

Visibility by design: Reliability of top domains in search engines for health-related queries

Natalie Tutzer^{a,b}

^a Pädagogische Hochschule Freiburg, Kunzenweg 21, Freiburg im Breisgau, 79117, Germany

^b TAKEPART Media + Science GmbH, Maria-Hilf-Str. 15, Cologne, 50677, Germany

1. Objective

This presentation introduces a study examining the search engine visibility of reliable health information websites in Germany across the most prevalent diagnoses encountered in general practice. Grounded in the concept of organisational health literacy, providers of high-quality, evidence-based health information should make accessibility of their contents through commercial search engines as easy as possible (M-POHL, n.d.). By systematically mapping which domains appear in search engine results pages (SERPs) for common health-related queries, this research aims to investigate the extent to which reliable sources compete effectively in the digital information environment.

2. Background

Search engines remain the dominant entry point through which people seek health information online. Finding and identifying reliable health information continues to present significant challenges for lay users. Research consistently shows that people struggle to evaluate the credibility of online health content and often cannot distinguish between high-quality, evidence-based resources and unreliable or primarily commercially motivated sources (Schaeffer et al., 2026).

At the same time, evidence-based health information providers frequently underperform in search engine results, also when their content meets recognised quality criteria. Research shows that websites with lower SEO scores tend to be rated as higher quality by independent evaluators, suggesting a structural disconnect between content quality and search engine discoverability (Schultheiß et al., 2022). This points to a systemic problem: the organisations best positioned to provide trustworthy health information are often not well equipped to ensure that their content reaches users through organic search. This may be due to a lack of technical infrastructure, resources, and expertise (Arnold et al., 2019).

Accessibility is an established criterion within quality frameworks for health information, yet in practice it is rarely operationalised to include search engine visibility. Developers and publishers of patient-oriented health information frequently lack both the institutional knowledge of how search engines function and the practical capacity to implement search engine optimisation (SEO) measures. As a result, high-quality content may be hard to find: medical patient guidelines are published in formats unsuited for indexing (such as PDF rather than HTML), lack optimised metadata, and are misaligned with the language and query patterns that users employ when searching.

This study frames this challenge through the lens of organisational health literacy. While health literacy research has traditionally focused on the capacities of individuals to access, understand, and use health information, the concept of organisational health literacy shifts the focus toward the responsibilities of institutions and information providers. They need to cater relevant information to its



target audiences, remove barriers, and design appropriate and modern communication tools. Applied to the digital information environment, this means that organisations providing health information need to understand how search engines work, why and how people use them, and how to structure and publish content in ways that improve its chances of being found by those who are to benefit from it. This study aims to show which types of providers of health information are visible for frequently searched health-related queries.

3. Methods

The study employs a software-assisted, large-scale approach to assess which domains appear in the top ten of search engine result pages – and as sources in AI overviews – for various keyword sets related to the most common diagnoses in German general practice.

Initially, keyword sets are compiled for each included diagnosis, such as sampled queries e.g. on *adipositas*, *diabetes*, *back pain*, *hypertension* and *depression*. The keyword sets are identified using software as the Query Sampler tool (Schultheiß et al., 2023) and SISTRIX (2025). Through this keyword research, dozens to hundreds of related keywords are sampled within one set per diagnosis.

In a second step, the top ten organic search results are retrieved for each individual query within these keyword sets. The software used for this is the result assessment tool developed by the HAW Hamburg (Sünkler et al., 2026). This produces a large dataset of domains appearing in SERPs across conditions and query types, enabling an analysis of the health information landscape as users may encounter it.

All domains retrieved from the SERPs are coded according to a classification scheme of publisher types. This classification allows for a structured analysis of which domains and types of publishers achieve visibility across the keyword sets. Main domain categories of health information providers ranking in top ten and being referenced in search engine AI overviews are:

1. scientific: universities / academia
2. journalistic: media outlets
3. commercial: advertisement-financed platforms / industry websites
4. healthcare providers: clinics and practices
5. health insurance providers: public and private insurance websites
6. non-governmental organisations such as foundations, medical associations and patient advocacy groups

Across these categories lies a cluster of pre-defined websites that have been labelled as reliable health information providers based on standardised criteria by non-commercial organisations such as the Stiftung Gesundheitswissen (Engler et al., 2025) and the German Network for Health Literacy (DNGK, 2026).

The combination of query sampling, SERP retrieval, and domain coding enables the study to address two related questions: first, how visible are reliable health information providers for commonly searched health queries? And second, does the distribution of publisher types in the SERPs shift between variables such as topic, search volume and query length?

4. Discussion and Outlook

Findings will identify which domains achieve visibility for the most common health conditions and assess to what degree the predefined reliable sources are represented in SERPs relative to other providers. The results are expected to reveal variation across conditions and domains. Since reliable sources may strategically choose not to create redundant content, the percentage of reliable domains in the overall findings may be low.

These findings will be discussed in the context of organisational health literacy. If high-quality health information providers are structurally underrepresented in search results, this raises the question of how laypeople can be supported in identifying reliable health information if high-quality websites



are not transparent or visible in search engine results – especially in light of health information provided by Large Language Models that lack many conventional markers of reliable information such as date of information, authorship and sources (DNGK, 2026).

5. References

Arnold, K., Breuing, J., Becker, M., Nothacker, M., Neugebauer, E., Schmitt, J., & Deckert, S. (2019). Entwicklung leitlinienbasierter Qualitätsindikatoren: Eine qualitative Studie zu Barrieren und förderlichen Faktoren aus der Sicht von S3-LeitlinienautorInnen. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 147–148, 34–44. <https://doi.org/10.1016/j.zefq.2019.09.001>

Bertelsmann Stiftung. (2023). *InfoCure: Qualität sichtbar machen*. Trusted Health Ecosystems. <https://trusted-health-ecosystems.org/infoq-qualitaet-sichtbar-machen/>

Deutsches Netzwerk Gesundheitskompetenz (DNGK). (2026). *Verlässliches Gesundheitswissen*. <https://dngk.de/verlaessliches-gesundheitswissen/>

Engler, A., Randig, J., Lindner, J., & Ströhlein, L.-M. (2025). *Wie erkenne ich gute Gesundheitsinformationen?* Stiftung Gesundheitswissen. <https://www.stiftung-gesundheitswissen.de/gesundheitsinformationen-verstehen/gesundheitsinformationen>

Schaeffer, D., Griesse, L., Singh, H., Ewers, M., & Hurrelmann, K. (2026). *Zusammenfassung Gesundheitskompetenz in Zeiten gesellschaftlicher Unsicherheiten: Ergebnisse des HLS-GER3*. Interdisziplinäres Zentrum für Gesundheitskompetenzforschung (IZGK), Universität Bielefeld. <https://doi.org/10.4119/unibi/3017049>

Schultheiß, S., Häußler, H., & Lewandowski, D. (2022). *Does search engine optimization come along with high-quality content? A comparison between optimized and non-optimized health-related web pages*. arXiv. <https://doi.org/10.48550/arXiv.2301.10105>

Schultheiß, S., Lewandowski, D., von Mach, S., & Yagci, N. (2023). Query sampler: Generating query sets for analyzing search engines using keyword research tools. *PeerJ Computer Science*, 9, e1421. <https://doi.org/10.7717/peerj-cs.1421>

SISTRIX. (2025). *SISTRIX Toolbox* [Software]. Retrieved April 30, 2026, from <https://www.sistrix.de>

Sünkler, S., Bilir, K., Kumar, T., Koop, O., Schultheiß, S., & Lewandowski, D. (2026). Result Assessment Tool (RAT): An open-source toolkit for conducting studies based on search results. In *Proceedings of the 2026 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '26)* (4 pages). ACM. <https://doi.org/10.1145/3786304.3787925>

WHO Action Network on Measuring Population and Organizational Health Literacy (M-POHL). (n.d.). *Organizational Health Literacy (OHL)*. Retrieved July 12, 2026, from <https://m-pohl.net/OHL>