Dividing the Number of Papers on Each Cancer Site (Published 2019–2023) by Japan's 2023 Deaths Gives Values 44.3 Times Apart Across Sites (Gallbladder/Bile Ducts 0.412 to Skin 18.269), and Dividing by Incidence Gives 30.7 Times (Stomach 0.195 to Brain/CNS 6.008) — the Gap Between the Two Divisions Is the Site's Deaths ÷ Incidence Itself; Prostate Is 8th of 18 Sites per Death and 17th per Incident Case [M113]

Papers per death = papers per incident case ÷ (deaths ÷ incidence) ── What separates them is whether you divide by deaths or by incidence

Yuuki YAMAGISHI (山岸 勇輝) ／ Independent Researcher
ORCID 0009-0006-8587-3450
September 2026 ／ CC BY 4.0

Abstract

For 18 cancer sites, this paper retrieved through E-utilities (27 September 2026) the number of PubMed (U.S. National Library of Medicine) records with the site's MeSH term published in 2019–2023, and divided it by Japan's 2023 deaths (Vital Statistics) and incidence (National Cancer Registry). Whether more or fewer papers is good or bad is not discussed. No new mathematical theorem and no new law are claimed.

First, **papers per death ranged from 0.412 for gallbladder/bile ducts to 18.269 for skin, 44.3 times apart across sites (standard deviation of the natural logarithm 0.981). Papers per incident case ranged from 0.195 for stomach to 6.008 for brain/central nervous system, 30.7 times apart (0.780)** (Section 4).

Second, and this is the core of this paper, **papers per death = papers per incident case ÷ (deaths ÷ incidence) holds exactly. The gap between the two divisions is the site's deaths ÷ incidence itself.**

Prostate (deaths ÷ incidence 0.132) is 8th of 18 sites per death and 17th per incident case; breast (0.152) moves from 3rd to 8th, and leukemia (0.651) from 6th to 2nd (Section 5).

Third, **the variance of the logarithm across the 18 sites split as: ratio per death 0.963 = ratio per incident case 0.608 + deaths ÷ incidence 0.444 − 2 × covariance 0.044. The rank correlation between the two orderings was 0.777** (Section 5).

Fourth, **the correlation with the number of papers (logarithm) was 0.523 with deaths and 0.626 with incidence for all papers, and 0.757 and 0.789 for papers with the MeSH term "Japan"; in both, incidence was higher** (Section 6).

Fifth, **pancreas has 21,472 papers, and because its deaths ÷ incidence is 0.845, close to 1, the two divisions give close values (0.534 per death and 0.452 per incident case)** (Section 5).

Sixth, what separates them is **whether you divide by deaths or by incidence**. When you hear that "this disease has few (or many) papers for the number of people who die of it," also check what happens when you divide by incidence.

Scope note: **This paper claims no new mathematical theorem and no new law** — the paper counts and the numbers of deaths and new cases are public, and this paper only divided them and lined them up. **This paper is a structural interpretation of the shape of numbers in public aggregate data; it is not medical advice, not a diagnosis, and not a treatment recommendation. The author is not a physician. Decisions about individual health should be discussed with a physician or a medical institution. This is a preprint that has not been peer reviewed. Whether more or fewer papers is good or bad is not discussed. No site is said to lack research. The quality or value of research is not discussed. No specific journal, publisher, author, or research institution is named.** The paper counts are worldwide, while deaths and incidence are Japan's for 2023. Because how common each cancer site is differs by country, these ratios are values of the combination "worldwide papers ÷ Japanese counts." The MeSH "… neoplasms" terms include benign tumors, so their scope does not match the statistics on malignant neoplasms. The paper counts are as of retrieval and change daily. **About the data**: Courtesy of the U.S. National Library of Medicine. The counts are as of 27 September 2026 and do not reflect the most current or accurate data from NLM. NLM does not endorse this paper. The National Cancer Registry and Vital Statistics were obtained from e-Stat and used under the Public Data License (Version 2.0). This paper's ratios, correlations, and variance decomposition are processed figures made from them. What this paper adds: **dividing the number of papers on each site by both deaths and incidence, and showing that the gap between the two divisions is the site's deaths ÷ incidence itself**.

1. Notation and question

- **Number of papers**: the number of PubMed records published in 2019–2023 ([dp]) with the site's MeSH term (for example, pancreas = pancreatic neoplasms[mh]).
- **Per death**: number of papers ÷ Japan's 2023 deaths (Vital Statistics, Table 5.24). **Per incident case**: number of papers ÷ 2023 incidence (National Cancer Registry, Table 1-A). The unit is papers.
- **Deaths ÷ incidence**: 2023 deaths ÷ new cases (made the same way as the MI ratio in M025).

- **Identity**: per death = per incident case ÷ (deaths ÷ incidence).

Question: Is the number of papers by site shaped more like the number of deaths or the number of new cases? When the division changes, which sites move in rank, and by how much?

---

2. Preview and design

Written honestly: the design was decided after lining up the numbers.

	Content
Data	PubMed E-utilities counts (18 sites × worldwide and with MeSH "Japan"), National Cancer Registry 2023 Table 1-A, Vital Statistics 2024 Table 5.24 (the same files as M025)
Noticed in preview	Divided by deaths, skin and breast rank high; divided by incidence, brain/CNS and leukemia rank high
Noticed in preview (2)	The ratio of the two divisions equals the site's deaths ÷ incidence (the MI ratio of M025) exactly
Noticed in preview (3)	Papers with MeSH "Japan" are few, only 7 for larynx
Compared	Max ÷ min and log standard deviation of the two divisions, ranks, decomposition of log variance, correlation with the number of papers

---

3. Data and checks

Calculation 3.1 The four files (the PubMed count JSON, the retrieval script, the cancer registry table, and the vital statistics table) were fixed by SHA-256, and the following were checked. The checker computes only from the saved JSON and tables and does not query PubMed.

- The file has 18 sites × 2 kinds (worldwide, with MeSH "Japan"). Retrieved on 27 September 2026, 01:25–01:26.
- PubMed's query translation contains the MeSH term, publication date 2019/01/01:2023/12/31, and (for Japan) "japan"[MeSH Terms] (all 36 queries), with 0 warnings.
- The 2023 totals for the 18 sites can be read from Table 1-A and Table 5.24 (the site mapping is the same as in M025).
- The identity (per death = per incident case ÷ deaths ÷ incidence) holds exactly for all 18 sites.

---

4. Values by site

Calculation 4.1 Number of papers by site and the two divisions (in order of papers per death)

Site	Papers	Deaths ÷ incidence	Per death (rank)	Per incident case (rank)
Skin	33,998	0.068	18.269 (1)	1.243 (4)
Brain/CNS	34,711	0.550	10.922 (2)	6.008 (1)
Breast	70,638	0.152	4.481 (3)	0.683 (8)
Uterus	23,722	0.232	3.324 (4)	0.772 (6)
Leukemia	30,792	0.651	3.120 (6)	2.030 (2)
Prostate	30,261	0.132	2.253 (8)	0.296 (17)
Liver	40,052	0.701	1.748 (10)	1.226 (5)
Colorectum	52,200	0.345	0.982 (14)	0.339 (16)
Lung	59,683	0.611	0.788 (15)	0.481 (11)
Pancreas	21,472	0.845	0.534 (16)	0.452 (14)
Stomach	20,495	0.370	0.529 (17)	0.195 (18)
Gallbladder/bile ducts	7,104	0.824	0.412 (18)	0.339 (15)

Observation 4.2: **Papers per death ranged from 0.412 for gallbladder/bile ducts to 18.269 for skin, 44.3 times apart; papers per incident case ranged from 0.195 for stomach to 6.008 for brain/CNS, 30.7 times apart.** The standard deviations of the natural logarithm were 0.981 and 0.780; dividing by incidence makes the sites more alike. The table shows 12 of the 18 sites; the other 6 (oral cavity/pharynx, esophagus, larynx, ovary, bladder, and lymphoma) are in the checker's output.

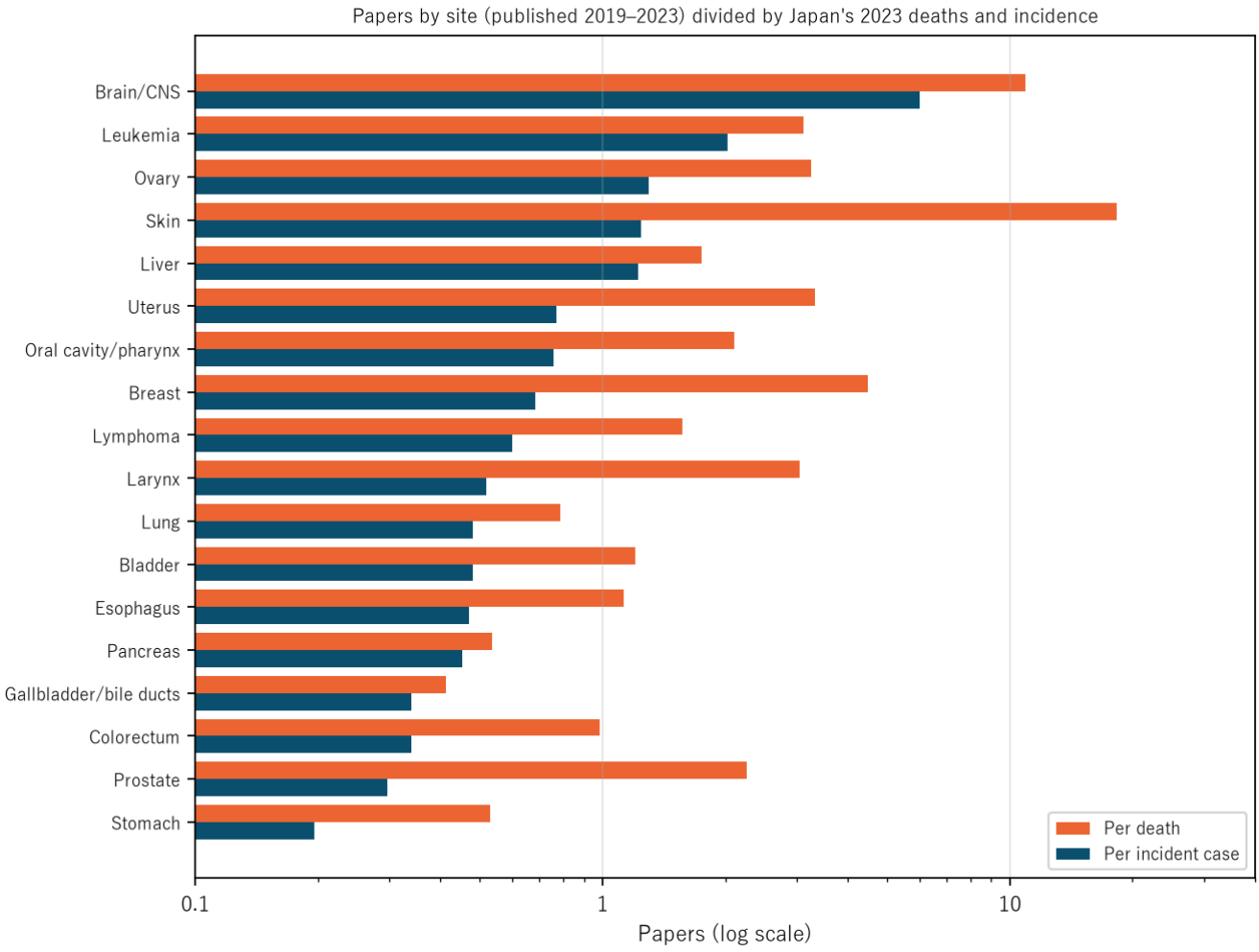


Figure 1: Number of papers by site (published 2019–2023) divided by Japan's 2023 deaths (orange) and incidence (navy). The horizontal axis is on a log scale. Sites are ordered by papers per incident case.

---

5. The core — the gap between the two divisions

Observation 5.1: **Because papers per death = papers per incident case ÷ (deaths ÷ incidence), the gap between the two divisions is the site's deaths ÷ incidence itself.** The smaller a site's deaths ÷

incidence (the fewer people who die of it relative to new cases), the larger its value per death.

Observation 5.2: **Prostate (deaths ÷ incidence 0.132) is 8th per death and 17th per incident case. Breast (0.152) moves from 3rd to 8th, and skin (0.068) from 1st to 4th. Conversely, leukemia (0.651) rises from 6th to 2nd, and liver (0.701) from 10th to 5th.** The rank correlation between the two orderings was 0.777.

Observation 5.3: **The variance of the logarithm across the 18 sites split exactly as: ratio per death 0.963 = ratio per incident case 0.608 + deaths ÷ incidence 0.444 − 2 × covariance 0.044.** The ratio per death varies more because deaths ÷ incidence differs greatly by site.

Observation 5.4: **Pancreas has deaths ÷ incidence of 0.845, close to 1, so the two divisions give close values: 0.534 papers per death and 0.452 per incident case.** Gallbladder/bile ducts (0.824) have the same shape.

---

6. Is the number of papers closer to deaths or to incidence?

Calculation 6.1 Correlation with the number of papers (natural logarithm, 18 sites)

How papers are counted	Correlation with deaths	Correlation with incidence
All papers worldwide	0.523	0.626
Papers with MeSH "Japan"	0.757	0.789

Observation 6.2: **For papers worldwide and for papers with MeSH "Japan," the correlation with the number of papers was higher with incidence than with deaths.** Papers with "Japan" correlate more strongly with Japan's deaths and incidence. However, papers with "Japan" are few (7 for larynx), and not every study conducted in Japan carries the MeSH term "Japan."

---

7. Relation to earlier papers

M025 showed that the year's deaths ÷ new cases (the MI ratio) is not a share of the same people. This paper showed that the same MI ratio is exactly the gap between dividing the number of papers by deaths and by incidence. M004 showed that the number of years in which papers double changes depending on what is counted in PubMed. This paper likewise counted how the ranking of "how many papers" changes depending on what you divide by.

---

8. Limitations

1. **Whether more or fewer papers is good or bad is not discussed. No site is said to lack research. The quality or value of research is not discussed.**
2. The paper counts are worldwide, while deaths and incidence are Japan's. Because how common each site is differs by country, the ratios are values of the combination "worldwide papers ÷ Japanese counts."
3. The MeSH "… neoplasms" terms include benign tumors, so their scope does not match the statistics on malignant neoplasms. The difference may be especially large for brain/CNS and skin.
4. Papers cover 5 years (2019–2023), while deaths and incidence cover one year (2023); the years do not match.
5. The counts are as of 27 September 2026 and change daily.

Conclusion

The number of papers on each cancer site differed across sites by 44.3 times when divided by deaths and by 30.7 times when divided by incidence. The gap between the two divisions is the site's deaths ÷ incidence itself: prostate and breast, where few people die relative to new cases, rank high when divided by deaths and low when divided by incidence. The number of papers was shaped more like incidence than like deaths.

What separates them is one thing — whether you divide by deaths or by incidence.

When you hear that "this disease has few (or many) papers for the number of people who die of it," **also check what happens when you divide by incidence, and that disease's deaths ÷ incidence.**

Data and reproduction

Data	Counts and query translations retrieved from PubMed (U.S. National Library of Medicine) E-utilities (esearch), 27 September 2026, 01:25–01:26. Courtesy of the U.S. National Library of Medicine. National Cancer Registry 2023, Table 1-A; Vital Statistics 2024 (final data), Volume 1, Table 5.24 (e-Stat)
Terms of use	NLM Terms and Conditions (free of charge; credit the source; do not imply NLM endorsement; state that the data are not the most current). e-Stat follows the Public Data License (Version 2.0). This paper's ratios, correlations, and variance decomposition are processed figures, not figures produced by the government
Coverage	18 cancer sites; papers published 2019–2023; deaths and incidence in Japan in 2023; both sexes combined
Verification	`検算/M113.py` (`--zu` also draws the figure). Every number in the text is taken from `検算/M113_output.txt`
Raw data	`data/pubmed_m113_20260927.json`, the retrieval script `data/fetch_pubmed.py`, `data/2023_1-a.xlsx`, `data/estat_000040316517.csv`, and `data/SOURCES.md` (terms of use, SHA-256, the MeSH query for each site, the site mapping)
Figure	`図/M113_zu1` (png, svg, English version)

References

- U.S. National Library of Medicine. PubMed (E-utilities, esearch). Courtesy of the U.S. National Library of Medicine (retrieved 27 September 2026). https://eutils.ncbi.nlm.nih.gov/entrez/eutils/

- U.S. National Library of Medicine. PubMed User Guide (accessed 27 September 2026). https://pubmed.ncbi.nlm.nih.gov/help/
- Ministry of Health, Labour and Welfare. National Cancer Registry: Cancer Incidence 2023, Table 1-A (in Japanese). e-Stat (accessed 26 September 2026). https://www.e-stat.go.jp/
- Ministry of Health, Labour and Welfare. Vital Statistics 2024 (final data), Volume 1, Table 5.24 (in Japanese). e-Stat (accessed 26 September 2026). https://www.e-stat.go.jp/
- Yamagishi, Y. (2026). The Year's Deaths Divided by Its New Cases (the MI Ratio) Is Not a Share of the Same People — Against 1 − 5-Year Survival It Is 0.97 Times for Pancreas and 3.07 Times for Prostate [M025]. Zenodo. https://doi.org/10.5281/zenodo.22968532
- Yamagishi, Y. (2026). "Medical Papers Double About Every 15 Years" Has a Different Answer Depending on What Is Counted [M004]. Zenodo. https://doi.org/10.5281/zenodo.22963542