iovs.08-3064>
- Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, Kivimäki M, Humphries SE (2010) Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ* 340:b4838. <https://doi.org/10.1136/bmj.b4838>
- van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, Hofman A, Sijbrands EJ, Janssens AC (2008) Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. *Diabetes* 57:3122–3128
- Visscher PM, Goddard ME (2019) From R.A. Fisher's 1918 paper to GWAS a century later. *Genetics* 211:1125–1130. <https://doi.org/10.1534/genetics.118.301594>
- Wald NJ, Old R (2019) The illusion of polygenic disease risk prediction. *Genet Med* 21:1705–1707. <https://doi.org/10.1038/s41436-018-0418-5>
- Yeh RW, Valsdottir LR, Yeh MW, Shen C, Kramer DB, Strom JB, Secemsky EA, Healy JL, Domeier RM, Kazi DS, Nallamotheu BK (2018) Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. *BMJ* 363. <https://doi.org/10.1136/bmj.k5094>
- Zwart H, Ter Meulen R (2019) Addressing research integrity challenges: from penalising individual perpetrators to fostering research ecosystem quality care. *Life Sci Soc Pol* 15:5. <https://doi.org/10.1186/s40504-019-0093-6>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.