The Precision Medicine Initiative announced by President Obama aims to be a novel approach to disease treatment and prevention by allowing for individual variability in genes, environment, and lifestyle. Although its immediate goal is to support clinical trials of targeted cancer therapies based on a tumor's molecular signature, about half of its $200+ million budget for this year is aimed at a new cohort study of 1 million people for research on etiology and prevention. Here, I focus on the relevance of genomic research to personalized prevention.

In his seminal article entitled “Sick individuals and sick populations,” Rose distinguished the 2 primary goals of epidemiology: discovering “the determinants of individual cases and the determinants of incidence rate.” He called the corresponding strategies for prevention as the “high-risk approach,” which seeks to protect susceptible individuals, and as the “population approach,” which seeks to control the underlying causes of incidence. In the genomics era, these correspond to using an individual’s genetic profile to target primary and secondary prevention versus classical public health strategies to clean up the environment or promote healthful behaviors so as to reduce the overall burden of disease. Population-wide interventions may provide large benefits to the population as a whole while offering little benefit to most individuals (the “prevention paradox”), so the trade-off depends in part on how much risk is confined to an identifiable minority.

This commentary discusses 2 main themes: the evaluation of targeted versus population-wide interventions and the need for causal inference methods to accomplish this. I will illustrate these points with examples from primary prevention (air pollution regulation and genetic targeting of smoking cessation) and from secondary prevention (colonoscopies, as discussed in the companion article).

THE COUNTER-FACTUAL NATURE OF PREVENTION RESEARCH

It is often argued that the importance of research on gene–environment (G × E) interactions is that it could lead to novel prevention strategies based on modifiable risk factors. The difficulty with using observational studies to evaluate prevention programs lies in the causal interpretation of the effect of an intervention. Even if an exposure is associated with disease
across individuals, it does not follow that changing exposure would change any individual’s disease risk. The general idea of counter-factual inference can be understood in terms of a hypothetical $2 \times 2$ table\(^{10}\) comparing the potential outcomes of the same individual under the alternative exposure scenarios. The outcomes of subjects on the main diagonal are unaffected by exposure (“doomed” or “immune”). The only ones for whom exposure has any effect (caused or prevented) are those in the off-diagonal cells. The net benefit of removing exposure is thus the difference in the size of these two cells. Unfortunately, it is only in a crossover trial that we observe individual outcomes under both exposure scenarios and these are only possible for acute responses and unethical for hazardous exposures. In observational studies, we see only the margins—the distribution of outcomes among exposed and among unexposed individuals—and these are different individuals. We therefore have to assume there is no uncontrolled confounding, that is, that the groups are comparable in terms of the distribution of other risk factors, or at least that all known confounders have been taken into account in the analysis. Only a randomized controlled trial can ensure that, and only in expectation. Causal inference\(^{11–15}\) changes the target of inference from association across individuals to the mean difference within individuals in their expected outcomes under exposed and unexposed scenarios (the “average causal effect”). Details of the various procedures differ, but one relatively easy to understand approach begins with a model for the “propensity” for exposure based on the available confounder information and then reweights subjects inversely to examine the effect of exposure in a hypothetically unconfounded population.

Hernán et al.\(^{16}\) have refined the potential outcomes approach based on inverse probability weighting, illustrating the different conclusions obtained in the case of AZT (zidovudine) treatment in HIV/AIDS survival and reconciling the conflicting results on hormone replacement therapy use and cardiovascular disease from observational epidemiology studies and the randomized Women’s Health Initiative trial.\(^{17}\) None of those examples bear directly on primary prevention; however, so I begin with an example applied to population-based prevention.

**Causal Inference for Population-wide Primary Prevention**

One of the most effective interventions for reducing all-cause mortality nationwide has been the Clean Air Act of 1970. A huge body of epidemiologic evidence has demonstrated associations between ambient air pollution and various health outcomes. But how can we assess the impact specifically of the Clean Air Act on mortality? Zigler et al.\(^{18}\) used causal inference to compare the counties of southwest United States that were in and out of compliance in 1990 in terms of changes in air pollution and mortality over the following decade. Since counties could differ in ways that might confound any comparisons, they developed a spatial hierarchical model for the propensity to be in compliance and then computed the expected pollution and mortality outcomes for each county under their observed compliance status and their counterfactual ones. They found that the average causal effect of regulation was a 2.7% reduction in mortality and an improvement in air quality attributable to regulation. However, only about half the change in mortality was directly attributable to improvements in ozone and PM\(_{10}\), whereas the rest was mediated by changes in other risk factors (possibly including other pollutants). This approach to evaluating policy could be helpful in many other areas besides air pollution.\(^{19}\)

**GENETIC TARGETING PREVENTION PROGRAMS**

Any prevention effort presumes that we have an effective intervention and that intervention changes outcomes. Next, would either be improved by targeting high-risk populations, say by genetics? This then presumes that we have a way of identifying high-risk individuals, assuming we can obtain genotypes or a surrogate like family history from the population at risk. Finally, is the intervention more effective in that group (gene–treatment interaction)?

**Individual Variability**

Although many people see some heavy smokers living to old age without lung cancer while some nonsmokers get lung cancer at an early age as evidence of interindividual variability in sensitivity to tobacco smoking (e.g.,\(^{21}\)), Peto\(^{21}\) argued that this observation is consistent with cancer simply being a stochastic process with a homogeneous baseline risk: “To ask why a particular individual failed to get cancer is probably as meaningless as asking why a particular uranium atom failed to decay.” Even if there were no interindividual variation, we would still see some smokers surviving to old age and other nonsmokers getting cancer at young ages. Of course, we have identified some differences in risk due to known genetic variants, health status, contextual factors, and so on, and family studies show that there must be more genetic variation still to be discovered.

The predictive value of a risk model is usually evaluated by its receiver operating characteristic curve. For Crohn’s disease and type I diabetes, the area under the curve (AUC) is substantial, but for most cancers, the predictive value of genetic risk scores is modest (AUCs in the range of 55%–70%),\(^{22–24}\) comparable to the AUCs obtained from models based only on family history or nongenetic factors, like the Gail models for breast\(^{25}\) or colorectal\(^{26}\) cancers. Furthermore, the addition of genetic markers to nongenetic risk factors is often modest, typically about a 5% improvement in the AUC.\(^{27,28}\) This will improve with larger consortia, but there may be a practical limit. For example, meta-analysis of height in 134,000 subjects has identified more than 100 variants, but they explain only about 7% of the total phenotypic variation; to raise that figure just to 15% would require a sample size of almost 500,000 individuals.\(^{29}\) Even the projected AUCs for all future single nucleotide polymorphisms will likely still be modest for most cancers.\(^{30,31}\)
Gene–Environment Interactions

G × E interactions have been studied for specific genes and specific exposures with limited success for several decades, the poster child being NAT2 and smoking for bladder cancer.\(^3\) Only recently have genome-wide interaction studies been done for specific exposures and metabolomic data used for environmental association studies\(^3\) and G × E-wide interaction studies.\(^3\)–\(^6\) More efficient statistical methods are being developed for such studies,\(^7\) but applications to date have been disappointing. Colorectal cancer is an obvious candidate, as there are 14 established environmental risk and protective factors, most of them modifiable. Based on these findings, several prevention trials have been launched,\(^8\)–\(^10\) aimed at prevention of further adenomas among individuals who have already had at least one. While not designed as genetically targeted trials, some postrandomization gene–treatment interaction analyses have been reported.\(^11\)–\(^13\) There have been numerous observational studies of G × E interactions with candidate genes,\(^14\)–\(^16\) but a comprehensive analysis of a genetic risk score composed of 27 genome-wide association study single nucleotide polymorphisms and these 14 environmental factors found no significant multiplicative interactions, and only weak additive interactions with height, processed meat, and hormone replacement therapy use. For other cancers, the yield of significant G × E interactions has been more rewarding: breast with body mass, age at menarche and parity,\(^17\) bladder with tobacco smoking,\(^18\) and lung with asbestos,\(^19\) as well as for other diseases like asthma,\(^20\) diabetes,\(^21\) and stroke.\(^22\)

Genetic Targeting for Primary Prevention

All smokers would benefit from quitting, but is that benefit any greater for those who are genetically at highest risk? A randomized placebo-controlled trial compared two smoking cessation treatments, the nicotine patch or varenicline, stratified by the nicotine metabolite ratio, a phenotypic assay of the activity level of the CYP2D6 enzyme that metabolizes nicotine and cotinine.\(^23\) They found that varenicline was more efficacious than the patch and had fewer side effects in normal metabolizers but not in slow metabolizers.

Genetic Targeting for Secondary Prevention

A recent perspective\(^24\) asked how much of the nearly 50% declines in colorectal cancer incidence and mortality since 1975 could be attributed to screening. Nine randomized controlled trials showed effectiveness of fecal occult blood testing and sigmoidoscopy, but they argued that much of the decline occurred before any effect of recent increases in screening. (Other simulations,\(^25\) however, have attributed about half of the decline to screening.) The discussed several possible explanations, including improvements in treatment and early detection for mortality and changes in risk factors for incidence. But neither article considered personalized prevention. Hsu et al,\(^26\) however, have provided a model for predicting the age at start of screening that would provide the same benefit conditional on gender, family history, and 27 genetic variants as would population-wide screening at age 50.

The complex dependence of an individual’s screening behavior on their own and family member’s screening histories could make screening look deleterious if high-risk individuals are more prone to get screened. In the companion article,\(^9\) I describe the application of causal inference methods to evaluating population-wide and targeted screening programs, using simulation and analysis of a large case–control study.\(^27\)–\(^29\) However, it is not clear that the predicted reductions in the number needed to screen by any of the targeted approaches compared with population-wide screening are enough to justify their greater complexity, which would require obtaining information on risk factors, family history, or genetics before deciding on an appropriate screening schedule. Although there may be a benefit in terms of reduced cancer incidence, it is not clear what the impact on mortality would be.\(^30\) This would require either a cohort study of cancer deaths in relation to screening (which would have the same problems of self-selection of screening behavior) or a randomized controlled trial. A full cost–benefit analysis would also have to address false negative and false positives, acceptability, the costs of targeting, and weigh the benefits of reduced incidence and mortality, all in relation to other screening modalities like occult blood and DNA-based tests.\(^31\)–\(^33\)

CONCLUSIONS

Although the genomics revolution has potential to transform prevention in addition to treatment, we still have a long way to go to more effectively identify who should be targeted. Environmental interventions like air pollution regulation cannot be targeted to any subgroup, genetic, or otherwise. Others, like antismoking campaigns, exercise, or diets, could in principle, but this might not be practical or cost efficient. The most efficient would be to identify those at high genetic sensitivity to avoidable exposures. But simply predicting genetic risks is not sufficient: we need evidence of G × E interaction. We should not let the enthusiasm for personalized medicine distract us from opportunities for classical public health approaches to prevention.

ABOUT THE AUTHOR

DUNCAN C. THOMAS is a Professor of Biostatistics and holds the Verna Richter Chair in Cancer Research at the USC/Norris Comprehensive Cancer Center. His research interests are in the development of statistical methods for genetic and environmental epidemiology.

REFERENCES

1. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015;372:793–795.
2. Hudson K, Lifton R, Patrick-Lake B, et al. The Precision Medicine Initiative Cohort Program—Building a Research Foundation for 21st Century Medicine. Bethesda, MD: National Institutes of Health; 2015.
3. Gillman MW, Hammond RA. Precision treatment and precision prevention: integrating “below and above the skin.” JAMA Pediatr. 2016;170:9–10.
4. Kensler TW, Spira A, Garber JE, et al. Transforming cancer prevention through precision medicine and immune-oncology. *Cancer Prev Res (Phila)*. 2016;9:2–10.

5. Khoury MJ, Evans JP. A public health perspective on a national precision medicine cohort: balancing long-term knowledge generation with early health benefit. *JAMA*. 2015;313:2117–2118.

6. Rebbeck TR. Precision prevention of cancer. *Cancer Epidemiol Biomarkers Prev*. 2014;23:2713–2715.

7. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 1985;14:32–38.

8. Rose G. The Strategy of Preventive Medicine. Oxford: Oxford University Press; 1992.

9. Thomas DC. Estimating the effect of targeted screening strategies on cancer outcomes: an application to colonoscopy and colorectal cancer. *Epidemiology*. 2016. [Epub ahead of print]

10. Greenland S. Randomization, statistics, and causal inference. *Epidemiology*. 1990;1:421–429.

11. Hernán MA, Robins J. Causal Inference. Boca Raton, FL: Chapman & Hall/CRC; 2013.

12. Pearl J. Causality: Models, Reasoning, and Inference. Cambridge, MA: Cambridge University Press, 2000.

13. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.

14. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med*. 2007;26:20–36.

15. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health*. 2005;95(suppl 1):S144–S150.

16. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11:561–570.

17. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed as randomized intermediate outcomes. *Biostatistics*. 2012;13:289–302.

18. Dominici F, Greenstone M, Sunstein CR. Science and regulation. *Int J Epidemiol*. 2011;35:506–514.

19. Thomas DC, Greenstone M, Sunstein CR. Estimating causal effects of air quality exposure on health outcomes. *Cancer Epidemiol Biomarkers Prev*. 2008;17:2625–2631.

20. Baron JA, Beach M, Mandal JS, et al. Calcium supplements and colorectal adenomas. Polyp Prevention Study Group. *Ann N Y Acad Sci*. 1999;889:138–145.

21. Baron JA, Beach M, Mandal JS, et al. Calcium supplements for the prevention of colorectal adenomas. *Calcium Polyp Prevention Study Group*. *N Engl J Med*. 1999;340:101–107.

22. Baron JA, Sandler RS, Haile RW, Mandal JS, Mott LA, Greenberg ER. Folate intake, alcohol consumption, cigarette smoking, and risk of colorectal adenomas. *J Natl Cancer Inst*. 1998;90:57–62.

23. Cole BF, Baron JA, Sandler RS, et al; Polyp Prevention Study Group. Folic acid for the prevention of colorectal adenomas: a randomized clinical trial. *JAMA*. 2007;297:2351–2359.

24. Martinez ME, O’Brien TG, Fultz KE, et al. Pronounced reduction in adenoma recurrence associated with aspirin use and a polymorphism in the ornithine decarboxylase gene. *Proc Natl Acad Sci U S A*. 2003;100:7859–7864.

25. Martinez ME, Marshall JR, Giovannucci E. Diet and cancer prevention: the roles of observation and experimentation. *Nat Rev Cancer*. 2008;8:694–703.

26. Martinez ME, Giovannucci E, Jiang R, et al. Folate fortification, plasma folate, homocysteine and colorectal adenoma recurrence. *Int J Cancer*. 2006;119:1440–1446.

27. Wertheim BC, Smith JW, Fang C, Alberts DS, Lance PA. Risk modification of colorectal adenomas by CYP7A1 polymorphisms and the role of bile acid metabolism in carcinogenesis. *Cancer Prev Res (Phila)*. 2012;5:197–204.

28. Martinez ME, Thompson P, Jacobs ET, et al. Dietary factors and biomarkers involved in the methylenetetrahydrofolate reductase genotype–colorectal adenoma pathway. *Gastroenterology*. 2006;131:1706–1716.

29. Korver ED, Hutter CM, Minnier J, et al. Gene-environment interactions, and risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev*. 2010;19:3167–3173.

30. Weider RC. Aspirin and NSAID chemoprevention, gene-environment interactions, and risk of colorectal cancer. *JAMA*. 2015;313:1111–1122.

31. Hutter CM, Chang-Claude J, Slattery ML, et al. Characterization of gene-environment interactions for colorectal cancer susceptibility loci. *Cancer Res*. 2012;72:2036–2044.
54. Figueiredo JC, Lewinger JP, Song C, et al. Genotype-environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. *Cancer Epidemiol Biomarkers Prev*. 2011;20:758–766.
55. Schoeps A, Rudolph A, Seibold P, et al. Identification of new genetic susceptibility loci for breast cancer through consideration of gene-environment interactions. *Genet Epidemiol*. 2014;38:84–93.
56. Figueroa JD, Han SS, Garcia-Closas M, et al. Genome-wide interaction study of smoking and bladder cancer risk. *Carcinogenesis*. 2014;35:1737–1744.
57. Liu CY, Stücker I, Chen C, et al. Genome-wide gene-asbestos exposure interaction association study identifies a common susceptibility variant on 22q13.31 associated with lung cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2015;24:1564–1573.
58. Wei S, Wang LE, McHugh MK, et al. Genome-wide gene-environment interaction analysis for asbestos exposure in lung cancer susceptibility. *Carcinogenesis*. 2012;33:1531–1537.
59. Meyers DA, Bleecker ER, Holloway JW, Holgate ST. Asthma genetics and personalised medicine. *Lancet Respir Med*. 2014;2:405–415.
60. Kauffmann F, Demenais F. Gene-environment interactions in asthma and allergic diseases: challenges and perspectives. *J Allergy Clin Immunol*. 2012;130:1229–1240.
61. Cornelis MC, Tchetgen EJ, Liang L, et al. Gene-environment interactions in genome-wide association studies: a comparative study of tests applied to empirical studies of type 2 diabetes. *Am J Epidemiol*. 2012;175:191–202.
62. Marceau R, Lu W, Holloway S, et al. A fast multiple-kernel method with applications to detect gene-environment interaction. *Genet Epidemiol*. 2015;39:456–468.
63. Lerman C, Schnoll RA, Hawk LW Jr, et al; PGRN-PNAT Research Group. Use of the nicotine metabolite ratio as a genetically informed biomarker of response to nicotine patch or varenicline for smoking cessation: a randomised, double-blind placebo-controlled trial. *Lancet Respir Med*. 2015;3:131–138.
64. Welch HG, Robertson DJ. Colorectal cancer on the decline—why screening can’t explain it all. *N Engl J Med*. 2016;374:1605–1607.
65. Zauber AG. The impact of screening on colorectal cancer mortality and incidence: has it really made a difference? *Dig Dis Sci*. 2015;60:681–691.
66. Hsu L, Jeon J, Brenner H, et al; Colorectal Transdisciplinary (CORECT) Study; Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO). A model to determine colorectal cancer risk using common genetic susceptibility loci. *Gastroenterology*. 2015;148:1330–1339.e14.
67. Brenner H, Chang-Claude J, Jansen L, Knebel P, Stock C, Hoffmeister M. Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. *Gastroenterology*. 2014;146:709–717.
68. Brenner H, Chang-Claude J, Jansen L, Seiler CM, Hoffmeister M. Colorectal cancers occurring after colonoscopy with polyp detection: sites of polyps and sites of cancers. *Int J Cancer*. 2013;133:1672–1679.
69. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. *N Engl J Med*. 2013;369:1106–1114.
70. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014;370:1287–1297.
71. Quintero E, Castells A, Bujanda L, et al; COLONPREV Study Investigators. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. *N Engl J Med*. 2012;367:1297–1297.
72. Lee JK, Lilis EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. *Ann Intern Med*. 2014;160:171.