AI-Powered Data Extraction and Categorization

A Test Case Systematic Review of Multivitamin Effects

Georgii Filatov

March 2025

MedInsightConsulting.com

Table of Contents

Disclaimer

This paper represents a test case for AI-assisted data extraction and categorization in the context of a systematic review. While it applies a systematic approach, it does not adhere to the full methodological rigor typically required for systematic reviews or meta-analyses conducted by human researchers.

The conclusions drawn in this paper should be interpreted with caution, as they are based on AI-extracted and AI-classified data with limited human verification. While AI offers efficiency in handling large volumes of literature, its reliability in complex categorization and nuanced interpretation remains an area for further refinement.

This work is first and foremost an evaluation of an AI-based data extraction approach, with the topic of multivitamin effects serving as a test case. Readers should be mindful that AI-driven methodologies are evolving, and findings presented here should not be taken as definitive evidence for or against the use of multivitamins in healthy individuals.

Introduction

A while back, I was working on a systematic review and found myself drowning in filtering and data extraction. That's when it hit me: why not automate the process using AI? At the time, I had picked up some basic Python, so, with a little help from ChatGPT, I got to work. A couple of months later (I was working full-time, don't judge me), my first app was up and running. The idea was simple: upload a document, ask a question and choose a response format – options include "Free-form", "Yes/No", "Comma-separated list", "Mean", "Median", "Range", etc. The app then sends the document and question to an LLM (specifically ChatGPT), retrieves the response and writes it into a .csv file. The result is a neatly organized table, with questions as columns and documents as rows (or vice versa). There's a bit more going on under the hood, but at its core, it's a straightforward tool. Since then, I have been using it in my own work – mostly for screening large volumes of literature, while continuing to make improvements.

Naturally, I had put my data extraction app through its paces, but then I had a thought: why not test it in the very scenario I originally built it for? A full systematic review, relying solely on my app for data extraction, seemed like the perfect way to uncover both its strengths and (as I would soon find out) its many weaknesses.

Around the same time, a friend sent me a YouTube video arguing against taking multivitamins. To oversimplify, it made three main points: i) most people don't have vitamin deficiencies (except for vitamin D), ii) taking too much of some vitamins can be harmful and iii) multivitamins are pushed by the pharma companies because they make them money. I was skeptical but open-minded. Plus, I distinctly remembered speaking to doctors who were both for and against taking multivitamins. I decided to put my app to the test by exploring what the literature had to say on the subject. In hindsight, this was probably one of the worst topics I could have picked (more on that later). So, with boundless enthusiasm that only comes from not knowing what you're getting yourself into, I dove in.

Methods

Initial Literature Search

The first step was to conduct a search. I chose the PubMed database because it is freely accessible and integrates well with Entrez, which I had some experience with. The search terms used were:

("multivitamin*"[Title/Abstract] OR "multi vitamin*"[Title/Abstract] OR "multiple vitamins"[Title/Abstract] OR "vitamin complex*"[Title/Abstract]) AND ((adaptiveclinicaltrial[Filter] OR classicalarticle[Filter] OR clinicalstudy[Filter] OR clinicaltrial[Filter] OR clinicaltrialphasei[Filter] OR clinicaltrialphaseii[Filter] OR clinicaltrialphaseiii[Filter] OR clinicaltrialphaseiv[Filter] OR comparativestudy[Filter] OR controlledclinicaltrial[Filter] OR equivalencetrial[Filter] OR evaluationstudy[Filter] OR meta-analysis[Filter] OR multicenterstudy[Filter] OR observationalstudy[Filter] OR pragmaticclinicaltrial[Filter] OR randomizedcontrolledtrial[Filter] OR twinstudy[Filter] OR validationstudy[Filter]) AND (humans[Filter]) AND (2015:2025[pdat]))

The search returned 301 entries, which I exported to EndNote.

The next step was to filter eligible articles. I exported the EndNote library into a series of .txt files – one for each search result. I then asked the following question via my app, setting the output to “Yes/No”:

Does this publication fulfill all of the following criteria?
1. Reports on a study (of any design, including retrospective or prospective, interventional or observational, randomized controlled trials, cohort studies, case-control studies or cross-sectional studies) or a meta-analysis;
2. Examines the effect of multivitamin supplementation (defined as preparations containing multiple vitamins, not single vitamins or single-nutrient interventions);
3. Focuses on healthy individuals (including the general population, elderly, children, or pregnant women, but excluding individuals with diagnosed medical conditions such as cancer, cardiovascular disease, diabetes, or other illnesses).

This initial screening yielded 117 eligible articles.

Expanded Literature Search

Unfortunately, I know from experience that PubMed filters, which I used to limit the number of results in the search above, are not particularly reliable – PubMed isn’t as well curated as MEDLINE is this regard. So, I conducted the same search again, but this time without the filters, which returned 1,898 entries. After excluding the 301 previously assessed articles, I had 1,597 additional entries to evaluate.

This is where AI-assisted screening became particularly useful – I applied the same eligibility question to this expanded dataset. While manually screening nearly 1,600 abstracts would be a tedious and time-consuming process, the app made it much more manageable. Although I forgot to time it, it really didn’t take that long with my app, and in the end, I had another 116 entries, for a total of 233 eligible articles.

Data Extraction

Since this was primarily a test case for AI-assisted data extraction, I chose to only include full-text articles that were freely available on PubMed Central (PMC). I used Python and Entrez to programmatically download these articles. For studies not available in full text, I extracted bibliographic data (title, abstract, publication year, journal, etc.) from PubMed and included these records in the dataset.

Next, I ran a series of structured extraction queries on all 233 eligible studies, exporting the responses into a .csv file. This dataset was then combined with bibliographic data from EndNote, and I used Microsoft Excel and pivot tables to quickly summarize the extracted results.

Results

Article Characteristics

There were 131 (56.22%) abstracts and 102 (43.78%) full-text articles. Among full years included in the search, the number of articles published per year ranged from 10 to 37, and there appeared to be a declining trend (Figure 1). Not surprisingly, *Nutrients* published the highest number of papers included in this analysis – 30 (12.88%). Most other top journal also were nutrition-themed (Table 1).

Study Characteristics

Most articles were classified by ChatGPT as prospective (n=161, 69.10%), with smaller percentages classified as retrospective (n=64, 27.47%) or cross-sectional (n=6, 2.58%). ChatGPT's response was "Information not available" for 2 articles (0.86%) – more on this in the Discussion (Figure 2).

"Randomized controlled trial" (RCT) was the most common type of study design (n=72, 30.90%), followed closely by "Cohort study" (n=71, 30.47%), according to ChatGPT's classification (Table 2).

USA was the most common location for the included studies (n=49, 21.03%), followed by China (n= 25, 10.73%), Australia (n=17, 7.30%) and, somewhat unexpectedly, Tanzania (n=9, 3.86%). In addition, there were 17 (7.30%) studies described only as "Multinational" (Table 3).

Figure 1. Number of articles published per year.

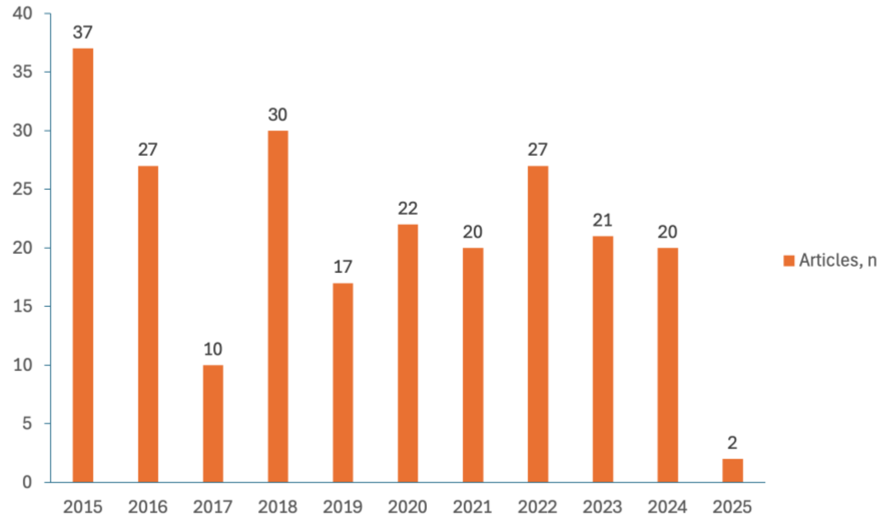

Table 1. Journals that published ≥4 of the included articles.

Journal	Articles, n	Articles, % of total
Nutrients	30	12.88%
J Nutr	9	3.86%
Am J Clin Nutr	9	3.86%
Matern Child Nutr	6	2.58%
Br J Nutr	5	2.15%
Front Nutr	5	2.15%
BMJ Open	5	2.15%
Eur J Nutr	4	1.72%
Neuro Endocrinol Lett	4	1.72%

Figure 2. Temporal classification of included articles.

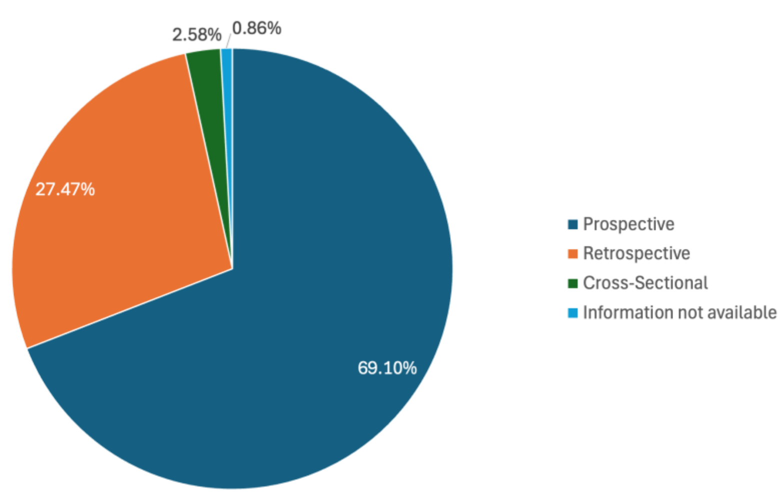

Table 2. Study designs represented by ≥2 included articles.

Study design	Articles, n	Articles, % of total
Randomized Controlled Trial	72	30.90%
Cohort Study	71	30.47%
Cross-Sectional Study	26	11.16%
Systematic Review	17	7.30%
Case-Control Study	9	3.86%
Meta-Analysis	7	3.00%
Observational Study	6	2.58%
Comparative Study	3	1.29%
Quasi-Experimental Study	2	0.86%
Mixed Methods Study	2	0.86%
Information not available	2	0.86%

Table 3. Countries where ≥3 studies were conducted.

Country	Articles, n	Articles, % of total
United States	49	21.03%
China	25	10.73%
Multinational	17	7.30%
Australia	17	7.30%
Tanzania	9	3.86%
Switzerland	8	3.43%
Denmark	7	3.00%
Canada	7	3.00%
Sweden	5	2.15%
Netherlands	5	2.15%
Japan	4	1.72%
Iran	4	1.72%
United Kingdom	4	1.72%
Norway	4	1.72%
Ethiopia	3	1.29%
Ukraine	3	1.29%
Spain	3	1.29%
Thailand	3	1.29%

Sample sizes ranged widely, from 16 to 2,019,862. Sample size information was not extracted in 20 articles (8.58%) (Table 4).

Mean age ranged from 4.0 years to 75.4 years. Most articles for which this mean age was available, focused on individuals aged 20.0–29.9 years (21, 9.01%) or 30.0–39.9 years (21, 9.01%). However, mean age could not be extracted from the majority of articles (Information not available: n=2, 0.86%; Not available: n=162, 69.53%) (Table 5).

The most common multivitamin category examined was "Multivitamin – composition not reported" (n=145, 62.23%), which was used when no details were given about its contents. Centrum Silver was the most common brand name mentioned (n=5, 2.15%). Notably, 9 (3.85%) articles were classified as "No multivitamin mentioned" – see Discussion for details (Table 6).

Outcome Categorization

After asking the question *"What primary (main) health outcomes or biomarkers were assessed in this study?"*, I realized that the variety of outcomes was significant. To deal with this, I decided to group outcomes into 10 categories: cognitive, reproductive health, mental health, cardiovascular, metabolic, nutritional, biomarker-based, cancer/oncology-related, general health and mortality. I then asked ChatGPT to rate the outcomes reported with multivitamins as "Beneficial", "Neutral" or "Harmful". To ensure that this categorization was accurate, I initially ran the same query three times for the first 3 outcome categories. After obtaining identical answers across all articles, I concluded that the likelihood of different responses was low enough to justify running the categorization query only once for the remaining outcomes.

No article reported more than 7 outcome categories, while a significant number of articles (n=44, 18.88%) did not evaluate any of the categories. Most articles reported 1 (n=69, 29.61%), 2 (n=74, 31.76%) or 3 (n=32, 13.73%) categories (Figure 3). The most commonly evaluated outcomes were nutritional outcomes (130 articles), followed by biomarker-based (68 articles) and reproductive health outcomes (56 articles).

For most outcomes, ChatGPT categorized the effect of multivitamins as "Beneficial" in the majority of articles. In fact, the percentage of articles reporting positive effects was ≥70% for 7 out of 10 categories evaluated. Cardiovascular (beneficial: n=11, 52.38%), general health (beneficial: n=7, 50.00%) and mortality outcomes (beneficial: n=3, 25.00%) constituted exceptions to this rule (Figure 4). Notably, some reproductive health (n=3, 5.36%), metabolic (n=1, 5.88%), nutritional (n=6, 4.62%) and biomarker-based outcomes (n=3, 4.41%) were categorized by ChatGPT as "Harmful" (Table 7). Among articles reporting beneficial outcomes, non-disclosure of conflicts of interest (COI) was common. However, it should be noted that many articles were represented only by their abstracts and bibliographic information, which frequently do not contain COI disclosures.

Table 4. Sample size groups.

Sample size group, n	Articles, n	Articles, % of total
1–49	19	8.15%
50–99	23	9.87%
100–199	29	12.45%
200–299	17	7.30%
300–499	18	7.73%
500–999	18	7.73%
1,000–2,999	21	9.01%
3,000–9,999	16	6.87%
10,000–19,999	15	6.44%
2,000–9,999	17	7.30%
10,000–49,999	12	5.15%
50,000–99,999	6	2.58%
100,000–2,000,000	1	0.43%
>2,000,000	1	0.43%
Not available	20	8.58%

Table 5. Mean age groups.

Mean age group	Articles, n	Articles, % of total
0–9.9 years	2	0.86%
10.0–19.9 years	5	2.15%
20.0–29.9 years	21	9.01%
30.0–39.9 years	21	9.01%
40.0–49.9 years	3	1.29%
50.0–59.9 years	6	2.58%
60.0–69.9 years	4	1.72%
≥70 years	7	3.00%
Information not available	2	0.86%
Not available	162	69.53%

Table 6. Multivitamin type.

Multivitamin type	Articles, n	Articles, % of total
Multivitamin - composition not reported	145	62.23%
Multivitamin - composition reported	41	17.60%
No multivitamin mentioned	9	3.86%
Centrum Silver	5	2.15%
LaVita	4	1.72%
Elevit	3	1.29%
Information not available	2	0.86%
Swisse Women 50+ Ultivite	2	0.86%
Supradyn	2	0.86%
Theravit	2	0.86%
Multivitamin - composition partially reported	2	0.86%

Figure 3. Number of outcome categories reported by the included articles.

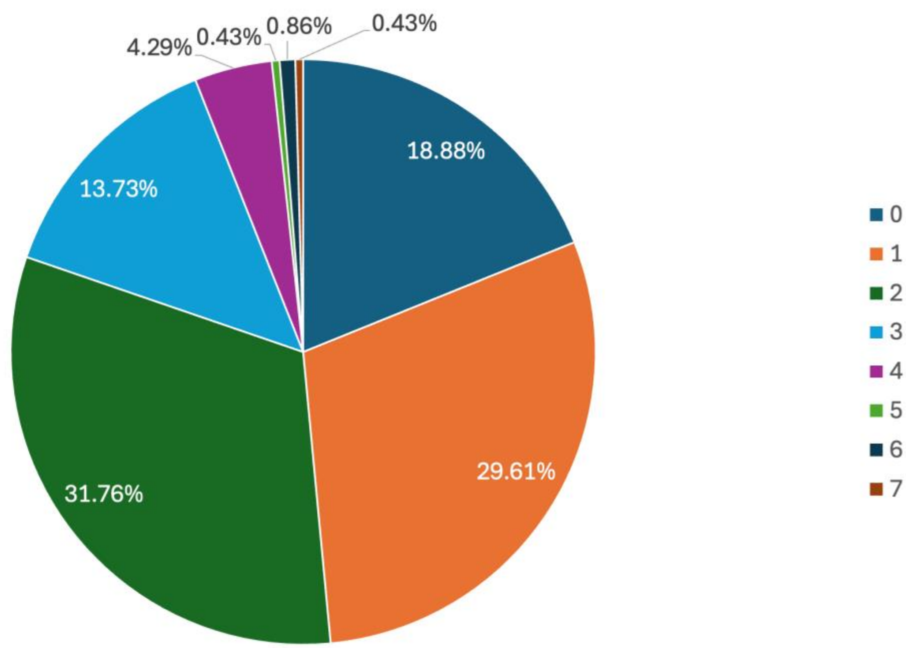

Figure 4. Outcome categorization and conflict of interest reporting.

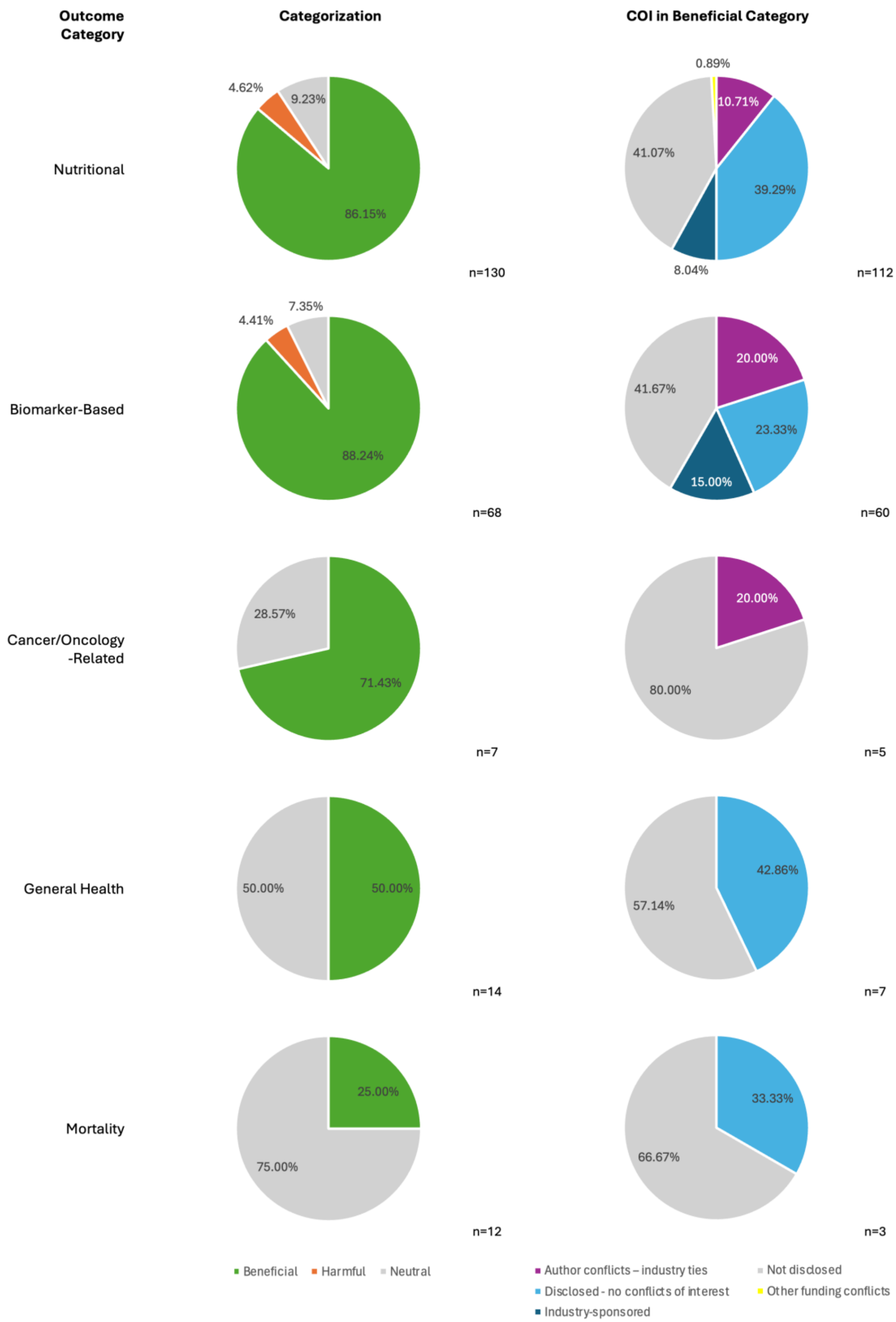
Outcome Category
Categorization
COI in Beneficial Category
Nutritional
4.62%
9.23%
86.15%
n=130
0.89%
10.71%
41.07%
39.29%
8.04%
n=112
Biomarker-Based
7.35%
4.41%
88.24%
n=68
20.00%
41.67%
23.33%
15.00%
n=60
Cancer/Oncology -Related
28.57%
71.43%
n=7
20.00%
80.00%
n=5
General Health
50.00%
50.00%
n=14
42.86%
57.14%
n=7
Mortality
25.00%
75.00%
n=12
33.33%
66.67%
n=3
Beneficial
Harmful
Neutral
Author conflicts – industry ties
Not disclosed
Disclosed - no conflicts of interest
Other funding conflicts
Industry-sponsored

Table 7. Articles reporting harmful outcomes.

Author	Source	Prospective or Retrospective	Study Design	Study Countries	Total Sample Size	Mean Age, years	Vitamin Type	Harmful Outcomes
Su, Gao et al. (2023)	Full-text	Prospective	Cohort Study	China	120,652	Not available	Elevit[a]	Reproductive health, nutritional
Lin, Chen et al. (2019)	Abstract	Prospective	Cohort Study	Taiwan	366	Not available	Multivitamin - composition not reported	Reproductive health, metabolic, nutritional
Yoshida, Takeuchi et al. (2020)	Full-text	Prospective	Cohort Study	Japan	98,787	31.6	Multivitamin - composition not reported	Reproductive health, nutritional
Christensen, Matthiessen et al. (2023)	Full-text	Retrospective[b]	Cross-Sectional Study	Denmark	499	Not available	Multivitamin - composition not reported	Nutritional
Jiang, Brumpton et al. (2015)	Full-text	Prospective	Cohort Study	Norway	16,952	39.4	Multivitamin - composition not reported	Nutritional
Pawlak, Vos et al. (2018)	Abstract	Cross-Sectional	Cross-Sectional Study	United States	74	Not available	Multivitamin - composition not reported	Biomarker-based
Ergon, Akil İ et al. (2018)	Abstract	Retrospective	Case-Control Study	Turkey	120	Not available	Multivitamin - composition not reported	Biomarker-based
Taghizadeh, Samimi et al. (2015)	Abstract	Prospective	Randomized Controlled Trial	Iran	70	Not available	Multivitamin - composition not reported	Biomarker-based

[a]While the article states that Elevit is the most common multivitamin on the Chinese market, the study did not ensure that participants received only Elevit.
[b]This study used data from the Danish National Survey of Diet and Physical Activity (DANSDA) 2011–2013, which is explicitly described as "cross-sectional" in the text. However, as part of DANSDA, face-to-face interview were conducted, during which participants were asked about dietary supplement intake over the previous year.

Discussion

When I set out to conduct this systematic review using my AI-based data extraction app, I expected challenges, but I may have underestimated just how many I would encounter. From unexpected study types to an overwhelming variety of outcomes, this review has been an excellent test of both the app and my own ability to adapt to a less conventional approach.

One of the most notable departures from a traditional systematic review was my decision to include systematic reviews and meta-analyses alongside primary studies. Normally, systematic reviews like this one focus exclusively on original research to avoid double-counting results. However, I wanted to ensure a comprehensive sample – partly to prove a point to a friend. (Lesson learned: methodological decisions should not be motivated by debates over YouTube videos.)

Another key choice I made was to include studies on all healthy individuals, which meant covering a wide range of populations, from pregnant women to the elderly. Typically, a systematic review would narrow its focus to a more homogenous group to facilitate clearer comparisons. While this broader inclusion allowed for a more comprehensive picture, it also had a side effect: a surprisingly large number of studies focused on reproductive outcomes. I did not anticipate how much of the literature on multivitamins in healthy individuals would center on pregnancy-related effects, which ended up skewing the overall representation of outcomes.

A third significant methodological decision was not applying selection criteria based on reported outcomes. Normally, systematic reviews have some outcome-related criteria, even if these are as simple as “reported efficacy” or “reported safety”. I, once again, wanted to have the most complete selection of articles I could get. As a result, I had to adopt a more exploratory approach and let ChatGPT categorize the outcomes. This resulted in a highly diverse set of reported effects, ranging from cognitive function to metabolic markers. The need to categorize outcomes after the fact placed a lot of reliance on the AI, which, while efficient, struggled with some ambiguous cases.

Challenges and AI Performance

Speaking of AI reliability, one of the most critical limitations of this review was the extent to which I verified the extracted data. In previous projects, my experience with this app suggested it was quite reliable, but I also knew it could struggle with complex questions. In this review, I only verified some of the extracted data, and while the app performed well overall, there were some issues. The first had to do with the type of multivitamin described in the article. I originally extracted this information using this question:

> *What type of multivitamin or vitamin supplement was used in this study? Provide the name, composition, or formulation. If not specified, state "Not reported."*

However, this produced comma-separated lists of interventions for many articles. Because I wanted to limit the amount of data processing, I decided to reword the question in such a way as to focus on multivitamins only. However, no matter how much I tried, at least some of the articles were classified as “No multivitamin mentioned”. In the end, I decided to cut my losses and use the wording that resulted in the smallest number of “No multivitamin mentioned” classifications.

In another illustrative example, ChatGPT classified the study by Taghizadeh, Samimi et al. (2015) as reporting harmful biomarker-based outcomes, but a closer look showed that this study actually compared multivitamin supplementation alone versus multivitamins plus minerals, with the latter group experiencing better results. This distinction is crucial but was lost in ChatGPT’s categorization. This highlights an important limitation: the app’s performance is dependent on how clearly the information is presented in the source material and how well it aligns with the question.

Interestingly, when I tested the consistency of outcome classification, running the same categorization query three times for cognitive, reproductive and mental health outcomes produced identical classifications across all studies. This suggests that ChatGPT is internally consistent when applied to structured classification tasks. However, consistency does not necessarily imply accuracy – if a misclassification occurred on the first attempt, repeating the query would not correct it.

All responses were “Information not available” for two articles, which on closer examination, turned out to be two Cochrane reviews: Peña-Rosas, De-Regil et al. (2015) and Suchdev, Jefferds et al. (2020). I believe that this was because both papers are extremely long (194 and 164 pages, respectively). However, I am still investigating.

Future Improvements

Given these limitations, a logical next step would be to conduct a similar review but with a more focused and uniform sample of articles. Instead of covering such a broad population and an extensive range of outcomes, a future review could concentrate on a specific subgroup, such as only RCTs in adults or only studies reporting cognitive outcomes. This narrower scope would allow for a higher degree of manual verification, improving data reliability while still benefiting from AI-assisted extraction. Additionally, refining the AI prompts and incorporating automated quality checks could further enhance accuracy. Such an approach would provide a more controlled test of the AI's capabilities while yielding results that are easier to interpret and compare.

Long-Term Potential

Looking further ahead, if the reliability issues of AI-based data extraction and classification can be fully addressed, and if statistical tools can be integrated into the process, this approach could enable near-instant systematic reviews and even live meta-analyses. Instead of taking weeks, months or years to synthesize evidence, AI-assisted reviews could provide real-time updates on the state of knowledge in a particular field. This could be especially valuable in fast-moving areas of medicine and science, where doctors and researchers need up-to-date information to inform decisions. A well-designed AI-driven system could automatically extract and analyze new findings as they are published, generating interactive dashboards or automated reports. While we are not there yet, the potential for transforming evidence synthesis in medicine is significant.

Final Thoughts

Despite its limitations, this AI-assisted systematic review was an interesting experiment. It demonstrated that AI can significantly accelerate data extraction and categorization, but it still requires human oversight, particularly for nuanced interpretation and quality control. For future iterations, I'd consider:

- Manually verifying a larger portion of the data
- Incorporating validation steps to catch errors and misclassifications earlier

In the end, was this the best topic to test my app on? Probably not. The sheer breadth of the literature and diversity of reported outcomes made this a challenging first case study. But that's the nature of experimentation: sometimes, you learn the most when things don't go as planned.

References

Christensen, C., J. Matthiessen, S. Fagt and A. Biltoft-Jensen (2023). "Dietary supplements increase the risk of excessive micronutrient intakes in Danish children." Eur J Nutr **62**(6): 2449-2462.

Ergon, E. Y., O. Akil İ, F. Taneli, A. Oran and B. C. Ozyurt (2018). "Etiologic risk factors and vitamin D receptor gene polymorphisms in under one-year-old infants with urolithiasis." Urolithiasis **46**(4): 349-356.

Jiang, L., B. Brumpton, A. Langhammer, Y. Chen and X. M. Mai (2015). "Intake of multivitamin supplements and incident asthma in Norwegian adults: the HUNT study." ERJ Open Res **1**(2).

Lin, C. Y., Y. J. Chen, S. H. Lee, C. P. Kuo, M. S. Lee and M. C. Lee (2019). "Uses of dietary supplements and herbal medicines during pregnancy in women undergoing assisted reproductive technologies- A study of taiwan birth cohort." Taiwan J Obstet Gynecol **58**(1): 77-81.

Pawlak, R., P. Vos, S. Shahab-Ferdows, D. Hampel, L. H. Allen and M. T. Perrin (2018). "Vitamin B-12 content in breast milk of vegan, vegetarian, and nonvegetarian lactating women in the United States." Am J Clin Nutr **108**(3): 525-531.

Peña-Rosas, J. P., L. M. De-Regil, H. Gomez Malave, M. C. Flores-Urrutia and T. Dowswell (2015). "Intermittent oral iron supplementation during pregnancy." Cochrane Database Syst Rev **2015**(10): Cd009997.

Su, J., S. Gao, R. Yan, R. Liu, S. Su, X. Nie, X. Liu, E. Zhang, S. Xie, J. Liu, Y. Zhang, W. Yue, C. Yin and X. Peng (2023). "Is the Tradeoff between Folic Acid or/and Multivitamin Supplementation against Birth Defects in Early Pregnancy Reconsidered? Evidence Based on a Chinese Birth Cohort Study." Nutrients **15**(2).

Suchdev, P. S., M. E. D. Jefferds, E. Ota, K. da Silva Lopes and L. M. De-Regil (2020). "Home fortification of foods with multiple micronutrient powders for health and nutrition in children under two years of age." Cochrane Database Syst Rev **2**(2): Cd008959.

Taghizadeh, M., M. Samimi, F. Kolahdooz, Z. Tabassi, M. Jamilian and Z. Asemi (2015). "Effect of multivitamin versus multivitamin-mineral supplementation on metabolic profiles and biomarkers of oxidative stress in pregnant women: a double-blind randomized clinical trial." J Matern Fetal Neonatal Med **28**(11): 1336-1342.

Yoshida, S., M. Takeuchi, C. Kawakami, K. Kawakami and S. Ito (2020). "Maternal multivitamin intake and orofacial clefts in offspring: Japan Environment and Children's Study (JECS) cohort study." BMJ Open **10**(3): e035817.

Appendix 1. Data Extraction Questions

General Study Characteristics

1. Prospective or Retrospective?

Indicate whether the study is Prospective, Retrospective, or Cross-Sectional. Provide only the type of study (e.g., “Prospective,” “Retrospective,” or “Cross-Sectional”).

2. Study Design

What is the study design? Provide only the type of study (e.g., "Randomized Controlled Trial," "Cohort Study," "Systematic Review").

3. Study Countries

In which country or countries was this study conducted? Provide a comma-separated list of exact countries. If only "multinational" is stated without details, enter "Multinational."

Population Characteristics

4. Vitamin Group Size

What was the number of participants who received vitamins? Provide participant count. + Integer parser

5. Mean Age

What was the mean age of participants? Provide a single numerical value. If not reported, state "Not reported." + Floating point parser

Intervention & Control Details

6. Vitamin Type

Identify the multivitamin supplement used in the study. For this question, a multivitamin is defined as a supplement containing two or more vitamins, with or without additional minerals. Follow these rules:

- If a brand name is provided, report only the brand name. Do not include additional details such as composition or dosage.
- If no brand name is provided, but the composition of a multivitamin is explicitly reported, state:
"Multivitamin – composition reported."
- If only part of the composition is reported (e.g., "folic acid-containing multivitamin"), state:
"Multivitamin – composition partially reported."
- If no brand name is provided and the composition is not reported, state:
"Multivitamin – composition not reported."
- If the study refers to multivitamins but does not evaluate a specific product (e.g., observational studies, meta-analyses, prevalence studies), classify according to the available information:
- "Multivitamin – composition reported" (if full composition given)
- "Multivitamin – composition partially reported" (if only some components are listed)
- "Multivitamin – composition not reported" (if no details are given about contents)
- If the study does not contain any reference to multivitamin use in any form, state:
"No multivitamin mentioned."

Each response should be concise and no longer than five words.

Outcome-Specific Categorization

7. Cognitive Outcomes:

If the study assessed cognitive outcomes, classify the effects of multivitamin supplementation only on cognitive function as 'Beneficial,' 'Neutral,' 'Harmful,' or 'Not reported,' based solely on the study findings. Ignore findings related to control conditions, individual nutrient or mineral supplements, or any interventions other than multivitamins. If the study did not assess the effect of multivitamin supplementation on cognitive outcomes, state 'Not reported.' Do not provide explanations—only the classification.

8. Reproductive Health Outcomes

If the study assessed reproductive health outcomes, classify the effects of multivitamin supplementation only on reproductive health as 'Beneficial,' 'Neutral,' 'Harmful,' or 'Not reported,' based solely on the study findings. Ignore findings related to control conditions, individual nutrient or mineral supplements, or any interventions other than multivitamins. If the study did not assess the effect of multivitamin supplementation on reproductive health, state 'Not reported.' Do not provide explanations—only the classification.

9. Mental Health Outcomes

If the study assessed mental health outcomes (e.g., mood, anxiety, depression), classify the effects of multivitamin supplementation only as ‘Beneficial,’ ‘Neutral,’ ‘Harmful,’ or ‘Not reported,’ based solely on the study findings. Ignore findings related to control conditions, individual nutrient or mineral supplements, or any interventions other than multivitamins. If the study did not assess the effect of multivitamin supplementation on mental health, state ‘Not reported.’ Do not provide explanations—only the classification.

10. Cardiovascular Outcomes

If the study assessed cardiovascular outcomes, classify the cardiovascular effects of multivitamin supplementation only as ‘Beneficial,’ ‘Neutral,’ ‘Harmful,’ or ‘Not reported,’ based solely on the study findings. Ignore findings related to control conditions, individual nutrient or mineral supplements, or any interventions other than multivitamins. If the study did not assess the effect of multivitamin supplementation on cardiovascular outcomes, state ‘Not reported.’ Do not provide explanations—only the classification.

11. Metabolic Outcomes

If the study assessed metabolic outcomes (e.g., weight, glucose regulation, diabetes risk), classify the metabolic effects of multivitamin supplementation only as ‘Beneficial,’ ‘Neutral,’ ‘Harmful,’ or ‘Not reported,’ based solely on the study findings. Ignore findings related to control conditions, individual nutrient or mineral supplements, or any interventions other than multivitamins. If the study did not assess the effect of multivitamin supplementation on metabolic outcomes, state ‘Not reported.’ Do not provide explanations—only the classification.

12. Nutritional Outcomes

If the study assessed nutritional outcomes (e.g., dietary intake, nutrient absorption, growth metrics), classify the nutritional effects of multivitamin supplementation only as ‘Beneficial,’ ‘Neutral,’ ‘Harmful,’ or ‘Not reported,’ based solely on the study findings. Ignore findings related to control conditions, individual nutrient or mineral supplements, or any interventions other than multivitamins. If the study did not assess the effect of multivitamin supplementation on nutritional outcomes, state ‘Not reported.’ Do not provide explanations—only the classification.

13. Biomarker-Based Outcomes

If the study assessed biomarker-based outcomes (e.g., vitamin levels, inflammatory markers), classify the biomarker findings of multivitamin supplementation only as ‘Beneficial,’ ‘Neutral,’ ‘Harmful,’ or ‘Not reported,’ based solely on the study findings.

Ignore findings related to control conditions, individual nutrient or mineral supplements, or any interventions other than multivitamins. If the study did not assess the effect of multivitamin supplementation on biomarkers, state ‘Not reported.’ Do not provide explanations—only the classification.

14. Cancer/Oncology-Related Outcomes

If the study assessed cancer/oncology-related outcomes (e.g., cancer risk, tumor markers), classify the cancer-related effects of multivitamin supplementation only as ‘Beneficial,’ ‘Neutral,’ ‘Harmful,’ or ‘Not reported,’ based solely on the study findings. Ignore findings related to control conditions, individual nutrient or mineral supplements, or any interventions other than multivitamins. If the study did not assess the effect of multivitamin supplementation on cancer-related outcomes, state ‘Not reported.’ Do not provide explanations—only the classification.

15. General Health Outcomes

If the study assessed general health outcomes (e.g., infections, hospitalization, overall well-being), classify the general health effects of multivitamin supplementation only as ‘Beneficial,’ ‘Neutral,’ ‘Harmful,’ or ‘Not reported,’ based solely on the study findings. Ignore findings related to control conditions, individual nutrient or mineral supplements, or any interventions other than multivitamins. If the study did not assess the effect of multivitamin supplementation on general health, state ‘Not reported.’ Do not provide explanations—only the classification.

16. Mortality Outcomes

If the study assessed mortality outcomes (e.g., all-cause mortality, disease-specific mortality), classify the mortality effects of multivitamin supplementation only as ‘Beneficial,’ ‘Neutral,’ ‘Harmful,’ or ‘Not reported,’ based solely on the study findings. Ignore findings related to control conditions, individual nutrient or mineral supplements, or any interventions other than multivitamins. If the study did not assess the effect of multivitamin supplementation on mortality, state ‘Not reported.’ Do not provide explanations—only the classification.

Conflicts of interest

17. Conflict of Interest Disclosure

Classify the study’s conflict of interest disclosure into one of the following categories based on the provided text:

- ‘Not disclosed’ – No conflict of interest statement is provided.

- ‘Disclosed - no conflicts of interest’ – The study explicitly states that the authors have no conflicts of interest.
- ‘Industry-sponsored’ – The study was funded by or conducted in collaboration with an industry entity (e.g., pharmaceutical, supplement, or food company).
- ‘Author conflicts – industry ties’ – At least one author has reported financial or professional ties to industry (e.g., consulting fees, stock ownership, grants).
- ‘Other funding conflicts’ – The study was funded by a non-industry entity with potential conflicts (e.g., government agency, advocacy group, foundation with vested interests).
- ‘Unclear disclosure’ – A conflict of interest statement is present, but the nature of conflicts is ambiguous or insufficiently detailed.

Select only one category based on the study’s disclosure. Do not summarize or provide explanations.