On the Origins of Type 1 Diabetes

An original hypothesis

STEPHEN SKOLNICK
JUN 30, 2023

Type 1 Diabetes is a weird disease.

Biologically, it's the classic picture of an autoimmune disorder, and this is simple enough: your immune system gets some bad intel, arrives at the conclusion that your pancreas is harboring terrorists, and starts doing drone strikes on little clumps of pancreatic cells called the "Islets of Langerhans". That's a shame, because these are the only cells in your body that make insulin.

This kind of thing mainly happens early on in life. After about age 15, it seems your immune system has a pretty good handle on what all the different kinds of cells in your body are supposed to look like, so that it's not easily duped into engaging in “friendly fire”.

The *what* and the *when* of T1D are not the weird part. Even the *how* is pretty well understood: the immune system churns out antibodies that bind to certain proteins produced in the pancreas, but not in other cell types. Eventually, the body identifies the islet cells as the source of these proteins, and targets them for destruction, because things that antibodies stick to are generally signs of a foreign invader. We have a pretty good idea of which proteins piss off the immune system—and even which *parts* of these proteins, since the antibodies in T1D patients’ blood mostly bind to certain specific chunks of them.

But looking at a disease at the molecular level can only get you so far in understanding *why* it happens. If we want to better understand the root cause, we have to zoom out, and look not at *how*, but at *where* it happens…and where it doesn’t.

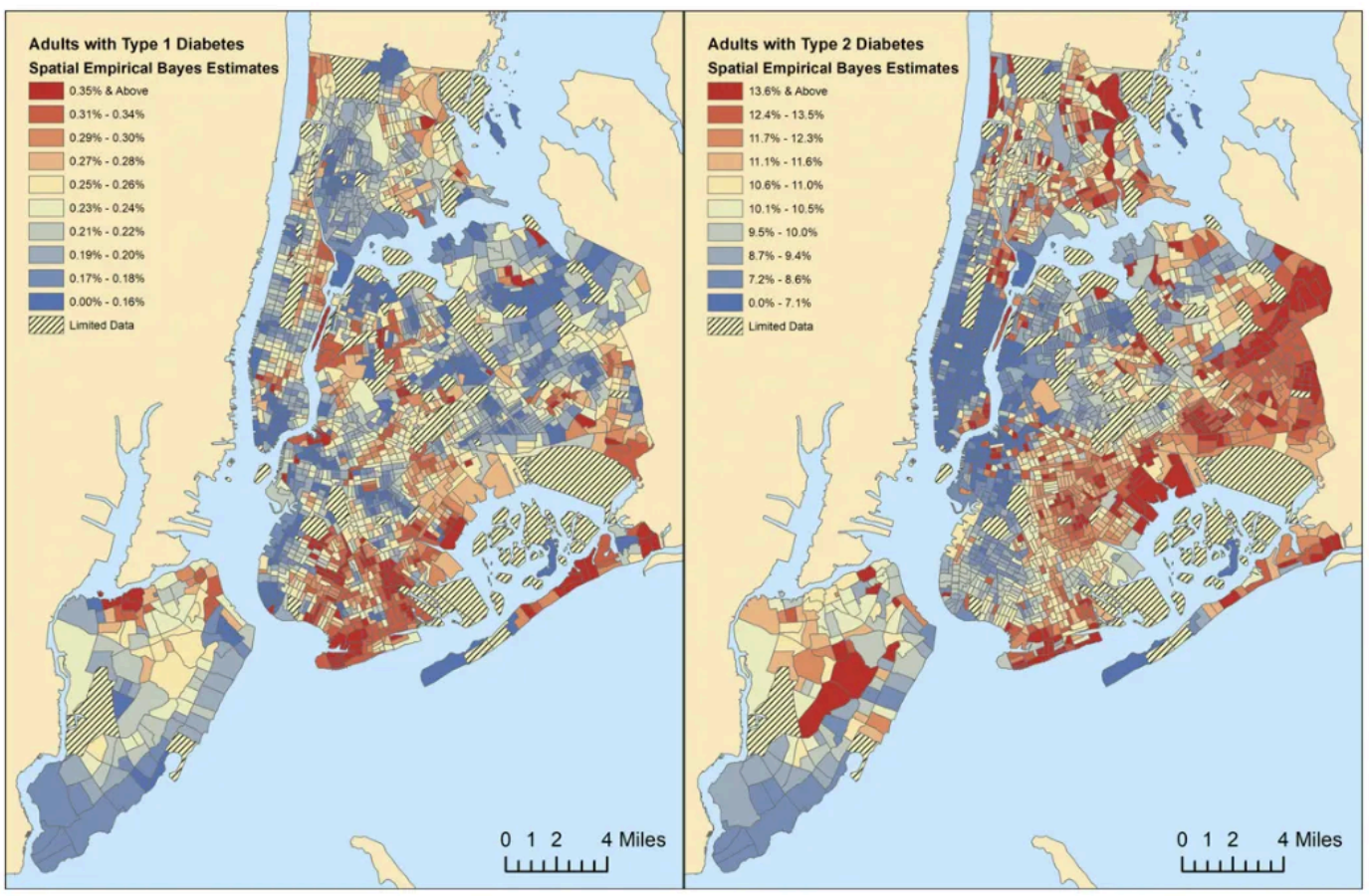
Some Quick Epidemiology

T1D has been around for a long time, with rare instances showing up in Egyptian medical texts as far back as 3000 years ago—but rates appear to be on the rise, increasing by as much as 5% per year worldwide. There *are* genetic risk factors—70% of people who get T1D have risk genes—but only 3% of people with those risk genes [1] go on to develop T1D.

The current best guess as to why is that those 3% are exposed to some bacterial or viral culprit, which has cellular components that look like pancreatic proteins—and that that this tricks the immune system into attacking an important organ. A few people’s cases can be chalked up to known pathogens like the *Rubella* virus, but these are generally a small fraction of the overall number of new T1D cases. For the vast majority of people with T1D, the cause remains unknown.

The Where

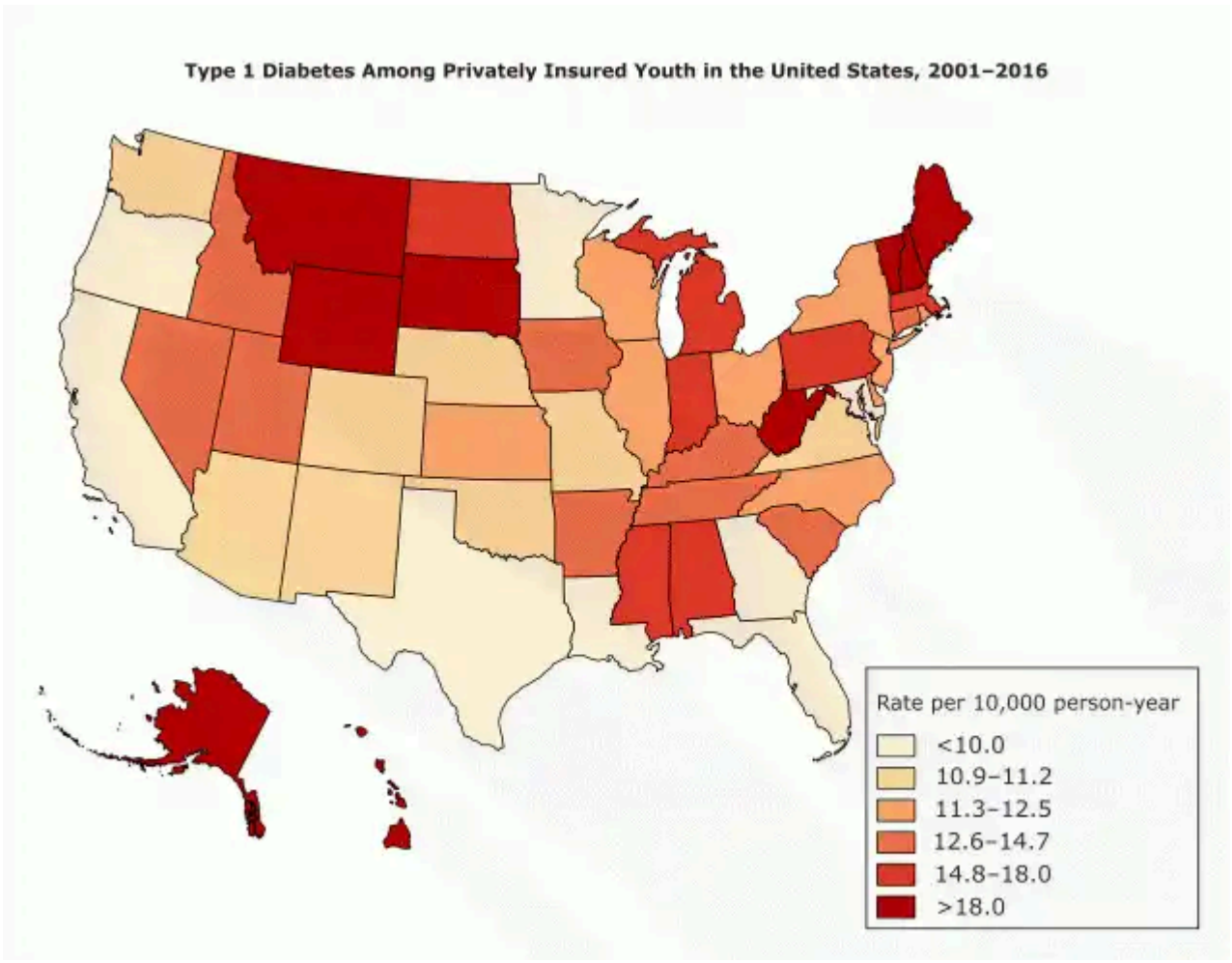
Each of the images below is a map of people with diabetes in New York City.



On the right is Type 2 Diabetes. The prevalence varies pretty significantly within the city and, although this is much more prevalent than Type 1, the incidence rate can be explained pretty neatly as a function of things like socioeconomic status and race. The bluest patches on that map tend to be wealthier and whiter; more Whole Foods per square mile, fewer ounces of soda consumed per capita.

But T1D, on the left, defies such explanations. There's still massive variation between low-incidence neighborhoods and high ones, but it doesn't track neatly with income, race, or any other known variable. If anything, the economic trend observed in T2D is reversed, with more affluent neighborhoods tending to have higher incidence rates. Swathes of blue are pocked by strange high-incidence enclaves in bright red.

If you zoom out to look at the entire United States, the picture becomes even stranger. There's a fifteen-fold difference in T1D rates between the lowest- and highest-incidence states.



Because America's healthcare system resembles nothing so much as feudalism, this data has the major caveat of only counting people on private insurance. Nevertheless, we do the best we can with what we've got. Source: CDC

The general trend is one of higher rates the further north you go, and this has led a lot of researchers down the path of vitamin D deficiency as a driver—but I don't buy it. For one thing, Mississippi and Alabama are among the highest-incidence states. Some have suggested that demographics can account for this: these states have more black residents, who have a harder time getting adequate vitamin D from sun exposure—but then why is Georgia (33% African American) a low-incidence state when Alabama (27%) is high? Why would Hawaii be in the same range as Alaska? [2]

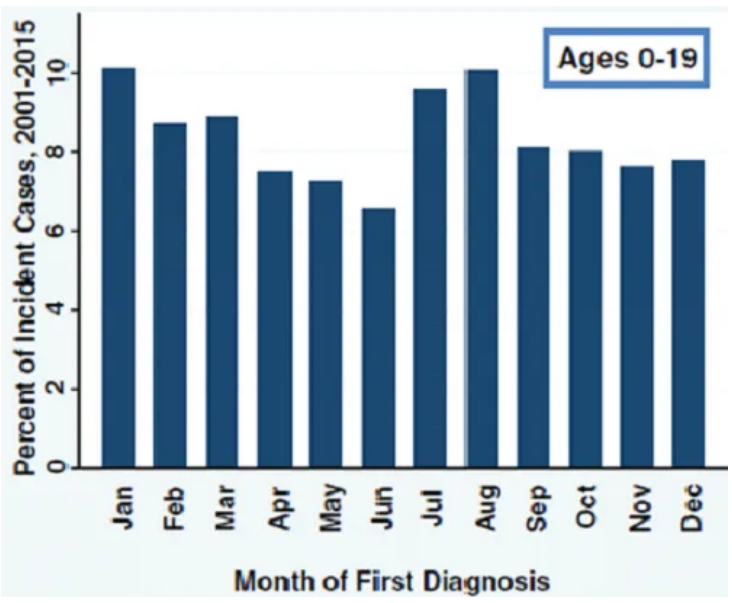
But zoom out even further, and things get stranger still. China has regions with rates as low as 0.1 cases per 100,000 people per year, while Finland and Sweden lead the pack with per-capita rates four to five *hundred* times that.

Position	Country	Incidence (per 100,000)
1	Finland	57.6
2	Sweden	43.1
3	Saudi Arabia	31.4
4	Norway	27.9
5	United Kingdom	24.5
6	USA	23.7
7	Australia	22.5
8	Kuwait	22.3
9	Denmark	22.2

Here, we again see a general pattern of increasing incidence the further you get from the equator…but again, certain exceptions throw a wrench in it. What are Kuwait and Saudi Arabia doing near the top of the list? Why is Venezuela near the bottom? Curiously enough, the pattern of higher incidence in wealthier places holds at an international level, as well.

Then there's the island of Sardinia, in the Mediterranean sea. It's part of Italy, and—by most cultural and geographic metrics, it's nearly identical to the rest of the country. But Sardinia's per-capita incidence of T1D is 3-5x mainland Italy's, placing it among the highest in the world.

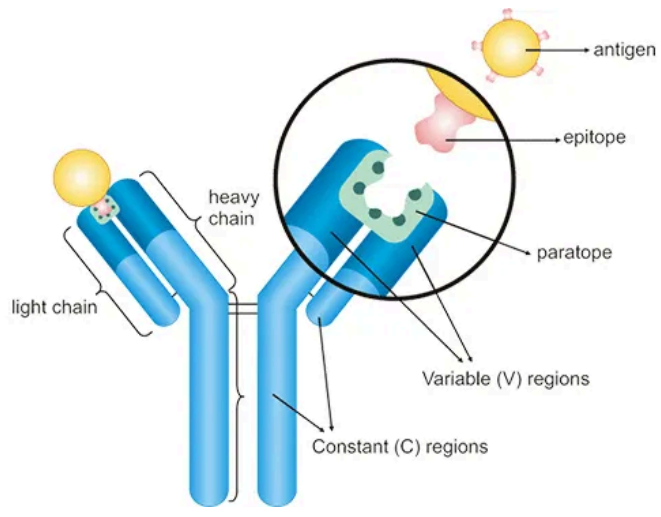
Oh, and just to add to the fun: I said earlier that *when* T1D happens isn't much of a mystery—but this is only true in one sense. There's also seasonal variation in Type 1 Diabetes diagnosis. In America, new diagnoses are lowest in June, highest in July and August, and then drop in the autumn before peaking again in January.



So: What the fuck is going on here?

I'm so glad you asked. I think I have an answer. I think something's in the water.

Earlier, I mentioned that the *auto-antibodies* produced by the immune system in T1D bind to specific bits of proteins that are made in the pancreas. In immunology, these bits are called "epitopes".



The antibody is like a lock, which the immune system designs to fit the "key" that is the epitope of the antigen. **Image** Source

In '98, a team from Australia did some very clever work to identify the most aggressively immune-system-offending epitopes in T1D, by breaking down a protein into its component bits, then checking which of those bits make the T cells from patients' blood go wild. In their writeup, they reported the protein sequence of a particularly egregious epitope from one of the three major antigenic proteins in T1D, tyrosine phosphatase IA2. The sequence goes: VIVMLTPLVEDGVKQC. [3]

This amino acid sequence bears some resemblance to a chunk of another pancreatic protein that's targeted in T1D called GAD65, [4] as well as to bits of proteins found in *Rotavirus,* and *Haemophilus influenzae*—a common gut bacterium.

Interestingly enough, *Rotavirus* infection is suspected to account for a small fraction of T1D cases…but again, only a small fraction. Still, this tracks nicely with the notion that an inflammation-inducing microbe might produce a protein that resembles pancreatic proteins, duping the immune system into attacking the pancreas.

But the match between that egregious epitope and the protein from *Haemophilus* wasn't very good. However, 1998 was a long time ago, and the databases have accumulated a lot more bacterial gene sequences since then—so in the spirit of fucking around and finding out, I decided to do my own search of bacterial genomes, to see if there are any other gut bacteria which might produce similar proteins.

Now, if you want to know whether a protein that's present in one organism is present in another, you can't just ctrl+f the genome for that sequence. For one thing, two proteins can differ by several amino acids and still perform the same function. On top of that, you'd expect a bacterial protein triggering an autoimmune response to look a lot—but not *exactly*—like the protein that it induces an immune response against. That's why we have BLAST, the Basic Local Alignment Search Tool. BLAST is the level 1 Bio-mancer spell; it's a messy search that lets you find similar regions in proteins, even if they're not a perfect match.

So I BLASTed the egregious epitope from human Tyrosine Phosphatase IA2 against bacterial genomes, and got a surprpising "hit" in a known pathogen.

It's not a human gut bacterium, though. Its name is *Legionella*.

```
Query  2    IVMLTPLVE--DGVK  14
            IVM+TPL+E  DGVK
Sbjct  212  IVMMTPLIEGLDGVK  226
```

Aligned region of similarity between the "egregious epitope" of human IA-2, top, and a tyrosyl-tRNA ligase enzyme conserved across a wide swath of the order *Legionellales*. This isn't a slam-dunk result—but it's an interesting lead.

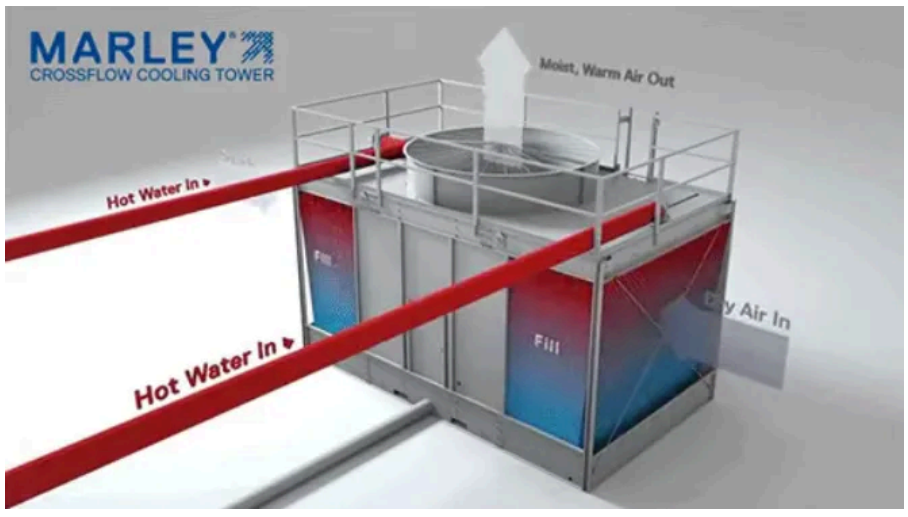
The Hit

If you've heard of *Legionella* before, it's likely in the context of its role in causing Legionnaire's disease: a relatively rare, noncontagious form of pneumonia that was only identified in 1976, when an outbreak killed a bunch of old people at an American Legion convention (hence the name). One major reason it was only discovered so recently is that *Legionella* is pretty weak, as far as pathogens go: you usually have to be seriously immunocompromised for a *Legionella* exposure to result in Legionnaire's

disease. Most of the time, the body can easily recognize it as a pathogen and clear it before it gets out of hand.

That's a good thing, because it's something that we're exposed to with surprising frequency. It's found all over the world, at low levels in the environment—soil, seawater, etc.—but it thrives in fresh water at a specific temperature range: 77℉ to 113℉. Much below this, and it won't grow. Much above it, and *Legionella* gets cooked like any other living thing. But anywhere in that magic window—anywhere there's hot water stagnating or circulating—*Legionella* can become a problem.

The culprit in the original 1976 outbreak was a hotel's rooftop A/C unit, which spewed warm, moist air from the building's roof as part of the process of cooling the air inside. In the heat of summer, this structure became a breeding ground for *Legionella*.



The kind of A/C units that provide central air conditioning for commercial and urban-residential buildings work partly by evaporative cooling, running air through warm water as it trickles through a porous layer, which creates the potential for bacteria growing in the warm water to be "aerosolized". This doesn't end up in the actual air supply to the building—rather, it can rain down on pedestrians below.

It's impossible to say how long the presence of the pathogen in that airflow went unnoticed, but when the American Legion convention brought a bunch of frail old veterans to the hotel, they began dropping like flies as the microbe insinuated itself into the cells of their lungs.

And while this was the prototype outbreak which alerted medicine to *Legionella*'s existence, it was soon realized that huge swaths of infrastructure harbored the potential to cause infections: not just air conditioners, but ice machines, hot tubs, showers, and even the pipes that carry drinking water.

New controls and standards were put in place—things like chlorinating the water in the A/C loop to prevevnt microbial growth. This works about as well as one can expect from a societal-scale disease prevention measure that hinges on your apartment building's super having his shit together. To complement this, a diligent system of reporting and tracking for Legionnaire's outbreaks was instituted in major cities to help identify and decontaminate sources of infection.

But that reporting system requires clusters of people to get sick with Legionnaire's disease. In most people, though, *Legionella* is easily fought off, producing only a mild and transient illness called *Pontiac fever*—something like a few-day flu.

Pontiac fever is almost certainly more common than Legionnaire's, but is pretty low-key by definition—meaning that it rarely receives a formal diagnosis. All this to say:

it's hard to estimate how much *Legionella* the average person is exposed to, but the occasional survey of things like public drinking water and hospital water systems have identified it as shockingly common: one Catalonian study found the bacterium in the plumbing of over 85% of hospitals tested, and more recent work has shown that it can persist over decades, even against continuous eradication and control measures. This is not a phenomenon unique to hospitals: that's just where the focus of study has been, because that's where you have a lot of the kind of people who might die of Legionnaire's disease (as well as the equipment necessary to test the water for obscure bacteria). Rates of *Legionella* in home water systems vary heavily by country and by study, but a few studies have reported figures in the 10% - 20% range. In a study of apartments in Bologna, Italy, about 4% of apartments had *Legionella* in their hot water lines—unless they were on a shared water-heating system with the rest of the building, in which case the prevalence was more than 40%.

So, just to spell it out, the hypothesis so far is this:
Legionella bacteria express proteins which resemble certain ones found in the human pancreas. They also produce proinflammatory molecules, by which the immune system learns to target them for destruction. In early life, while the immune system is still learning to differentiate "self" from "other", transient or chronic exposure to *Legionella* can cause the development of antibodies which are active against the bacterium, but which also cross-react with human proteins such as GAD65, IA2, and insulin, especially in genetically susceptible individuals. This results in autoimmune destruction of pancreatic beta cells, which in turn results in Type 1 Diabetes.

The Fit

I'll be upfront and say that I knew only a little about the epidemiology of *Legionella* when I found that BLAST hit. But the more I learn about it, the stronger the fit seems to be.

If we want to pin the blame for a substantial portion of T1D on any one classical pathogen, it's got to be a relatively ubiquitous one, and most infections have to go unreported—otherwise we'd likely have noticed the association by now. This already makes *Legionella* a solid candidate: If you're old enough to get Legionnaire's, you're probably past the risk zone for T1D, and we know that a lot of *Legionella* exposure goes undetected, especially in the young.

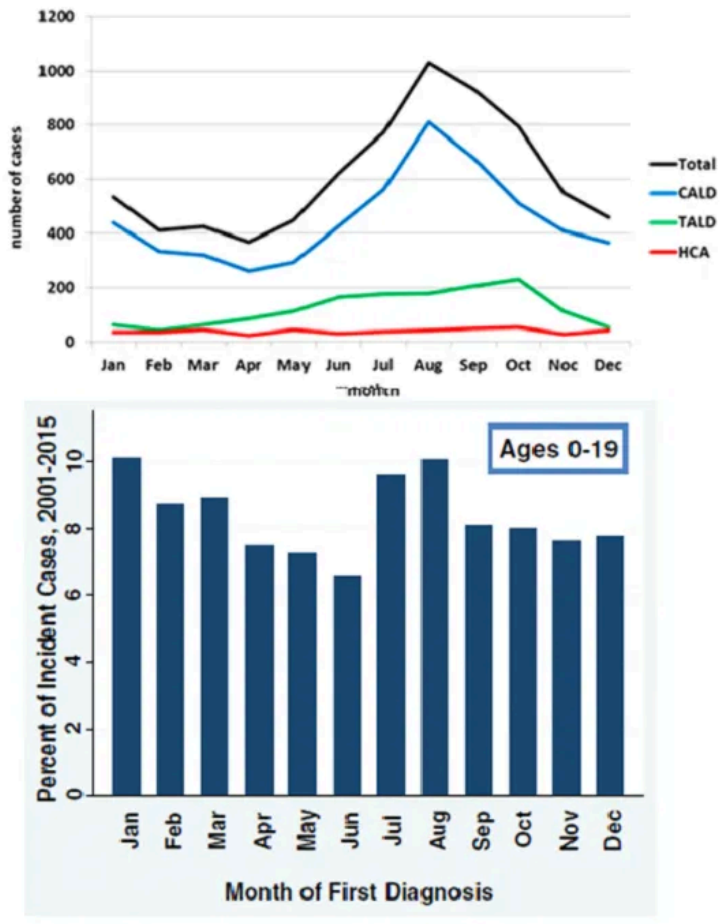
We wouldn't necessarily expect to see strong geographic associations between T1D and Legionnaire's disease incidence, either: if anything, you might expect *lower* incidence of Legionnaire's in places where people are frequently exposed to a little bit of *Legionella*. After all, if a person has enough antibodies to a pathogen that some of them could hit pancreatic cells, presumably it would at least provide protection against the more acute/severe forms of disease that could result from exposure to a bad air conditioner.

On the note of air conditioners: residential units (the big fan-boxes you find outside many suburban homes) aren't much of a *Legionella* risk by virtue of how they're designed, and the same goes for window units. While anyone who's spent a summer day walking the streets of a big city has been annoyed by dribbles of water leaking

from window units overhead, this is just condensation—water molecules crashing out of the air as it passes over the unit's cooling coils, then quickly running down.

But in this, there's an interesting potential explanation for the unusual prevalence of T1D in high-income neighborhoods: In a city like New York, who has central air? Office buildings, hotels, and the newer luxury high-rises nearest them—while window units are the hallmark of the brownstone and barrio.

Another point in favor of *Legionella* as a potential driver? Seasonal variation. Like Type 1 diabetes diagnosis, exposure to *Legionella* hits a peak in the hottest months of the year, thanks to poorly maintained A/C units across the world kicking on.



Top: month-to-month variation in Legionnaire's disease incidence, in a German cohort. Bottom: The graph from earlier for comparison, charting T1D diagnoses throughout the year in a cohort of American youths. Note that the "summer spike" in T1D diagnoses is paralleled by an increased incidence of Legionnaire's.

Hot & Cold

But how do we square this with the fact that, overall, T1D incidence is highest in colder climes—northern countries like Finland and Sweden, where air conditioners are a rarity? Or the fact that, in these colder countries, T1D rates rise as the minimum annual temperature in population centers drops?

The answer is that A/C units aren't the only source of *Legionella* in our lives, and likely aren't even the most common: Hot and cold water pipes can play host to the bacterium. Remember: *Legionella* likes to be within a dozen degrees of body temperature on either side, and it only dies after prolonged exposure to temperatures over ~140℉.

Most hot-water systems are designed to raise water to this kind of pathogen-killing temperature, but in places with the bitterest winters, standard heating systems might not be enough to do the job. Water that leaves the boiler at 140℉ can cool substantially before it reaches the end of the pipe, creating zones that don't get heat-sanitized the way they're supposed to.

When you finish your morning shower and turn the tap off, the water in the hot pipe starts cooling—and might spend several hours in the temperature range that provides optimal growth conditions for *Legionella,* giving the microbe time to reproduce and form biofilms on the interior of the pipe. The next time you turn the shower on, the spray of that first water leaving the faucet can splash up the bacterium the same way that an A/C unit does—part of why hospital showers are regularly a major vector for outbreaks of Legionnaire's.

According to a German study from 2011, cold water lines may be even worse, perhaps because they never get the scalding heat-treatment that the hot-water pipes do. Instead, the authors point out, cold lines often run directly adjacent to hot ones in the same shafts, and this has the potential to bring the temperature up above 77℉ in large portions of the system.

So this makes it seem like subclinical *Legionella* exposure is a multi-variable equation: How much time does your hot water spend below 113℉? How much time does your cold water spend above 77℉? Do you have a hot tub, that beloved vat of vaguely mildewy people-broth, which spends all winter bubbling away in the perfect ~100℉ temperature range? Do you live in a building with central air, and is the system well taken care of? How humid is your town in the summer? *Legionella* spewed out of A/C units is most infectious on humid days and days following heavy rainfall, when it takes longer for the miniscule water droplets that carry the cells to evaporate.

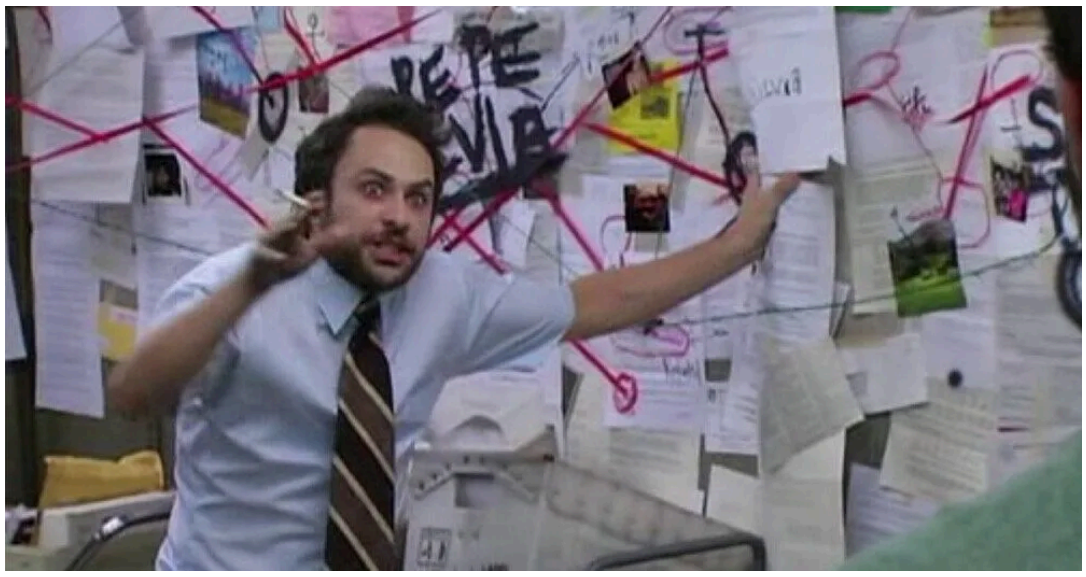
You can see evidence of this complexity in the German month-by-month graph of Legionnaire's above: there's a remarkable spike during the hottest and coldest months of the year, but only in one of the cohorts studied, presumably due to differences in things like A/C use and humidity. In Japan, there's a winter spike in Legionnaire's rates, thanks to the culture's enthusiasm for *onsen* hot spring bathing.

But in this daunting complexity, there's also a neat explanation for the weird geographic exceptions that plague the data on T1D. In northern states and countries, water heating systems might be the main culprit, while in places with unusually high T1D rates despite their proximity to the equator, a combination of central air use and "cold" tap water stagnating above 77℉ could be major drivers.

While I haven't fact-checked this, something tells me that Venezuela's uniquely low rate of T1D is accompanied by a similarly unique deficit in terms of the percentage of the population that has a hot tub, an apartment with central air, or indeed hot running water. At the other end of the T1D incidence spectrum is Saudi Arabia—where much of the country drinks from water coolers that perpetually sit out in the sun, due to a public water supply system that functions only periodically. Kuwait, another country with a high T1D rate that defies the general cold-temperature trend, has a documented problem with *Legionella* in both water and air conditioning systems.

The (central air) x (rainfall) effect might also help explain why e.g. Hawaii, Alabama, and Mississippi are hotbeds of T1D, while nearby states like Texas aren't.

I could go on, but at a certain point, more examples probably stops adding weight to the message and starts to sound like this…



…so I'll leave you with the thing that, for me, clinched this as a hypothesis worth reporting: Sardinia.

This is the island in the Mediterranean we talked about earlier. One of the highest rates of T1D of any place on Earth, but it doesn't really fit either mode of the pattern we've drawn here so far. Summers are rarely hot enough to warrant A/C, or heat the cold tap above the critical temperature. Nor are winters so cold that your average home's water heater would struggle to do its job.

But a little digging turns up an interesting facet of Sardinia's epidemiological landscape: Ferries.

Roughly 25% of the traffic to and from the island occurs by boat, thanks to its proximity to the shores of a number of European countries. But the journey can take up to 15 hours, so these are less like "water taxis" and more like shoddy cruise ships: They have little coffee bars and private rooms, and the fancier of these include beds, sinks, and showers. The fresh water for these is heated and treated right there on the ship, and—wouldn't you know it? If you plug "Sardinia + *Legionella*" into Google Scholar, the first page is littered with studies reporting alarmingly high rates of contamination in the water systems of these boats.

A 2004 study sampled water from the showers and sinks of seven ferries docked at Sardinia's southern port, and found *Legionella* in the water systems of **six out of the seven**. Taking a shower in a random room on one of these seven boats, you'd have a 42% chance of being exposed to the bacterium, which researchers attribute to overtaxed boilers and a rather relaxed attitude on the subject of chlorine treatment. The picture hasn't gotten any better in recent years, either: similar reports were published in 2008, and then again in 2017. Nobody seems to be too motivated to do anything about it, I suspect in part because the incidence of proper Legionnaire's in Sardinia is curiously low. But what if that's only one manifestation of this pathogen?

What If?

It seems like a relatively easy hypothesis to gather some data around: check whether the T-cells reactive against pancreatic proteins like GAD65 and IA2 are also active against proteins from *Legionella*—or else see if people recently diagnosed with Type 1 Diabetes have more antibodies against *Legionella* in their blood than healthy controls.

A little more challenging is to prove causality. Typically, you satisfy Koch's postulates by infecting an animal with a pathogen to create the disease state, then treating the

infection and doing away with the disease. But if T1D truly works the way we think it does, even the right pathogen may not be able to produce the syndrome in animals, unless you engineered human immune and pancreatic genes into them first.

Unless we want to try infecting kids with *Legionella* to see if it gives them diabetes, maybe the most conclusive way to prove this out would be to go straight to the source: Create an updated map of T1D in a major city, then go out and sample water systems and cooling towers, to see if contamination rates differ depending on whether you're in a high- or low-T1D neighborhood. If they do, you could implement active *Legionella* remediation in a subset of the high-incidence ones, and see if the rate of new cases goes down in those places over the next few years, relative to the ones you didn't fix.

I'm not an immunologist (or an epidemiologist, for that matter) so I'm sure there are a million more clever ways to investigate this that I can't even fathom—so consider this an open call for collaborators. If you have expertise in *Legionella*, T1D, or immunology generally, I'd like to hear from you: comment below or reach out by replying to any of this blog's emails, and let's turn this into a grant app.

Got questions or got a killer counterexample? Equally welcome. And if you're someone who happens to have a lot of money, and you'd like to see this hypothesis tested without subjecting it to the whims and vagaries of the usual funding bodies, my inbox is always open.

If you're none of those things but like the idea, please share with your friends and subscribe below for more dives into the wild frontiers of biology. This blog is free, and I plan to keep it that way.

'til next time.

—🖖💩

1 I mean "risk alleles" but "risk genes" is common parlance now, don't @ me

2 More generally, if you've spent any time reading epidemiology literature, you're absolutely sick of hearing about vitamin D; it's been suggested as an explanation for everything from schizophrenia to autism, and never seems to hold up.

3 If you're not a biologist: Each of these letters stands for an individual amino acid, like glutamine or tryptophan. All proteins are made of long strings of amino acids, which—due to their various electromagnetic properties—naturally fold up into sheets, tubes, and other shapes. All of these structural elements interact to form what are essentially origami machines at a molecular scale. Just for example, the enzyme lactase (which digests milk sugars) is nearly 2000 amino acids long.

4 GAD65 is a neat enzyme that makes the molecule GABA, which—in addition to being the primary inhibitory neurotransmitter in your brain—is also a major regulator of insulin release in the pancreas. GAD65 is the pancreas-local version of the enzyme, different from the one that makes GABA in the brain.