Counting Clinical Trials by Phase and Adding Them Up Gives More Than the Number of Trials — ClinicalTrials.gov's 461,397 Interventional Studies Carry 485,973 Phase Labels, and 0.260 of Phase 1 Labels Also Carry Phase 2 (a Registrant Selects Only One Phase, but Phase 1/Phase 2 Becomes Two Labels). For 18 Cancer Sites the Sum of Labels Is 1.0778–1.1585 Times the Interventional Studies, and 11 Sites Have More Studies with Unknown Status Than Studies Recruiting Now [M114]

Interventional studies = sum of phase labels − studies with two labels + studies with no phase ── What separates them is whether you count studies or phase labels

Yuuki YAMAGISHI (山岸 勇輝) ／ Independent Researcher

ORCID 0009-0006-8587-3450

September 2026 ／ CC BY 4.0

Abstract

Using the API of the clinical trial registry ClinicalTrials.gov (U.S. National Library of Medicine), this paper counted studies by type, phase, and status for all registered studies and for 18 cancer sites (the same list as M113) (retrieved 27 September 2026; the data were processed by ClinicalTrials.gov on 2026-09-25T09:00:04). Whether trials are good or bad, or whether a disease has more or fewer trials, is not discussed. No new mathematical theorem and no new law are claimed.

First, and this is the core of this paper, **there were 461,397 interventional studies in all, and counting their phase labels (Early Phase 1, Phase 1, Phase 2, Phase 3, Phase 4, and not applicable) and adding them up gave 485,973 labels, 1.0533 times the interventional studies. Of the 65,822 Phase 1 labels, 0.260 belonged to studies that also carry a Phase 2 label** (Section 5).

Second, **a registrant selects only one phase ("Select only one."), but choosing "Phase 1/Phase 2" or "Phase 2/Phase 3" gives two labels in the API. Interventional studies = sum of labels − Phase 1 and Phase 2 (17,095) − Phase 2 and Phase 3 (7,606) + no phase (125) held exactly in all 19 cases (all registered studies and 18 sites)** (Sections 3 and 5).

Third, **for the 18 cancer sites, the sum of labels was 1.0778 (breast) to 1.1585 (leukemia) times the interventional studies, larger than for all registered studies at every site. The share without phases (such as device or behavioral studies) was 0.113–0.376, smaller than for all studies (0.515), and even among studies with a drug phase, the share with two labels was 0.1236–0.1876, larger than for all studies (0.1104)** (Section 5).

Fourth, **the number registered and the number seeking participants now are different. Of the 5,089 registered pancreas studies, 896 (0.176) were recruiting, 679 were recruiting interventional studies, and 40 of those had a location in Japan** (Section 6).

Fifth, **studies with unknown status (studies whose status has not been verified for more than 2 years after their completion date, which ClinicalTrials.gov treats as closed) numbered 96,838 in all, 1.489 times the 65,044 recruiting studies. At 11 of the 18 cancer sites, unknown status outnumbered recruiting** (Section 6).

Sixth, what separates them is **whether you count studies or phase labels**. When you hear "there are so many Phase 1 trials," check whether Phase 1/Phase 2 trials are included. When you hear "there are so many trials for this disease," also check how many of them are recruiting now.

Scope note: **This paper claims no new mathematical theorem and no new law** — the numbers of studies are published by ClinicalTrials.gov, and this paper only counted and compared them. **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. It is not material for deciding whether to join a trial. The quality or value of trials, and whether a disease has more or fewer trials, are not discussed. No disease is said to lack trials. Studies with unknown status are not discussed as the responsibility of the people or organizations who ran them. No specific study, sponsor (company or research institution), or researcher is named.** Sites were counted with ClinicalTrials.gov searches using MeSH headings, and one study can match several sites. **About the data**: The source is ClinicalTrials.gov (U.S. National Library of Medicine). ClinicalTrials.gov processed the data on 2026-09-25T09:00:04, and this paper retrieved them on 27 September 2026. ClinicalTrials.gov is updated daily, so **check ClinicalTrials.gov for the latest numbers**. This paper does not reproduce study records; it uses only the counts returned by the API and computed their sums, differences, and ratios (these are all the modifications made). The U.S. Government and NLM make no warranties about the data. What this paper adds: **showing by an identity how counting phase labels gives more than the number of studies, and comparing cancer sites**, and **lining up the number registered with the numbers recruiting now and with unknown status**.

---

1. Notation and question

- **Interventional study**: a study whose type on ClinicalTrials.gov is INTERVENTIONAL. Other types include observational studies.
- **Phase label**: a label in the phase field of an interventional study (`phases` in the API): EARLY_PHASE1, PHASE1, PHASE2, PHASE3, PHASE4, and NA (not applicable). A registrant

selects one phase, but "Phase 1/Phase 2" becomes two labels, PHASE1 and PHASE2, and "Phase 2/Phase 3" becomes PHASE2 and PHASE3.

- **No phase**: an interventional study with no label in the phase field.
- **Identity**: interventional studies = sum of labels − Phase 1 and Phase 2 − Phase 2 and Phase 3 + no phase.
- **Recruiting**: studies with status RECRUITING. **Unknown status**: studies with status UNKNOWN (definition in Section 6). **Location in Japan**: studies with Japan among their location countries.

Question: How different is the sum of counts by phase from the number of studies? Is the difference the same across diseases? Of the registered studies, how many are seeking participants now?

---

2. Preview and design

Written honestly: the design was decided after a trial count for pancreas.

	Content
Data	ClinicalTrials.gov API v2 counts (all studies and 18 sites × 15 kinds, 285 queries; no phase, 19 queries), and copies of the original text of the terms of use and the definitions
Noticed in preview	Against 4,011 interventional pancreas studies, the counts by phase added up to 4,584
Noticed in preview (2)	On the definitions page, a registrant selects only one phase ("Select only one."), but the choices include "Phase 1/Phase 2" and "Phase 2/Phase 3"
Noticed in preview (3)	For all studies, 125 were still unaccounted for after subtracting studies with two labels. Counting interventional studies with an empty phase field gave exactly 125
Compared	Sum of labels ÷ interventional studies, share of Phase 1 labels that also carry Phase 2, share without phases, shares recruiting, unknown, and in Japan

---

3. Data and checks

Calculation 3.1 Six files — the two count JSON files, the two retrieval scripts, and the copies of the terms of use and the definitions — were fixed by SHA-256, and the following were checked. The checker computes only from the saved JSON files and does not query ClinicalTrials.gov.

- 285 main queries (27 September 2026, 01:43:23–01:52:02) and 19 additional queries. API version 2.0.5.
- The data processing date (dataTimestamp) was 2026-09-25T09:00:04 in all 4 checks before and after the main and additional retrievals, so no update came in between.
- The query parameters (site query, interventional, recruiting, Japan) are as designed.
- The identity (interventional studies = sum of labels − studies with two labels + no phase) holds with a gap of 0 in all 19 cases.
- All subset relations hold (recruiting interventional in Japan ≤ recruiting interventional ≤ recruiting ≤ all, and so on).

---

4. Phase labels

Calculation 4.1 Phase labels (interventional studies; in order of sum of labels ÷ interventional studies)

Scope	Interventional studies	Sum of labels ÷ interventional	Phase 1 labels also Phase 2	Share without phases
All registered studies	461,397	1.0533	0.260	0.515
Leukemia	8,191	1.1585	0.349	0.113
Lymphoma	9,678	1.1435	0.346	0.118
Pancreas	4,011	1.1429	0.362	0.238
Gallbladder/bile ducts	874	1.1327	0.409	0.203
Stomach	3,316	1.1218	0.380	0.234
Lung	11,374	1.1191	0.374	0.213
Oral cavity/pharynx	4,859	1.1029	0.254	0.183
Prostate	6,109	1.0969	0.361	0.328
Breast	13,144	1.0778	0.350	0.376

Observation 4.2: **The 461,397 interventional studies in all carried 485,973 phase labels (1.0533 times). Of the 65,822 Phase 1 labels, 17,095 (0.260) belonged to studies that also carry a Phase 2 label.** Of the 49,929 Phase 3 labels, 7,606 (0.152) also belonged to studies with a Phase 2 label. The table shows 9 of the 18 sites; the rest are in the checker's output.

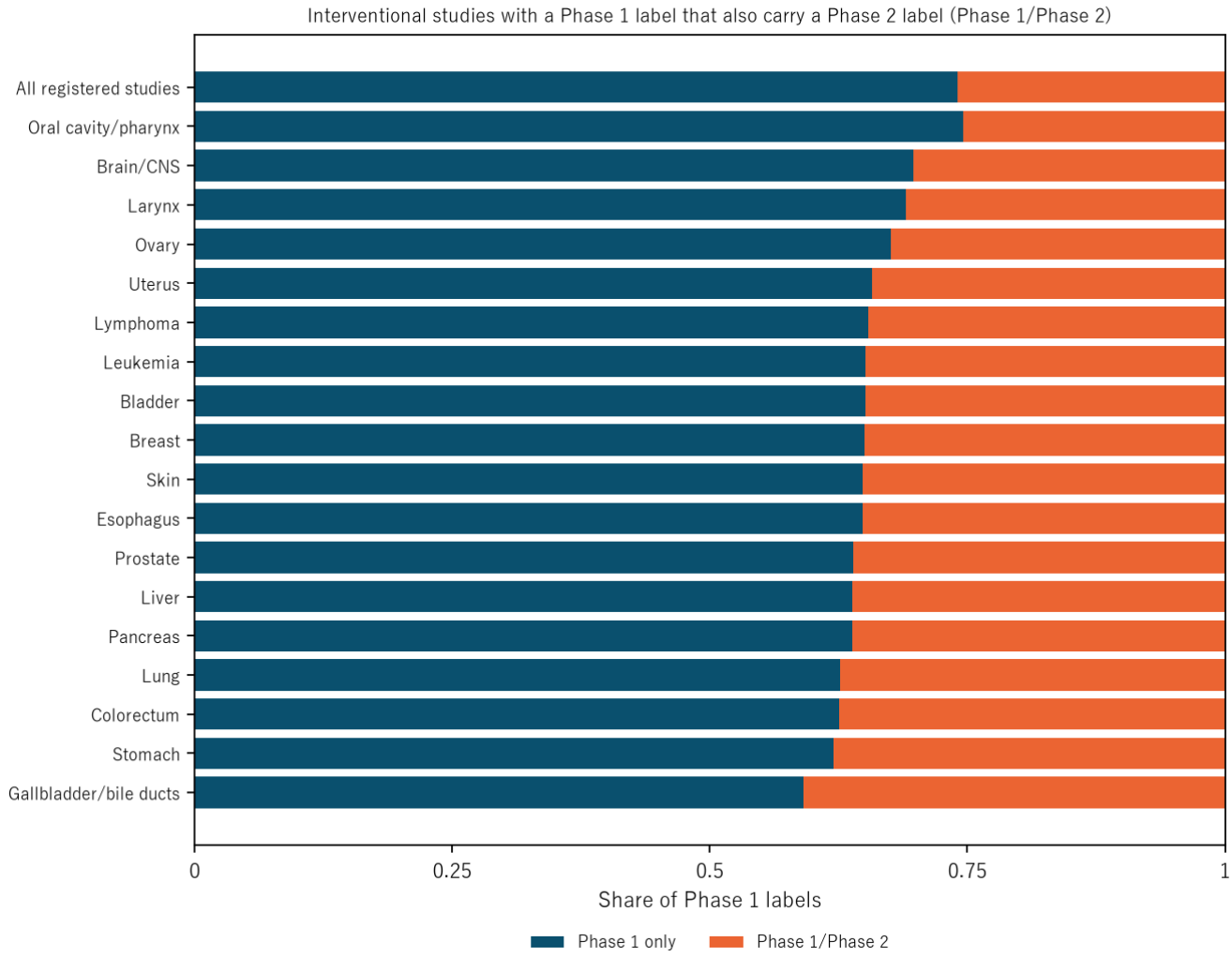


Figure 1: Among interventional studies with a Phase 1 label, the share that are Phase 1 only (navy) and Phase 1/Phase 2 (orange). The top row is all registered studies (all diseases).

---

5. The core — counting studies or counting phase labels

Observation 5.1: **A registrant selects only one phase, but "Phase 1/Phase 2" and "Phase 2/Phase 3" become two labels. So the counts by phase add up to more than the number of studies, and the difference equals exactly the number of studies with two labels (17,095 + 7,606 = 24,701 in all) minus the 125 with no phase.** No other combination, such as Phase 1 and Phase 3, appeared in the difference (the identity had a gap of 0 in all 19 cases).

Observation 5.2: **For the 18 cancer sites, the sum of labels ÷ interventional studies ranged from 1.0778 for breast to 1.1585 for leukemia, larger than for all registered studies (1.0533) at every site.**

The share of Phase 1 labels that also carry Phase 2 ranged from 0.254 for oral cavity/pharynx to 0.409 for gallbladder/bile ducts, and 17 sites were above all studies (0.260).

Observation 5.3: **One difference is the share without phases (NA).** Studies without phases include device and behavioral studies and carry only one label. For all registered studies, 0.515 of interventional studies have no phase, but for the 18 cancer sites the share ranged from 0.113 for leukemia to 0.376 for breast, smaller at every site. **Even comparing only interventional studies with a drug phase, excluding those without phases, the share with two labels was 0.1104 for all studies and 0.1236 for larynx to 0.1876 for pancreas at the 18 sites, larger at every site.**

---

6. The number registered and the number recruiting now

Calculation 6.1 Registered, recruiting, unknown status, and Japan (selected sites, in order of share with unknown status)

Scope	Registered	Recruiting (share)	Unknown status (share)	Location in Japan	Recruiting interventional in Japan
All registered studies	604,566	65,044 (0.108)	96,838 (0.160)	9,348	983
Stomach	4,330	726 (0.168)	1,095 (0.253)	225	40
Liver	5,448	884 (0.162)	1,271 (0.233)	145	14
Colorectum	10,292	1,581 (0.154)	2,034 (0.198)	212	47
Gallbladder/bile ducts	1,092	207 (0.190)	192 (0.176)	69	14
Pancreas	5,089	896 (0.176)	819 (0.161)	147	40
Breast	16,853	2,452 (0.145)	2,488 (0.148)	326	63
Larynx	713	75 (0.105)	85 (0.119)	5	0
Prostate	7,593	1,229 (0.162)	863 (0.114)	148	38

Observation 6.2: **The number registered and the number seeking participants now are different. The share recruiting at the 18 sites ranged from 0.105 for larynx to 0.190 for gallbladder/bile ducts. Of the 5,089 registered pancreas studies, 896 were recruiting, 679 were recruiting**

interventional studies, and 40 of those had a location in Japan (0.059 of recruiting interventional studies).

Observation 6.3: **In the ClinicalTrials.gov glossary, unknown status identifies a study whose last known status was recruiting, not yet recruiting, or active, not recruiting, but that has passed its completion date and whose status has not been verified within the past 2 years; such studies are considered closed. In all, 96,838 studies had unknown status, 1.489 times the 65,044 recruiting studies. At 11 of the 18 sites (oral cavity/pharynx, esophagus, stomach, colorectum, liver, larynx, lung, skin, breast, uterus, and ovary), unknown status outnumbered recruiting.** The share with unknown status ranged from 0.114 for prostate to 0.253 for stomach. The counts do not tell why status was not verified.

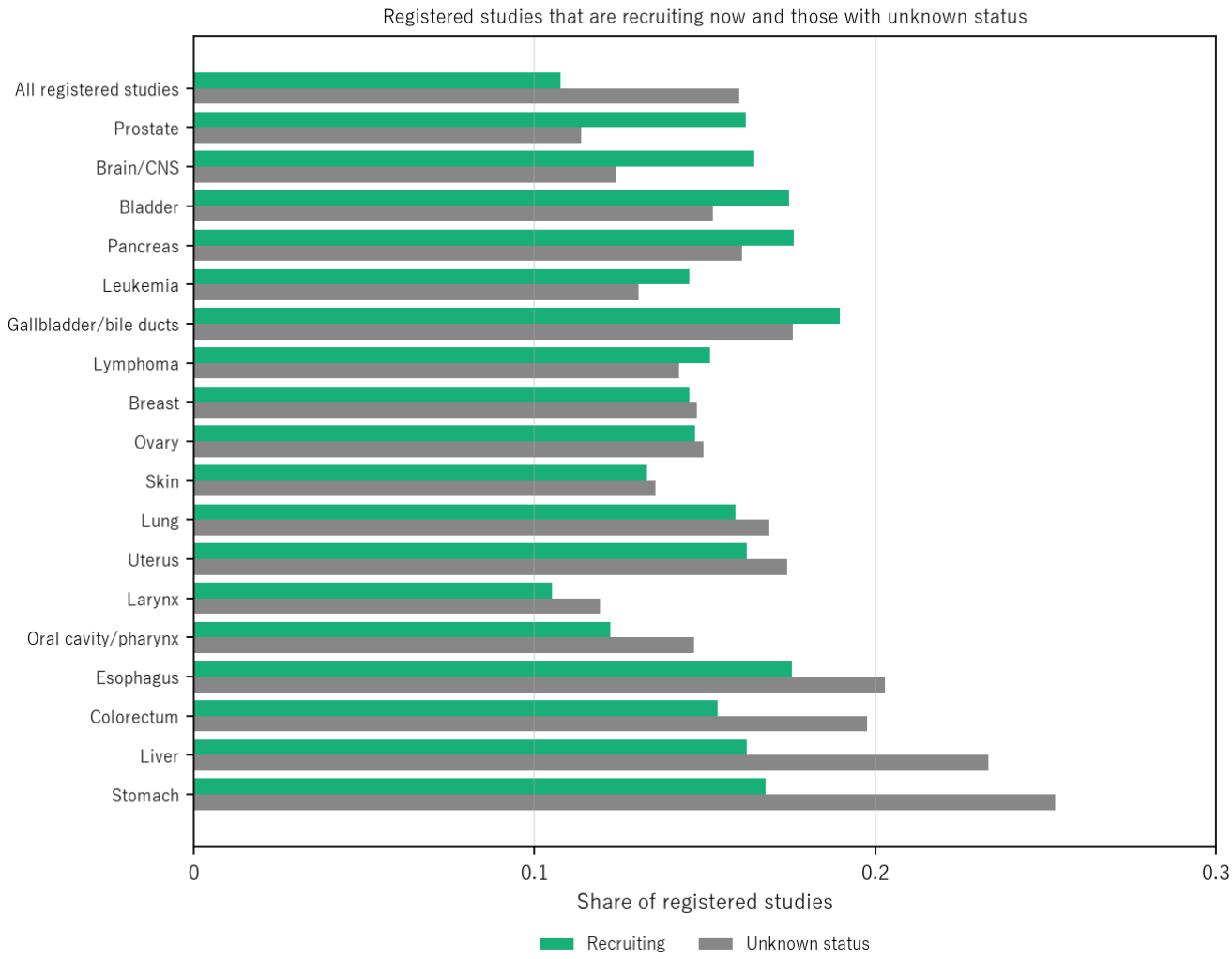


Figure 2: Share of registered studies that are recruiting now (green) and with unknown status (gray). The top row is all registered studies (all diseases).

7. Relation to earlier papers

M004 showed that the number of years in which papers double changes depending on what is counted in PubMed. This paper turned the same kind of question to a clinical trial registry and showed that the number changes depending on whether studies or phase labels are counted. M113 divided the number of papers on the same 18 sites by deaths and by incidence and showed that the ranking moves with the division.

8. Limitations

1. **The quality or value of trials, and whether a disease has more or fewer trials, are not discussed. Studies with unknown status are not discussed as the responsibility of the people or organizations who ran them.**
2. Sites were counted with ClinicalTrials.gov searches using MeSH headings. One study can match several sites, and the counts for the 18 sites do not add up to all registered studies.
3. The phase field is the value selected by the registrant. This paper did not check whether those selections are correct.
4. The counts are as of retrieval on 27 September 2026 (data processed on 2026-09-25T09:00:04), and ClinicalTrials.gov is updated daily.
5. The registry covers trials worldwide, and not every trial in Japan is necessarily registered on ClinicalTrials.gov. Japan also has other registries.

Conclusion

In a clinical trial registry, adding up the counts by phase gave more than the number of studies. A registrant selects one phase, but Phase 1/Phase 2 studies carry two labels, and the difference equals exactly the number of studies with two labels minus the studies with no phase. At cancer sites, where

studies without phases are fewer and studies with two labels more common, the difference was larger. And of the registered studies, only about 1 to 2 in 10 were seeking participants now, and at some sites studies with unknown status were more numerous.

What separates them is one thing — whether you count studies or phase labels.

When you hear "there are so many Phase 1 trials," **check whether Phase 1/Phase 2 trials are included.** When you hear "there are so many trials for this disease," **also check, with the latest numbers on ClinicalTrials.gov, how many are recruiting now and how many have unknown status.**

Data and reproduction

Data	Counts from the ClinicalTrials.gov (U.S. National Library of Medicine) API v2: 285 main queries and 19 additional queries (27 September 2026, from 01:43:23; API version 2.0.5). ClinicalTrials.gov processed the data (dataTimestamp) on 2026-09-25T09:00:04
Terms of use	ClinicalTrials.gov Terms and Conditions (attribute the source as ClinicalTrials.gov; display the date the data were processed; state modifications). This paper used only counts and computed their sums, differences, and ratios. The data are not always current, so check ClinicalTrials.gov for the latest numbers
Coverage	All registered studies and 18 cancer sites (MeSH headings in query.cond). Counts by type, phase, status, and location in Japan
Verification	`検算/M114.py` (`--zu` also draws the figures). Every number in the text is taken from `検算/M114_output.txt`
Raw data	`data/ctgov_m114_20260927.json` and `data/ctgov_m114_missing_20260927.json` (parameters, URL, count, retrieval time, processing date), the retrieval scripts `data/fetch_ctgov.py` and `data/fetch_ctgov_missing.py`, copies of the terms of use and definitions, and `data/SOURCES.md` (terms of use, SHA-256, site queries)
Figures	`図/M114_zu1` and `図/M114_zu2` (png, svg, English versions)

References

- U.S. National Library of Medicine. ClinicalTrials.gov REST API (API v2; data processed 2026-09-25T09:00:04; retrieved 27 September 2026). https://clinicaltrials.gov/api/oas/v2
- U.S. National Library of Medicine. ClinicalTrials.gov Terms and Conditions (accessed 27 September 2026). https://clinicaltrials.gov/about-site/terms-conditions
- U.S. National Library of Medicine. Protocol Registration Data Element Definitions for Interventional and Observational Studies (accessed 27 September 2026). https://clinicaltrials.gov/policy/protocol-definitions
- U.S. National Library of Medicine. ClinicalTrials.gov Glossary (accessed 27 September 2026). https://clinicaltrials.gov/study-basics/glossary
- 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
- Yamagishi, Y. (2026). 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]. Zenodo (unpublished).