BY KATHARINE MILLER

BIG DATA FEEDS UNDERSTANDING OF OBESITY

How data about genes, physical activity, and the environment are yielding insights about obesity

Despite extensive efforts, a clear understanding of the obesity epidemic remains elusive. Scientists have implicated specific causal genes (40 to 60 percent of obesity is considered heritable); epigenetic effects (environmental changes to genes); and the microbiome (though this is controversial). Behavioral causes such as diet and exercise clearly matter, but behavior is hard to change; many people lose weight only to regain it, or begin an exercise regimen only to quit.

With so many variables at play and no one-size-fits-all solution, some researchers are turning to big data. "Obesity lends itself well to big-data approaches that can bring out the real signals despite noise and heterogeneity," says **Elizabeth Speliotes**, **MD**, **PhD**, **MPH**, associate professor of internal medicine and computational medicine & bioinformatics, University of Michigan Medical School.

Big data is now beginning to cut through some of obesity's opacity, including clarifying roles played by genetics, physical activity, and the environment. But relating all of the different and complex variables will require more data and, in particular, more integrated data.

Obesity and Genes

Body mass index (BMI—weight over height squared) is considered a useful way to measure overall obesity, while waist-to-hip ratio (WHR) captures obesity that is centered in the abdomen. In two papers published simultaneously in *Nature* in 2015, Speliotes and her colleagues conducted genome-wide association meta-analyses of BMI in nearly 340,000 people, and of WHR (adjusted for BMI) in nearly 225,000 individuals. They found that a different set of gene variants predispose a person to a high BMI compared to those associated with a high WHR. The genes linked to high BMI were largely based

in the nervous system whereas those linked to high WHR related to adipose differentiation—i.e., the development of fatty tissue. High WHR is also more closely linked to health-related complications such as diabetes and cardiovascular disease. "There may be a more protective effect of depositing the fat subcutaneously rather than in the belly," Speliotes says. Researchers are currently exploring what the discovered genes do, how they do it, and how they are connected to each other.

Her team is also finding that the complications of obesity

Different genetic loci are associated with different measures of obesity, as shown in this Venn diagram, which identifies genetic loci associated with BMI, body fat percentage, waist-hip ratio adjusted for BMI (WHRadjBMI), visceral adipose tissue (VAT), subcutaneous adipose tissue (SAT), the VAT/SAT ratio and extremes of BMI and WHR. Reprinted with permission from Fall T, Mendelson M, Speliotes E, Recent Advances in Human Genetics and Epigenetics of Adiposity: Pathway to Precision Medicine? Gastroenterology 152:7:1695-1706 (May 2017).
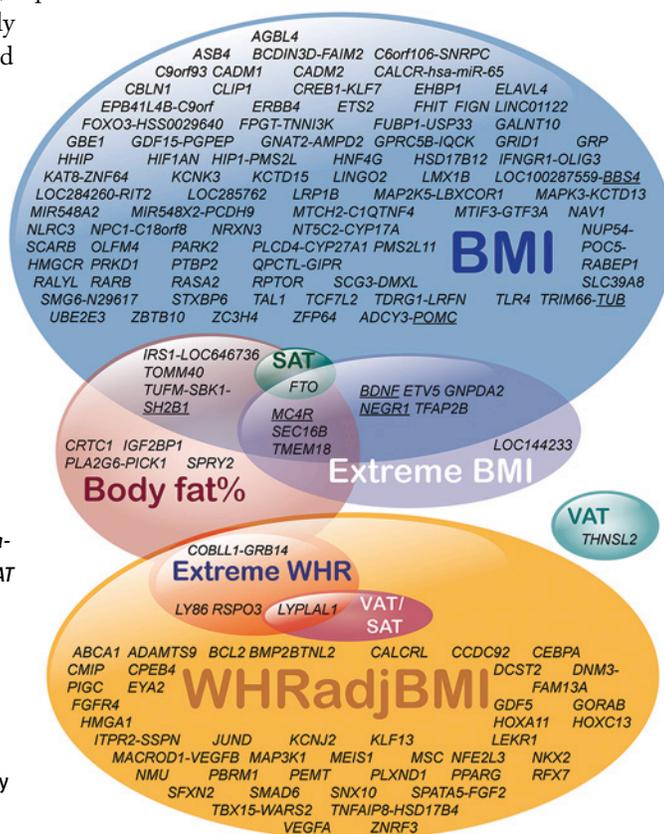
don't always follow from a predisposition to obesity. Other genes may sometimes intervene to provide either a protective or detrimental effect. With more and bigger datasets will come a better sense of how these interactions work, Speliotes says. "The combination of genes will be more and more predictive."

Obesity and Physical Activity

Historically, most studies of physical activity and obesity were relatively small;

relied on individuals' self-reports, which aren't entirely reliable; and reported the respondents' average activity levels—resulting in a further loss of information.

Wearable activity monitors have changed all that. Sample sizes have blossomed and detailed information is abundant. Researchers at Stanford University's

Mobilize Center recently analyzed 68 million days of physical activity data for 717,527 people in 111 countries collected from a smartphone fitness application (Azumio's). The study, published in *Nature* in 2017, found that the disparity of physical activity distribution within a country, which they dubbed "activity inequality," is a better predictor of a country's obesity prevalence than average activity levels.

"The activity inequality concept really

jumped out at us," says **Tim Althoff**, a PhD candidate in computer science at Stanford University. When the team plotted individuals' steps per day in each location or country, a routine calculation done just to look for outliers, he says, "some of these curves were much wider than others." In Saudi Arabia, for example, there are fewer people on the low end of the norm, while in Japan, the curves are narrower with relatively few people walking way more or way less per day than typical. And the biggest

revelation: "We found that in countries that are more unequal in activity, activity in females is reduced disproportionately." When they modeled possible interventions, it turned out that "if you just lifted women's activity levels to the level of their male counterparts, you would cut activity inequality by half," Althoff says.

Activity data can also be connected to genetics. For example, **Misa Graff, PhD**, research assistant professor of epidemiology at the University of North Carolina Gillings School of Global

Public Health, recently did a meta-analysis of results from 60 genome-wide association studies (GWAS) covering more than 200,000 adults, looking at the interaction between obesity genes and physical activity. Although her data for physical activity was the old-fashioned self-reported kind, she did identify genes that are influenced by physical activity and, by controlling for physical activity, new obesity-related genes.

Still, she says she's "not very satisfied, actually…. We could find more if we had better measures of physical activity." Graff would love to use data from smartphones to look at how active people really are and coordinate that with eating and sleeping habits. "Without data that is harmonized across individuals, you can't get at a lot of the questions about genes," she says.

Obesity and the Environment

Big data could be useful in understanding obesity-environment connections at multiple scales: the epigenetic level (the ways that genes are turned on and off in response to environmental influences); the microbiome level; and even at the urban infrastructure level.

When it comes to how the microbiome and the epigenome affect obesity, the
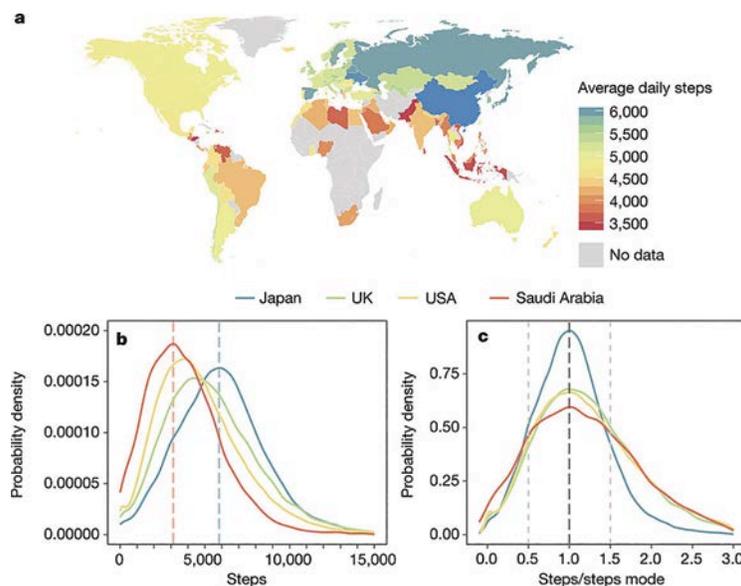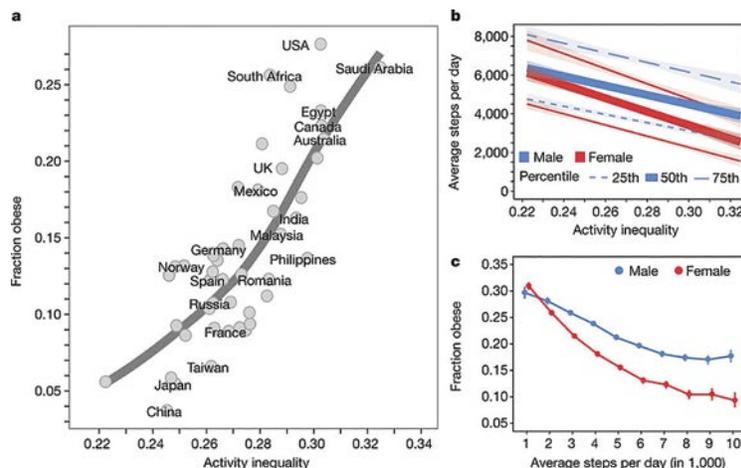

Figure 1

Figure 2

jury is still out. In a recent meta-analysis of the gut microbiome, researchers at the University of Michigan found only a weak association between the microbial communities found in human feces and obesity status. And although epigenetic changes might influence obesity risk, the specifics are still unclear, Speliotes says. Moreover, she notes, obesity itself might have consequences for epigenetics. Indeed, in a 2017 *Nature* paper, an epigenome-wide association study of over 10,000 people found that BMI is associated with widespread changes in DNA methylation (a common epigenome marker), and that these changes predict future development of type 2 diabetes.

On the infrastructure front, however, the evidence is clearer: Certain neighborhood attributes—such as high poverty and crime, a lack of physical activity amenities, and a lack of transportation and recreation infrastructure—are associated with lower physical activity and higher obesity. But Graff found that, in a cohort of nearly 8,000 adolescents, physical activity helped reduce BMI regardless of the neighborhood-level factors. "Genes influence the BMI at the given activity level regardless of where you live," Graff says. While this result might seem unsurprising, it reinforces the value of higher levels of physical activity in obesogenic neighborhoods—a policy promoted by

Althoff and his colleagues found that higher walkability scores (from WalkScore.com) for 69 US cities are associated with lower activity inequality (a). Highly walkable cities show greater spikes in weekday walking during commute and lunch hours (b) as well as on weekends during daytime hours (c). Higher walkability is also associated with more daily steps across age, gender and BMI groups (d). Bars show the steps gained per day for each point increase in walkability score for 24 cities in the US. Reprinted by permission from Macmillan Publishers Ltd: Althoff T, Sosič R, Hicks JL, King AC, Delp SL, & Leskovec J,, Large-scale physical activity data reveal worldwide activity inequality, Nature 547, 336–339 (2017).

whether or not a walkable city actually causes higher physical activity levels, only that the two are related. To get at the causality question, Althoff will look at people in his dataset who move between cities to see if activity and obesity status change in response to the walkability of the new location. In future work, Althoff

regarding how physical activity changes with all kinds of factors," he says.

The Big Data Difference
One-third of the world is now obese; consumer wearables are collecting activity and diet information globally; and genetic and genomic datasets are explod-

One-third of the world is now obese; consumer wearables are collecting activity and diet information globally; and genetic and genomic datasets are exploding as well. Can all of these sources of big data converge to help us make sense of the various causes and potential solutions to this complex condition called obesity?

former first lady Michelle Obama.

Althoff also considered infrastructure in his *Nature* paper. He and his colleagues found that physical activity levels were higher (and activity inequality was lower) in more walkable cities. However, their analysis cannot reveal

also wants to look at food and nutrition patterns using data from people who track that on their devices; as well as how weather or changes to a transportation network affect activity levels. "Sensor-equipped smartphones and watches allow us to get into more detail

ing as well. Can all of these sources of big data converge to help us make sense of the various causes and potential solutions to this complex condition called obesity? It remains to be seen, but these data sources are certainly feeding an insatiable appetite for understanding. □