

SQA skillset, AI and Advancing medical research

Tracy T. Burr ridge

tburridge@gmail.com

Abstract

Software professionals have a social responsibility to utilize our skillset to help advance medical/scientific research. SQA discipline and skillset is uniquely positioned, as we do not build tools, we utilize tools to push boundaries, to verify declared limits; to discover software failings that coders and designers did not imagine, love the phrase “why would a user do that?” We, SQA, reside outside of the box, we do not just want to break things, and we want to describe what we did, document our steps and show that it’s repeatable.

Arguably, Y Combinator is one of the best known and most respected incubators. In February 2024, YC published a list of “Start-ups we want to fund” innovative entrepreneur requests. On the list is “Cure cancer, the technology we’re talking about is an MRI. Modern MRIs are sensitive enough to detect cancer masses as small as a millimeter.” Kaggle has been offering bounties for a few years now. Analyzing data, approaching all publicly provided data with all skepticism - exception boundary testing, edge cases - SQA/SQE professionals know how to do this better than most other professions.

ChatGPT, Knowledge Graphs and other AI tools, can be used with publicly available medical datasets to advance medical challenges such as cancer or autoimmune diseases such as multiple sclerosis -MS. What is fascinating about MS is that we, in 2024, still have no idea what causes MS. There are theories, there are pharmaceutical efforts that pause or delay the progression but are very very expensive.

What if I use my SQA skillset, AI tools, and available datasets to establish a set of repeatable steps to establish new methodology for early detection of MS?

Biography

Tracy Burr ridge has 20 plus years in software quality engineering from start-ups and enterprise software companies. She has earned an MBA and MAT, experience in developing software requirements, SQA frameworks and training/executing tests in environments from Gaming, databases and UI heavy applications. Currently, at TriMet as a Fare Systems Analyst and has worked on an assortment of projects including smartcard transportation upgrade, research and requirements gathering to update internal POS systems, Federal Transportation IMI Grant and the updating of all TriMet customer facing systems - app, website, Inspection and ticket machine software.

1 Introduction:

A good software quality engineering has an understanding of a successful test cycle. Not just finding a defect, but the ability to drill down and communicate the finding. When a bug report is submitted, it will contain steps of repeatability, common identifiable factors, and patterns.

A person, a human being, is for all intent and purpose an extremely sophisticated, complicated program. Like any program, there are defects, issues that are sometimes explainable and other times unexplainable. In either instance if there is a defect, we can or should be able to identify overlapping environmental factors, steps to reproduce the defect, rate of repeatability and predictable or expected outcomes.

We evaluate a defect for causation vs. correlation. As SQA we sometimes, often times, do this in a vacuum; we spend a lot of time explaining how we found the defect and how the average person on the street “just needs to follow the step-by-step instructions and they would be able to generate the same defect.” Sometimes this is a showstopper, often times there are work arounds.

Today, however, we have new tools at our fingertips – a near personal assistant or eager college intern, technologies such as ChatGPT 4.0, WikiData and Google AI. We can introduce the defect to these tools, use traditional SQA/E practices to assist in advancing research in ALS and MS prevention/detection/repair.

2 Approach:

At the start of this endeavor, I was quite uncomfortable making the necessary shift of Person to Software Program and Disease to Defect. If I want to use my SQA skills to address the very human disease, the adjustment was necessary, so onward ho!

- Evaluate and select disease:
 - Cancer: I chose not to evaluate Cancer as it is broad and has an exceptional number of research efforts.
 - **Multiple Sclerosis MS: MS is predominantly a female disease, 1:300 persons are afflicted (1)**
 - Lou Gehrig’s disease ALS: ALS is predominantly a male disease; 1:400 persons are afflicted. (2) It is a sclerosis like MS, but it has the advantage of an amazing speech made by a beloved athlete:
 - Lou Gehrig “For the past two weeks you have been reading about the bad break I got. Yet today I consider myself the luckiest man on the face of this earth.” (2)
- Tools: Mindmeister, ChatGPT 4.0 and Google AI - to narrow down large datasets or organize dis-organized information

- Define the defect: Description, environment, expected and actual behavior. Rate of repeatability.
 - Incorporate the call out, data organizations, such as Kaggle and the incubator Y Combinator, have started making the call to software professionals to start engaging with science research to use publicly available data/datasets to help find cures to diseases like Cancer, MS, ALS, etc. Countries like the U.K. have already integrated the two professions. These vetted and solid resources for reviewing analysis are underway, allowing for comparisons and peer reviews.
 - Observations and recommendations
- In this situation, the objective of a mind map is a way to assemble documentation to create Requirements – remember, we are QA and if we do not have requirements, we make requirements.

3 Define the disease (defect) that will be used as a Use case

A customer has reported “Bug”

- a. What is the: Defect” - Multiple Sclerosis / ALS (Lou Gehrig’s disease)
 - i. MS affects women
 - ii. ALS affects men
 - iii. Both are a Sclerosis - with unknown causes
- b. What is the process to isolate and describe the bug? Develop and fill out a Bug report (draw from “best practices”) See appendix
- c. Select and define the QA methodology that will be used to approach reported bug

I will take on the role of “disease tester”, I have been given a program, the human body; I have been informed that the program is very complicated and for the most part very reliable and very predictable. While the human body is a solid program it will, on occasion, turn on itself – the ordinarily reliable program, will, without predictors, progressively stop working. The defect has been given the name sclerosis.

In QA, if there is documentation, ideally a detailed requirement, the first thing I would do is read through and develop a plan defining test cases, scenarios, user flows and user stories. In this case, I know what the “ideal” human looks like – disease free, immune system not gone array, a software program in which everything works perfectly.

Because of the complexity of diseases listed above, I chose to focus on MS for the following reasons:

- There are many theories that Vitamin D deficiency is a potential contributing factor
- Vitamin D deficiency is an established condition
- Vitamin D can be researched
- MS is a disease, in comparison to ALS, with theories

As a Tester, I am given a program and told that it will:

- Not run optimally in 1 out of 300 installs

- Once the poor performance starts the program can never be reset
- There are tools that help the program to run as close to optimal as possible – still not optimal

My approach is not going to be that of an SQA Lead (planning/proactive), it will be as a tester engineer (reactive/ all types of testing). The project is underway:

Tester characteristics defined as (3):

- Reactive to a program that has been assign
- Focus on finding bugs
- Conduct a variety of test techniques, both manual and automated
- Support the development process

Table to create understanding of ideal human vs defect in human:

Ideal/Expected	Defect
Absorbs Vitamin D	Does not absorb Vitamin D from sun or diet
Immune system does not attack person	Immune system over protects, attacks the person.
Muscles grow and regenerate	Muscles degrade and not repaired
Vision is assisted with glasses, not pharma	Vision degrades and cannot be repaired
Walks with ease until a senior citizen or injury	Ability to walk becomes a struggle and only gets worse

4 No documentation, how do I trigger the defect?

If I am not provided documentation, I will need to use **ad-hoc or exploratory testing** to first understand the defect. The tricky part with this Human program turns out to be creating these User scenarios as the documentation is sparse, and what is available is very theoretical and there seems to be a lot of inconsistency.

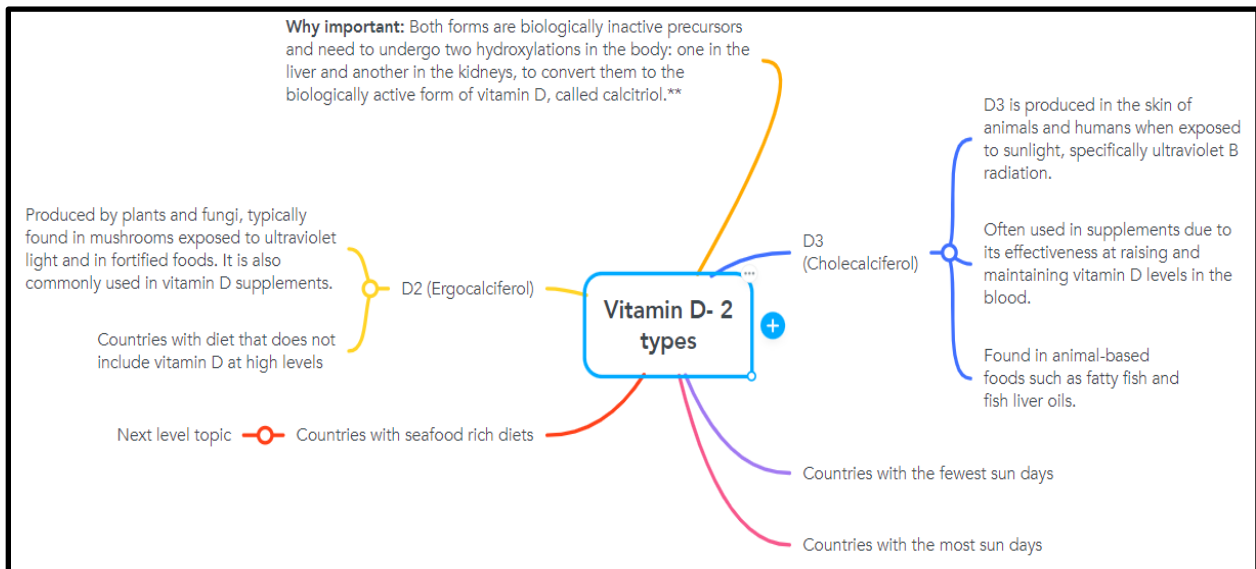
- Uniquely, this defect we are taking on comes with Use ChatGPT to help get through large datasets that are focused on:
 - a. MS
 - i. Autoimmune disease
 - ii. MRI's
 - b. In addition, potentially:
 - i. Recent Studies
 1. International
 2. Domestic
- Use Knowledge Graphs to parse the datasets:
 - a. Pull/Graph the known:
 - i. Validation of current theory
 - ii. Challenge current theory
 - iii. Find the gaps in current theory
 - b. Pull/Graph new findings and theories
 - c. Develop new theories

5 Create Mind maps: Disease – Multiple Sclerosis



Figure 1: Multiple Sclerosis at a high level , <https://mm.tt/app/map/3333723184?t=k8HaPIbe4x>

5.1 Vitamin D mind map:



Develop a mind map in order to have a quick reference. Here, Vitamin D is outlined as having two distinct components, D2 and D3.

D2: Absorbed through diet and D3: Absorbed through Sun exposure, Both forms of Vit D can also be found in supplements.

5.2 Countries with high levels of Vitamin D3 deficiency, ChatGPT 4 vs. Google AI

Source: ChatGPT 4.0

<p> Vitamin D deficiency is a global health issue, but certain regions and countries are more affected due to factors like geographic location, lifestyle, skin pigmentation, and cultural practices. Here are some of the countries and regions with the highest levels of Vitamin D deficiency:</p> <p>1. Middle East and North Africa (MENA)</p> <ul style="list-style-type: none"> Saudi Arabia, UAE, Egypt, Iran: Despite abundant sunlight, cultural practices like clothing that covers most of the body and limited outdoor activities contribute to high deficiency rates. Some studies show deficiency rates as high as 60-90% in these regions. <p>2. South Asia</p> <ul style="list-style-type: none"> India, Pakistan, Bangladesh: High population density, pollution, and darker skin pigmentation reduce the synthesis of Vitamin D. Deficiency rates in urban areas can exceed 70%. <p>3. Northern Europe</p> <ul style="list-style-type: none"> Norway, Sweden, Finland, UK: Limited sunlight during the long winters and higher latitudes contribute to lower Vitamin D synthesis. Despite fortification efforts, deficiency is still a concern, especially during the winter months. <p>4. East Asia</p> <ul style="list-style-type: none"> China, Japan, South Korea: Urbanization, indoor lifestyles, and air pollution contribute to Vitamin D deficiency. Studies in these countries report deficiency rates ranging from 40-60%. <p>5. Africa</p> <ul style="list-style-type: none"> Morocco, Tunisia, Algeria: Similar to the MENA region, despite high sunlight, cultural clothing and indoor lifestyles result in high deficiency rates, particularly among women. 	<p>6. North America</p> <ul style="list-style-type: none"> Canada, Northern United States: In Canada and the northern parts of the U.S., the long winters and limited sunlight exposure lead to lower Vitamin D levels, particularly in the winter months. <p>7. Australia and New Zealand</p> <ul style="list-style-type: none"> Despite being sunny regions, cultural practices like avoiding the sun to prevent skin cancer contribute to Vitamin D deficiency, especially in Australia. <p>8. South America</p> <ul style="list-style-type: none"> Argentina, Chile: Higher latitudes and colder climates in southern parts of these countries contribute to higher deficiency rates. <p>Key Factors Contributing to Vitamin D Deficiency:</p> <ul style="list-style-type: none"> Latitude: Countries farther from the equator have less sunlight, especially in winter. Cultural Practices: Clothing that covers most of the skin, particularly in Muslim-majority countries. Urbanization: Increased time spent indoors and air pollution blocking UVB rays. Skin Pigmentation: Darker skin synthesizes less Vitamin D from sunlight compared to lighter skin. Diet: Inadequate intake of Vitamin D-rich foods or lack of fortified foods. <p>Addressing Vitamin D deficiency often requires a combination of public health measures, including supplementation, fortification of foods, and education about safe sun exposure.</p>
---	--

Source: Google AI search results

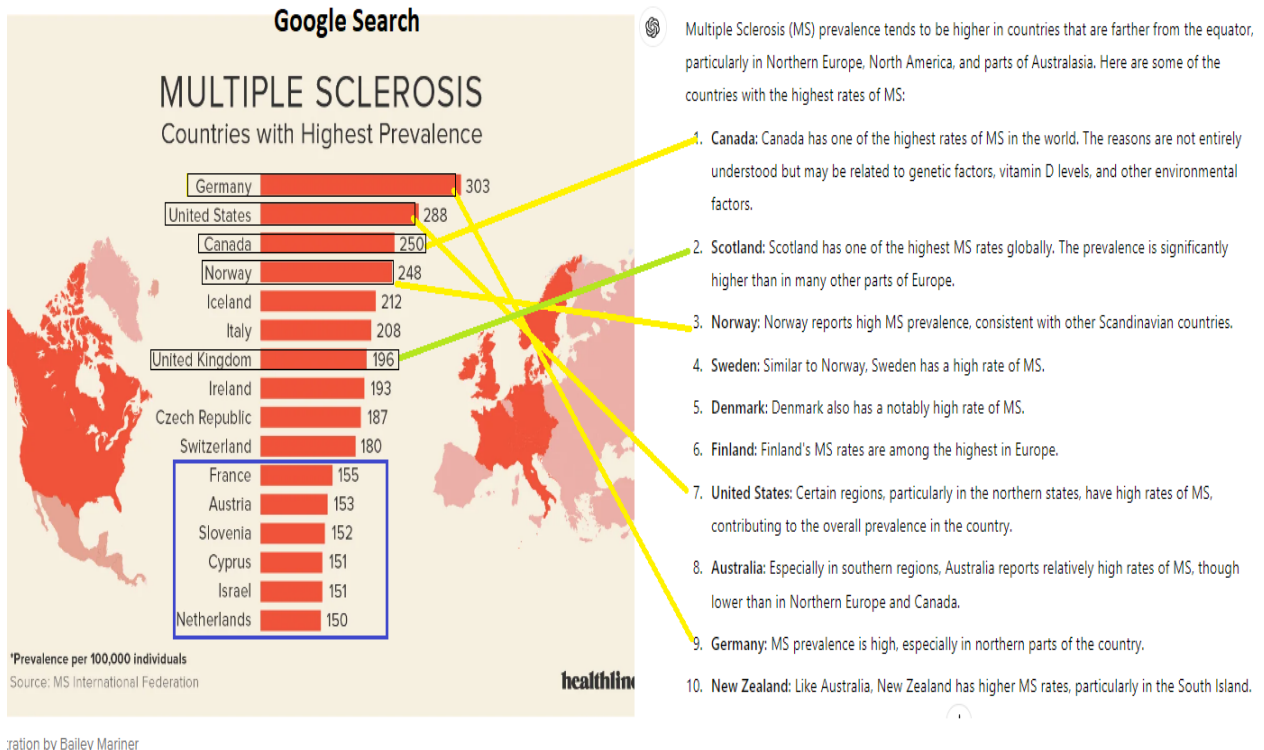
Search Labs | AI Overview

Some countries with high levels of vitamin D deficiency include:

- Pakistan: Has the highest prevalence of vitamin D deficiency in adults, at 73%
- Bangladesh: Has a prevalence of 67%
- India: Has a prevalence of 67%
- Nepal: Has a prevalence of 57%
- Sri Lanka: Has a prevalence of 48%
- Afghanistan: Has a high prevalence of vitamin D deficiency, with 45–65% of the population having levels below 20 nmol/L
- Iran: Has a high prevalence of severe vitamin D deficiency in infants, at 86%
- Turkey: Has a high prevalence of severe vitamin D deficiency in infants, at 51%

Some reasons for vitamin D deficiency include diet, sun exposure, and darker skin pigmentation.


5.3 Geography: Where is MS found on the world?



Alignment of Regions/Countries with Vitamin D deficiency and High levels of MS

Country/Region with Vitamin D deficiency		Country with High Levels of MS
Saudi Arabia, UAE, Egypt, Iran		
S.Asia: India, Pakistan, Bangladesh		
Northern Europe: Norway, Sweden, Finland, UK		Norway, UK, Germany
East Asia: China, Japan, South Korea		
Africa: Morocco, Tunisia, Algeria		
N.A.: Canada, Northern United States		United States, Canada
Australia and New Zealand		
S.A.: Argentina, Chile		

5.4 Where is MS found in the United States:



Multiple Sclerosis (MS) prevalence can vary significantly by region, but generally, higher rates are observed in areas farther from the equator. In the United States, certain states have been identified as having higher prevalence rates of MS. Some of these states include:

1. **Minnesota:** Known for having one of the highest rates of MS in the U.S.
2. **Vermont:** Also reports high MS prevalence.
3. **New Hampshire:** Similar to Vermont, New Hampshire has a higher-than-average rate of MS.
4. **Maine:** High prevalence of MS, consistent with trends in other northern states.
5. **Wisconsin:** Another state with a significant MS population.
6. **Iowa:** Reports higher rates of MS compared to the national average.
7. **Washington:** Has one of the higher prevalence rates on the West Coast.
8. **Oregon:** Similar to Washington, Oregon reports higher MS rates.

These states are generally located in the northern part of the United States, consistent with the global pattern where MS is more common in regions farther from the equator. However, it's important to note that MS prevalence data can vary over time and with new research, and access to healthcare services and diagnostic practices can also impact reported rates.

5.5 Vitamin D deficiency in the US:

Vitamin D deficiency in the United States is more common in certain regions and populations due to factors such as geographic location, climate, lifestyle, and demographics. The prevalence of Vitamin D deficiency tends to be higher in states with less sunlight exposure, especially during the winter months. Below are the regions and states generally known to have higher levels of Vitamin D deficiency:

Northeastern States	These states experience long, cold winters with limited sunlight, which reduces the natural production of Vitamin D through skin exposure to UVB rays.
Midwestern States	Like the Northeast, the Midwest has a significant period during the year when sunlight is scarce, leading to higher risks of Vitamin D deficiency.
Pacific Northwest	These states often have overcast weather and fewer sunny days, which can contribute to lower levels of Vitamin D.
Northern Plains States	The combination of high latitude and cold weather in these states can result in reduced sunlight exposure, leading to a higher prevalence of Vitamin D deficiency.
Alaska	Due to its extreme northern latitude, Alaska has very little sunlight for several months of the year, making Vitamin D deficiency more common.

Alignment of Regions/States with Vitamin D deficiency and High levels of MS

Regions/States with Vitamin D deficiency	Regions/States with High Levels of MS
4.1.1 Northeastern States Maine, Vermont, New Hampshire Massachusetts, New York	Vermont, New Hampshire, Maine
4.1.2 Midwestern States Michigan, Wisconsin, Minnesota Illinois, Ohio, Indiana	Minnesota, Wisconsin, Iowa
4.1.3 Pacific Northwest ● Washington ● Oregon	Washington, Oregon
4.1.4 Northern Plains States ● North Dakota, South Dakota, Montana	
Alaska	

Table to create understanding of ideal human vs defect in human:

Ideal/Expected	Defect
Absorbs Vitamin D	Does not absorb Vitamin D from sun or diet
Immune system does not attack person	Immune system over protects, attacks the person.
Muscles grow and regenerate	Muscles degrade and not repaired
Vision is assisted with glasses, not pharma	Vision degrades and cannot be repaired
Walks with ease until a senior citizen or injury	Ability to walk becomes a struggle and only gets worse

6 Create a Defect Report that pulls in the disease:

Date		Disease		Dataset	
Stage/Status		Duration		Severity	
Reproducibility		Expected Result		Actual Result	

Environment					
	Age	Gender	Diet	Geography	
Description					

7 Pulling it together: Bug -> Data -> Possible solutions

As I gathered data from publicly available datasets covering MS, I wanted to be able to isolate and develop a list of causations or early identifiers for MS.

- How to detect early?
- How to pause or stop progression?
- How to reverse the damage?

The list above was my ambition at the start of this citizen scientist project and paper. However, as I dove into the research of the disease using available tools, the need to make adjustments set in. My original goal was to use Chat GT4.0 to help me analyze large datasets, using my QA skillset I would look for specific patterns, assessing what is a causes vs correlations, unfortunately, I hit a brick wall. I discovered that unlike software engineering, software development, medical research is a much tighter almost closed community.

I discovered that to get access to large data sets you had to be part of an organization- academic or research. Most of the data in Multiple Sclerosis out there is very tightly held. If I compare the medical research community to the software development, I feel spoiled. In Software we love thinking aloud, we push software programs in various stages – beta, beta 2, pre-production we are not afraid of failure. We are not afraid of someone “saying hey you did that wrong, I tested this and look I found these issues.”

Software development and Medical Research are two different industries, that is a given. However, one could easily say, software engineering is all about research and advancement. A major difference is our ability to embrace failure as part of the process of finding the answer.

With the adjustment made, I pivoted and became pretty good at prompt engineering. This still allowed me to use my QA skills to ask better questions. As a QA person, have a tendency to not trust everything I hear/read. I narrowed my toolset to ChatGPT 4 and Google AI Search Labs. I focused on addressing Multiple Sclerosis and the theory that vitamin D deficiency is a contributing factor to developing MS

Am I done? No, not at all. Here is my bug reply:

Date	9/1/2024	Disease	Multiple Sclerosis	AI resources	ChatGPT 4.0 Google AI
Stage/Status	Pre-	Duration	Unknown	Severity	HIGH
Reproducibility	.33% (1/300)	Expected Result	Person w/Vitamin deficiency DOES develop MS	Actual Result	Person w/Vitamin deficiency DOES NOT develop MS.
Environment:					
	Age	Gender	Diet	Geography	
	20 –40 yrs old	Female	Unknown	Pacific Northwest	
Description	<p>It is found that persons who develop MS also have a deficiency in Vitamin D. It is therefore theorized that if a person has low levels of Vitamin, they are at risk of developing Multiple Sclerosis.</p> <p>Based on an evaluation of the sources of Vit D – geography and diet. Initial focus on Geography, as it is a factor that is easier to evaluate on its own. Contrast to diet where there is great variety (vegetarian, vegan, pescatarian) in the US and the world.</p> <p>Findings: More correlation over causation. Looking at the common countries/regions from Chat vs Google, the Countries and Regions with high levels of Vit D deficiency are not nations/regions with high levels of MS.</p> <p>In the US, there is a greater, almost confirmable correlation that those who live in areas with high levels of Vit D deficiency and MS rates are in direct correlation, possible causation.</p>				
Recommendation	<p>Vitamin D deficiency occurs in everyone who is diagnosed with MS.</p> <p>Not everyone who has a Vitamin D deficiency develops MS.</p> <p>Based on publicly available data, the Vitamin D deficiency theory is at best a correlation, not a cause.</p> <p>If this were a software defect, I would remove Vitamin D deficiency from my list of possible cause and continue testing. I would look for a possible cause with a much greater reproducibility rate, .33% is not enough.</p>				

8 References

Lawlor, Megan. 2021 “1 in 400: How Many People Will Get ALS?,” ALS Therapy Development Institute, entry posted July 19, 2021, <https://www.als.net/news/1-in-400-how-many-people-will-get-als/#:~:text=Blog:%20in%20400:%20How,ALS%20Therapy%20Development%20Institute> , (accessed 6/29/2024)

Author and entry date not provided, “Luckiest Man”, National Baseball Hall of Fame, <https://baseballhall.org/discover-more/stories/baseball-history/lou-gehrig-luckiest-man> (accessed 6/29/2024)

O, Alona. 2023 “Software Quality Assurance, Quality Control, and Testing – the Basics of Software Quality Management,” TestFort, entry posted October 10, 2023 <https://testfort.com/blog/qa-qc-testing-the-basics-of-quality-management> (accessed 8/19/2024)

Parciak, Tina, et al, 2023. “Introducing a core dataset for real-world data in multiple sclerosis registries and cohorts: Recommendations from a global task force,” National Library of Medicine” entry posted December 23, 2023 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10935622/> (accessed 3/3/2024)

Wallin, Mitchell, et al, 2019. “The prevalence of MS in the United States A population-based estimate using health claims data” entry posted March 5, 2019, <https://www.neurology.org/doi/10.1212/WNL.0000000000007035> (accessed 3/3/2024)

Collazo, Iris Marin. 2023 “ [Vitamin D and MS: Is there any connection?](https://www.mayoclinic.org/diseases-conditions/multiple-sclerosis/expert-answers/vitamin-d-and-ms/FAQ-20058258?p=1)” Mayo Clinic, entry posted April 19, 2023, <https://www.mayoclinic.org/diseases-conditions/multiple-sclerosis/expert-answers/vitamin-d-and-ms/FAQ-20058258?p=1> (accessed 4/17/2024)

Yadav, Yash, 2022. “Identifying plants and it’s medicinal properties., “ entry posted, posted 2022, <https://www.kaggle.com/search?q=Vitamin+D+deficiency> (accessed 6/20/2024) <https://medlineplus.gov/multiplesclerosis.html>