

# A Common Language? Analyzing the Use of Health-Related Vocabulary Between Laypeople and Medical Professionals

Lucien Baumgartner (lucien.baumgartner@philos.uzh.ch)  
University of Zurich, Department of Philosophy, 8008 Zurich, Switzerland

Kevin Reuter (kevin.reuter@gu.se)  
University of Gothenburg, Department of Philosophy, Linguistics and Theory of Science, 405 30 Gothenburg, Sweden

Somogy Varga (varga@cas.au.dk)  
Aarhus University, Department of Philosophy and History of Ideas, 8000 Aarhus C, Denmark

## Abstract

The meaning of being healthy is widely debated, with many suggesting it is a multidimensional concept encompassing key dimensions such as the absence of disease, the presence of well-being, and a healthy lifestyle. While recent studies indicate that lifestyle may be a dominant dimension, it remains unclear whether this holds true across populations or if significant differences exist, particularly between laypeople and healthcare professionals. Our studies reveal a difference, but surprisingly, in the opposite direction of what the literature would predict: medical professionals are substantially more likely than laypeople to frame discussions of “healthy” and “unhealthy” in lifestyle contexts. This result challenges prevailing assumptions about the biomedical focus of healthcare professionals and has implications for improving health communication.

**Keywords:** health; corpus analysis; health-related terminology; comparison laypeople–medical professionals.

## Introduction

Health-related terms such as “healthy”, “ill”, and “sick” play a crucial role in both clinical interactions between patients and health professionals and in public health communication. While effective clinical communication and public health messaging rely on shared understandings of these terms, their meanings are neither universally agreed upon nor is it clear whether different groups, such as laypeople and health professionals, use them differently.

For instance, in the literature, health is often conceptualized as a multidimensional concept encompassing various aspects, including the absence of disease, the presence of well-being, and adherence to a particular lifestyle involving specific dietary and behavioral habits and exercising (Herzlich, 1973; Williams, 1983; Millstein & Irwin, 1987; Fugelli & Ingstad, 2001; Hughner & Kleine, 2004; Blaxter, 2010; Bishop & Yardley, 2010; Downey & Chang, 2013; Varga & Latham, 2024; Reuter, Latham, & Varga, 2025). However, it remains unclear whether laypeople and health professionals emphasize these aspects differently, highlighting certain dimensions depending on the context. While evidence is inconclusive, some extant research suggests that medical professionals tend to adopt a biomedical perspective, defining health primarily as the absence of disease, whereas laypeople may take a broader, lifestyle-oriented view (e.g., Huber et al., 2016). Such potential misalignment between professionals and laypeople could impact health communication, hinder adherence to medical advice, and contribute to poorer health

outcomes, underscoring the need for further research to better understand these differences and their implications.

In this paper, we seek to complement conceptual, qualitative, and experimental research on this topic by identifying broad patterns in the usage of key health-related terms and providing a deeper understanding of the multidimensional nature of the concept of health. To achieve this, we conduct a computational corpus analysis using Natural Language Processing (NLP) methods—an approach that is novel in the literature—to systematically examine health-related discourse.<sup>1</sup> Although this approach sacrifices the control offered by direct manipulation of stimuli, it provides relatively unbiased access to a vast amount of data. Specifically, we will analyze a subreddit designed for laypeople to engage with healthcare professionals (r/AskDocs), offering large-scale, naturally occurring interactions that provide insights into how these groups conceptualize health.

We conduct two experiments to examine patterns in health-related language: Experiment 1 tests whether laypeople and medical professionals use health-related vocabulary in systematically different ways. Experiment 2 investigates whether laypeople and medical professionals use the terms “healthy” and “unhealthy” primarily in disease- or lifestyle-related contexts.

## Data

Our initial corpus comprises 122,589 Reddit submissions and 315,017 comments from the subreddit r/AskDocs, spanning November 2022 to March 2023. Reddit is one of the world’s biggest online forums and consists of many different communities called “subreddits”, which serve vastly different purposes. In the subreddit we are interested in—r/AskDocs—laypeople seek advice on medical issues from medical professionals verified by moderators. These professionals include a diverse group, such as clinical doctors, nurses, and psychiatrists. We only retained submissions made by laypeople and comments made by verified medical professionals.

Submissions and comments were all treated as individual documents, split into sentences, preprocessed, and filtered to retain only those containing a set of pre-defined target terms. The pre-defined health-related vocabulary we examine in this

<sup>1</sup>For related computational corpus analyses employing NLP techniques, see, e.g., Baumgartner, 2022; Baumgartner & Kneer, 2025; Reuter & Baumgartner, 2025.

paper includes the adjectives “healthy”, “ill”, “sick”, “unhealthy”, “unwell”, “well”, and the nouns “disease”, “dysfunction”, “illness”, “health”, “disorder”. After preprocessing and filtering, we retained only 8,986 submissions (from laypeople) and 1466 comments (from medical professionals).<sup>2</sup>

## Experiment 1

In the first experiment, we test whether laypeople and medical professionals use health-related terms in similar semantic contexts.<sup>3</sup> For this, we compare the similarity of contextualized token-level embeddings of our target terms between the two populations. Token-level embeddings are high-dimensional vector representations of natural language tokens in their immediate context of use—in our case, single sentences—which we generate using a pre-trained transformer model.<sup>4</sup> The pairwise similarity between tokens of the same type is calculated using cosine distance between their embedding vectors. We then use permutation tests to assess whether the average observed similarity scores between laypeople and medical professionals differ significantly from a random distribution of means, generated by randomly permuting the group labels (i.e., shuffling the assignment of laypeople and medical professionals as the sentences’ authors).

We suspect that the education of medical professionals fosters a specialized usage of health-related terms, which stands in contrast to the usage of laypeople. This disparity suggests systematic differences, leading us to propose the following hypothesis:

$H_1$ : The observed similarities between the use of our target terms by laypeople and medical professionals differ systematically from a random distribution, suggesting that laypeople and medical professionals use target terms differently.

The observed mean similarity  $\bar{S}_{\text{observed}}$  for a given target term is calculated as:

$$\bar{S}_{\text{observed}} = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^m \text{cosine\_similarity}(\mathbf{L}_i, \mathbf{P}_j),$$

where  $\mathbf{L}_i$  are the embeddings of target tokens in sentences written by laypeople,  $\mathbf{P}_j$  are the embeddings of target token in sentences written by medical professionals,  $N$  is the total number of possible comparisons (pairs of sentences) between the two groups, and  $n$  and  $m$  are the number of sentences written by laypeople and medical professionals, respectively.

For each permutation  $k$ , we shuffle the group labels and calculate the mean similarity for the permuted data. The permuted mean similarity  $\bar{S}_{\text{permuted}}^{(k)}$  for a given target term is calculated as:

$$\bar{S}_{\text{permuted}}^{(k)} = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^m \text{cosine\_similarity}(\mathbf{L}_i^{(k)}, \mathbf{P}_j^{(k)}),$$

where  $\mathbf{L}_i^{(k)}$  and  $\mathbf{P}_j^{(k)}$  are the original embeddings but with the permuted group assignment. The permutation of the group labels in the permutation step is intended to create a random distribution of similarities that would occur if there were no true difference between the groups (laypeople vs. medical professionals). By randomly swapping the group labels (while keeping the individual token embeddings intact), we break any actual relationship between the group and the similarity scores. This allows us to assess whether the observed similarity scores are significantly higher than what would be expected by chance, based on this random reassignment.

After performing  $K$  permutations, the  $p$ -value is calculated as the proportion of permuted mean similarities that are greater than or equal to the observed mean similarity:

$$p = \frac{\sum_{k=1}^K \mathbb{1}(\bar{S}_{\text{permuted}}^{(k)} \geq \bar{S}_{\text{observed}})}{K},$$

where  $\mathbb{1}(\bar{S}_{\text{permuted}}^{(k)} \geq \bar{S}_{\text{observed}})$  is an indicator function that equals 1 if the permuted mean similarity is greater than or equal to the observed mean, and 0 otherwise, and  $K$  is the total number of permutations.

In addition to the permutation tests, we will perform a Welch’s t-test to compare the average semantic similarity within the two groups. This will tell us whether one of the groups has a significantly higher internal semantic coherence in the use of terms than the other. It might appear intuitive to expect medical professionals to have a significantly higher internal semantic cohesion than laypeople, because they use a more technical and restricted language, whereas laypeople use health-related expressions more loosely. However, it is, conversely, possible that while laypeople use the target terms more loosely, they do so consistently in the same way. Hence, we refrain from postulating a directed hypothesis:

$H_2$ : There are significant differences in intra-group semantic similarity between laypeople and medical-professionals.

<sup>2</sup>It is worth noting that of the total number of comments in r/AskDocs for this time period only 12.72% were actually by health-care professionals. This partly explains the low number of occurrences for that population.

<sup>3</sup>The scripts and results for all experiments are available on OSF at <https://osf.io/fznmq/>. The data cannot be shared due to privacy policies. However, we provide the scripts to reproduce the corpus using the Reddit bulk-data freely available from Academic Torrents: <https://academictorrents.com/details/ba051999301b109eab37d16f027b3f49ade2de13>.

<sup>4</sup>We use Google’s ELECTRA model (Clark, Luong, Le, & Manning, 2020), available at: <https://huggingface.co/google/electra-base-discriminator>. We selected the model based on a small set of  $2 \times 5$  curated sentences exemplifying the use of “health” and “healthy” in clearly disease- or lifestyle-related contexts (i.e. 5 per group). A one-sided Welch’s t-test confirmed that the semantic similarity within the lifestyle group was significantly higher than between lifestyle- and disease-related uses,  $t = 2.0488$ ,  $p = 0.0491$ . The within-group similarity for disease-related uses was also significantly higher than the between-groups similarity,  $t = 2.1296$ ,  $p = 0.0457$ . None of the other models we tested passed this sanity check.

A Kolmogorov-Smirnov (KS) test is used as a robustness check for the findings of the Welch’s t-test.<sup>5</sup>

## Results

Figure 1 shows the distribution of the observed pairwise similarity scores between token embeddings of contributions made by laypeople and those made by medical professionals, by target token.

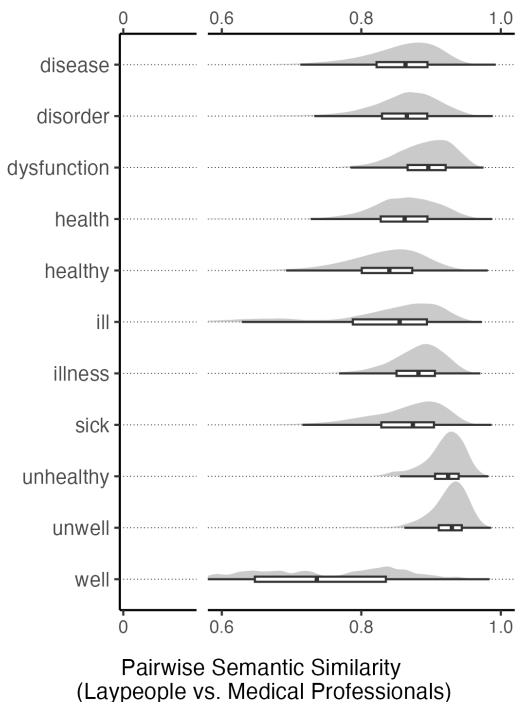


Figure 1: Observed pairwise cosine similarity between token-level embeddings of laypeople’s submissions and medical professionals in r/AskDocs.

Table 1 details the summary statistics of the permutation tests, with 10,000 permutations. It includes the observed average similarity per target term, and the  $p$ -values indicating whether the similarity score differs significantly from a null distribution. The results show that there are no significant differences in the distribution of observed cosine similarities and the permuted ones, for neither one of our target terms (on 0.05-alpha level). Hence, based on the permutation tests, the null hypothesis cannot be rejected in favor of  $H_1$ . This implies that laypeople and medical professionals use the target terms in similar ways, on average.

<sup>5</sup>If the t-test shows a significant difference (in means) but the KS test does not (no significant distributional difference), it could indicate that while the central tendencies of the groups differ, their distributions might still be similar overall, possibly with different variances or shapes. Conversely, if the KS test rejects the null hypothesis but the t-test does not, this might suggest that there are significant distributional differences (such as variability or skewness) that are not captured by the t-test, which only looks at the mean.

Table 1: Permutation test statistics for pairwise cosine similarity of sentence embeddings with permuted group labels.

Term	$\bar{S}_{\text{observed}}$	$p$
disease	0.8437	0.5236
disorder	0.8457	0.3515
dysfunction	0.8914	0.0975
health	0.8494	0.8575
healthy	0.8330	0.9882
ill	0.8138	0.1655
illness	0.8539	0.9987
sick	0.8539	0.8247
unhealthy	0.9194	0.9544
unwell	0.9208	0.4555
well	0.7355	0.7135

Table 2 shows the results of the Welch’s t-tests for each target term, comparing the internal semantic cohesion in the two groups (laypeople vs. medical professionals). The results indicate that there are significant differences regarding the semantic cohesion between the groups for most target terms, on 0.05-alpha level, except for “well” ( $t = 1.6681$ ,  $p = 0.0985$ ). The KS tests largely converge on the same findings (see Table 3 in Appendix A). Hence,  $H_2$  cannot be rejected for the majority of target terms, with the aforementioned exception. The average semantic similarity within the group is generally higher for medical professionals compared to laypeople, rather than the other way around, except for “illness”, “sick”, “unhealthy”, and “well”. For the latter, the semantic cohesion is, perhaps surprisingly, higher for laypeople than medical professionals. Unfortunately, the results from Experiment 1 do not allow any inferences as to why this is the case, though.

Table 2: Welch’ t-tests for the comparison of intra-group semantic similarity between laypeople ( $L$ ) and medical professionals ( $P$ ).

Term	$\bar{L}_{\text{observed}}$	$\bar{P}_{\text{observed}}$	$t$	$p$
disease	0.8423	0.8545	-21.8139	<0.0001
disorder	0.8415	0.8628	-13.4190	<0.0001
dysfunction	0.8844	0.9121	-26.1540	<0.0001
health	0.8555	0.8632	-3.8448	0.0001
healthy	0.8366	0.8373	-4.7599	<0.0001
ill	0.8025	0.8399	-14.0387	<0.0001
illness	0.8788	0.8422	10.3446	<0.0001
sick	0.8587	0.8568	2.8386	0.0045
unhealthy	0.9253	0.9151	4.9219	<0.0001
unwell	0.9187	0.9294	-9.2268	<0.0001
well	0.7467	0.7248	1.6681	0.0985

## Discussion

On the one hand, Experiment 1 suggests that the observed similarity between laypeople’s and professionals’ use of the

target terms is not systematically different from chance, which means that the two groups use the terms similarly on average. On the other hand, we find that medical professionals use most terms more consistently compared to laypeople, with some exceptions where it is the other way around. At first glance, this might appear contradictory. We perhaps intuitively expect that if one group is significantly more consistent in their word use (higher intra-group similarity), it should negatively impact the average inter-group similarity. However, inter-group similarity reflects shared uses and is not directly affected by how varied the rest of a group's uses are. Intra-group cohesion only negatively affects inter-group similarity if one group uses the term in a way that never overlaps semantically with the other group. Our results, though, suggest exactly the opposite.

## Experiment 2

In the introduction, we highlighted a central issue concerning the meaning of the concept of health: While some scholars argue that health is merely the absence of disease, others contend that it is a broader, more positive state linked to lifestyle factors such as nutrition and physical activity. In all likelihood, when people talk about health, they can invoke either one or both dimensions of the concept. On the one hand, individuals may be referring to the *absence of disease*—a negative definition that focuses on the elimination or prevention of illness. On the other hand, they might be alluding to a positive state that encompasses *lifestyle factors* such as proper nutrition and physical activity.

For example, consider a conversation where someone mentions that they are “trying to be healthy”. Depending on the context, this statement could imply that the individual is vigilant about avoiding diseases through regular medical check-ups or vaccinations. Alternatively, it might indicate that the person is adopting a balanced diet and engaging in regular physical exercise. Often, both sets of actions are intertwined, making it challenging to isolate one dimension from the other. This complexity is further compounded when considering interactions across diverse communities: while healthcare professionals often use specific norms and cues to discuss health, these may not be as familiar or accessible to lay audiences.

St Claire et al. (1996) provide some initial evidence that there may be notable differences in the way laypeople and medical professionals apply “healthy” and “unhealthy” within disease- and lifestyle-related contexts. In contrast, our Experiment 1 shows that laypeople and professionals use many of our target terms similarly overall. Hence, we expect the following:

*H*<sub>3</sub>: The distribution of disease- and lifestyle-related sentences containing “healthy” or “unhealthy” does not differ from the expected distribution (i.e., equal proportions for each category).

*H*<sub>4</sub>: There is no significant difference between laypeople and medical professionals in the likelihood of the terms

“healthy” and “unhealthy” being related to a disease context versus a lifestyle context.

In order to examine how the terms “healthy” and “unhealthy” are used to describe both disease-related aspects as well as lifestyle properties, we drew a random sample of 100 sentences from our corpus containing “healthy” and “unhealthy”. These sentences were manually annotated by the authors, classifying each as disease-related, lifestyle-related, or unclear.<sup>6</sup> We also drew a random sample of 100 sentences featuring the terms “sick” and “ill”.

For “healthy” and “unhealthy”, intercoder reliability was notably high, with a Fleiss'  $\kappa$  of 0.8632. The true labels for the annotation were determined using majority voting. In contrast, none of the 100 sentences containing “sick” or “ill” were categorized as lifestyle-related. Instead, nearly all were clearly disease-related, with a few too short or vague to be definitively classified. As a key preliminary finding, we can conclude:<sup>7</sup>

1. In health-related contexts, “healthy” and “unhealthy” are ambiguous, exhibiting at least two distinct interpretations: a lifestyle-related reading and a disease-related reading.
2. In health-related contexts, (almost) all uses of “sick” and “ill” indicate the presence of disease.

Due to the extreme class imbalances in the manually annotated sample for “sick” and “ill”, we excluded them from the training sample. Accordingly, the hypotheses only focus on “healthy” and “unhealthy”.

To provide readers with a clearer understanding of how sentences were categorized as lifestyle-related or disease-related, we present three illustrative examples for each category below. These examples are drawn from the corpus but are slightly modified here to preclude the possibility of identifying individuals.

### Lifestyle-Related Uses of “Healthy” and “Unhealthy”

- “Background on myself: 38 years old, white female, generally healthy although could lose a couple of lbs.”
- “I’m a 21 year old healthy man, I spend a lot of time doing sports.”
- “I live a fairly healthy lifestyle, I smoke occasionally, and all I take in terms of medicine is Zyrtec for allergies.”

### Disease-Related Uses of “Healthy” and “Unhealthy”

- “No baseline medical information provided; the following assumes you are in good health.”
- “I gave the full history because I’ve been very ill like this for 8 years now after being healthy.”

<sup>6</sup>Sentences coded as “unclear” by the majority of coders were excluded from the training sample.

<sup>7</sup>Given that these findings are based on the manual annotation of the training set with an arbitrary number of sentences, we refrain from statistical testing to support these conclusions.

- “I thought of myself as completely healthy until jan 2021 when I went to the emergency room for heart palpitations.”

### Disease-Related Uses of “Ill” and “Sick”

- “I tend to eat food as quickly as possible, and often feel quite sick afterward—no matter what I’ve eaten.”
- “My dada was really sick and had diarrhea for like 10 days.”
- “I experienced brain fog, which often happens when I’m sick.”

For the annotation of the disease- vs lifestyle-related contexts, we trained a few-shot classifier using SetFit (Tunstall et al., 2022), a lightweight few-shot learning framework that fine-tunes a sentence transformer model using contrastive learning on a small labeled dataset.<sup>8</sup> After filtering out unclear cases, the sample was randomly split into training and validation subsamples (80–20%). Validation accuracy is high at 0.9412 (with identical Recall, Precision, and F1).<sup>9</sup> The classifier is then used to annotate the remaining sentences containing ‘healthy’ or ‘unhealthy’ ( $n = 2,115$ ).

$H_3$  is tested with a simple  $\chi^2$ -test. For  $H_4$ , we use a mixed-effects logistic regression model with the predicted context label (disease vs lifestyle) as the dependent variable and population (laypeople vs medical professionals) as the independent variable. We additionally include posting time and sentence word counts as controls, and random intercepts for the target term (“unhealthy” vs “healthy”) and the syntactic dependency tag of the target term in the sentence.

### Results

A  $\chi^2$ -test examined whether disease- and lifestyle-related contexts were distributed differently in the annotated dataset. The results show no significant difference,  $\chi^2(1) = 0.9991$ ,  $p = 0.3175$ , indicating that our target terms occurred with similar frequencies in both contexts. Hence, we cannot reject  $H_3$ .

The mixed-effects logistic regression model results provide evidence against  $H_4$ , which posits no significant difference between laypeople and medical professionals regarding the likelihood of “healthy” and “unhealthy” being used in a disease versus lifestyle context. Specifically, the population variable (laypeople vs medical professional) shows a significant effect ( $\beta = 0.5959$ ,  $Z = 4.8511$ ,  $p < 0.0001$ ), with an odds ratio of 1.815. This indicates that, relative to laypeople, medical professionals are approximately 81.5% more likely to use “healthy” and “unhealthy” in lifestyle-related contexts. Among the control variables, the effect of time is

<sup>8</sup>For the sentence transformer, we used Microsoft’s paraphrase-mpnet-base-v2, a widely used model pre-trained on sentence similarity tasks. The model is available on Huggingface at <https://huggingface.co/sentence-transformers/paraphrase-mpnet-base-v2>.

<sup>9</sup>The classifier is openly available on Huggingface: [https://huggingface.co/lucienbaumgartner/lifestyle\\_disease\\_classifier](https://huggingface.co/lucienbaumgartner/lifestyle_disease_classifier)

non-significant ( $\beta = -1.0179$ ,  $Z = -0.3894$ ,  $p = 0.6970$ ); neither is the one for word count ( $\beta = 0.0919$ ,  $Z = 1.9166$ ,  $p = 0.0553$ ), on 0.05-alpha level.<sup>10</sup> Overall, these findings suggest that medical professionals are significantly more likely than laypeople to use “healthy” and “unhealthy” in lifestyle-related contexts.

### Discussion

The results from Experiment 2 provide important insights into the use of the terms “healthy” and “unhealthy” in disease- and lifestyle-related contexts. The non-significant result of the  $\chi^2$ -test for  $H_3$  suggests that the annotated corpus does not exhibit a notable imbalance in the distribution of disease-versus lifestyle-related contexts, when we pool both populations together. However, the mixed-effects logistic regression model offers compelling evidence against  $H_4$ , highlighting a significant difference between the two populations under study. Specifically, medical professionals appear substantially more likely to frame discussions of “healthy” and “unhealthy” in lifestyle contexts compared to laypeople. Taken together, these findings underline the nuanced ways in which health-related terms are contextualized differently by distinct populations.

### General Discussion

Our studies accomplish two main objectives: they highlight general trends regarding the use of target terms (including, among others, “healthy”, “ill”, “sick”, “disease”, “disorder”, “dysfunction”) and they provide more detail about the multi-dimensional nature of the concept of health. While our results reinforce some key issues highlighted in previous research, there are at least two findings that diverge from prevalent views in current research, which could offer new insights.

First, the finding that laypeople use *certain* target terms (even if it is not the majority of terms) more consistently than medical professionals is perhaps initially surprising, as medical professionals, given their training, might be expected to use a more uniform terminology across all target terms. However, while laypeople typically rely on less nuanced understandings that may allow for greater consistency in their use, medical professionals straddle several environments where the target terms can take on different meanings depending on the patient, condition, or perhaps even the medical field. For instance, “unhealthy” might refer to abnormal findings in a clinical test, or an imbalanced diet in another, suggesting variability in term usage. Thus, upon closer consideration, the relative inconsistency is not particularly surprising and reflects a flexibility in usage.

Second, while our findings affirm earlier research that identified *absence of disease* and *lifestyle* as key dimensions in the

<sup>10</sup>The random effect variance for syntactic dependency tags is 0.5115,  $SD = 0.7152$ , indicating moderate variability in the classification outcome based on the syntactic roles of the terms “healthy” and “unhealthy”. By contrast, the random effect variance for the target terms (“healthy” vs “unhealthy”) is larger at 1.8061,  $SD = 1.3439$ .

concept of health, the finding that medical professionals are more likely than laypeople to use “healthy” and “unhealthy” in lifestyle terms is unexpected. Indeed, given their training within the biomedical paradigm that centers on diagnosis and treatment and less on prevention via lifestyle, health professionals might be expected to use “healthy” primarily in the sense of “the absence of disease”. In fact, much criticism of contemporary medicine holds that the biomedical paradigm is an obstacle to a more holistic, patient-centered care that operates with a more inclusive understanding of health and acknowledges a fuller range of needs. While such an interpretation is admittedly speculative, our findings might be taken to suggest that a change towards a more inclusive understanding of health is already occurring in health professionals. Perhaps this is in part because the benefits of lifestyle modifications have become more evident and public health measures have become more prevalent.

However, a more mundane explanation may lie in the communicative context and pragmatic dynamics of the subreddit we investigated. Laypeople seeking advice typically provide a detailed account of their medical situation, while medical professionals tend to offer a concise assessment based on the information shared. If much of this information points to poor lifestyle choices, it is unsurprising that medical professionals frequently emphasize lifestyle-related advice. Alternatively, such advice may be perceived as a more medically prudent and legally cautious response—minimizing professional liability and potential harm if the layperson’s self-reports are inaccurate. Since our methods do not explicitly account for these communicative dynamics, we cannot rule out their influence on our findings.

Overall, then, when subjecting the two surprising findings to closer scrutiny, one of them turns out not to be overly unexpected, and the other, while perhaps still surprising, can be seen as a sign of positive development, more mundane explanations notwithstanding.

### Limitations

While our study offers valuable insights into how health-related terms are used, it is subject to several limitations. First, the interaction format within our dataset is rigid: laypeople pose questions, and medical professionals respond with relatively short comments. This structure provides little insight into how these groups use terms such as “healthy” in conversations among themselves or in clinical settings. Future research could examine complementary dialogues, such as peer-to-peer discussions in online health forums or doctor-patient exchanges in clinical settings.

Second, although our findings illuminate broad contextual patterns in the use of “healthy” and “unhealthy”, they do not allow us to draw firm conclusions about their meaning. Experiment 2 demonstrated that both disease- and lifestyle-related aspects are relevant to how these terms are used, but their semantic scope likely extends further. For example, “unhealthy” might also be used to describe relationships, mindsets, or even aesthetic features (e.g., “sickly-looking

skin”). A systematic exploration of these additional dimensions would help us to develop a more comprehensive semantic map of health-related vocabulary.

Third, our study focused on a narrow set of target terms, which, as our data suggest, are not particularly frequent in the r/AskDocs subreddit. A broader analysis incorporating a wider range of health-related expressions across different online communities could provide a richer and more representative picture. Expanding the vocabulary through bottom-up data-driven methods, such as nearest neighbor search in word embeddings, could help identify related terms that naturally cluster within same semantic space.

Taken together, these limitations point to several promising directions for future research. A more fine-grained analysis of communicative contexts, a broader conceptual exploration of health-related language, and a data-driven expansion of vocabulary could all contribute to a more nuanced understanding of how people conceptualize and communicate about health.

### Conclusion

This study identifies general trends in the use of health-related terms and examines differences in how laypeople and health-care professionals conceptualize health. Two key findings reveal differences between laypeople and medical professionals in both the consistency of their use of health-related terms and how they conceptualize health—either as a matter of lifestyle or the absence of disease. Further research is needed to investigate these differences in broader contexts beyond the specific communicative setting of the subreddit.

### Acknowledgments

The research of Lucien Baumgartner was funded by the Swiss National Science Foundation (SNSF), grant number PZ00P1\_201737.

### References

- Baumgartner, L. (2022). Why Are Reckless Socks Not (More Of) A Thing? Towards An Empirical Classification Of Evaluative Concepts. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44, pp. 2869–2875).
- Baumgartner, L., & Kneer, M. (2025). The meaning of ‘reasonable’: evidence from a corpus-linguistic study. In K. P. Tobia (Ed.), *The Cambridge Handbook of Experimental Jurisprudence* (pp. 440–463). Cambridge University Press.
- Bishop, F., & Yardley, L. (2010). The development and initial validation of a new measure of lay definitions of health: The wellness beliefs scale. *Psychology and Health, 25*(3), 271–287.
- Blaxter, M. (2010). *Health*. Polity.
- Clark, K., Luong, M.-T., Le, Q. V., & Manning, C. D. (2020). *ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators*. doi: 10.48550/arXiv.2003.10555

- Downey, C. A., & Chang, E. C. (2013). Assessment of everyday beliefs about health: the lay concepts of health inventory, college student version. *Psychology & Health*, 28(7), 818–832.
- Fugelli, P., & Ingstad, B. (2001). Health–people’s perspective. *Tidsskrift for den Norske laegeforening: tidsskrift for praktisk medicin, ny raekke*, 121(30), 3600–3604.
- Herzlich, C. (1973). *Health and illness*. London: Academic Press.
- Huber, M., van Vliet, M., Giezenberg, M., Winkens, B., Heerkens, Y., Dagnelie, P. C., & Knottnerus, J. A. (2016). Towards a ‘patient-centred’ operationalisation of the new dynamic concept of health: a mixed methods study. *BMJ Open*, 6(1), e010091.
- Hughner, R. S., & Kleine, S. S. (2004). Views of health in the lay sector: A compilation and review of how individuals think about health. *Health*, 8(4), 395–422.
- Millstein, S. G., & Irwin, C. E. (1987). Concepts of health and illness: different constructs or variations on a theme? *Health Psychology*, 6(6), 515.
- Reuter, K., & Baumgartner, L. (2025). Conspiracy Theories Are Not Theories: Time to Rename Conspiracy Theories. In M. Isaac, S. Koch, & K. Scharp (Eds.), *New Perspectives on Conceptual Engineering*. Springer.
- Reuter, K., Latham, A., & Varga, S. (2025). Concept(s) of Health: Lifestyle at the Heart of Modern Health. *Erkenntnis*.
- St Claire, L., Watkins, C. J., & Billinghamurst, B. (1996). Differences in meanings of health: An exploratory study of general practitioners and their patients. *Family Practice*, 13(6), 511–516.
- Tunstall, L., Reimers, N., Jo, U. E. S., Bates, L., Korat, D., Wasserblat, M., & Pereg, O. (2022). Efficient few-shot learning without prompts. *arXiv*.
- Varga, S., & Latham, A. (2024). Is health the absence of disease? *Inquiry*, 1–18.
- Williams, R. (1983). Concepts of health: an analysis of lay logic. *Sociology*, 17(2), 185–205.

## Appendix

### Appendix A

Table 3: KS test statistics for the comparison of intra-group semantic similarity between laypeople and medical professionals.

Term	<i>D</i>	<i>p</i>
disease	0.1690	<0.0001
disorder	0.1095	<0.0001
dysfunction	0.3036	<0.0001
health	0.1715	<0.0001
healthy	0.0247	<0.0001
ill	0.2085	<0.0001
illness	0.1289	<0.0001
sick	0.0828	<0.0001
unhealthy	0.1416	0.0001
unwell	0.1065	0.0002
well	0.1834	0.0045