



Cystic fibrosis heterozygosity: Carrier state or haploinsufficiency?

David Fisman^{a,1}

Cystic fibrosis (CF) is a common genetic disorder, caused by mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. The CFTR gene encodes a transmembrane chloride channel, which is important for key physiological functions, such as production of sweat and mucus, as well as mucociliary clearance in the lungs (1). Affected individuals are homozygous for mutated copies of CFTR and are at elevated risk for a variety of diseases, including bronchiectasis and repeated pulmonary infection; gastrointestinal disorders, including malabsorption and nutritional deficiency states; and pancreatitis and diseases of the hepatobiliary system (1). While management of CF has improved drastically in recent decades, individuals with CF experience reduced life expectancy relative to the general population (2). The prevalence of carrier state of mutated CFTR genes is high [greater than 3% in some subpopulations (3)], leading to the suggestion that the carrier state must be positively selected for due to positive health effects, analogous to protection against malaria conferred by carrier states in sickle cell anemia (4, 5).

CF has been regarded as a classical autosomal recessive disorder, with no adverse health effects associated with the carrier state. I use the past tense, because in PNAS Miller et al. (6) provide convincing evidence that CF heterozygosity may represent a haploinsufficiency state, analogous to that seen with thalassemia, where individuals with a single copy of the disease-causing allele do suffer adverse health effects, presumably due to production of the gene's product at lower levels than would be seen in noncarriers. Beta-thalassemia is an autosomal recessive disorder of hemoglobin, but the carrier state is often characterized as "thalassemia minor" with individuals experiencing anemia that is less severe than homozygotes (who experience thalassemia major or thalassemia intermedia) (7). The contrast with CF relates to the multisystemic nature of CF, and the numerous disease states linked to CF, compared to beta-thalassemia as a hematological disease.

Miller et al. (6) were able to link genetic testing information to diagnostic codes using a very large,

commercial health analytics database built on insurance claims data [the Truven MarketScan Database (8)]. They evaluated the risk of 59 diseases that occur with higher frequency in individuals with cystic fibrosis in a cohort of 19,802 CF carriers matched by age, gender, and duration of enrollment to 99,010 controls. Remarkably, they found that individuals with the CF carrier state were at significantly increased risk of nearly all of the diseases evaluated (57/59). In a second analysis, they calculated odds ratios for the same 59 diseases in individuals with CF compared to matched controls and found strong correlation in odds ratios (OR) for disease states between CF carriers and individuals with CF. In other words, the higher the OR of the disease in individuals with CF, the higher the OR in CF carriers.

Their results were robust when restricted to individuals who might be presumed not to have undergone CF genetic testing (mothers of CF cases during the time period prior to the pregnancy) and also when they removed individuals in whom CF genetic testing might have been performed due to occurrence of a disease that suggested underlying CF. They also performed simulations to determine whether false discovery might explain their results and found this unlikely to be the case. Their results have important implications not only for how CF is conceptualized, but also for the potential burden of disease associated with the CF gene at the population level. The study also raises important ethical questions related to how we use the ever-increasing volume of electronic clinical records linked to genetic (or genomic) data.

In Miller et al.'s (6) study, the relative elevations in risk of CF-linked disease states are lower in carriers than in individuals with CF, but as Miller et al. (6) note, the high prevalence of CFTR mutation heterozygosity in the general population means that any elevation in risk has important implications for burden of disease. To understand this, we need to explore the concept of attributable risk or the etiological fraction of disease due to a particular cause (in this case either CF carrier state or CF disease). Epidemiologists typically consider two different types of attributable risk: attributable

^aDivision of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T3M7, Canada

Author contribution: D.F. wrote the paper.

The author declares no competing interest.

Published under the [PNAS license](#).

See companion article on page 1621 in issue 3 of volume 117.

¹Email: david.fisman@utoronto.ca.

First published January 28, 2020.

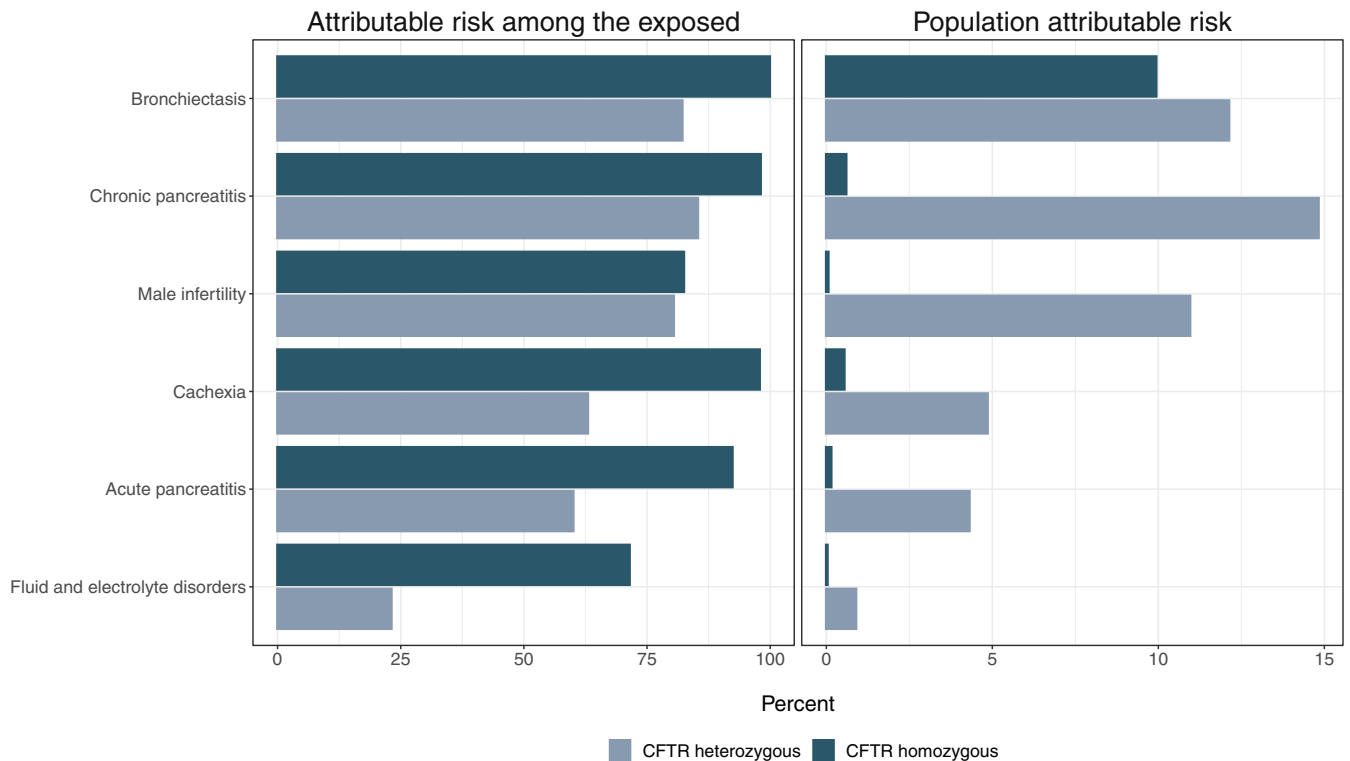


Fig. 1. Bar graph illustrating the difference in attributable risk percent (Left) and population-attributable risk percent (Right) for individuals heterozygous (light gray) and homozygous (dark gray) for mutated CFTR alleles, for selected CF-linked diseases. It can be seen that attributable risk for a given disease is higher in those with cystic fibrosis than in CF gene carriers. However, because of the higher prevalence of CFTR heterozygosity than homozygosity, the PAR% for these diseases is higher (generally, far higher) for CF gene carriers than for those with cystic fibrosis. An exception is bronchiectasis, where an extraordinarily high odds ratio in those with cystic fibrosis (~920) results in similar PAR% estimates (12% for CF carriers vs. 9% for individuals with cystic fibrosis). Data for this bar graph are derived from Miller et al. (6). Prevalence estimates are as noted in the text.

risk percent among exposed individuals (AR%) and population-attributable risk percent (PAR%) (9). AR% represents the fraction of disease in exposed individuals that is due to their exposure, as opposed to background risk.

For example, in the study of Miller et al. (6), the OR for acute pancreatitis in CF carriers is 2.5. Assuming the disease is rare, such that the OR approximates a relative risk (9), and confronted with a CF carrier with acute pancreatitis, we would say that the AR% for this individual is $(OR - 1)/OR$, which is $(2.5 - 1)/2.5$ or 60% (9). This approach acknowledges that there is baseline risk of acute pancreatitis even in the absence of CF carrier state and considers only risk to the individual. The OR for acute pancreatitis is much higher in those with CF [around 13 in Miller et al. (6)], so the AR% for CF in an individual with CF and pancreatitis would be $(13 - 1)/13$ or 92%.

By contrast, PAR% considers both prevalence of exposure in the population and relative elevation of risk and serves as an estimate of total population-level disease burden that could be modified by eliminating risk. Again, assuming OR approximates a relative risk, we can estimate PAR% as $[pe(OR - 1)]/[pe(OR - 1) + 1]$, where pe is the prevalence of exposure (9). If prevalence of mutated CFTR heterozygosity in the population is 3% (3), the PAR% for pancreatitis is ~4%. In 1990, it was estimated that there were 30,000 individuals with CF living in the United States (10), at a time when the US population was 250 million persons. Thus, the

prevalence of CF would have been ~12/1,000,000 at that time. Based on this prevalence, the PAR% for CF is only 0.1% or 1/40th of that attributable to CF carrier state. I explore this idea further in Fig. 1.

In addition to forcing us to rethink the relationship between CF heterozygosity and burden of disease, Miller et al.'s (6) study should flag the emerging likelihood that common genetic variants will be linked to elevated risk of disease occurrence, as genetic testing data, and genomic sequencing data, become more widely available. Are we at risk for pathologizing widespread and important genetic variation within human populations? Will existing legislation that forbids stigmatization by insurers and employers, based on genetic information, prove sufficiently robust to protect carriers (11)? Could fear of stigma result in a decline in preconception genetic testing and undermine efforts to prevent CF? These questions are critically important for, but transcend, CF. The rapid growth in availability of genetic and genomic information means that we will come to see disease pathways more clearly, and differently, in the near future. We need a firm ethical footing to be able to handle this new knowledge.

Acknowledgments

I thank Dr. Ashleigh Tuite for her guidance on graphical representation of attributable risks, and Dr. Jennifer Brooks for her guidance on issues related to legal protections for those who choose to undergo genetic screening.

1 J. S. Elborn, Cystic fibrosis. *Lancet* **388**, 2519–2531 (2016).

2 P. R. Burgel et al.; ERS/ECFS Task Force on Provision of Care for Adults with Cystic Fibrosis in Europe, Future trends in cystic fibrosis demography in 34 European countries. *Eur. Respir. J.* **46**, 133–141 (2015).

- 3 C. M. Strom *et al.*, Cystic fibrosis testing 8 years on: Lessons learned from carrier screening and sequencing analysis. *Genet. Med.* **13**, 166–172 (2011).
- 4 C. Wiuf, Do delta F508 heterozygotes have a selective advantage? *Genet. Res.* **78**, 41–47 (2001).
- 5 A. Ferreira *et al.*, Sickle hemoglobin confers tolerance to Plasmodium infection. *Cell* **145**, 398–409 (2011).
- 6 A. C. Miller *et al.*, Cystic fibrosis carriers are at increased risk for a wide range of cystic fibrosis-related conditions. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1621–1627 (2020).
- 7 C. S. Johnson, C. Tegos, E. Beutler, Thalassemia minor: Routine erythrocyte measurements and differentiation from iron deficiency. *Am. J. Clin. Pathol.* **80**, 31–36 (1983).
- 8 I. B. M. Watson Health, Truven Health Analytics. <https://www.ibm.com/watson-health/learn/truven-health-analytics>. Accessed 4 January 2020.
- 9 M. Szklo, J. Nieto, “Measuring associations between exposures and outcomes” in *Epidemiology: Beyond the Basics* (Jones and Bartlett, Burlington, MA, 2019), pp. 87–118.
- 10 S. C. FitzSimmons, The changing epidemiology of cystic fibrosis. *J. Pediatr.* **122**, 1–9 (1993).
- 11 National Human Genome Research Institute, Genetic discrimination. <https://www.genome.gov/about-genomics/policy-issues/Genetic-Discrimination>. Accessed 4 January 2020.