



JAN 6, 2016 AT 6:00 AM

You Can't Trust What You Read About Nutrition

We found a link between cabbage and innie bellybuttons, but that doesn't mean it's real.

By Christie Aschwanden

Filed under Nutrition



Photographs by Anna Maria Barry-Jester

As the new year begins, millions of people are vowing to shape up their eating habits. This usually involves dividing foods into moralistic

categories: good/bad, healthy/unhealthy, nutritious/indulgent, slimming/fattening — but which foods belong where depends on [whom you ask](#).

The [U.S. Dietary Guidelines Advisory Committee](#) recently released its [latest guidelines](#), which [define a healthy diet](#) as one that emphasizes vegetables, fruits, whole grains, low- or nonfat dairy products, seafood, legumes and nuts while reducing red and processed meat, refined grains, and sugary foods and beverages.¹ Some cardiologists recommend a [Mediterranean diet rich in olive oil](#), the American Diabetes Association gives the nod to both [low-carbohydrate and low-fat diets](#), and the [Physicians Committee for Responsible Medicine](#) promotes a [vegetarian diet](#). Ask a hard-bodied CrossFit aficionado, and she may champion a “[Paleo](#)” diet based on foods our Paleolithic ancestors (supposedly) ate. My colleague [Walt Hickey](#) swears by the [keto diet](#).

Who's right? It's hard to say. When it comes to nutrition, everyone has an opinion. What no one has is an airtight case. The problem begins with a lack of consensus on what makes a diet healthy. Is the aim to make you slender? To build muscles? To keep your bones strong? Or to prevent heart attacks or cancer or keep dementia at bay? Whatever you're worried about, there's no shortage of diets or foods purported to help you. Linking dietary habits and individual foods to health factors is easy — ridiculously so — as you'll soon see from the little experiment we conducted.



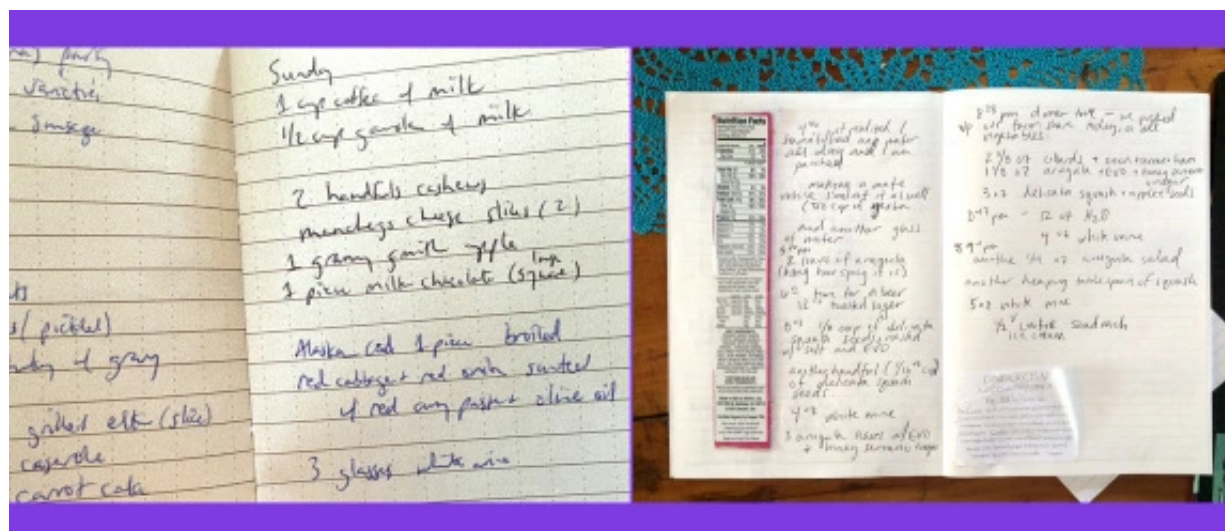
Our foray into nutrition science demonstrated that studies examining how foods influence health are inherently fraught. To show you why, we're going to take you behind the scenes to see how these studies are done. The first thing you need to know is that nutrition researchers are studying an incredibly difficult problem, because, short of locking people in a room and carefully measuring out all their meals, it's hard to know exactly what people eat. So nearly all nutrition studies rely on measures of food consumption that require people to remember and report what they ate. The most common of these are food diaries, recall surveys and the food frequency questionnaire, or FFQ.

Several versions of the FFQ exist, but they all use a similar technique: Ask people how often they eat particular foods and what serving size they usually consume. But it's not always easy to remember everything you ate, even what you ate yesterday. People are prone to underreport what they consume, and they may not fess up to eating certain foods or may miscalculate their serving sizes.

“The bottom line here is that doing dietary assessment is difficult,” said [Torin Block](#), CEO of NutritionQuest, a company that conducts FFQs and was founded by his mother, [Gladys Block](#), a pioneer in the field who began

developing food frequency questionnaires at the National Cancer Institute. “You can’t get away from it — there’s error involved.” Still, there’s a pecking order in terms of completeness, he said. Food diaries rank high and so do 24-hour food recalls, in which an administrator sits the subject down for a guided interview to catalog everything eaten in the past 24 hours. But, Block said, “you really need to do multiple administrations to get an assessment of someone’s usual long-term dietary intake.” For study purposes, researchers are not usually interested just in what people ate yesterday or the day before, but in what they eat regularly. Studies that use 24-hour recalls tend to under- or overestimate nutrients people don’t eat every day, since they record only a small and perhaps unrepresentative snapshot.

When I tried keeping a seven-day food diary, I discovered how right Block was — it’s surprisingly difficult to capture a record that reflects normal eating patterns when you collect only a few days’ worth of data. It so happened that I was traveling to a conference during my diary week, so I ate packaged snacks and restaurant meals far different from the foods I usually eat from my garden at home. My diary showed that before dinner one day, I’d eaten only a doughnut and two snack packs of potato chips. And what did I have for dinner? I can tell you that it was a delicious Indonesian seafood curry, but I couldn’t possibly begin to list all its ingredients.



Pages from the author’s and a colleague’s food diaries.

Another lesson from my short stint keeping a food diary is that the sheer act of keeping track can change what you eat. When I knew I had to write it

down, I paid far greater attention to how much I ate, and that sometimes meant that I opted not to eat something because I felt too lazy to write it down or else realized, nah, I didn't really want a second doughnut (or else didn't want to admit to eating it).

It's not easy to circumvent the human instinct to fib about what we eat, but the FFQ aims to overcome the unrepresentativeness of short-term food records by assessing what people consume over a longer period. When you read a headline saying something like “[blueberries prevent memory loss](#),” the evidence usually comes from some version of the FFQ. The questionnaire typically asks about what the survey-taker ate during the last three, six or 12 months.

In order to get a sense of how these surveys work and how reliable they might be, we hired Block to administer his company's six-month FFQ to me, my colleagues [Anna Barry-Jester](#) and [Walt Hickey](#), and a group of reader volunteers.²

Some questions — how often do you drink coffee? — were straightforward. Others confounded us. Take tomatoes. How often do I eat those in a six-month period? In September, when my garden is overflowing with them, I eat cherry tomatoes like a child devours candy. I might also eat two or three big [purple Cherokees](#) drizzled with balsamic and olive oil per day. But I can go November until July without eating a single fresh tomato. So how do I answer the question?

Questions about serving sizes perplexed us all. In some cases, the survey provided weird but helpful guides — for example, it depicted what a half-cup, one cup or two cups of yogurt looked like with photographs of bowls filled with various amounts of wood chips. Other questions seemed absurd. “Who on this planet knows what a cup of salmon or two cups of ribs looks like?” Walt asked.

Although the questionnaire was meant simply to measure our food intake, at times it felt judgmental — did we take our milk full fat, low fat or fat free? I

noticed that when I was offered three choices of serving sizes, my inclination was to pick the middle one, regardless of what my actual portion might be.

Despite these challenges, Anna, Walt and I did our best to answer completely and honestly. Afterward, we compared our results. The questionnaire deemed “cheese, full fat” and some version of alcohol as our top sources of calories.³

From there, our diets diverged. Walt has lost 50 pounds on a ketogenic diet, Anna eats relatively little protein and, according to the FFQ, I devour almost twice the calories as either of them.

Could these results be correct? Anna and I are virtually the same height and weight; we could probably share clothes. How could I eat more than twice the calories she does?⁴ Block acknowledged that it’s difficult to get an accurate count of calories, especially without a long-term food record, and when you start looking at individual nutrients it gets even trickier. He pointed me to a [1987 study](#) concluding that to estimate a true average calorie count, it takes an average of 27 days of daily intake data for men and 35 days for women. Some nutrients required even longer — 474 days on average to measure vitamin A intake for women, for example. This suggests our reports might be correct, but they might also contain lots of errors.



Sure, memory-based measures have limitations, said [Brenda Davy](#), a professor of human nutrition at Virginia Tech, “but most of us in the nutrition world still believe they have value.” Calories are probably the

Science & Health

Share on Facebook

[people underreport foods deemed unhealthy](#), the high fat foods or sugary snacks. “But that doesn’t mean that *everything* is underreported. It doesn’t mean that fiber intake or calcium intake is problematic.”

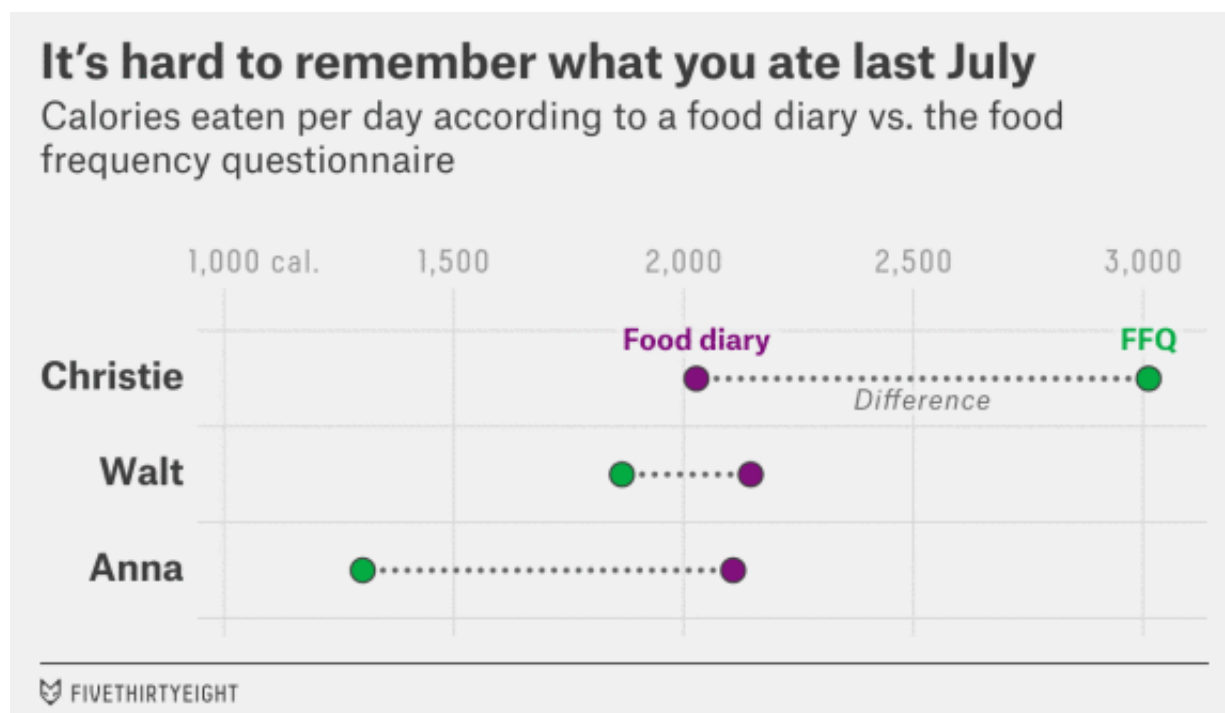
Developers of the surveys recognize that answers are imperfect, and they correct for this with validation studies that check FFQ results against those

obtained via other methods, [usually a 24-hour food recall](#) or longer food diary. The results of such validation studies, Block said, allow researchers to account for variability in daily intake.

Critics of FFQs, such as [Edward Archer](#), a computational physiologist at the University of Alabama's Nutrition Obesity Research Center in Birmingham, say that these validations are nothing more than [circular reasoning](#). "You're taking one type of subjective report and validating it with another form of subjective report," he said.

Recording what you eat is harder than it might seem, said Tamara Melton, a registered dietitian and [spokesperson](#) for the Academy of Nutrition and Dietetics in Atlanta. Among other things, it's almost impossible to measure ingredients and portion sizes when you dine out. "It's cumbersome. If you're out at a business lunch, you can't whip out your measuring cup."

When Anna, Walt and I compared the caloric intakes that our FFQs had spit out with the ones that we calculated from our seven-day food diaries,⁵ they didn't match up. We ran into trouble estimating portions in the FFQ, too, and who's to say which was more accurate?



Although concerns about self-reported dietary intakes have been around for decades, the debate has come to a head in recent years, said [David Allison](#),

director of the University of Alabama's Nutrition Obesity Research Center in Birmingham. Allison was an author of a [2014 expert report](#) from the Energy Balance Measurement Working Group that called it “unacceptable” to use “decidedly inaccurate” methods of measurement to set health care policies, research and clinical practice. “In this case,” the researchers wrote, “the adage ‘something is better than nothing’ must be changed to ‘something is worse than nothing.’”

The problems with food questionnaires go even deeper. They aren't just unreliable, they also produce huge data sets with many, many variables. The resulting cornucopia of possible variable combinations makes it easy to [p-hack your way to sexy \(and false\) results](#), as we learned when we invited readers to take an FFQ and answer a few other questions about themselves. We ended up with 54 complete responses and then looked for associations — much as researchers look for links between foods and dreaded diseases. It was silly easy to find them.

Our shocking new study finds that ...

EATING OR DRINKING	IS LINKED TO	P-VALUE
Raw tomatoes	Judaism	<0.0001
Egg rolls	Dog ownership	<0.0001
Energy drinks	Smoking	<0.0001
Potato chips	Higher score on SAT math vs. verbal	0.0001
Soda	Weird rash in the past year	0.0002
Shellfish	Right-handedness	0.0002
Lemonade	Belief that “Crash” deserved to win best picture	0.0004
Fried/breaded fish	Democratic Party affiliation	0.0007
Beer	Frequent smoking	0.0013
Coffee	Cat ownership	0.0016
Table salt	Positive relationship with Internet service provider	0.0014
Steak with fat trimmed	Lack of belief in a god	0.0030
Iced tea	Belief that “Crash” didn't deserve to win best picture	0.0043

Bananas	Higher score on SAT verbal vs. math	0.0073
Cabbage	Innie bellybutton	0.0097

SOURCE: FFQ & FIVETHIRTYEIGHT SUPPLEMENT

The FFQ we used produced 1,066 variables, and the additional questions we asked sorted survey-takers according to 26 possible characteristics (left- or right-handed, for example). This vast data set allowed us to do 27,716 regressions in just a few hours. (You can see the full results on [GitHub](#).) With that many possibilities to examine, we were guaranteed to find some “[statistically significant](#)” correlations that aren’t real, said [Veronica Vieland](#), a statistician who directs the [Battelle Center for Mathematical Medicine](#) at Nationwide Children’s Hospital in Columbus, Ohio. Using a [p-value](#) of 0.05 or less as the metric for statistical significance (as is common) equates to an error rate of 5 percent, Vieland said. And with 27,716 regressions, that means we should expect about 1,386 false positives.⁶

But false positives aren’t the only issue. It was also very likely that we’d discover real correlations that are scientifically useless, Vieland said. For instance, our experiment found that people who trim the fat from their steaks were more likely to be atheists than those who ate the fat that god had provided for them. It’s possible that there’s a real correlation between cutting the fat from meat and being an atheist, Vieland said, but that doesn’t mean that it’s a causal one.

A preacher who advised parishioners to avoid trimming the fat from their meat, lest they lose their religion, might be ridiculed, yet nutrition epidemiologists often make recommendations based on similarly flimsy evidence. A few years back, [Jorge Chavarro](#), a nutritional epidemiologist at the Harvard School of Public Health, advised that women trying to conceive [consider swapping low-fat dairy foods for high-fat dairy products such as ice cream](#), based on FFQ data from [an ongoing study of nurses](#). He and his colleague [Walter Willett](#) also wrote a book [promoting a “fertility diet”](#) based on the results. When I reached Chavarro this week to ask how confident he was in the link between dairy intake and fertility, he said that “of all the associations we found, this is the one we had the least confidence in.” It’s

also, of course, the one that made headlines.

Nearly every nutrient you can think of has been linked to some health outcome in the peer-reviewed scientific literature using tools like the FFQ, said [John Ioannidis](#), an [expert on the reliability of research findings](#) at the [Meta-Research Innovation Center at Stanford](#). In a [2013 analysis](#) published in the American Journal of Clinical Nutrition, Ioannidis and a colleague selected 50 common ingredients at random from a cookbook and looked for studies evaluating each food's association to cancer risk. It turned out that studies had found a link between [80 percent of the ingredients](#) — including salt, eggs, butter, lemon, bread and carrots — and cancer. Some of those studies pointed to an increased risk of cancer, others suggested a decreased risk, but the size of the reported effects were “implausibly large,” Ioannidis said, while the evidence was weak.

Foods that may or may not give you cancer

Risk estimates for 20 foods (each studied at least 10 times) from a 2012 meta-analysis



One outlier study not shown (corn, risk estimate of 19.43).

FIVETHIRTYEIGHT

SOURCE: AMERICAN JOURNAL OF CLINICAL NUTRITION

But the problems weren't just statistical. Many of the reported findings were also biologically improbable, Ioannidis said. For instance, a 2013 study found that people who ate [three servings of nuts per week had a nearly 40 percent reduction in mortality risk](#). If nibbling nuts really cut the risk of dying by 40 percent, it would be revolutionary, but the figure is almost certainly an overstatement, Ioannidis told me. It's also meaningless without context. Can a 90-year-old get the same benefits as a 60-year-old? How many days or years must you spend eating nuts for the benefits to kick in,

and how long does the effect last? These are the questions that people really want answers to. But as our experiment demonstrated, it's easy to use nutrition surveys to link foods to outcomes, yet it's difficult to know what these connections mean.

FFQs “aren’t perfect,” said Harvard’s Chavarro, but at the moment there are few other options. “It may be that we have reached a limit of current methodology for nutritional assessments and it’s going to require a major shift to do something better,” he said.

Current studies suffer another fundamental problem: We expect far too much from them. We want to answer questions like, what’s healthier, butter or margarine? Can eating blueberries keep my mind sharp? Will bacon give me colon cancer? But observational studies using memory-based measures of dietary intake are tools too crude to provide answers with this level of granularity.

One reason is that single nutrients like saturated fat or an antioxidant seem to produce only trivial differences in the absolute risk of disease, Ioannidis said. (His [conclusion](#) comes from more rigorous randomized trials.) This is why headlines so often report *relative* risks — how many people got cancer in the group who ate the most bacon compared with those who ate none. Relative risks are almost always much more extreme than absolute risk, but absolute risk (your [risk of getting cancer if you consume bacon](#), for instance) is what we really care about. If, say, 1 out of 10,000 people who ate the most bacon got cancer, compared with 3 out of 10,000 who ate none, that’s a threefold difference. But the difference in absolute risk — a 0.01 percent chance of cancer versus 0.03 percent — is tiny and probably not enough to change anyone’s eating habits.

The tendency to report results as more precise and important than they are also explains why we get so many back-and-forth headlines about things like coffee. “Big data sets just confer spurious precision status to noise,” Ioannidis wrote in [his 2013 analysis](#).

So we’re left with our original question: What is a healthy diet? We know the basics — we need sufficient calories and protein to keep our bodies alive. We

need nutrients like vitamin C and iron. Beyond that, we may be overthinking it, said Archer, the Nutrition Obesity Research Center physiologist. “We have cultures that eschew fruits and vegetables that were perfectly healthy for thousands of years,” he said. Some populations today thrive [on very few vegetables](#), while others subsist almost entirely on plant foods. The takeaway, Archer said, is that our bodies are adaptable and pretty good at telling us what we need, if we can learn to listen.

Even so, I doubt we'll give up looking for secret health elixirs in our pantries and refrigerators. There's a reason the media and the public gobble up these studies, and it's the same reason that researchers spend billions of dollars doing them. We live in a world where scary diseases constantly strike people around us, sometimes out of the blue. The natural reaction when someone has a heart attack or is diagnosed with cancer is to look for a way to protect yourself from a similar fate. So we turn to food to regain a modicum of control. We can't direct what's going on inside our cells, but we can control what we put into our bodies. Science has yet to find a magic vitamin or nutrient that will allow us to stay healthy forever, but we seem determined to keep trying.

CORRECTION (Jan. 6, 1:10 p.m.): A previous version of this article incorrectly described the affiliation of the Energy Balance Measurement Working Group, which wrote a report on obesity research methods. It is not affiliated with the National Cancer Institute, although there is another group with a similar name that is affiliated with the institute.

The problem with nutrition studies

***Christie Aschwanden** reported and wrote this story and discovered two new foods — hush puppies and [cheese straws](#) (her new obsession) — in the process. **Anna Maria Barry-Jester** contributed reporting and photography. She also learned how hard it is to calculate the calories in gyro meat. **Andrew Flowers** p-hacked the hell out of our data, against his better judgment. For our survey, **Walt Hickey** identified important unanswered questions about the relationship between certain foods and bellybuttons, weird rashes and opinions of the movie “Crash.”*

Footnotes

1. The guidelines stirred immediate [controversy](#). An editorial in the medical journal BMJ concluded that they [lacked rigorous evidence](#), a claim that committee members [disputed](#). [^]
2. Our “study” was just a rough experiment to explore how typical study methods work. We won’t be attempting to publish it in any scientific journal. [^]
3. Here are our top three sources of calories, as measured by the FFQ.
 - Christie: cheese (full fat), both red and white wine, oatmeal.
 - Anna: cheese (full fat), beer, mac & cheese/cheese dishes.
 - Walt: cheese (full fat), liquor/cocktails, peanuts, and other nuts and seeds.

[^]

4. Sure, our levels of exercise may vary, and other things like body composition may factor in,

but the level of disparity here shocked us both. [^]

5. Our calculations were akin to those you might get from a calorie-counting phone app. For packaged foods, we referenced the nutrition facts labels. For foods prepared at home or in restaurants, we referenced a variety of online sources, including [WolframAlpha](#) and [myfitnesspal](#), either for calories in similar recipes or the calories of individual ingredients. Studies involving food diaries typically follow a more rigorous process that includes detailed interviews to help participants remember and estimate what they ate and a more complicated nutrient analysis, often based on data from the U.S. Department of Agriculture [National Nutrient Database for Standard Reference](#). Our numbers do not represent a comprehensive analysis of the calories we consumed. [^]
6. This error rate reflects only false positives due to sampling variability; it doesn't say anything about other sources of error, such as inaccurate data. Statistician David Colquhoun has estimated that, in practice, [a p-value of 0.05 yields at least a 30 percent error rate](#), and usually more. [^]

Christie Aschwanden is FiveThirtyEight's lead writer for science. | [✉](#) | [@cragcrest](#)

FILED UNDER FOOD, HEALTHY EATING, NUTRITION