

Dylan Lawless@kispi.uzh.ch  
Department of Intensive Care and Neonatology  
University Children's Hospital Zürich  
University of Zürich.

July 1, 2025

Dear Editors,

In July 1908, G. H. Hardy wrote to *Science* [1], presenting what he described as a “very simple” piece of mathematics. As a pure mathematician, Hardy regarded the problem as straightforward enough to solve in conversation with Reginald Punnett during a cricket match. His insight—that under random mating, genotype frequencies remain constant across generations in the absence of selection, mutation, migration or drift—became fundamental to modern population genetics. Independently, Wilhelm Weinberg reached the same conclusion in Germany [? ], establishing what is now known as the Hardy-Weinberg equilibrium (HWE).

The pathway to this result, however, was neither immediate nor trivial. Following the rediscovery of Mendelian genetics in 1900, biologists debated how discrete genetic inheritance could explain continuous traits such as height. Udny Yule (1902) argued that Mendelian inheritance could not account for such variation and worried that dominant alleles might inevitably increase in frequency [? ]. William E. Castle (1903) demonstrated that, without selection, genotype frequencies remain stable [? ]. Karl Pearson (1903) explored equilibrium states, including the case where allele frequencies equal 0.5 [? ]. Punnett, unable to resolve Yule's objections alone, sought Hardy's assistance. The result was a brief note that became a cornerstone of genetic theory.

Hardy demonstrated that if the proportions of pure dominants ( $AA$ ), heterozygotes ( $Aa$ ), and pure recessives ( $aa$ ) in one generation are  $p : 2q : r$ , then after random mating, the next generation's frequencies are:

$$(p + q)^2 : 2(p + q)(q + r) : (q + r)^2.$$

Equilibrium is achieved when  $q^2 = pr$ , ensuring genotype proportions remain stable across generations.

While Hardy's calculation clarified how dominant traits do not inevitably spread through populations, it left unanswered a separate, crucial question. HWE describes how often genotypes are expected to occur, but it does not indicate whether a specific genotype causes disease in an individual.

Over the twentieth century, medical genetics advanced substantially, yet a gap remained between population-level expectations and individual clinical interpretation. Researchers and clinicians developed systems to classify variants as pathogenic, benign or of uncertain significance. These categorical labels, however, did not quantify the probability that a particular variant causes disease in a patient, nor did they incorporate the possibility of pathogenic variants that remain undetected due to technical limitations or gaps in current knowledge.

For more than a century, a central question persisted: given a candidate variant in a patient, what is the probability that it is truly causal, and how certain can we be in light of possible alternative, unobserved causes?

Our work addresses this question directly. We extend the principles laid down by Hardy and Weinberg into clinical genomics by combining classical population genetics with empirical variant data and Bayesian inference. We integrate HWE-derived genotype expectations with allele frequencies and pathogenicity annotations from large databases such as gnomAD and ClinVar. Bayesian modelling then merges these priors with individual patient sequencing data, explicitly incorporating the possibility of unobserved pathogenic variants.

This framework yields credible intervals (CrIs) that quantify both the probability that an observed variant is causal and the residual probability attributable to potential unobserved variants. It shifts variant interpretation from categorical labelling to quantitative assessment, enabling clinicians and researchers to assign precise probabilities to variant pathogenicity, grounded in theoretical genetics and empirical data.

We demonstrated this approach in the context of inborn errors of immunity (IEI), calculating variant-level probabilities across 557 disease-associated genes and validating predictions against national patient cohorts. Although illustrated in IEI, the method is applicable to any genetic disease, offering a rigorous quantitative basis for variant interpretation, clinical decision-making and future data-driven analyses in genomics.

Hardy’s “very simple” calculation proved to be the beginning rather than the conclusion of this scientific path. Resolving the question of variant causality has required statistical innovation, large-scale genomic resources and advances in clinical genetics. Our work represents the next step in this continuum, addressing a problem left unresolved in quantitative terms since the earliest days of modern genetics.

Yours sincerely,

Dylan Lawless, PhD  
on behalf of all co-authors

- [1] G. H. Hardy. Mendelian Proportions in a Mixed Population. *Science*, 28(706):49–50, 1908.