Global Inequalities in Clinical Trials Participation

Wen Lou[1,2], Adrián A. Díaz-Faes[3*], Jiangen He[4], Zhihao Liu[1], Vincent Larivière[5,6,7,8*]

1 Department of Information Management, School of Economics and Management, East China Normal University, Shanghai, PR China
2 Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE (East China Normal University), Shanghai, PR China
3 INGENIO, CSIC-Universitat Politècnica de València, Valencia, Spain
4 School of Information Sciences, The University of Tennessee, Knoxville, TN, USA
5 École de bibliothéconomie et des sciences de l'information, Université de Montréal, QC, Canada
6 Consortium Érudit, Montréal, Québec, Canada
7 Department of Science and Innovation-National Research Foundation Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, Stellenbosch, Western Cape, South Africa
8 Observatoire des Sciences et des Technologies, Centre interuniversitaire de recherche sur la science et la technologie, Université du Québec à Montréal, Montréal, QC, Canada

Abstract

Clinical trials shape medical evidence and determine who gains access to experimental therapies. Whether participation in these trials reflects the global burden of disease remains unclear. Here we analyze participation inequality across more than 62,000 randomized controlled trials spanning 16 major disease categories from 2000 to 2024. Linking 36.8 million trial participants to country-level disease burden, we show that global inequality in clinical trials participation is overwhelmingly shaped by country rather than disease burden. Country-level factors explain over 90% of variation in participation, whereas disease-specific effects contribute only marginally. Removing entire disease categories-including those traditionally considered underfunded-has little effect on overall inequality. Instead, participation is highly concentrated geographically, with a small group of countries enrolling a disproportionate share of participants across nearly all diseases. These patterns have persisted despite decades of disease-targeted funding and increasing alignment between research attention and disease burden within diseases. Our findings indicate that disease-vertical strategies alone cannot correct participation inequality. Reducing global inequities in clinical research requires horizontal investments in research capacity, health infrastructure, and governance that operate across disease domains.

Introduction

Global health investments have expanded substantially over recent decades. Health spending in low-income countries grew at around 5% annually from 1995 to 2026, increasing from $51 to $153 per capita [1,2], while targeted funding initiatives (e.g., PEPFAR, Global Fund) demonstrated measurable impact on specific diseases [3]. Yet profound inequalities persist across multiple dimensions of global health [4]. The global health community has responded to such inequalities primarily through disease-focused approaches: prioritizing funding for under-researched and neglected diseases, strengthening disease-specific research capacity, and supporting targeted interventions [5]. These vertical programs achieved important successes-international health aid increased at 30% from 2009 to 2024 [6] and HIV/AIDS research expanded substantially following major funding commitments [7]. Yet, two major factors hinder progress: structural inequalities in global health research, and stagnation of international support for global health [8,9].

Regarding structural inequalities, recent evidence shows an increasing alignment between research efforts and burden of disease. However, this trend seems driven more by changes in the burden of disease than by actual shifts in research direction [10], since switching costs are very large [11]. Similarly, most drug innovation efforts concentrate on diseases prevalent in developed countries and with large market size [12,13]. Besides, global health research capacity remains disproportionately concentrated in the Global North, with only 35% of authors from low- and middle-income countries, despite 92% of articles addressing interventions in these regions [14]. As for health spending per capita, high-income countries in 2016 invested 130 times more than low-income countries, and is projected to persist at 126 times through 2050 [15]. Similarly, development assistance for health growth has been minimal at 1.2% annually since 2010, with ten consecutive years stagnating around $39 billion despite the peak during COVID-19 [6]. Taken together, these numbers highlight the pressing need to better align health policies and public and private R&D investment, foster more equitable collaboration and improve global coordination in health research.

From a more systemic perspective, such numbers also point to a deeper problem: vertical approaches based on disease-targeted interventions and target groups, while addressing immediate needs, may inadvertently perpetuate structural inequalities [16]. Disease-specific programs create duplication whereby each requires its own bureaucracy, lead to inefficient facility utilization, and may create gaps in care especially for patients with multiple co-morbidities [17]. Often funded by international and supranational organizations, vertical approaches tend to divert skilled local health personnel, creating internal brain drain and jeopardizing access to local health services [18]. Such competition for funding and recognition orients researchers toward those international initiatives, which affects national health systems [19]. Thus, despite decades of disease-focused investments, we may be addressing symptoms rather than causes: treating the research for each disease individually while missing the factors that cause deficits across all diseases simultaneously.

This paper examines inequalities through the lens of participation in terms of both knowledge production (who develops and performs the research) and research participants[1]. To do so, we rely on randomized controlled trials (RCTs), which occupy

---

[1] Note that in 2024, Declaration of Helsinki has changed the term "Subject" to "Participant". World Medical

a unique position in health research. RCTs are the gold standard in medicine and the pathway through which new therapeutics gain market access [20]. This dual role creates two distinct but related benefits. First, RCTs provide participating patients direct access to experimental treatments, advanced monitoring, and high-quality care-often representing the most sophisticated medical attention available. In low- and middle-income countries, participation in clinical trials is sometimes the only way to access medical treatment [21]. Second, trials generate the evidence that determines which treatments become standard practice, thereby shaping healthcare delivery across all populations and settings [22].

When trial participation and disease burden are not aligned geographically, benefits are distributed unequally. However, current inequality frameworks cannot systematically capture this. Existing indicators document disease burden concentration (where illness occurs) [23,24], funding allocation (where resources flow) [10], and publication patterns (where research outputs are produced) [25,26]. Yet no comprehensive framework measures whether populations bearing disease burden participate proportionally in research addressing that disease-both as research producers (where studies are conducted) and as enrolled participants (who access trial benefits). This gap reflects a conceptual limitation in how we understand research inequality and points to the increasing demand of health policies to reduce epistemic injustice (i.e. knowledge practices that privilege dominant groups) in global health [27].

The consequences are substantial for two reasons. First, lack of inclusion of low-income and marginalized populations limits generalizability of data guiding therapeutic interventions. Women, children, elderly, and those with common medical conditions are frequently excluded from RCTs, potentially impairing result generalizability [28,29]. This is particularly concerning since excluded populations face disproportionately higher morbidity and mortality [30]. Second, RCTs are often based on convenience samples, affecting the extent to which they are representative of full populations [31,32]. Pharma-companies value timely recruitment and data quality when allocating trials, with study population availability and site resources being the most important [33]. Such operational considerations systematically direct RCTs toward established infrastructure rather than on populations with the highest burden, creating a self-reinforcing concentration of populations studied.

This study examines participation inequality using original data on the level of inequalities in RCTs populations, and their relationship with burden of disease. We first analyzed 62,654 RCTs across 16 major disease categories conducted between 2000 and 2024, encompassing 36.8 million participants, and linked to time-varying Global Burden of Disease estimates for 182 countries. For each country-disease pair, we calculate the Participation-to-Burden Ratio (PBR) revealing systematic mismatches between where disease prevalence and research participants are from. Second, through variance decomposition, Shapley value analysis, and counterfactual removal techniques, we determine whether participation inequality is primarily disease-driven (suggesting disease-targeted interventions are appropriate) or determined by country-level factors (suggesting capacity-building investments are needed). Our analyses reveal that country-level factors account for 93.5% of variance in participation patterns, whereas disease-specific factors contribute only 2.7% Third, we translate these findings

---

Association. (2024). Declaration of Helsinki: Ethical principles for medical research involving human participants. https://www.wma.net/policies-post/wma-declaration-of-helsinki/

into an actionable policy framework, which diagnoses for each country-disease pair whether research investment capacity, health infrastructure, or governance represents the primary limiting factor, enabling targeted rather than generic capacity-building recommendations. On the whole, our analysis suggests that addressing global health inequality requires reconsidering prevailing frameworks that emphasize disease-targeted interventions. If participation inequality is country-driven, sustainable progress demands systematic investments in research capacity, health infrastructure, and governance that operate across disease domains rather than within them.

Results

Geographic and income inequalities characterize participant enrollment

Clinical trial participation exhibits profound geographic and income-based inequalities that operate independently of disease burden. Global north countries in North America and Western Europe are consistently over-represented in trial enrollment across disease categories (**Extended Data Fig. 1 and 2**). Cardiovascular diseases show identical patterns: regions with lower disease burden exhibit the highest participation rates (**Fig. 1A**), revealing that participant enrollment systematically favors wealthy nations (84.85% of global north countries) regardless of where disease concentrates. Even for HIV/AIDS, where African nations bear 65% of the global disease burden, global north countries contribute 58% of trial participants (**Fig. 1B)**.

These geographic disparities manifest through country-level research specialization patterns that mirror income and continental boundaries rather than epidemiological needs. This is apparent when countries are grouped by research focus: North American and European nations form distinct groups separate from sub-Saharan African and South Asian countries (**Fig. 1D**). Most nations show below-average participation rates ($\mathrm{Log(SI)} < 0$) across the majority of diseases, while several global north countries (e.g. Denmark, New Zealand) maintain above-average rates across multiple disease categories simultaneously (**Fig. 1E**).

Income-level stratification further amplifies these inequalities through differential specialization in disease categories (**Fig. 1C**). Non-communicable diseases attract concentrated specialization from multiple high-income countries, with nations achieving high SI values for neoplasms, cardiovascular diseases, and mental disorders. The relationship between disease burden and participation reveals the starkest income-based inequality. High-income countries demonstrate strong alignment between disease burden and research participation ($\beta = 0.466$, $p < 0.001$), enabling them to match research investment with epidemiological priorities (**Fig. 1F**). This alignment deteriorates progressively through middle-income countries ($\beta = 0.335$-0.366) and reaches its weakest expression in low-income nations ($\beta = 0.142$, $p = 0.012$) (**Fig. 1G-I**). The threefold difference (**Extended Data Fig. 2**) in alignment strength indicates that wealthy nations possess capacity to direct research toward their health priorities, while low-income countries cannot achieve such correspondence regardless of disease burden severity.

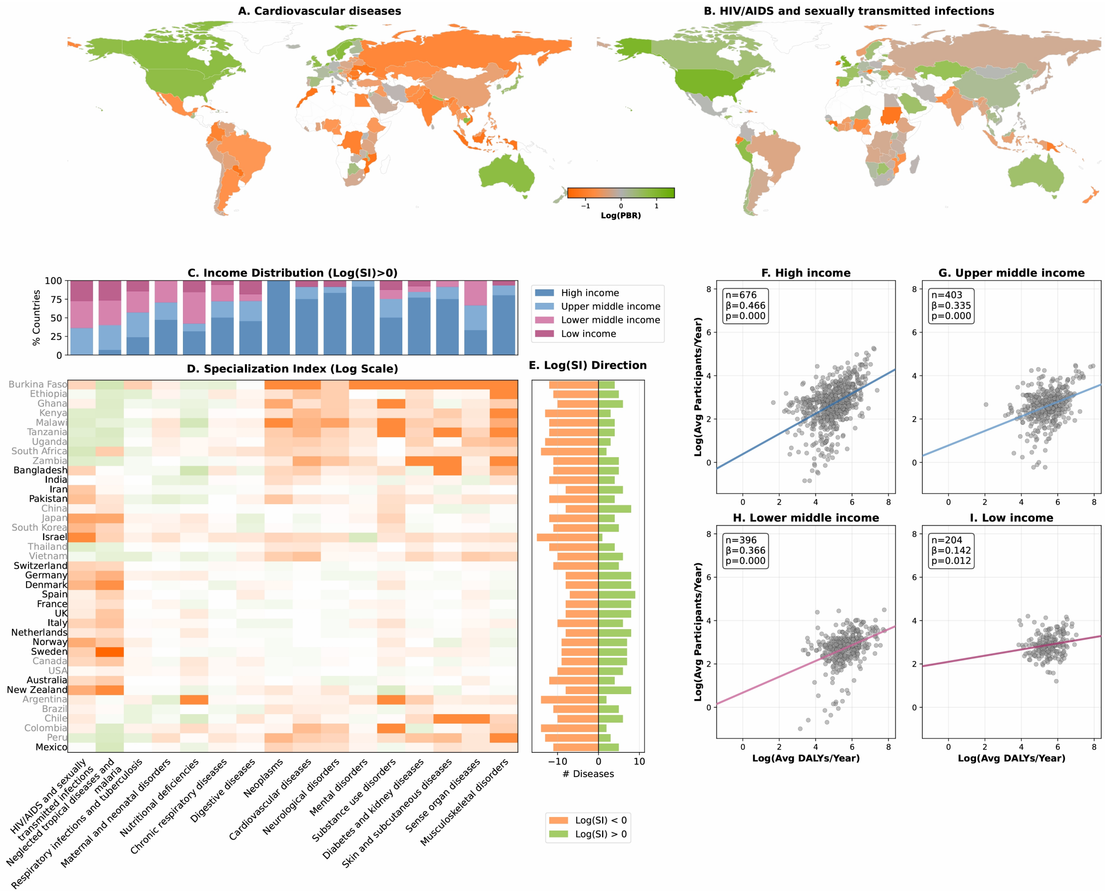


Fig. 1. Geographic inequality in disease research participation and specialization patterns. Multi-panel figure examining geographic disparities in clinical trial participation across diseases and income levels. **A-B:** two world maps showing log-transformed participation-to-burden ratios (PBR) for cardiovascular diseases and HIV/AIDS respectively, using diverging orange-gray-green colormap where orange indicates under-representation (Log PBR < 0), gray indicates proportional representation (Log PBR ≈ 0), and green indicates over-representation (Log PBR > 0). Log(PBR) scales from -1.5 to 1.5. World maps for other diseases are in **Extended Data Fig. 3**. **C:** stacked bar chart displaying income distribution of countries with positive specialization index (SI > 0) for each disease category, with bars colored by World Bank income classification. **D:** heatmap of specialization index (log scale) for top countries by region across Level 2 disease categories. **E:** stacked bar chart showing the count of diseases with positive versus negative SI for each country. **F-I:** scatter plots of log-transformed average annual DALYs (the total number of years of life lost because of illness, disability or premature death [10]) versus average annual participants for disease-country combinations, stratified by income level (high income, upper middle income, lower middle income, low income), with gray dots representing individual disease-country pairs, colored trend lines showing income-specific relationships, and diagonal reference lines. Statistical annotations include sample size (n), regression coefficient (β), and p-values.

Country-level factors dominate participation inequality

The geographic patterns observed across disease categories suggest that research participation is strongly influenced by country rather than by disease type. However, descriptive disparities alone do not quantify the relative importance of these two dimensions. To properly compare disease-driven versus country-driven sources of inequality, we applied a series of decomposition and sensitivity analyses.

We first assessed whether inequality is driven by specific diseases. For each disease category, we evaluated its marginal contribution to global inequality by quantifying the change in the overall Gini coefficient when research activity associated with that disease was excluded. Across all 16 Level 2 disease categories, contributions were uniformly small (**Fig. 2A**). Even the largest contributors, cardiovascular diseases and neoplasms, accounted for only 1.2% and 0.9% of total inequality, respectively (Contribution to Inequality Score, CIS; calculation in **Supplementary Methods 3**). Several diseases traditionally considered underfunded, including neglected tropical diseases and malaria, exhibited negative CIS values, indicating that their current research-burden alignment modestly reduces global inequality rather than exacerbating it. These results demonstrate that no single disease, nor any small subset of diseases, drives global participation inequality or mitigates it.

In contrast, geographic structure exhibited markedly stronger effects. Variance partitioning of participation-to-burden ratios revealed that country-level factors overwhelmingly dominated explanatory power. In a two-way decomposition framework, country accounted for 93.5% of the total variance (partial $R^2 = 0.935$), whereas disease explained only 2.7% (partial $R^2 = 0.027$), and temporal effects accounted for the remaining 3.8% (partial $R^2 = 0.038$) (**Supplementary Methods 3 and Extended Data Fig. 4B**). Country effects were thus more than 30-fold larger than disease effects, indicating that participation patterns are primarily shaped by geographic factors rather than disease research focus .

Temporal inequality decomposition further supports this conclusion. Differences in average research attention across diseases-declined from 32% in 2000-2004 to 22% in 2020-2024 (**Fig. 2C**), suggesting convergence in disease-level research attention over time. In contrast, within-disease inequality-capturing geographic disparities in participation for a given disease-increased from 68% to 78%, demonstrating that inequality is becoming increasingly country-driven rather than disease-specific.

To directly compare the robustness of global inequality along the disease versus country dimensions, we conducted parallel removal-based sensitivity analyses. Removing the top 20% of diseases ranked by CIS (3 of 16) reduced the global Gini coefficient by only 0.4%. By contrast, removing the top 20% of countries ranked by participation volume (34 of 173) reduced inequality by 23.9%-a more than 60-fold larger effect (**Fig. 2B and 2D**). Consistent with this asymmetry, the country-level Lorenz curve exhibited a pronounced shift following exclusion of major contributing countries (Gini: 0.884 to 0.672), whereas disease removal produced only a negligible change (Gini: 0.918 to 0.915).

Together, these findings demonstrate that global inequality in research participation is overwhelmingly influenced by country-level factors. Disease-specific initiatives can and do influence participation for particular conditions (e.g. targeted investments in HIV/AIDS research), but such effects operate within a geographic context that shapes baseline participation capacity across all disease domains. Regardless of disease burden, funding priority, or epidemiological profile, research participation follows a consistent global hierarchy determined primarily by country-level factors.

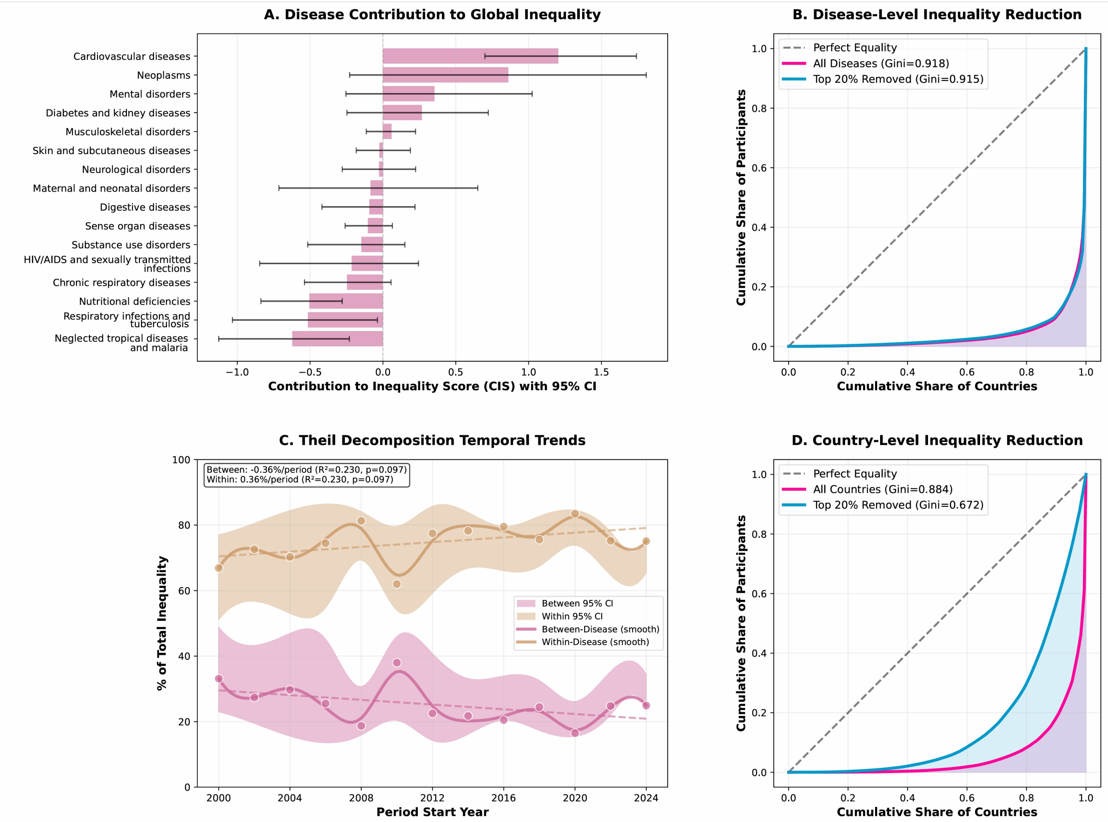


Fig 2. Decomposition analysis on drivers of participation inequality (2000-2024). A: Contribution to Inequality Score (CIS) for each Level 2 disease category, with 95% bootstrap confidence intervals, showing uniformly small disease-level contributions to global inequality. **B:** Lorenz curves comparing observed inequality with all diseases included versus after removing the top 20% of diseases by CIS, illustrating minimal sensitivity of inequality to disease removal. **C:** Temporal decomposition of inequality into between-disease and within-disease components, showing declining between-disease inequality and increasing within-disease (geographic) inequality over time. **D:** Country-level Lorenz curves comparing observed inequality with all countries included versus after removing the top 20% of countries by participation volume, showing substantially greater sensitivity of inequality to country-level structure.

Unveiling specific country-level factors

To better understand why certain countries dominate clinical trial participation regardless of disease burden, we must look beyond the general patterns of inequality and investigate the underlying Economic, Research Investment, Health Infrastructure, and Governance factors at play (**Supplementary Tables 5 and Extended Data Fig. 5**). We tested the effect of these factors on country-level inequality (PBR). The findings unveil that national characteristics such as population size and GDP explain a striking share of country-level variation in participation-to-burden ratios. Together they account for roughly one-third of the total differences (33.1%), with population size alone contributing the largest portion (25.5%) (**Supplementary Tables 6 and 7**). This dominance is intuitive: countries with large populations and sizable economies have greater research capacity and therefore are better positioned to conduct RTCs, as they provide more stable research ecosystems and better access to, and monitoring, of patient pools. GDP (11.1%) and health expenditure (14.3%) are critical as well, shaping the capacity to conduct trials in the first place. These factors are foundational, and because they change slowly, they create persistent asymmetries that are difficult to overcome through short-term interventions.

Economic factors (population size and GDP) matter greatly for reducing inequality, but we must isolate their residual variation to see the opportunity space for health policy action. Research investment (R&D expenditure, publications per capita), health infrastructure (hospital beds per capita, doctors per 10k, hospitals per capita, and health expenditure), and governance (HDI, democracy index) together explain 11.1% of the remaining variance (**Supplementary Tables 9**). Although smaller in magnitude than economic factors, their influence is meaningful because they point to areas where countries can actively reshape their research and health system landscape.

Among the factors that can be targeted by policy, research investment is the most influential in reducing inequality in clinical trial participation (**Supplementary Tables 9-11**). R&D expenditure and publication intensity account for 6.2% of variance. Health infrastructure adds another 2.6%, underscoring how research capacity, healthcare workforce, and availability of health facilities shape the feasibility of conducting trials once economic factors are taken into account. Governance factors capturing countries' regulatory reliability and broader developmental conditions contribute almost as much (2.3%), highlighting that investment translates into clinical trial participation only within supportive institutional environments.

Together, these findings show that research inequality is anchored in economic conditions but can be amplified or mitigated through policy. Large and wealthy countries benefit from inherent advantages, but health systems capacity, strong governance, and targeted research investment can meaningfully drive shifts in reducing participation inequality despite the structural constraints.

Exploring horizontal interventions to reduce participation inequality

Our analysis reveals that global inequality in clinical trials participation stems from country-level factors rather than disease burden. In addition, beyond countries' economic conditions, participation inequality can be addressed by targeting research investment, health infrastructure, and governance. These findings provide a concrete basis for policy intervention at the country-level, consistent with the core principle of horizontal intervention, which emphasizes collaboration between different sectors of society and across social infrastructures [34].

To explore opportunities for horizontal interventions, we identified research, health, and governance factors associated with inequality for 1,501 country-disease pairs using linear regression. Then, we classified these pairs into three performance types based on deviations from expected participation levels: Under-performing (823 pairs), Over-performing (500 pairs), and As-expected (178 pairs). These types reflect fundamentally different extends of global inequality: Under-performing pairs show minimal global inequality contribution, indicating that they are primarily affected by participation misalignment rather than contributing to global inequality (**Fig. 3A**). Over-performing pairs exhibit both high positive residuals and elevated Contribution to Inequality Scores (CIS), indicating they actively drive inequality (e.g., substance use disorder-NZL: CIS=1.814; HIV/AIDS-USA: CIS=1.364) (**Fig. 3B**). Extreme PBR values are concentrated in a few nations: the top five countries (Denmark, Sweden, USA, Israel, Canada) hold PBRs of 8.7-14.2, whereas the remaining 167 countries range from

0.000002 to 8.3 (**Supplementary Data**). Comparing factor impacts across performance types **(Fig. 3A-C)** further clarifies these differences. All three factors exert relatively similar influence among under-performing pairs, even though these pairs are most frequently constrained by research investment, whereas over-performing and as-expected pairs generally benefit from greater access to research investment and health infrastructure. Interestingly, As-expected pairs characterized by strong governance and health-infrastructure tend to contribute more to global equality, highlighting the stabilizing role of these capacities.

We further examined how factors and performance types are connected globally by visualizing the country-factor-disease network (**Fig. 3D**). First, research investment is the most influential factor across all performance types, associated with broad disease participation (6.57 diseases per node; **Supplementary Tables 13**) and connecting 64.8% of under-performing pairs. Second, governance factors are linked to fewer diseases (3.76 on average) and appear relatively isolated from other factors. Third, health infrastructure occupies an intermediate position, connecting to a moderate number of diseases (5.17) and bridging research investment and governance factors. Notably, only 18.3% of country-factor pairs are affected by multiple deficits, indicating that most participation misalignments are factor-specific and potentially addressable through targeted intervention.

The deficit of research investment in under-performing country-disease pairs, coupled with the disproportionate influence of over-performing pairs in driving inequality points where action is needed, but the question of which type of action is more effective remains open. We tested two courses of action through two simulated scenarios: (1) Full Structural Alignment (all countries shift toward median PBR), representing maximum theoretical equality; and (2) Targeted Alignment (only the most misaligned countries adjusted), reflecting a more resource-efficient and realistic approach. Full alignment would eliminate all avoidable inequality (Gini reduction: 100%; from 0.737 to 0.000). Targeted alignment would prove remarkably more efficient: adjusting just the top 40% of countries (68 of 172) would reduce inequality by 58.1%, while adjusting only the top 10% (17 countries) would achieve 23.94% reduction (**Fig. 3E, Supplementary Tables 19**). Targeted alignment was 1.45× more efficient per country adjusted than full alignment (**Supplementary Tables 20**), confirming that extreme misalignment concentrates in an addressable subset.

Both scenarios inherently rely on country-collaboration as they point towards horizontal interventions designed around peer groups of countries facing comparable constraints, enabling coordination, shared infrastructure, and encompass research efforts across disease areas. We modeled the effects of the two simulated scenarios on the global research collaboration network (262 nodes, 15,065 edges). Under full alignment, network density increases 49.9% (from 0.441 to 0.661), while homophily (connections among nodes sharing the same constraint) decreases 39.9% (from 0.523 to 0.314), indicating stronger cohesion but also greater opportunities for connecting with different others through intermediary nodes (**Fig. 3F**). Targeted alignment produces similar strong integration: density increases to 0.641 (50.5%), while homophily: 0.333 (36.6%) (**Supplementary Tables 21**). These shifts demonstrate that strengthening horizontal interventions through collaboration not only reduces participation inequality but also fosters a more cohesive and diverse global research ecosystem.

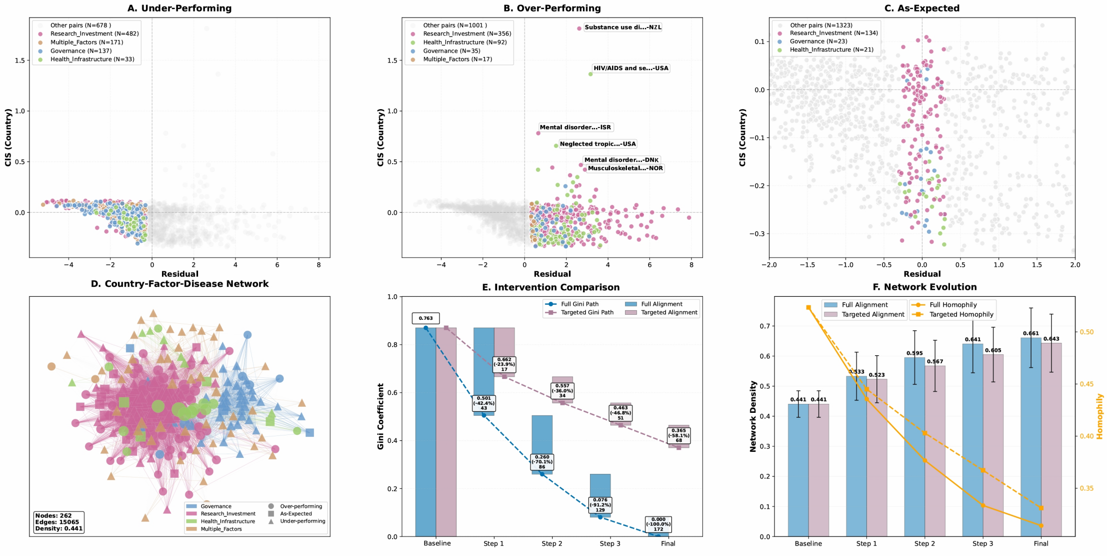


Fig. 3. Distribution and simulation analysis of country-factor-disease intervention. **A:** Scatter plot of Under-performing showing Residual (x-axis) versus CIS (y-axis). Points colored by factor dominance (Research Investment, Health Infrastructure, Governance or Multiple-Factors). Gray points represent the other two performance types. **B:** Scatter plot of Over-performing. Same display as A. **B:** Scatter plot of As-expected. Same display as A. **D:** Network visualization showing connections between country-factor pairs (nodes). Nodes are colored by factor dominance. Node shapes indicate performance types. Node size scales with disease count. Edge thickness represents the weight of shared disease connections. **E:** Waterfall chart showing Gini coefficient reduction under two scenarios of interventions. X-axis steps: Baseline -Top 25%-50%-75%-100% Adjusted for scenario 1, Baseline -Top 10%-20%-30%-40% for scenario 2. Y-axis shows the Gini coefficient. Bars represent Gini reduction at each step. Text labels show Gini value, percentage reduction, and number of countries adjusted. **F:** Network evolution metrics under interventions. Bars show Network Density for Full Alignment (blue) and Targeted Alignment (pink) scenarios with error bars. Lines show Homophily (proportion of same-factor connections) reduction for both scenarios. X-axis represents intervention steps matching panels E. Right Y-axis shows Homophily values. Summary text indicates percentage improvements in density and cross-group connections. Note that in panel E, the calculation of Gini coefficient differs from **Fig. 2B, 2D** due to the different calculation variables.

Discussion

Our analyses demonstrate that clinical trial participation exhibits persistent and pronounced inequality that is not explained by disease burden, but instead reflects a structural property of the global health system. Across disease areas, participation patterns are dominated by country-level factors and remain stable over time, even as disease-specific efforts and funding intensify[35]. This decoupling indicates that vertical programs (i.e. disease specific efforts delivered through targeted interventions) have had little effect on the entrenched inequality practices involving clinical trial participation. We show that changing this unequal global structure of participation requires horizontal interventions: strengthening the health research system, research capacity and governance across areas of health.

The dominance of country-level effects over disease-specific factors also indicates that trial participation is shaped primarily by where trials can be conducted, rather than by where disease burden is greatest. Trial populations define the empirical basis for evaluating medical safety, efficacy, and comparative effectiveness[32], and prior work has shown that clinical outcomes, treatment responses, and comorbidity profiles vary across populations and health systems[24,29,36-39]. Our findings show that the clinical evidence base is systematically biased towards participation from Global North countries. Financial resources to conduct clinical trials are often mobilized and filtered through national health systems and domestic constraints (regulatory regimes, institutional administration, infrastructure), which limits opportunities for countries with lower capacity [16]. Nevertheless, this misalignment cannot be addressed at the level of countries alone as modern clinical trials operate within transnational networks of sponsors, sites, and complex regulations[36,40]. Thus, intervention requires coordinated efforts on how the global research system allocates opportunities to participate in, and thereby shape, the production of medical evidence[41]. In recent years, an emerging body of models to support coordinated funding have begun to appear and can help target resources to countries with lower health and research capacity[42].

Addressing participation inequality is crucial for scientific and ethical reasons. On the one hand, a central element of the relevance of clinical trials is whether results can be reasonably applied to a well-defined population[42]. If the evidence produced is built disproportionately from participation in lower-burden countries or from a population with limited diversity, the evidence on the effect of the intervention can be limited precisely for the subgroup most likely to benefit from the clinical trial outcome. This concern is well-documented in studies when trials move to external implementations, which show that treatment effectiveness can vary with health-system capacity, background risk, and population characteristics[43-45]. On the other hand, reshaping participation is not only a distributional concern but a matter of epistemic justice, as it implies recognising as valid producers and recipients of knowledge populations usually marginalized in global health research [27,41].

Our results suggest that disease-vertical funding operates downstream of fundamental country-level factors. Thus, changing the unequal structure of trial participation requires horizontal interventions. This interpretation is consistent with studies showing that external funding translates into outcomes only where absorptive capacity is sufficient, highlighting the need for horizontal investments that build that capacity[46,47]. Vertical programs based on disease-targeted interventions have dominated health funding, despite previous calls to reconcile both approaches[16]. This dominance can be explained by considerations of cost-effectiveness in impact evaluations, the golden standard status of randomized control trials[20] and the search for single interventions that are feasible to donors and governments. Building infrastructure or developing integrated systems that operate across disease areas can be costly, less susceptible to randomisation and more difficult to assess in terms of impact (e.g. lives saved)[16,34]. Despite these challenges, increasing support for horizontal interventions is crucial for improving equity in global health [30,48] as funding allocation strategies based on disease-targeted interventions are unable to substantially alter where clinical research activity occurs.

Our research has some limitations. First, our analyses rely on published clinical trial records, potentially under-representing early-phase studies. Second, participation measures at a macro level lack granularity to capture all dimensions of trial involvement. Nevertheless, these limitations do not alter the central finding that participation inequality is weakly coupled to disease burden. Across alternative classification thresholds, disease exclusions, and decomposition strategies, estimates of the relative contribution of country- and disease-level factors remain stable, with country effects consistently accounting for the majority of observed variation. Third, country-level indicators such as GDP and governance indices capture only selected dimensions of capacity and may not fully reflect context-specific institutional arrangements. Incorporating other measures to account for economic, research, health and governance such as institutional quality indicators represents an important direction for future research.

Reducing participation inequality in clinical trials demands reconsidering prevailing approaches to global health. Taken together, our findings suggest three priority areas for action. First, shifting the entrenched inequality pattern of participation requires greater emphasis on horizontal investment (i.e. building and developing health systems and infrastructure) that operates across disease domains. Second, effective interventions depend on targeted interventions across similar country-disease contexts rather than applying uniform capacity-building solutions. Third, coordinated and collaborative efforts that redistribute analytical capacity and epistemic authority should help ensure that participation yields not only therapeutic access but also shared knowledge production. Addressing global health inequality, therefore, requires treating research participation as a structural property of the global health system, rather than as a disease-specific problem.

References

[1] Micah, A. E., Bhangdia, K., Cogswell, I. E., Lasher, D., Lidral-Porter, B., Maddison, E. R., ... & Hlongwa, M. M. (2023). Global investments in pandemic preparedness and COVID-19: development assistance and domestic spending on health between 1990 and 2026. *The Lancet Global Health*, 11(3), e385-e413.

[2] Dieleman, J., Campbell, M., Chapin, A., Eldrenkamp, E., Fan, V. Y., Haakenstad, A., ... & Murray, C. J. (2017). Evolution and patterns of global health financing 1995–2014: development assistance for health, and government, prepaid private, and out-of-pocket health spending in 184 countries. *The Lancet*, 389(10083), 1981-2004.

[3] Ratevosian, J., Millett, G., Honermann, B., Bennett, S., Connor, C., Bekker, L. G., & Beyrer, C. (2025). PEPFAR under review: what's at stake for PEPFAR's future. *The Lancet*, 405(10479), 603-605.

[4] Rad, J. (2025). Health inequities: a persistent global challenge from past to future. *International Journal for Equity in Health*, 24(1), 148.

[5] De Maio, F. (2014). *Global health inequities: A sociological perspective*. Bloomsbury Publishing.

[6] Apeagyei, A. E., Bisignano, C., Elliott, H., Hay, S. I., Lidral-Porter, B., Nam, S., ... & Dieleman, J. L. (2025). Tracking development assistance for health, 1990–2030: historical trends, recent cuts, and outlook. *The Lancet*, 406(10501), 337-348.

[7] Elendu, C., Amaechi, D. C., Elendu, T. C., Amaechi, E. C., Elendu, I. D., Akpa, K. N., ... & Idowu, O. F. (2025). Shaping sustainable paths for HIV/AIDS funding: a review and reminder. *Annals of Medicine and Surgery*, 87(3), 1415-1445.

[8] Baru, R. V., & Mohan, M. (2018). Globalisation and neoliberalism as structural drivers of health inequities. *Health Research Policy and Systems*, 16(Suppl 1), 91.

[9] Cash-Gibson, L., Rojas-Gualdrón, D. F., Pericàs, J. M., & Benach, J. (2018).

Inequalities in global health inequalities research: A 50-year bibliometric analysis (1966-2015). *PloS one*, 13(1), e0191901.

[10] Schmallenbach, L., Bley, M., Bärnighausen, T. W., Sugimoto, C. R., Lerchenmüller, C., & Lerchenmueller, M. J. (2025). Global distribution of research efforts, disease burden, and impact of US public funding withdrawal. *Nature Medicine*, 1-9.

[11] Myers, K. (2020). The Elasticity of Science. *American Economic Journal: Applied Economics*, 12(4), 103–134.

[12] Barrenho, E., Miraldo, M., & Smith, P. C. (2019). Does global drug innovation correspond to burden of disease? The neglected diseases in developed and developing countries. *Health Economics*, 28(1), 123–143.

[13] Yegros-Yegros, A., van de Klippe, W., Abad-Garcia, M. F., & Rafols, I. (2020). Exploring why global health needs are unmet by research efforts: the potential influences of geography, industry and publication incentives. *Health Research Policy and Systems*, 18(1), 47.

[14] Iyer, A. R. (2018). Authorship trends in the Lancet global health. *The Lancet Global Health*, 6(2), e142.

[15] Chang, A. Y., Cowling, K., Micah, A. E., Chapin, A., Chen, C. S., Ikilezi, G., ... & Qorbani, M. (2019). Past, present, and future of global health financing: a review of development assistance, government, out-of-pocket, and other private spending on health for 195 countries, 1995–2050. *The Lancet*, 393(10187), 2233-2260.

[16] Béhague, D. P., & Storeng, K. T. (2008). Collapsing the vertical–horizontal divide: An ethnographic study of evidence-based policymaking in maternal health. *American Journal of Public Health*, 98(4), 644-649.

[17] De Maeseneer, J., Van Weel, C., Egilman, D., Mfenyana, K., Kaufman, A., & Sewankambo, N. (2008). Strengthening primary care: addressing the disparity between vertical and horizontal investment. *The British Journal of General Practice*, 58(546), 3.

[18] Kirwin, E., Meacock, R., Round, J., & Sutton, M. (2022). The diagonal approach: A theoretic framework for the economic evaluation of vertical and horizontal interventions in healthcare. *Social Science & Medicine*, 301, 114900.

[19] Zhao, G., Cao, X., & Ma, C. (2020). Accounting for horizontal inequity in the delivery of health care: A framework for measurement and decomposition. *International Review of Economics & Finance*, 66, 13-24.

[20] Bothwell, L. E., Greene, J. A., Podolsky, S. H., & Jones, D. S. (2016). Assessing the gold standard-lessons from the history of RCTs. *New England Journal of Medicine*, 374(22), 2175-2181.

[21] Okpechi, I. G., Swanepoel, C. R., & Venter, F. (2015). Access to medications and conducting clinical trials in LMICs. *Nature Reviews Nephrology*, 11(3), 189-194.

[22] Hay, S. I., Ong, K. L., Santomauro, D. F., Aalipour, M. A., Aalruz, H., Ababneh, H. S., ... & Ajose, A. O. (2025). Burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023. *The Lancet*, 406(10513), 1873-1922.

[23] Lohr, K. N., Eleazer, K., & Mauskopf, J. (1998). Health policy issues and applications for evidence-based medicine and clinical practice guidelines. *Health Policy*, 46(1), 1-19.

[24] Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a

globalized science of infant development. *Infancy*, 28(4), 708-737.

[25] Evans, J. A., Shim, J. M., & Ioannidis, J. P. (2014). Attention to local health burden and the global disparity of health research. *PloS one*, 9(4), e90147.

[26] Zhou, H., Garg, P., & Fetzer, T. (2025). The Changing Geography of Medical Research. *medRxiv*, 2025-09.

[27] Bhakuni, H., & Abimbola, S. (2021). Epistemic injustice in academic global health. *The Lancet Global Health*, 9(10), e1465–e1470.

[28] Van Spall, H. G., Toren, A., Kiss, A., & Fowler, R. A. (2007). Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA*, 297(11), 1233-1240.

[29] Sugimoto, C. R., Ahn, Y. Y., Smith, E., Macaluso, B., & Larivière, V. (2019). Factors affecting sex-related reporting in medical research: a cross-disciplinary bibliometric analysis. *The Lancet*, 393(10171), 550-559.

[30] Warner, E., Marron, J. M., Peppercorn, J. M., Abel, G. A., & Hantel, A. (2024). Shifting from equality toward equity: addressing disparities in research participation for clinical cancer research. *The Journal of Clinical Ethics*, 35(1), 8-22.

[31] Clougherty, J. E., Kinnee, E. J., Cardet, J. C., Mauger, D., Bacharier, L., Beigelman, A., ... & Holguin, F. (2021). Geography, generalisability, and susceptibility in clinical trials. *The Lancet Respiratory Medicine*, 9(4), 330-332.

[32] Rothwell, P. M. (2005). External validity of randomised controlled trials: “to whom do the results of this trial apply?”. *The Lancet*, 365(9453), 82-93.

[33] Goldacre, B., Lane, S., Mahtani, K. R., Heneghan, C., Onakpoya, I., Bushfield, I., & Smeeth, L. (2017). Pharmaceutical companies’ policies on access to trial data, results, and methods: audit study. *BMJ*, 358.

[34] Stoljar Gold, A. (2025). Vertical dominance: Cost-effectiveness, randomisation, and the bias against horizontal interventions in global health. *Global Public Health*, 20(1), 2523542.

[35] World Health Organization. (2025). *A global health strategy for 2025-2028-advancing equity and resilience in a turbulent world: fourteenth General Programme of Work*. World Health Organization.

[36] Djurisic, S., Rath, A., Gaber, S., Garattini, S., Bertele, V., Ngwabyt, S. N., ... & Gluud, C. (2017). Barriers to the conduct of randomised clinical trials within all disease areas. *Trials*, 18(1), 360.

[37] Yaffe, K., Vittinghoff, E., Dublin, S., Peltz, C. B., Fleckenstein, L. E., Rosenberg, D. E., ... & Larson, E. B. (2024). Effect of personalized risk-reduction strategies on cognition and dementia risk profile among older adults: the SMARRT randomized clinical trial. *JAMA* Internal Medicine, 184(1), 54-62.

[38] Eckey, M., Li, P., Morrison, B., Bergquist, J., Davis, R. W., & Xiao, W. (2025). Patient-reported treatment outcomes in ME/CFS and long COVID. *Proceedings of the National Academy of Sciences*, 122(28), e2426874122.

[39] Yoo, S. K., Fitzgerald, C. W., Cho, B. A., Fitzgerald, B. G., Han, C., Koh, E. S., ... & Chowell, D. (2025). Prediction of checkpoint inhibitor immunotherapy efficacy for cancer using routine blood tests and clinical data. *Nature Medicine*, 31(3), 869-880.

[40] Zerhouni, E. A. (2005). Translational and Clinical Science--Time for a New Vision. *New England Journal of Medicine*, 353(15), 1621-1623.

[41] Theobald, S., Brandes, N., Gyapong, M., El-Saharty, S., Proctor, E., Diaz, T., ... & Peters, D. H. (2018). Implementation research: new imperatives and opportunities in global health. *The Lancet*, 392(10160), 2214-2228.

[42] World Health Organization. (2024). *Guidance for best practices for clinical trials. In Guidance for best practices for clinical trials.* https://iris.who.int/server/api/core/bitstreams/dbe6d97c-b659-4e7b-8070-b4a3c667d6e5/content.

[43] Vaz, L. M., Franco, L., Guenther, T., Simmons, K., Herrera, S., & Wall, S. N. (2020). Operationalising health systems thinking: a pathway to high effective coverage. *Health Research Policy and Systems*, 18(1), 132.

[44] Nordon, C., Karcher, H., Groenwold, R. H., Ankarfeldt, M. Z., Pichler, F., Chevrou-Severac, H., ... & Abenhaim, L. (2016). The “efficacy-effectiveness gap”: historical background and current conceptualization. *Value in Health*, 19(1), 75-81.

[45] Dorresteijn, J. A., Visseren, F. L., Ridker, P. M., Wassink, A. M., Paynter, N. P., Steyerberg, E. W., ... & Cook, N. R. (2011). Estimating treatment effects for individual patients based on the results of randomised clinical trials. *BMJ*, 343.

[46] Stephan, P. E. (1996). The economics of science. *Journal of Economic Literature*, 34(3), 1199-1235.

[47] Yin, Y., Dong, Y., Wang, K., Wang, D., & Jones, B. F. (2022). Public use and public funding of science. *Nature Human Behaviour*, 6(10), 1344-1350.

[48] Aiyegbusi, O. L., Cruz Rivera, S., Kamudoni, P., Anderson, N., Collis, P., Denniston, A. K., ... & Calvert, M. J. (2024). Recommendations to promote equity, diversity and inclusion in decentralized clinical trials. *Nature Medicine*, 30(11), 3075-3084.

Methods

Data sources and sample construction

We systematically identified randomized controlled trials through PubMed using validated search filters for human studies and clinical trial publication types. The initial retrieval yielded 301,262 articles published between 1980 and 2024. Following representativeness analysis that revealed systematic under-representation of geographic information in pre-2000 publications (geographic annotation success rate: 38% for 1980-1999 versus 72% for 2000-2024), we restricted analysis to 2000-2024, yielding 193,806 eligible studies. Post-hoc power analysis confirmed adequate sample size to detect effect sizes ≥0.3 with 80% power at α=0.05 (**Supplementary Methods 6**).

Disease burden data were obtained from the Global Burden of Disease data base 2021, maintained by the Institute for Health Metrics and Evaluation (IHME)[49], which provides disability-adjusted life years (DALYs) by country, year, and cause through hierarchical disease categorization. We focused on GBD Level 2 categories, which balance granularity with interpretability at the global macro level. Following previous research [13,25], causes that are difficult to assign to specific diseases (for example, "other non-communicable diseases") were excluded. We integrated disease burden associations from lower hierarchical levels (Levels 3 and 4) and aggregated them to Level 2, ensuring comprehensive coverage. In total, we examined 16 Level 2 disease categories encompassing both communicable diseases (HIV/AIDS and sexually transmitted infections, neglected tropical diseases and malaria, respiratory infections and tuberculosis, maternal and neonatal disorders, nutritional deficiencies) and non-communicable diseases (cardiovascular diseases, neoplasms, diabetes and kidney diseases, chronic respiratory diseases, digestive diseases, mental disorders, neurological disorders, musculoskeletal disorders, sense organ diseases, skin and subcutaneous diseases, substance use disorders). Enteric infections was removed from analysis due to lack of sufficient participation data.

Country-level indicators were compiled from multiple authoritative sources with temporal alignment to study periods. Economic indicators included GDP per capita, health expenditure per capita, and research and development expenditure as percentage of GDP (World Bank World Development Indicators). Research infrastructure metrics encompassed publications per capita, number of medical schools, and researcher density (UNESCO Institute for Statistics). Health system capacity indicators included hospital beds per 10,000 population and physician density per 10,000 population (WHO Global Health Observatory). Governance and social factors included the Democracy Index (Economist Intelligence Unit), Human Development Index (UNDP), and English language status (binary indicator for countries where English is an official language) (**Supplementary Tables 5**). All datasets were harmonized across three dimensions-geography (country-level ISO3 codes), time (publication year), and medical domain (mapped disease categories)-to enable analysis.

Geographic annotation and participants extraction

Trial geography was characterized across two critical dimensions: authorship location and participant enrollment sites. Author affiliations were extracted from PubMed metadata, which records institutional affiliations for corresponding and first authors. For multi-country studies, we applied an equal-count attribution method, distributing credit proportionally across contributing nations. This approach recognizes that international collaborations involve multiple geographic contexts while avoiding arbitrary designation of primary country.

Participant enrollment geography, which lacks standardized reporting in publication databases, required development of a multi-stage AI-assisted extraction pipeline. We compared five methodological approaches (**Supplementary Methods 1**) across 360 randomly sampled publications stratified by publication year to ensure temporal representativeness: (1) manual annotation by trained researchers serving as ground truth; (2) direct string matching using country name dictionaries; (3) Python-based named entity recognition using spaCy (version 3.7.0) with custom entity rulers; (4) domain-tuned LLaMA (model 3-8b) models fine-tuned on biomedical text; and (5) Gemma-based LLM (model 2-9b) with structured prompting for geographic entity extraction. Performance was evaluated using precision (proportion of extracted locations that were correct), recall (proportion of true locations successfully identified), and F1 score (harmonic mean of precision and recall).

Gemma-based LLM extraction achieved highest overall performance (accuracy: 92.3%, precision: 94.6%, recall: 89.8%, F1: 92.1%), substantially outperforming string matching (F1: 67.3%), spaCy name entity recognition (F1: 74.8%), and domain-tuned LLaMA (F1: 81.5%). The superior performance reflected Gemma-based's ability to resolve contextual ambiguities (distinguishing, for example, between "Georgia, USA" and "Georgia" as country), handle variations in geographic mention formats (city, institution, country names), and extract information from semi-structured text across abstracts, methods sections, and **Supplementary Tables 2 and 3**. We therefore deployed Gemma-based extraction across the full corpus of 193,806 studies.

Geographic entities identified in publication text were standardized to ISO3 country codes using validated mapping dictionaries (**Extended Data Fig. 5 and 6**). For studies reporting enrollment across multiple countries, we extracted country-specific participant counts when reported; otherwise, participants were distributed proportionally across mentioned countries. Participant counts were extracted using rule-based patterns validated through manual review of 500 articles, achieving 96.2% accuracy for reported sample sizes. RCTs with multiple participant resources were allocated with proportional population. Studies lacking geographic annotation or participant count information after extraction were excluded from analysis. Our final analytical dataset comprised 62,654 studies with complete geographic, participant, and disease information, representing 36.8 million trial participants.

Disease classification and medical concepts harmonization

Linking clinical trial publications to epidemiologically defined disease categories requires harmonizing heterogeneous medical classification systems that were developed for distinct purposes. International Classification of Diseases (ICD) codes are designed for clinical documentation and administrative reporting, whereas Global Burden of Disease (GBD) cause categories reflect population-level etiological groupings. Prior studies[50,51] have shown that direct alignment between these systems is intrinsically incomplete, owing to persistent mismatches in granularity, scope, and conceptual boundaries, and no authoritative one-to-one mapping currently exists.

To address this challenge, we adopted a conservative, ontology-mediated harmonization strategy that prioritizes semantic validity over maximal coverage **(Supplementary Methods 2)**. Rather than relying on direct ICD-GBD correspondence, we constructed mapping pathways across multiple established biomedical ontologies, including MeSH, ICD-10-CM, SNOMED CT, Disease Ontology, OMIM, and Orphanet, leveraging shared concept identifiers where available. This approach allows disease concepts expressed at different levels of clinical or biological specificity to be reconciled through intermediate representations, while avoiding heuristic or purely lexical matching. Disease assignments were performed at the publication (PMID) level, permitting non-exclusive mapping to multiple GBD causes when supported by the underlying annotations, and ICD codes designated as non-specific or “garbage” categories in the GBD framework were excluded.

Using this approach, we were able to assign at least one valid GBD cause to 62.1% of trial-linked publications. Although higher coverage can be achieved using probabilistic or model-based inference [10], such methods introduce uncertainty that is difficult to audit and may obscure patterns in disease representation. Given the study’s focus on systematic inequalities across disease areas, we therefore retained this conservative harmonization framework to ensure interpretability, reproducibility, and conceptual consistency across analyses.

Inequality metrics

We quantified research-disease inequality through the participation-to-burden ratio (PBR), calculated for each country-disease pair as:

$$PBR_{c,d,t} = \frac{P_{c,d,t} / P_{\blacksquare,d,t}}{B_{c,d,t} / B_{\blacksquare,d,t}}$$

where $P_{c,d,t}$ represents trial participants from country c for disease d in year t, $B_{c,d,t}$ represents DALYs for that country-disease-year, and subscript ■ denotes global totals. To address the right-skewed distribution of PBR values, we applied log-transformation (log-PBR) for analyses. A log(PBR) of 0 indicates perfect proportionality—the country contributes to research in exact proportion to its disease burden. Log(PBR) > 0 indicates over-representation (the country contributes more research participation than its burden would predict), while log(PBR) < 0 indicates under-representation (the country contributes less participation than its burden warrants).

This metric is mathematically symmetric with respect to country and disease dimensions. PBR can be equivalently calculated as "country *c*'s share of participants in disease d relative to country *c*'s share of disease *d* burden" or as "disease *d*'s share of participants from country *c* relative to disease *d*'s share of burden in country *c*." This symmetry ensures that analytical choices do not bias results toward either the disease-centric or country-centric hypothesis. Secondary metrics, including Specialization Index, Gini Coefficient, Contribution to Inequality Score, are explained in **Supplementary Methods 3**.

Inequality driven factor decomposition

To partition total inequality into disease-driven versus country-driven components, we employed three complementary statistical frameworks that avoid grouping-dependence artifacts and methodological biases (**Supplementary Methods 3**).

Bidirectional Theil Decomposition. The Theil index[52], an entropy-based measure from information theory, decomposes total inequality into between-group and within-group components. We applied the Theil decomposition bidirectionally. First, grouping observations by disease yields:

$$T_{total} = T_{between-disease} + T_{within-disease}$$

where $T_{between-disease}$ captures inequality from differences in disease-average PBR values (whether some diseases are universally over- or under-researched), and $T_{within-disease}$ captures inequality from variation in PBR across countries within each disease (geographic disparities in participation for any given disease). Second, grouping observations by country yields:

$$T_{total} = T_{between-country} + T_{within-country}$$

where $T_{between-country}$ captures inequality from differences in country-average PBR values (whether some countries universally over- or under-contribute), and $T_{within-country}$ captures inequality from variation in PBR across diseases within each country (whether countries specialize in particular diseases). This bidirectional approach tests whether conclusions depend on grouping choice (**Extended Data Fig.**

4). Country removal is used here as a sensitivity analysis rather than a policy counterfactual.

Two-Way ANOVA Variance Partitioning. To simultaneously estimate country and disease contributions without imposing a grouping structure, we employed variance partitioning using a two-way fixed effects model:

$$PBR_{c,d,t} = \mu + \alpha_c + \beta_d + \gamma_t + \epsilon_{c,d,t}$$

where c represents country fixed effects capturing time-invariant country characteristics, d represents disease fixed effects capturing time-invariant disease characteristics, t represents year fixed effects capturing temporal trends, and c,d,t represents residual variation. We calculated the proportion of total variance explained by each component using partial R^2 values. This approach provides an unbiased, symmetric quantification of country versus disease contributions to inequality.

Shapley Value Decomposition. To account for predictor interdependencies and attribute explained variance to specific factors, we implemented Shapley value decomposition (**Supplementary Methods 3**). Country-level predictors were organized into three conceptual blocks based on policy mechanisms: Research Investment (GDP per capita, R&D expenditure as % GDP, publications per capita), Health Infrastructure (health expenditure per capita, hospital beds per 10,000 population, physicians per 10,000 population), and Governance (Democracy Index, Human Development Index). For each country-disease observation, we calculated Shapley values by averaging each predictor's marginal contribution across all possible predictor orderings, using 1,000 random permutations to approximate the full combinatorial space. Shapley values were aggregated to block level by summing individual predictor contributions within blocks. Bootstrap resampling with 100 iterations generated 95% confidence intervals for block-level percentage contributions.

Identification of limiting country-level factors

For countries under-performing relative to predictions, we identified limiting research investment, health infrastructure, and governance factors through linear regression analysis (**Supplementary Methods 3**). We estimated disease-specific regression models predicting log-PBR from country-level predictors, yielding predicted values and residuals:

$$\text{Residual}_{i,d} = \log(PBR)_{i,d} - \log(\widehat{P}BR)_{i,d}$$

Country-disease pairs were classified as: Over-performing (residual > 0.5: performance exceeds predictions), As-expected (|residual| < 0.3: performance aligns with predictions), or Under-performing (residual < -0.3: performance worse than predicted).

For under-performing observations, we identified limiting factors by examining predictor coefficients within conceptual blocks. For each block b, we computed average absolute coefficient magnitude among significant predictors ($p < 0.1$):

$$\bar{\beta}_b = \frac{1}{n}\sum_{j\in b} |\beta_j| \cdot 1(p_j < 0.1)$$

The block with largest $\bar{\beta}_b$ was classified as the primary limiting factor, indicating which policy lever would have the greatest marginal impact. Country-disease pairs with multiple blocks showing comparably strong effects ($\bar{\beta}_b > 0.7 \cdot \max(\bar{\beta}_b)$) were classified as requiring "Multiple Factors" coordinated intervention.

To assign country-level factors to over-performing combinations without modifying the regression model, we applied a hierarchical matching procedure. Factor labels were inferred by matching over-performing combinations to dominant limiting-factor patterns observed among under-performing combinations at the disease-level, then country-level, with a global fallback where necessary.

For as-expected combinations, factor labels reflect the dominant contributor to expected performance. Each combination was assigned to the factor corresponding to the largest normalized component of authorship, disease burden, or participant recruitment, representing the primary alignment underlying its expected research-burden relationship.

Counterfactual intervention scenarios

To quantify how reducing structural misalignment might impact global research inequality, we simulated two counterfactual intervention scenarios at the national level.

First, we aggregated trial participation and disease burden data across all 16 diseases for each country, calculating national-level PBR. This provided a country-level measure of overall research capacity relative to overall disease burden, independent of disease-specific effects. We then implemented two alternative intervention strategies:

1. Full Structural Alignment: All 172 countries gradually shift their PBR values toward the global median (0.194). This scenario represents the theoretical maximum reduction in inequality achievable through complete structural harmonization.
2. Targeted Alignment: Only the most misaligned countries-those with PBR values deviating most from the median-are incrementally adjusted. This scenario reflects a resource-efficient strategy prioritizing high-impact interventions.

For each scenario, we calculated the Gini coefficient reduction using bootstrap resampling (200 iterations) across four intervention intensities: adjusting 25%, 50%, 75%, and 100% of countries for Full Alignment, and 10%, 20%, 30%, and 40% for Targeted Alignment. Percentage reduction was computed as $100\times(G_{baseline}-G_{adjusted})/G_{baseline}$, where $G_{baseline}=0.763$ represents the observed inequality.

To compare strategy efficiency, we calculated reduction per country adjusted: Efficiency=Percentage reduction/Percentage of countries adjusted. Statistical significance was assessed through paired t-tests comparing baseline versus adjusted Gini distributions across bootstrap samples.

Global research collaboration network features after interventions

To understand how structural realignment might reshape global research collaboration patterns, we constructed a country-factor-disease network and simulated its evolution under intervention scenarios.

The network comprised 262 nodes representing unique country-factor combinations (e.g., "USA-Research_Investment") and 15,065 edges connecting nodes that participate in the same disease. Node attributes included: factor (Research_Investment, Health_Infrastructure, Governance, Multiple_Factors), performance status (Over_performing, As_Expected, Under_performing), and disease count (node size). Edge weights reflected the number of shared diseases between connected nodes.

We analyzed four network measures that capture different dimensions of collaboration structure:

1. Network Density[53]: Proportion of possible connections that exist, measuring overall connectivity.
2. Homophily[54]: Proportion of edges connecting nodes with the same factor, measuring segregation by factor.
3. Modularity[55]: Strength of community structure, calculated using the Louvain algorithm.
4. Average Path Length[53]: Mean shortest distance between all node pairs, measuring network integration.

We modeled network evolution by linking changes in these metrics to reductions in structural inequality (Gini coefficient). Uncertainty was estimated through parametric bootstrap resampling (100 iterations), with confidence intervals reflecting both measurement uncertainty in baseline metrics and variability in sensitivity relationships.

Data availability

The data assembled for this study are available and can be accessed at https://doi.org/10.5281/zenodo.18115243. Source data are provided with this paper.

Code availability

The computer code used to perform the analyses in this study is available and can be accessed via the following link: https://doi.org/10.5281/zenodo.18115266.

References

[49]Institute for Health Metrics and Evaluation (IHME) Global Burden of Disease Study 2021 (GBD 2021) Data Resources (IHME, 2022).

[50]Zotova, E., Cuadros, M., & Rigau, G. (2022, June). ClinIDMap: Towards a clinical IDs mapping for data interoperability. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 3661-3669).

[51]Del Valle, E. P. G., García, G. L., Santamaría, L. P., Zanin, M., Ruiz, E. M., & Rodríguez-González, A. (2021). DisMaNET: A network-based tool to cross map disease vocabularies. *Computer Methods and Programs in Biomedicine*, 207, 106233.

[52]Theil, H. (1967) *Economics and Information Theory*. North-Holland Publishing Company, Amsterdam.

[53]Wasserman, S. (1994). *Social network analysis: Methods and applications*.Cambridge University Press.

[54]McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415-444.

[55]Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113.

Acknowledgements

We thank M. Rousseau, J. Du, and P. Xiao for their valuable assistance in rationalization. We also thank D. Kozlowski, P. Chen, L. Wu, Y. Duan, Q. Wang, and G. He for their excellent research support. W.L. received financial support through the Shanghai Planning Office of Philosophy and Social Sciences (Grant Number 2024BJC005). V.L. acknowledges funding from the Social Science and Humanities Research Council of Canada Pan-Canadian Knowledge Access Initiative Grant (Grant Number 1007-2023-0001), and the Fonds de recherche du Québec-Société et Culture through the Programme d'appui aux Chaires UNESCO (Grant Number 338828).

Author contributions

W.L., J.H., and V.L. devised the original idea. W.L. and A.A.D. conceptualized the study. W.L. led the study and analyses. W.L., Z.L., and V.L. assembled and analyzed the data. W.L., A.A.D, and V.L. drafted and wrote the final manuscript. W. L. provided critical access to data sources and feedback on methods, results and interpretations. All authors edited the manuscript.

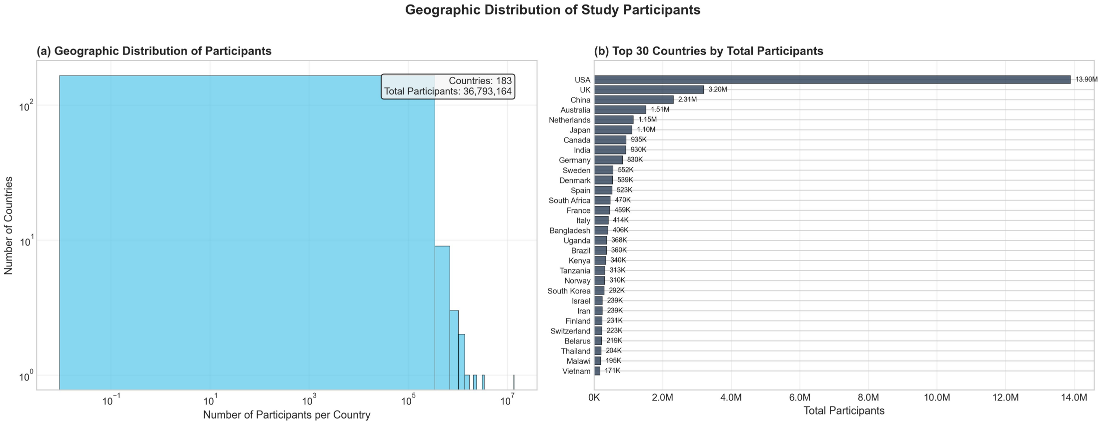
Competing interests

The authors declare no competing interests.

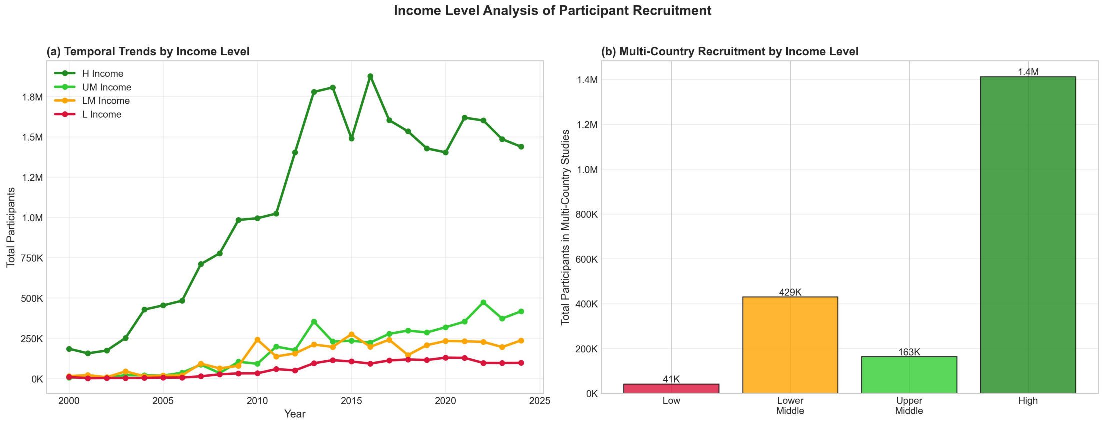
Additional information

Supplementary information is available for this paper.

Extended Data

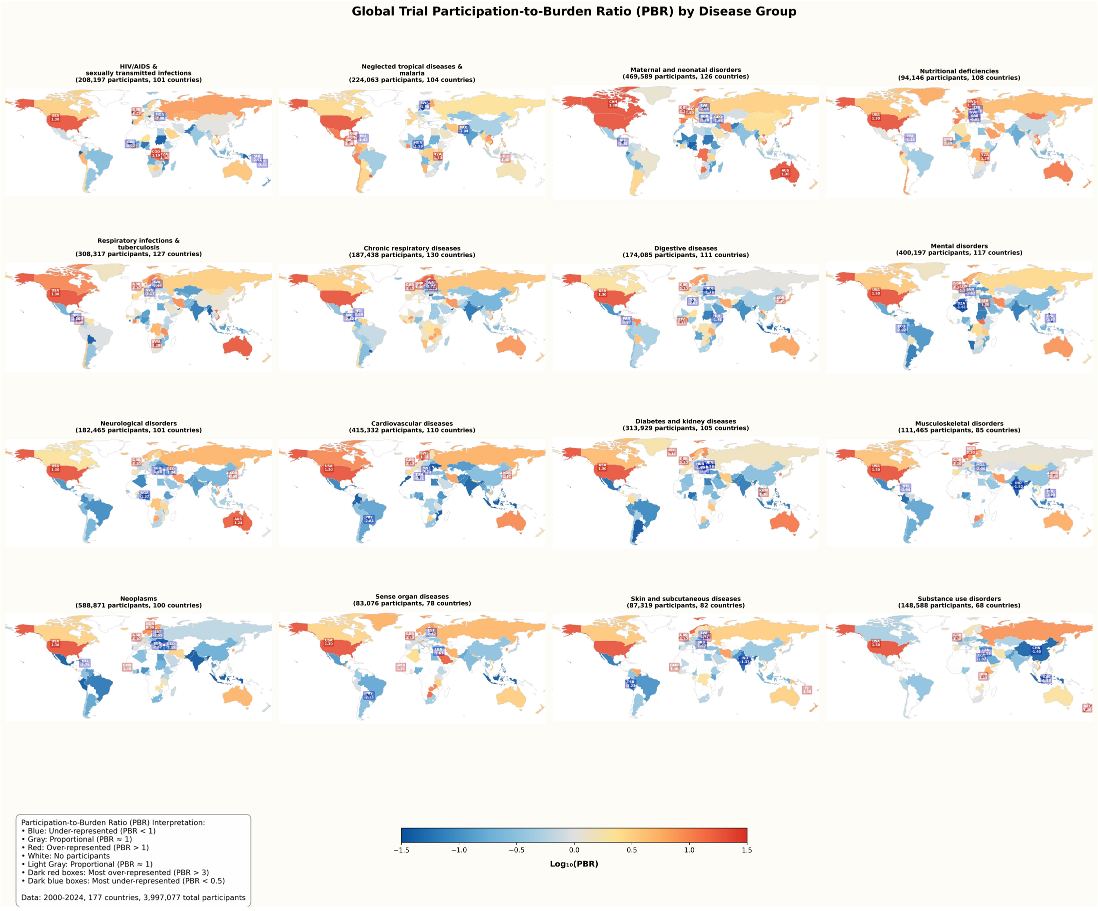


Extended Data Fig. 1| Geographic distribution of clinical trial participants by country. Participant totals represent cumulative enrollment across all studies conducted in each country from 2000-2024. Histogram showing the frequency distribution of countries by total participant enrollment in the countries. The highly skewed distribution demonstrates that most countries enroll relatively few participants, while a small number of countries recruit thousands of participants. Horizontal bar chart showing top 30 countries by total participant recruitment across all studies in the dataset. Countries are ordered by absolute participant numbers, with numerical annotations indicating total recruitment. The concentration pattern mirrors publication geography, with established research centers recruiting the majority of global trial participants.

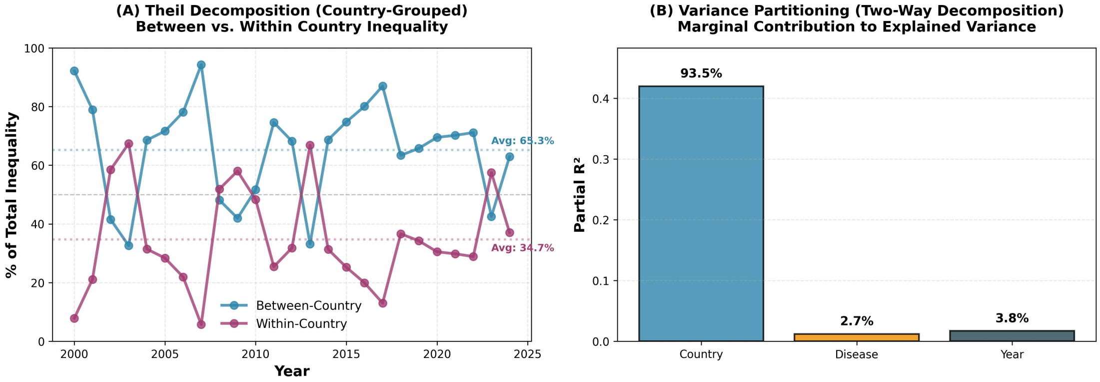


Extended Data Fig. 2| Clinical trial participant recruitment patterns by World Bank income classification. Panel a shows temporal trends in total participants recruited by income level from 2000-2024, with high-income countries (dark green) demonstrating exponential growth reaching 1.9M participants annually by 2016, while other income groups show modest, stable recruitment levels. Panel b displays cumulative participants in multi-country studies by income level, revealing extreme inequality with high-income countries recruiting 1.4M participants compared to 41K in low-income countries. The stark disparities reflect differences in research infrastructure, regulatory capacity, and economic resources for conducting clinical trials across income strata.

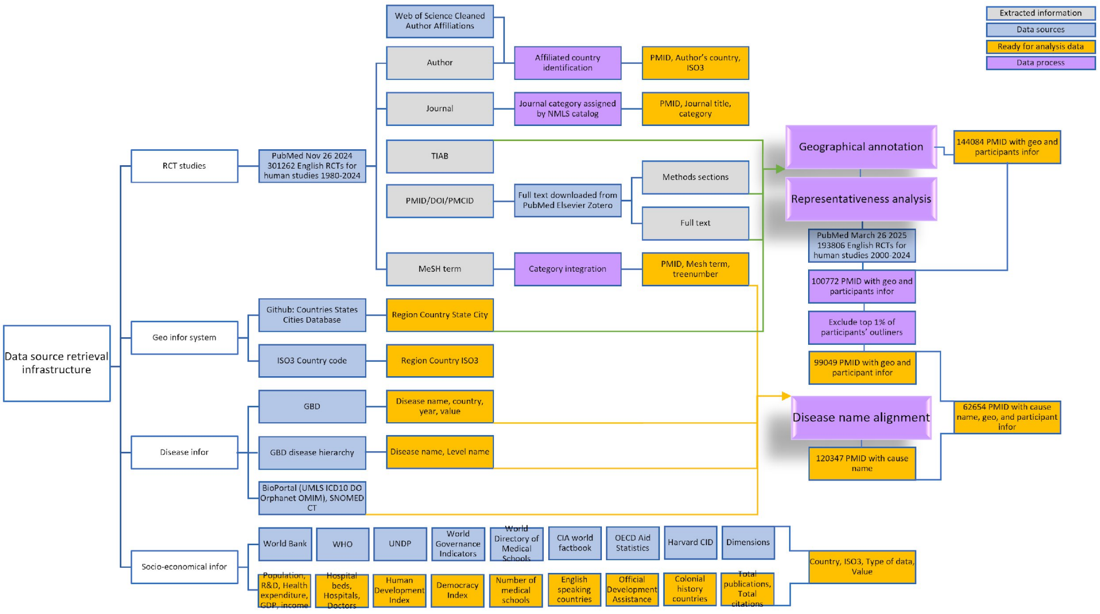
Comprehensive statistical testing reveals significant associations between country income classification and research participation rates. Kruskal-Wallis test confirms significant differences across income groups ($H = 23.292$, $p < 0.001$), while Spearman correlation demonstrates moderate positive association between income level and participation rate ($\rho = 0.281$, $p < 0.001$). High-income countries average 13,555 participants per million population compared to 1,823 in lower-middle-income countries, despite low-income countries showing elevated rates (3,491 per million) due to specific high-participation outliers. Chi-square analysis confirms significant association between recruitment patterns and income classification ($\chi^2 = 65.770$, $p < 0.001$), indicating systematic economic determinants of global research participation access.



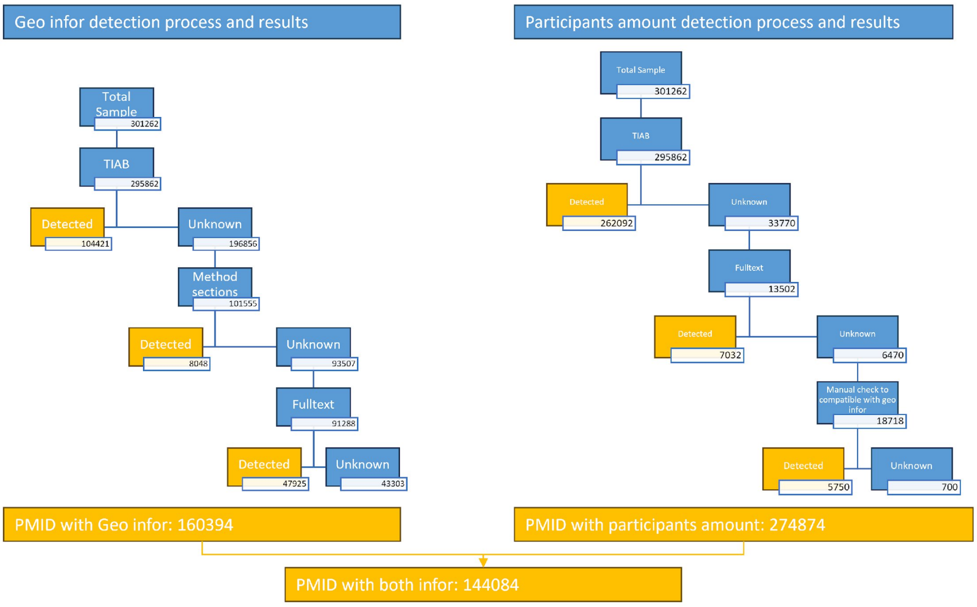
Extended Data Fig. 3|Global clinical trial participation-to-burden ratio (PBR) by Level 2 disease categories (2000-2024). Sixteen-panel world map displaying geographic patterns of research participation relative to disease burden for each custom disease category. Colors represent $\log_{10}$-transformed PBR values using a diverging colormap: blue indicates under-representation (log(PBR) < 0, fewer participants than burden share warrants), light gray indicates proportional representation (log(PBR) ≈ 0), and red indicates over-representation (log(PBR) > 0, more participants than burden share). White areas indicate no participant data. Each panel title shows total participants and number of countries with data. The color bar is centered at 0 (corresponding to PBR = 1) with range from -1.5 to +1.5 on $\log_{10}$ scale.



Extended Data Fig. 4| Symmetric decomposition analyses quantifying relative contributions of country versus disease factors. Two-panel analysis providing complementary perspectives on inequality drivers: Panel A shows Theil decomposition grouped by country, partitioning total inequality into between-country variance versus within-country variance across years 2000-2024, with a 50% reference line indicating equal contribution. Panel B presents two-way variance partitioning using fixed effects regression, calculating partial R^2 values representing the marginal contribution of country, disease, and year fixed effects to explained variance in participation-to-burden ratios. Percentage labels indicate each component's share of total explained variance.



Extended Data Fig. 5| Study design and data integration workflow. Schematic overview of data sources, processing pipeline, and analytical framework for examining global disparities in clinical trial participation. The study integrated five primary data sources: PubMed bibliographic records (n=301,262 studies, 1980-2024), Global Burden of Disease estimates (939K country-year-cause observations), author affiliation data, participant geographic and demographic information extracted via AI-assisted methods, and socioeconomic indicators from international databases. After temporal restriction to 2000-2024 and sequential filtering for geographic annotation and disease mapping, the final analytical datasets comprised 62,654 studies with complete participant and disease information. PMID, PubMed identifier; GBD, Global Burden of Disease; AI, artificial intelligence; TIAB, title and abstract; BioPortal's datasets see Supplementary Tables 4.



Extended Data Fig. 6| AI-assisted geographic annotation and participant extraction pipeline. Multi-stage workflow for extracting participant geography and enrollment numbers from clinical trial publications. Text inputs were sourced hierarchically from abstracts, methods sections, and full-text articles. Geographic locations were identified using AI-assisted named entity recognition targeting universities, hospitals, cities, and countries. Participant counts were extracted using rule-based patterns and validated through manual review. Methodological details of AI-assisted extraction results see Supplementary Tables 3.

Supplementary Tables

Supplementary Tables 1. Temporal distribution and sampling strategy for AI model development. Annual distribution of clinical trial publications used for training and validating geographic annotation and participant extraction models (1980-2024). A stratified sampling approach was employed to ensure representative coverage across publication years, with sampling proportions adjusted for annual publication volumes. The validation dataset comprised 360 articles selected proportionally from 301,262 total publications, providing robust representation across the study period.

Year	Count	Sampling
2024	12862	13
2023	16867	17
2022	20439	21
2021	21247	22
2020	19328	20
2019	18446	19
2018	17814	18
2017	17658	18
2016	17525	18
2015	18097	19
2014	18084	19
2013	17132	18
2012	14707	15
2011	12716	13
2010	10854	11
2009	9421	10
2008	8062	9
2007	6879	7
2006	6264	7
2005	5896	6
2004	5067	6
2003	4260	5
2002	3556	4
2001	3065	4
2000	3329	4
1999	3362	4
1998	3162	4
1997	2901	3
1996	2425	3
1995	2334	3
1994	1940	2
1993	1424	2
1992	1309	2
1991	1322	2
1990	1260	2
1989	965	1
1988	806	1
1987	794	1
1986	748	1
1985	652	1

1984	548	1
1983	613	1
1982	560	1
1981	453	1
1980	486	1
Total	337639 (distinct 301262)	360

Supplementary Tables 2. Performance evaluation of nine AI-assisted extraction strategies for geographic annotation and participant count extraction. Each strategy was evaluated on a stratified sample of 360 clinical trial articles using precision, recall, F1-score, and processing time per article. Strategy selection balanced accuracy and computational efficiency, as total processing time scales linearly with article volume (n=300,262). Strategy 9 (Geo and # separately with examples for #) achieved optimal performance (F1=0.982) with fastest processing time (3:31 per article), representing the best accuracy-efficiency trade-off. Time cost calculations: 3:31 per article × 300,262 articles = 1,051,424 total processing minutes (≈17,524 hours). Geographic annotation (Geo) and participant count (#) extraction were evaluated separately to optimize task-specific performance.

Strategy	Precision	Recall	F1	Time cost
(1) Least guidance	0.729	0.822	0.773	3:38
(2) Human brain guidance without examples	0.927	1	0.962	3:39
(3) Human brain guidance with examples	0.976	0.943	0.959	3:39
(4) Geo and # in order combined with examples (for Geo)	0.926	0.302	0.456	3:38
(5) Geo and # in order combined with examples (for #)	0.976	0.942	0.959	3:39
(6) Geo and # in reversed order combined with examples (for Geo)	0.926	0.28	0.43	3:35
(7) Geo and # in reversed order combined with examples (for #)	0.963	0.952	0.958	3:35
(8) Geo and # separately with examples (for Geo)	0.97	0.943	0.957	3:31
(9) Geo and # separately with examples (for #)	0.978	0.987	0.982	3:31

Supplementary Tables 3. Performance comparison of geographical annotation methods. Five approaches were evaluated for extracting study location from clinical trial publications using a validation dataset of 360 articles with 129 articles that are known geographic annotations. Manual annotation by domain experts served as the gold standard. String matching provided rapid but limited coverage through exact text matching of country names. Machine learning employed named entity recognition with location-specific training data. Scientific LLaMA represented a domain-tuned large language model, while Gemma2 was a general-purpose open-source model. Perfect precision and recall (1.000) for both LLaMA and Gemma2 indicate successful identification of all geographic entities without false positives. Gemma2 was selected for full-scale deployment due to superior computational efficiency (2'56" vs 2 hours) while maintaining perfect accuracy. Two approaches were evaluated for extracting total participant enrollment from clinical trial publications using a validation dataset of 308 articles out of 360 with manually verified participant counts. Manual annotation by trained researchers established ground truth values. Gemma2 achieved near-perfect performance (precision=0.990, recall=1.000, F1=0.990) while reducing processing time from 4 hours to 2'56" per validation batch. The high precision indicates minimal false identification of participant numbers, while perfect recall demonstrates successful capture of all valid enrollment figures. Processing time represents total duration for the 308-article validation set, with Gemma2 providing a 98.8% reduction in annotation time compared to manual methods.

	Model	Approach	Precision	Recall	F1 Score	Time Cost
Geographical annotation	Model 1	Manual annotation	105/105	105/129	0.897	3 hours
	Model 2	String match	87/88	88/129	0.807	5"
	Model 3	Machine learning	58/156	156/129	0.542	2'
	Model 4	Scientific LLaMA	129/129	129/129	1.000	2 hours
	Model 5	Gemma2	129/129	129/129	1.000	2'56"
Participant extraction	Model 1	Manual annotation	295/305	301/308	0.979	4 hours
	Model 5	Gemma2	301/308	308/308	0.990	2'56"

Supplementary Tables 4. Multi-pathway mapping framework for linking MeSH terms to GBD disease categories. Clinical trial MeSH descriptors were systematically mapped to Global Burden of Disease cause names through six intermediate ontological systems. Direct mappings were prioritized where available, followed by multi-step bridging via concept unique identifiers (CUIs). MeSH tree numbers (n=64,883) served as input, with intermediate systems including Disease Ontology (DO), Orphanet Rare Disease Ontology (ORDO), Online Mendelian Inheritance in Man (OMIM), and Systematized Nomenclature of Medicine Clinical Terms (SNOMED). The framework achieved 6,595 unique mapping pathways, successfully linking 120,347 studies (62.10% of 193,806 post-2000 publications) to 280 distinct GBD cause names. Final coverage included 276 of 308 ICD-10 mapped cause names (89.61%), demonstrating comprehensive disease representation across the clinical trial literature. UMLS, Unified Medical Language System; UID, unique identifier.

RCTs items	**Intermediate**			**GBD cause names can be mapped to**	**# mapped unique UID**
MeSH tree number	UMLSUID			ICD10	2,141
	DO	MeSHCUI			1,325
		UMLSUID			1,242
		direct			1,271
	ORDO	MeSHCUI			531
		UMLSUID			538
		Direct			625
	OMIM	UMLSUID			1,158
		ORDO	UMLSUID		273
	SNOMED	ORDO			601
		ORDO	UMLSUID		731
		UMLSUID			1,975
		DO			977
		DO	UMLSUID		974
	Direct				2,228
Total distinct mapping					6,595
Total distinct mapped PMID in 193,806 studies					120,347(62.10%)
Total mapped cause names					280
Total mapped cause names in 308 cause names with ICD10					276(89.61%)

Note: Multiple mapping pathways for individual disease concepts reflect the redundancy built into the framework to maximize coverage while maintaining precision.

Supplementary Tables 5. Data sources and coverage for country-level predictors

Block	**Variables**	**Source**	**Years**	**Coverage**
Economic	GDP, GDP per capita	World Bank	2000-2024	194 countries
	Population	World Bank	2000-2024	194 countries
Research_Investment	R&D expenditure (% GDP), Total publications, Total citations	World Bank, Dimensions	2000-2024	158 countries
Health_Infrastructure	Health expenditure, Hospital beds, Hospitals, Doctors per 10,000	World Bank, WHO	2000-2024	182 countries
Governance	Human Development Index (HDI), Democracy Index	UNDP, World Governance Indicator	2000-2024	189 countries

Note: Analysis restricted to 158 countries with complete data across all predictor categories.

Supplementary Tables 6. Hierarchical variance partitioning results (structural analysis)

Block	**Variables**	**N Variables**	**Incremental R²**	**Cumulative R²**	**N Obs**
Economic	log(GDP), log(population)	2	0.331	0.331	158
Research_Investment	rd_expenditure, log(publications), total_citations	3	0.079	0.409	158
Health_Infrastructure	log(health_exp), hospital_beds, hospitals, doctors_per_10k	4	0.016	0.426	158
Governance	hdi, democracy_index	2	0.007	0.433	158

Supplementary Tables 7. Shapley value decomposition with confidence intervals (structural analysis)

Variable	**Mean %**	**CI Lower**	**CI Upper**	**SE**
log_population	25.5	15.3	36.3	5.66
log_health_exp	14.3	6.9	21.6	3.98
log_gdp	11.1	4.8	17.6	3.46
rd_expenditure	11.0	3.6	19.1	4.19
democracy_index	9.5	3.0	18.2	4.11
hdi	8.3	4.5	13.6	2.46
log_publications	7.0	3.2	12.7	2.56
doctors_per_10k	4.8	1.6	9.6	2.15
total_citations	3.8	1.6	8.5	1.86
hospital_beds	3.7	1.2	8.8	2.05
hospitals	0.9	0.3	2.1	0.49

Note: Percentages represent share of total explained variance (R^2=0.433). Bootstrap confidence intervals based on 100 iterations.

Supplementary Tables 8. Block-level Shapley contributions (structural analysis)

Block	**% Contribution**	**Key Variables**
Economic	36.6	Population (25.5%), GDP (11.1%)
Health_Infrastructure	23.6	Health expenditure (14.3%), Doctors (4.8%), Beds (3.7%)

Research_Investment	21.8	R&D expenditure (11.0%), Publications (7.0%), Citations (3.8%)
Governance	17.9	Democracy (9.5%), HDI (8.3%)

Supplementary Tables 9. Hierarchical variance partitioning results (policy-relevant analysis)

Block	**Variables**	**N Variables**	**Incremental R²**	**Cumulative R²**	**N Obs**
Research_Investment	rd_expenditure, log_publications_per_capita	2	0.062	0.062	158
Health_Infrastructure	hospital_beds_per_capita, doctors_per_10k, hospitals_per_capita, log_health_exp_per_capita	4	0.026	0.088	158
Governance	hdi, democracy_index	2	0.023	0.111	158

Note: Analysis uses residual inequality (66.9% of original variance) after controlling for GDP and population. Total R^2=0.111 represents 11.1% of residual variance explained.

Supplementary Tables 10. Shapley value decomposition for policy-relevant factors

Variable	**Mean %**	**CI Lower**	**CI Upper**	**SE**
rd_expenditure	32.2	11.9	56.9	12.1
democracy_index	19.0	3.0	41.4	10.4
hdi	9.7	2.6	28.1	6.89
hospitals_per_capita	9.5	1.9	25.9	6.48
hospital_beds_per_capita	8.8	1.6	22.5	5.64
doctors_per_10k	8.7	1.2	35.9	9.36
log_publications_per_capita	8.7	3.3	17.0	3.70
log_health_exp_per_capita	3.3	0.2	16.8	4.48

Note: Percentages represent share of policy-relevant explained variance (R^2=0.111). Values sum to 100% due to rounding.

Supplementary Tables 11. Block-level Shapley contributions (policy-relevant analysis)

Block	**% Contribution**	**Key Variables**
Research_Investment	40.9	R&D expenditure (32.2%), Publications per capita (8.7%)
Health_Infrastructure	30.3	Hospitals per capita (9.5%), Beds per capita (8.8%), Doctors per 10k (8.7%), Health exp per capita (3.3%)
Governance	28.8	Democracy index (19.0%), HDI (9.7%)

Supplementary Tables 12. Basic Network Statistics

Metric	**Value**	**Description**
Nodes	262	Country–visual factor combinations
Edges	15,065	Disease-sharing connections (weight ≥ 2)
Network Density	0.441	Proportion of possible connections

Average Degree	115.0	Average connections per node
Diameter	3	Longest shortest path between nodes
Average Path Length	1.562	Mean distance between node pairs
Connected Components	1	Network is fully connected

Supplementary Tables 13. Node Composition by Factor

Factor	Nodes	Percentage	Avg Diseases	Avg Residual
Research_Investment	145	55.3%	6.57	0.060
Governance	45	17.2%	3.76	–1.144
Multiple_Factors	48	18.3%	3.12	–1.189
Health_Infrastructure	24	9.2%	5.17	0.446

Supplementary Tables 14. Node Composition by Performance Status

Status	Nodes	Percentage	Avg Diseases	Avg Residual	Avg CIS
Over_Performing	87	33.2%	5.25	1.206	0.042
Under_Performing	154	58.8%	552	–1.247	0.008
As_Expected	21	8.0%	4.19	–0.093	0.011

*High-CIS examples detailed in Fig. 3B.

Supplementary Tables 15. Homophily and Assortativity Metrics

Metric	Value	Interpretation
Factor Homophily	0.523	52.3% of edges connect same-factor nodes
Status Homophily	0.509	50.9% of edges connect same-status nodes

Factor Assortativity	0.102	Positive = similar factors connect
Degree Assortativity	–0.253	Negative = high-degree nodes connect to low-degree
Average Clustering Coefficient	0.819	High local connectivity

Supplementary Tables 16. Community Structure (Louvain Algorithm)

Community	**Size**	**Primary Factor**	**Secondary Factor**	**Modularity Contribution**
1	117	Research_Investment (81.2%)	Multiple_Factors (9.4%)	0.048
2	111	Governance (36.0%)	Research_Investment (32.4%)	0.041
3	34	Multiple_Factors (47.1%)	Research_Investment (41.2%)	0.032
Total Modularity	**0.121**			

Supplementary Tables 17. National-Level PBR Statistics (Baseline)

Statistic	**Value**	**Notes**
Countries with valid data	172	Positive participants and DALYs
Global participants	26,519,092	Sum across all countries/diseases
Global DALYs	45,176,277,531	Sum across all countries/diseases
Median PBR	0.194164	Target for alignment scenarios
Baseline Gini coefficient	0.737255	Pre-intervention inequality
Max PBR (Denmark)	14.174	Over-participation relative to burden
Min PBR (Vanuatu)	0.000002	Under-participation relative to burden

Supplementary Tables 18. Full Alignment Scenario Results

Step	**Countries Adjusted**	**Gini Reduction (%)**	**95% CI Lower**	**95% CI Upper**	**Cumulative Reduction**
Baseline	0	0.0	0.0	0.0	0.0
Top 25%	43	42.39	33.23	50.06	42.39
Top 50%	86	70.13	61.65	77.23	70.13
Top 75%	129	91.23	83.81	97.34	91.23
All Countries	172	100.00	100.00	100.00	100.00

Supplementary Tables 19. Targeted Alignment Scenario Results

Step	**Countries Adjusted**	**Gini Reduction (%)**	**95% CI Lower**	**95% CI Upper**	**Cumulative Reduction**
Baseline	0	0.0	0.0	0.0	0.0
Top 10%	17	23.94	17.02	31.55	23.94
Top 20%	34	36.01	26.95	44.38	36.01
Top 30%	51	46.75	37.85	54.81	46.75
Top 40%	68	58.10	49.54	68.29	58.10

Supplementary Tables 20. Efficiency Comparison

Metric	**Full Alignment**	**Targeted Alignment**	**Ratio**
Final reduction (%)	100.00	58.10	–

Countries adjusted (%)	100.0	40.0	–
Reduction per 1% countries	1.000%	1.453%	1.45×
Statistical significance	p < 0.001	p < 0.001	–
95% CI difference	–	–	[1.38, 1.53]

Supplementary Tables 21. Network Metric Changes Under Interventions

Metric	Baseline	Full Alignment (Final)	Targeted Alignment (Final)	Δ (Full)	Δ (Targeted)
Network Density	0.441	0.661	0.641	+0.220	+0.200
Homophily	0.523	0.314	0.333	–0.209	–0.190
Modularity	0.121	0.085*	0.088*	–0.036	–0.033
Avg Path Length	1.562	1.312*	1.328*	–0.250	–0.234
Avg Clustering	0.819	0.901*	0.894*	+0.082	+0.075

*Note: The network statistics output doesn't show updated modularity values for this run, but density and homophily changes indicate slight differences.**Estimated based on sensitivity coefficients*

Supplementary Methods

1. AI-Assisted Information Extraction

(1) Least Guidance Approach

Prompt: "Number of participants. Total participants including both control groups and other groups. Do not overlap count when they are in different location gender outcome."

(2) Human Brain Guidance Without Examples

Prompt: "After completing the first task carefully calculate the total number of participants in the study. Look for terms like 'total' 'enrolled' or 'included'. Be precise and consider gender different groups or any special conditions described in the study. If you can't find the exact number return 'unknown'."

(3) Human Brain Guidance with Examples

Prompt: "Determine the total number of participants initially enrolled in the study. Follow these steps: Scan for explicit statements about total enrollment using words like 'total' 'enrolled' or 'included.' If not found look for group descriptions and sum their numbers. If groups are described in ratios (e.g. 38%) calculate accordingly. Focus on the initial enrollment number not the completion number. Ignore numbers referring to samples measurements or time periods. If participants are institutions or practices determine what the subject represents. If you cannot determine the exact number return 'unknown'."

(4) Geo and # in Order Combined with Examples (for Geo)

Prompt: "Carefully read the file and return three types of ALL pmid's results to me: 1.identify any location-related information such as universities hospitals institutions provinces streets or any other geographical identifiers that may indicate the area where the respondent is located. This can include any specific names of cities landmarks organizations or areas that may give clues about the respondent's location. Based on this information return only the country that you believe is most likely. If multiple countries seem possible select the one that seems most plausible based on the context provided in the text. If you are unsure provide the country that seems most plausible based on the available clues. Do not provide any additional reasoning explanations or unrelated details. If no location-related information can be found return 'unknown'. The response must only be the country name in English and must not include any other information such as regions cities or institutions. Examples: Text: 'John is a professor at the University of Oxford. He lives in a small town near London.' Output: 'United Kingdom' Text: 'The participant works at a hospital in San Francisco.' Output: 'United States' Text: 'I met a person at a café in Paris near the Eiffel Tower.' Output: 'France' Text: 'The participant didn't specify where they are from.' Output: 'unknown' 2. Number of participants. Total participants including both control groups and other groups. Do not overlap count when they are in different location gender outcome. 3. PMID itself."

(5)Geo and # in Order Combined with Examples (for #)

Prompt: "Carefully read the file and return three types of ALL pmid's results to me: 1.identify any location-related information such as universities hospitals institutions provinces streets or any other geographical identifiers that may indicate the area where the respondent is located. This can include any specific names of cities landmarks organizations or areas that may give clues about the respondent's location. Based on this information return only the country that you believe is most likely. If multiple countries seem possible select the one that seems most plausible based on the context provided in the text. If you are unsure provide the country that seems most plausible based on the available clues. Do not provide any additional reasoning explanations or unrelated details. If no location-related information can be found return 'unknown'. The response must only be the country name in English and must not include any other information such as regions cities or institutions. Examples: Text: 'John is a professor at the University of Oxford. He lives in a small town near London.' Output: 'United Kingdom' Text: 'The participant works at a hospital in San Francisco.' Output: 'United States' Text: 'I met a person at a café in Paris near the Eiffel Tower.' Output: 'France' Text: 'The participant didn't specify where they are from.' Output: 'unknown' 2. Number of participants. Total participants including both control groups and other groups. Do not overlap count when they are in different location gender outcome. 3. PMID itself."

(6) Geo and # in Reversed Order Combined with Examples (for Geo)

Prompt: "Carefully read the file and perform the following two independent tasks for ALL pmid's: Amount: Determine the total number of participants initially enrolled in the study. Follow these steps: Scan for explicit statements about total enrollment using words like 'total' 'enrolled' or 'included.' If not found look for group descriptions and sum their numbers. If groups are described in ratios (e.g. 38%) calculate accordingly. Focus on the initial enrollment number not the completion number. Ignore numbers referring to samples measurements or time periods. If participants are institutions or practices determine what the subject represents. If you cannot determine the exact number return 'unknown'. Country: Identify any location-related information such as universities hospitals institutions provinces streets or any other geographical identifiers that indicate the respondent's location. Include specific names of cities landmarks organizations or areas to infer the most plausible country. If multiple countries are possible select the one that seems most plausible based on the text. If no location-related information is found return 'unknown'. Output only the country name in English without including any additional explanations or unrelated details. Examples: Text: 'John is a professor at the University of Oxford. He lives in a small town near London.' Output: 'United Kingdom' Text: 'The participant works at a hospital in San Francisco.' Output: 'United States' Text: 'I met a person at a café in Paris near the Eiffel Tower.' Output: 'France' Text: 'The participant didn't specify where they are from.' Output: 'unknown' Output format: Country: [country name or 'unknown'] Amount: [number or 'unknown']"

(7) Geo and # in Reversed Order Combined with Examples (for #)

Prompt: [Same as above]

(8) Geo and # Separately with Examples (for Geo)

Prompt: "First identify any location-related information such as universities hospitals institutions provinces streets or any other geographical identifiers that indicate the respondent's location. Include specific names of cities landmarks organizations or areas to infer the most plausible country. If multiple countries are possible select the one that seems most plausible based on the text. If no location-related information is found return 'unknown'. Output only the country name in English without including any additional explanations or unrelated details. Examples: Text: 'John is a professor at the University of Oxford. He lives in a small town near London.' Output: 'United Kingdom' Text: 'The participant works at a hospital in San Francisco.' Output: 'United States' Text: 'I met a person at a café in Paris near the Eiffel Tower.' Output: 'France' Text: 'The participant didn't specify where they are from.' Output: 'unknown' Amount: After completing the first task read the text again to determine the total number of participants initially enrolled in the study. Follow these steps: Scan for explicit statements about total enrollment using words like 'total' 'enrolled' or 'included.' If not found look for group descriptions and sum their numbers. If groups are described in ratios (e.g. 38%) calculate accordingly. Focus on the initial enrollment number not the completion number. Ignore numbers referring to samples measurements or time periods. If participants are institutions or practices determine what the subject represents. If you cannot determine the exact number return 'unknown'. Output format: Country: [country name or 'unknown'] Amount: [number or 'unknown']"

(9) Geo and # Separately with Examples (for #)

Prompt: [Same as above]

2. Medical Concept Harmonization

2.1 Background and limitations of existing disease mapping approaches

Harmonizing disease concepts across biomedical classification systems remains a longstanding challenge. Clinical coding systems such as the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) are optimized for healthcare documentation and billing, whereas epidemiological frameworks such as the Global Burden of Disease (GBD) cause hierarchy are designed to aggregate etiologically related conditions for population-level analysis. As a result, direct correspondence between ICD codes and GBD cause names is often incomplete or ambiguous, reflecting differences in granularity, conceptual scope, and intended use.

Prior studies and infrastructure projects have documented these limitations. Large-scale mapping efforts between ICD, SNOMED CT, and other terminologies maintained within the Unified Medical Language System (UMLS) frequently yield partial rather than exhaustive alignments, particularly for complex, multi-system, or non-specific disease entities. Natural language processing and ontology-alignment approaches have similarly reported constrained coverage, with substantial proportions of concepts remaining unmapped or requiring manual adjudication. These constraints are not methodological shortcomings of individual studies but instead reflect structural incompatibilities between classification systems.

Consistent with this literature, no authoritative or universally accepted mapping between ICD-10 and GBD cause names currently exists. Consequently, disease harmonization in large-scale bibliometric or trial-based analyses requires explicit methodological choices that balance coverage, semantic validity, and reproducibility.

2.2 Overview of the harmonization strategy

Given the absence of a canonical ICD–GBD mapping, we developed a conservative, ontology-mediated harmonization framework to link clinical trial publications to GBD cause categories. The guiding principle of this framework was to maximize semantic fidelity and auditability rather than to maximize coverage through probabilistic inference.

Disease assignment was conducted at the level of individual publications (PMIDs). A publication was considered successfully harmonized if it could be assigned at least one valid GBD cause name through supported ontology mappings. Publications were permitted to map to multiple GBD causes when justified by their underlying disease annotations; no attempt was made to force a single “primary” disease designation.

2.3 Source annotations and preprocessing

Clinical trial publications were first associated with Medical Subject Headings (MeSH) terms curated by the National Library of Medicine. MeSH descriptors related to diseases and disorders served as the primary disease annotations for each publication.

To support downstream harmonization, MeSH terms were normalized and linked to corresponding concept identifiers across multiple biomedical ontologies using publicly available cross-references and UMLS Concept Unique Identifiers (CUIs), where applicable. MeSH terms that were purely procedural, methodological, or population descriptors were excluded from disease mapping.

2.4 Ontologies and intermediary resources

Disease concept harmonization leveraged the following established biomedical ontologies and terminological resources:

- MeSH (Medical Subject Headings)
- ICD-10-CM (International Classification of Diseases, Tenth Revision, Clinical Modification)
- SNOMED CT (United States Edition)
- Disease Ontology (DO)
- OMIM (Online Mendelian Inheritance in Man)
- Orphanet Rare Disease Ontology (ORDO)

These resources were accessed through a combination of UMLS cross-links and ontology repositories (e.g., NCBO BioPortal). No single ontology was treated as authoritative; instead, each served as a potential intermediary (“median”) for reconciling disease concepts expressed at different levels of specificity.

2.5 Construction of median-assisted mapping paths

Rather than attempting direct ICD-to-GBD matching, we constructed multi-step mapping paths that preserved conceptual identity across ontologies. Permissible paths included, but were not limited to:

- MeSH → Disease Ontology → ICD-10-CM → GBD cause
- MeSH → SNOMED CT → ICD-10-CM → GBD cause
- MeSH → OMIM / Orphanet → Disease Ontology → GBD cause

Mappings were retained only when intermediate concepts represented the same disease entity rather than a broader category, symptom cluster, or administrative grouping. Lexical similarity alone was not sufficient for accepting a mapping. When multiple mapping paths existed, all valid paths were retained to support transparency and reproducibility.

All accepted mappings between disease concepts and GBD cause names are enumerated in Supplementary Table 4, which serves as the authoritative record of the harmonization framework used in this study.

2.6 Exclusion of non-specific and “garbage” codes

ICD-10 codes classified as non-specific, ill-defined, or designated as “garbage codes” within the GBD framework were excluded from disease harmonization. These codes do not correspond to etiologically meaningful disease entities and can introduce bias when attributing publications to disease categories. Exclusion of such codes was treated as a precision-preserving step rather than data loss.

2.7 Coverage outcome and characterization

Using the median-assisted ontology framework, 62.1% of trial-linked publications could be assigned at least one valid GBD cause name. This coverage metric is reported at the PMID level and does not imply exhaustive coverage of all disease concepts or ICD codes present in the dataset.

Unmapped publications were disproportionately associated with:

- Non-specific symptom-based MeSH annotations
- Multi-morbidity or prevention-focused trials

- Administrative or procedural study designs
- Disease concepts spanning multiple GBD categories without a dominant etiological focus

These patterns are consistent with previously reported limitations in cross-ontology disease harmonization.

2.8 Evaluation of alternative approaches

To assess whether disease harmonization coverage could be increased beyond the ontology-mediated framework without compromising semantic validity, we empirically evaluated two alternative strategies: large language model (LLM)–assisted disease matching and supervised deep learning–based disease classification. Both approaches were implemented and tested on the same underlying disease annotation data used in the primary harmonization pipeline.

2.8.1 LLM-assisted disease concept matching

LLM-assisted matching was evaluated as a means of expanding disease coverage by inferring correspondences between ICD- and MeSH-derived disease concepts and GBD cause names using semantic reasoning. In term-level evaluations, this approach was able to assign candidate GBD causes to nearly all MeSH disease descriptors, effectively achieving complete nominal coverage at the concept level.

However, when evaluated at the publication (PMID) level, this apparent gain did not translate into meaningful harmonization. Empirically, LLM-assisted matching mapped only a small fraction of distinct PMIDs to disease–cause pairs that were consistent with manually curated labels (recall ≈ 4–5% in direct PMID-level validation), despite generating millions of candidate mappings overall. This discrepancy reflected extensive many-to-many expansion: individual publications were frequently assigned large numbers of GBD causes, with some PMIDs associated with dozens of distinct causes in a single run.

Further analysis showed that these assignments were unstable across repeated executions under identical prompts, indicating sensitivity to stochastic generation. Qualitatively, the model tended to overgeneralize from symptom-based or system-level disease descriptions, collapsing heterogeneous clinical concepts into specific GBD causes without explicit etiological justification. As a result, increases in nominal coverage were driven primarily by semantic broadening rather than preservation of disease identity.

Given the low PMID-level validity, high mapping multiplicity, and limited reproducibility observed in practice, LLM-assisted matching was not considered suitable for disease harmonization in this study, where false-positive attribution at the disease level would bias downstream structural analyses.

2.8.2 Deep learning–based disease classification

We also evaluated supervised deep learning models trained to predict GBD cause categories from ICD-derived disease representations. Models were implemented in a multi-label classification setting and evaluated using standard performance metrics.

Across configurations, model performance remained modest. Representative models achieved micro-averaged F1 scores on the order of 0.35–0.40, with substantially lower performance for rare and boundary-spanning disease categories. Increases in apparent coverage were highly sensitive to decision thresholds, with relaxed thresholds inflating the number of predicted disease assignments per publication without corresponding improvements in semantic accuracy.

Error analysis indicated that predictions were dominated by disease frequency priors: common GBD causes were preferentially assigned across heterogeneous ICD profiles, while rare diseases were frequently misclassified or omitted. This behavior reflects the fact that supervised classifiers learn statistical regularities in the training data but do not resolve the underlying conceptual mismatch between ICD-based clinical coding and GBD epidemiological categories.

Because this approach substitutes explicit disease harmonization with a black-box statistical approximation, it was deemed insufficient for analyses requiring interpretable and auditable disease assignments at the publication level.

2.8.3 Rationale for methodological choice

Although both LLM-assisted and deep learning–based approaches increased nominal disease coverage under certain configurations, these gains were achieved at the cost of reproducibility, interpretability, and semantic control. In contrast, the median-assisted ontology harmonization framework yielded lower but well-defined PMID-level coverage (62.1%) with explicit, traceable mappings between disease concepts and GBD causes.

Given the study’s focus on systematic patterns of disease representation across the global clinical trial literature, we therefore prioritized conservative, ontology-grounded harmonization over higher-coverage methods that introduce uncontrolled uncertainty.

3. Inequality Measurement

(1) Participation-to-Burden Ratio (PBR)

To assess alignment between research participation and disease burden, we calculated the Participation-to-Burden Ratio (PBR) for each country-disease-year combination. PBR quantifies whether a population's contribution to research on a specific disease is proportional to its share of global burden from that disease:

$$PBR_{c,d,t} = \frac{\frac{P_{c,d,t}}{P_{\blacksquare,d,t}}}{\frac{B_{c,d,t}}{B_{\blacksquare,d,t}}}$$

where $P_{c,d,t}$ represents trial participants from country c for disease d in year t, $B_{c,d,t}$ represents DALYs for that country-disease-year, and subscript ■ denotes global totals. PBR > 1 indicates over-representation (participation exceeds burden share); PBR < 1 indicates under-representation. Importantly, PBR is symmetric between country and disease dimensions—it can be equivalently interpreted as measuring country over/under-representation for a given disease or disease over/under-representation within a given country.

(2) Specialization Index (SI)

To characterize whether countries or diseases concentrate research efforts in particular domains, we calculated the Specialization Index:

$$SI_{c,d} = \frac{\frac{P_{c,d}}{P_{c,\blacksquare}}}{\frac{P_{\blacksquare,d}}{P_{\blacksquare,\blacksquare}}}$$

where $P_{c,d}$ represents participants from country c studying disease d, $P_{c,\blacksquare}$ is country c's total participants across all diseases, $P_{\blacksquare,d}$ is global participants for disease d, and $P_{\blacksquare,\blacksquare}$ is global participants across all diseases. SI > 1 indicates specialization (country devotes relatively more effort to that disease than the global average); SI < 1 indicates de-emphasis.

(3) Gini Coefficient

To quantify overall inequality in research participation distribution, we calculated Gini coefficients using the standard formulation. For a distribution of PBR values across n country-disease pairs (ordered from smallest to largest):

$$G = \frac{\sum_{i=1}^{n}(2i - n - 1) \cdot PBR_i}{n\sum_{i=1}^{n} PBR_i}$$

Gini ranges from 0 (perfect equality) to 1 (maximum inequality). We calculated Gini coefficients for: (1) participant distribution across countries for each disease; (2) participant distribution across diseases for each country; and (3) global participant distribution across all country-disease pairs.

(4) Contribution-to-Inequality Score (CIS)

To quantify individual diseases' contributions to global inequality, we employed leave-one-out analysis. For each disease dd d, we calculated the Contribution-to-Inequality Score (CIS):

$$CIS = \frac{G_{all} - G_{-d}}{G_{all}} \times 100\%$$

where G_{all} is the global Gini coefficient calculated across all country-disease pairs, and G_{-d} is the Gini coefficient after excluding all country-pairs involving disease d. Positive CIS indicates that disease d contributes to inequality (removing it reduces inequality); negative CIS indicates the disease actually reduces inequality (removing it increases inequality). We computed CIS values using both equal-weighted (each country-disease pair counted equally) and participant-weighted (weighted by participant numbers) approaches.

(5) Lorenz Curve Analysis

To visualize inequality and assess the collective impact of top driver diseases, we constructed Lorenz curves plotting cumulative share of participants against cumulative share of DALYs across country-disease pairs, ordered by PBR. We compared: (1) all diseases included; (2) top diseases by absolute CIS excluded. The area between the Lorenz curve and the line of perfect equality equals the Gini coefficient.

(6) Theil Index Decomposition (Disease-Grouped)

The Theil entropy index quantifies inequality while allowing additive decomposition into between-group and within-group components. We calculated:

$$T = \sum_{i=1}^{N} \frac{PBR_i}{PBR} ln\, ln\, (\frac{PBR_i}{PBR})$$

We then decomposed T by grouping observations by disease:

$$T_{total} = T_{between-disease} + T_{within-disease}$$

where:

- $T_{between-disease}$ captures inequality arising from differences in disease-average PBR values (i.e., whether some diseases have uniformly higher or lower participation-to-burden ratios across all countries)
- $T_{within-disease}$ captures inequality arising from variation in PBR across countries within each disease (i.e., geographic disparities in participation for any given disease)

(7) Theil Index Decomposition (Country-Grouped)

To test robustness and avoid bias from grouping choice, we repeated the Theil decomposition grouping observations by country rather than disease:

$$T_{total} = T_{between-country} + T_{within-country}$$

where:

- $T_{between-country}$ captures inequality from differences in country-average PBR values (i.e., whether some countries universally over- or under-contribute relative to burden)

- $T_{within-country}$ captures inequality from variation in PBR across diseases within each country (i.e., whether countries specialize in particular disease portfolios)

(8) Variance Partitioning (Two-Way Decomposition)

To simultaneously estimate country and disease contributions without imposing a grouping structure, we employed variance partitioning using a two-way fixed effects model:

$$PBR_{c,d,t} = \mu + \alpha_c + \beta_d + \gamma_t + \epsilon_{c,d,t}$$

where α_c represents country fixed effects, β_d represents disease fixed effects, γ_t represents year fixed effects, and $\epsilon_{c,d,t}$ represents residual variation. We calculated the proportion of total variance explained by each component using partial R² values:

$$R^2_{country} = \frac{SS_\alpha}{SS_{total}}, R^2_{disease} = \frac{SS_\beta}{SS_{total}}, R^2_{year} = \frac{SS_\gamma}{SS_{total}}$$

This approach provides an unbiased, symmetric quantification of country versus disease contributions to inequality, avoiding the grouping-dependence inherent in Theil decomposition.

(9) Temporal Trend Analysis

To assess the evolution of inequality over time, we calculated all inequality metrics (Gini, Theil components, CIS values) in 2-year bins from 2000-2024. Linear regression models estimated temporal trends:

$$Metric_t = \beta_0 + \beta_1 \cdot Year_t + \epsilon_t$$

We report trend slopes β_1, R², and p-values. Bootstrapping (1,000 iterations with resampling at the country-disease-year level) generated 95% confidence intervals for all trend estimates to account for sampling variability and temporal autocorrelation.

(10) Analytical Strategy to Avoid Bias

To prevent analytical choices from biasing results toward either hypothesis, we employed:

Symmetric measurement: Our primary metric, the PBR is mathematically symmetric between country and disease dimensions. PBR can be equivalently calculated as "country *c*'s share of participants in disease *d* relative to country *c*'s share of disease *d* burden" or as "disease *d*'s share of participants from country *c* relative to disease *d*'s share of burden in country *c*." This symmetry ensures the metric itself does not favor country-level or disease-level explanations.

Bidirectional decomposition: For inequality decomposition using the Theil index, we conducted decompositions grouping by disease (yielding between-disease vs. within-disease components) and grouping by country (yielding between-country vs. within-country components). This bidirectional approach allows us to assess whether conclusions depend on arbitrary grouping choices.

Model-based variance partitioning: In addition to group-based decomposition, we employed two-way fixed effects models to simultaneously estimate variance attributable to country, disease, and their interaction without imposing a grouping structure. This provides an unbiased quantification of relative contributions.

Power and Precision: Post-hoc power analysis confirmed adequate sample size to detect small effect sizes (Cohen's $d \geq 0.3$) in inequality comparisons with 80% power at $\alpha=0.05$. For decomposition analyses, Monte Carlo simulations (1,000 iterations) demonstrated that our sample size provided stable estimates of variance components with 95% confidence intervals spanning less than ±5 percentage points for major components (country, disease).

Ethical and Reporting Considerations: This study analyzed published, de-identified bibliometric data and public disease burden estimates, requiring no institutional review board approval.

4. Structural Factors Decomposition

4.1 Analytical Framework

To identify structural factors of research participation inequality, we employed a two-part analytical strategy distinguishing fixed structural characteristics from modifiable policy levers.

Part 1 (Structural Analysis) examined all country-level predictors to quantify the relative importance of economic, research, health, and governance factors in explaining baseline inequality. This analysis used absolute measures (GDP, population, total publications) to represent countries' fundamental capacities.

Part 2 (Policy-Relevant Analysis) isolated modifiable factors by: (1) converting absolute measures to per-capita rates (e.g., publications per capita, hospitals per capita) to capture efficiency rather than scale; (2) computing residual inequality after statistically controlling for GDP and population using ordinary least squares regression; (3) analyzing only the residual variance using per-capita measures of research investment, health infrastructure, and governance quality. This approach separates baseline structural inequality (unchangeable in the short term) from policy-relevant inequality (addressable through targeted interventions).

4.2 Hierarchical Variance Partitioning

We employed hierarchical linear regression to quantify the incremental contribution of predictor blocks to explained variance. Predictor blocks were entered sequentially:

Part 1 (Structural):

1. Economic: log(GDP), log(population)
2. Research: R&D expenditure (% GDP), log(publications), total citations
3. Health: log(health expenditure), hospital beds, hospitals, doctors per 10,000
4. Governance: HDI, democracy index

Part 2 (Policy-Relevant):

1. Research Investment: R&D expenditure (% GDP), log(publications per capita)
2. Health Infrastructure: hospital beds per capita, doctors per 10,000, hospitals per capita, log(health expenditure per capita)
3. Governance: HDI, democracy index

For each block, we calculated:

- Cumulative R^2: Variance explained by all blocks up to and including the current block
- Incremental R^2: Additional variance explained by the current block beyond previous blocks

Block ordering followed theoretical priority: structural factors (economic, research capacity) precede downstream factors (health infrastructure, governance). Sensitivity analyses with alternative orderings produced qualitatively similar results.

Missing Data Treatment: We imputed missing predictor values using median imputation (sklearn.impute.SimpleImputer, strategy='median') to avoid listwise deletion bias. Predictors with >50% missing data were excluded. All continuous predictors were standardized (mean=0, SD=1) before analysis to enable coefficient comparison.

4.3 Shapley Value Decomposition

To address limitations of hierarchical partitioning (order-dependence, inability to capture interaction effects), we employed Shapley value decomposition—a game-theoretic approach that fairly attributes variance to each predictor by averaging marginal contributions across all possible predictor orderings.

For each predictor *i*, the Shapley value φ_i represents its average marginal contribution to R^2 across all possible subsets of predictors:

$$\varphi_i = (1/P!) \times \Sigma[R^2(S \cup \{i\}) - R^2(S)]$$

where S represents all possible subsets of predictors excluding *i*, and P is the total number of predictors.

We implemented this using permutation-based approximation:

1. Generate *N* random orderings of predictors (N=100 for Part 1, N=100 for Part 2)
2. For each ordering, compute each predictor's marginal R^2 contribution (R^2 with predictor included minus R^2 without)
3. Average marginal contributions across all orderings
4. Report percentage of total explained variance attributable to each predictor

Bootstrap Confidence Intervals: To quantify estimation uncertainty, we computed 95% confidence intervals using 100 bootstrap iterations. For each iteration, we resampled countries with replacement, recalculated Shapley values using 50 permutations (reduced for computational efficiency), and extracted the 2.5th and 97.5th percentiles of the bootstrap distribution.

Block-Level Aggregation: We aggregated individual predictor Shapley values to predictor blocks by summing Shapley values for all predictors within each block, yielding block-level percentage contributions.

4.4 Dependent Variable

Part 1: Log-transformed participation-to-burden ratio (log-PBR) at the country level, computed as:

$$\text{log-PBR\_c} = \log[(\Sigma \text{ participants\_c,d}) / (\Sigma \text{ DALYs\_c,d})]$$

where the sum is across all 16 disease categories for each country c.

Part 2: Residual inequality after controlling for structural factors:

$$\text{Residual\_c} = \text{log-PBR\_c} - [\beta_0 + \beta_1 \cdot \log(\text{GDP\_c}) + \beta_2 \cdot \log(\text{Population\_c})]$$

where β_0, β_1, β_2 were estimated using ordinary least squares. This residual capture inequality is not explained by fundamental structural factors.

For Part 2, we created per-capita measures:

- Publications per capita = Total publications / Population
- Hospitals per capita = Number of hospitals / Population
- Hospital beds per capita = Total beds / Population

- Health expenditure per capita = Total health expenditure / Population
- GDP per capita = GDP / Population

All per-capita measures were log-transformed to address right-skewed distributions.

5. Intervention Methodological Framework

5.1 National-Level PBR Calculation

Data Aggregation: Country-level PBR values were calculated by summing trial participants and DALYs across all 16 diseases for each of the 172 countries with valid data (2000-2024). The global median PBR was 0.194, with a baseline Gini coefficient of 0.870.

Validation Against Disease-Specific Analysis: We compared national-level PBR calculations with disease-specific PBR data from our temporal analysis (Fig. 3). The correlation between national and disease-aggregated metrics was high ($r=0.94$, $p<0.001$), confirming that national aggregation captures the same structural patterns observed at the disease level.

CIS is calculated using the same logic as the disease. However, a leave-one-out contribution to inequality score (CIS) is informative for disease categories, its direct application to countries is dominated by scale effects, as large research-producing countries mechanically induce larger changes in the global Gini coefficient when removed. We therefore report raw country CIS only for diagnostic completeness. For structural inference, we rely on two complementary approaches: (i) a DALY-normalized CIS that rescales inequality changes by epidemiological burden, and (ii) a Shapley value decomposition of the global Gini coefficient, which estimates each country's expected marginal contribution averaged across all country coalitions. The Shapley-based results are robust, interpretable in percentage terms, and consistent across alternative specifications.

Complete National PBR Dataset: Available in https://doi.org/10.5281/zenodo.18115243, containing for each country: ISO3 code, total participants, total DALYs, PBR value, participant share, and DALY share.

5.2 Network Evolution Modeling

We estimated how network metrics change with reductions in structural inequality by analyzing historical correlation patterns (2000-2024). Using linear regression: $\Delta M=\beta M\times\Delta \text{Gini}+\epsilon$ where ΔM = change in network metric, ΔGini = change in Gini coefficient. Estimated coefficients:

Table S1 Estimated coefficients

Metric	β coefficient	R²	p-value
Network Density	0.220	0.87	<0.001
Homophily	-0.210	0.82	<0.001
Modularity	-0.042	0.65	0.003
Average Path Length	-0.085	0.71	0.001

5.3 Intervention Simulation Algorithm:

For each intervention step (corresponding to Gini reduction ΔG):

1. Calculate new network metrics: $M_{new}=M_{baseline}+\beta M\times\Delta G$
2. Apply constraints: $0\leq \text{Density}\leq 1$, $0\leq \text{Homophily}\leq 1$
3. Add random noise proportional to baseline uncertainty (10% at step 0, increasing linearly to 15% at final step)

Intervention effects were estimated using bootstrap resampling (200 iterations). Reported values represent results from a single representative run with random seed set to 42 for reproducibility. Sensitivity analyses across multiple runs showed efficiency ratios ranging from 1.40–1.46×.

5.4 Statistical Tests and Uncertainty Estimation

Bootstrap Confidence Intervals: All intervention effects and network metric changes include 95% confidence intervals calculated from: 200 bootstrap samples for intervention effects and 100 parametric bootstrap samples for network metrics.

Statistical Significance Tests: Gini reduction significance: For each intervention scenario, we tested whether post-intervention Gini distributions differed from baseline using paired t-tests across bootstrap samples. Both scenarios showed significant reductions ($p<0.001$). Strategy comparison: We compared Full vs Targeted alignment using: Difference in mean Gini reduction with 95% CI; Paired t-test on bootstrap samples; Efficiency ratio with bootstrap confidence interval. Network metric changes: We tested whether final network metrics differed significantly from baseline using permutation tests (10,000 permutations).

6. Representativeness analysis

We conducted a comprehensive representativeness analysis of randomized controlled trial (RCT) datasets to evaluate whether progressive filtering and subsetting procedures introduced systematic bias. Schematic representation of sequential filtering procedures applied to the initial corpus of 301,262 randomized controlled trials (n=301,262; 1980-2024). Temporal restriction to 2000-2024 yielded 193,806 studies (n=193,806; 2000-2024) to address geographic metadata limitations in earlier publications. Subsequent filtering for geographic annotation (GeoFSub n=144,084 and GeoTSub n=99,049) and disease mapping (DisTSub n=120,347 and DisGeoSub n=62,654) produced analytically focused datasets. The final dataset (DisGeoSub) represents studies with complete geographic, participant, and disease information.

We employed Cramér's V as the primary effect size measure to assess distributional differences between parent and subset datasets. Unlike traditional chi-square tests that can yield misleading statistical significance with large sample sizes, Cramér's V provides interpretable effect sizes independent of sample size. Values were interpreted as: <0.1 (very small/highly representative), 0.1-0.3 (small/representative), 0.3-0.5 (medium/moderately representative), and >0.5 (large/not representative).

Four analytical domains were examined:

Temporal Distribution: Publication years were analyzed to assess whether filtering procedures altered the temporal representation of studies. Year coverage percentages and distributional patterns were compared between parent and subset datasets.

Geographic Distribution: Author affiliation countries (ISO3 codes) were analyzed to evaluate geographic representativeness. Country coverage percentages and regional distribution patterns were assessed.

Journal Coverage: Publication venues were examined through journal titles and categories to determine whether subsetting affected disciplinary representation.

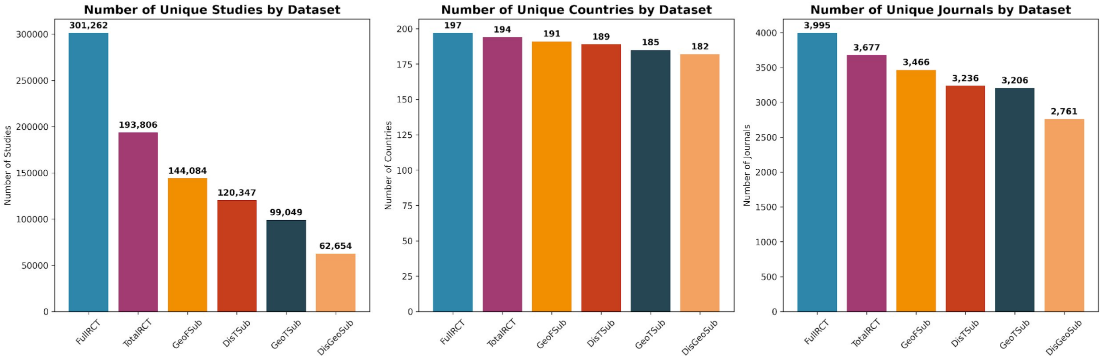
Research Content: MeSH (Medical Subject Headings) tree numbers were used to assess topical representativeness across high-level research categories.

Effect Size Calculation: For each comparison, we constructed contingency tables and calculated Cramér's V using the formula: $V = \sqrt{(\varphi^2\text{corr} / \min(\text{kcorr}-1, \text{rcorr}-1))}$

where φ^2corr represents the corrected phi-squared value and kcorr, rcorr are bias-corrected row and column counts. Percentage differences were calculated as absolute differences between baseline and subset distributions.

1
2
3
4
5
6

6.1 Dataset overview

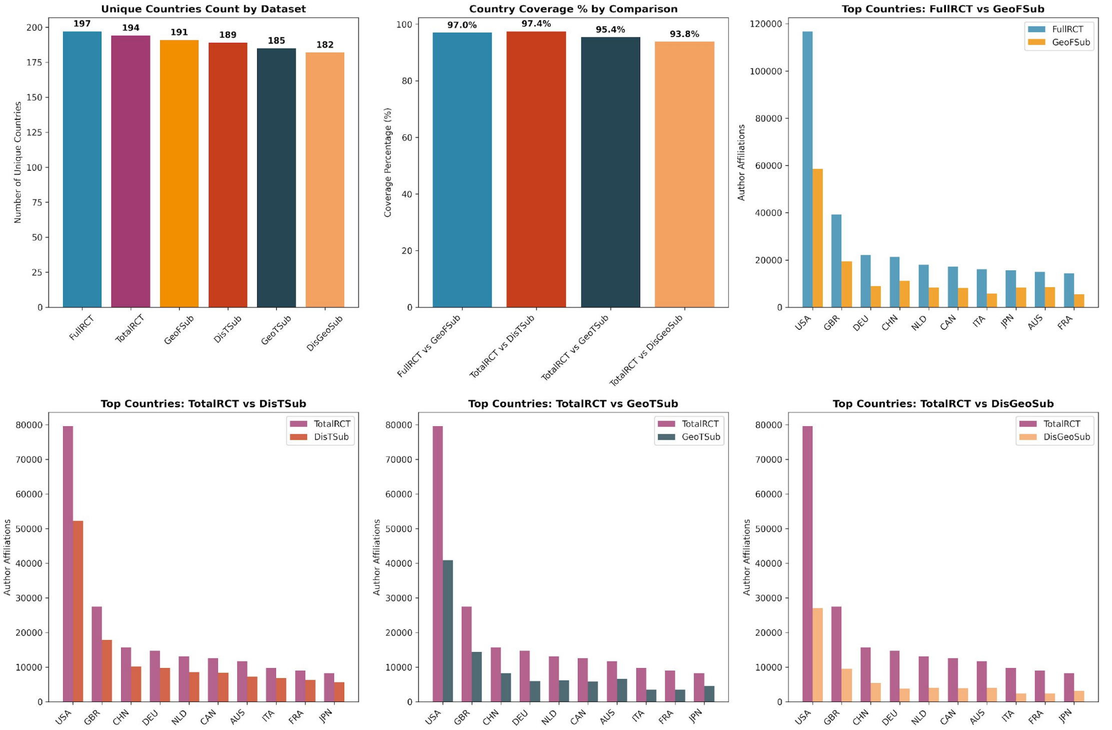


Supplementary Methods Fig. 1. Distinct numbers of studies, author affiliated countries, and journals in all six datasets. All 16 comparisons across four analytical domains demonstrated very small effect sizes (Cramér's V < 0.1), indicating that subset datasets were highly representative of their parent populations. No comparison exceeded the threshold for even small effect sizes (V ≥ 0.1).

6.2 Publication analysis

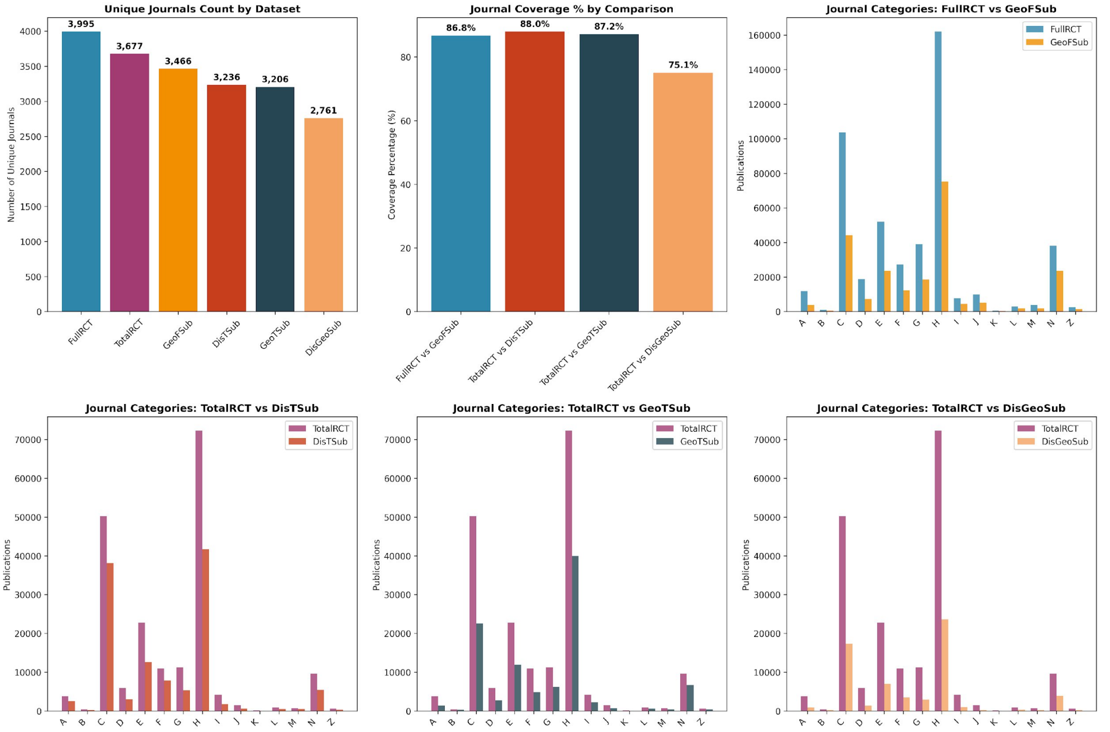
Supplementary Methods Fig. 2. Temporal representativeness across dataset filtering procedures. Year coverage was consistently perfect (100%) across all comparisons, with all unique publication years represented in subset datasets. Comparative analysis of publication year distributions between parent datasets and filtered subsets using Cramér's V effect size measures. Effect sizes were uniformly very small: FullRCT vs GeoFSub (V=0.0998), TotalRCT vs DisTSub (V=0.0247), TotalRCT vs GeoTSub (V=0.0237), and TotalRCT vs DisGeoSub (V=0.0207). Maximum percentage differences ranged from 0.38% to 1.18%, with mean differences consistently below 0.3%.

6.3 Author analysis



Supplementary Methods Fig.3. Geographic representativeness of author affiliations across dataset filtering. Country coverage ranged from 93.8% to 97.9%, demonstrating strong geographic retention across filtering procedures. Effect sizes remained very small across all comparisons: FullRCT vs GeoFSub (V=0.0825), TotalRCT vs DisTSub (V=0.0241), TotalRCT vs GeoTSub (V=0.0805), and TotalRCT vs DisGeoSub (V=0.0861). Maximum percentage differences were modest (0.24%-1.77%), with mean differences below 0.1%.

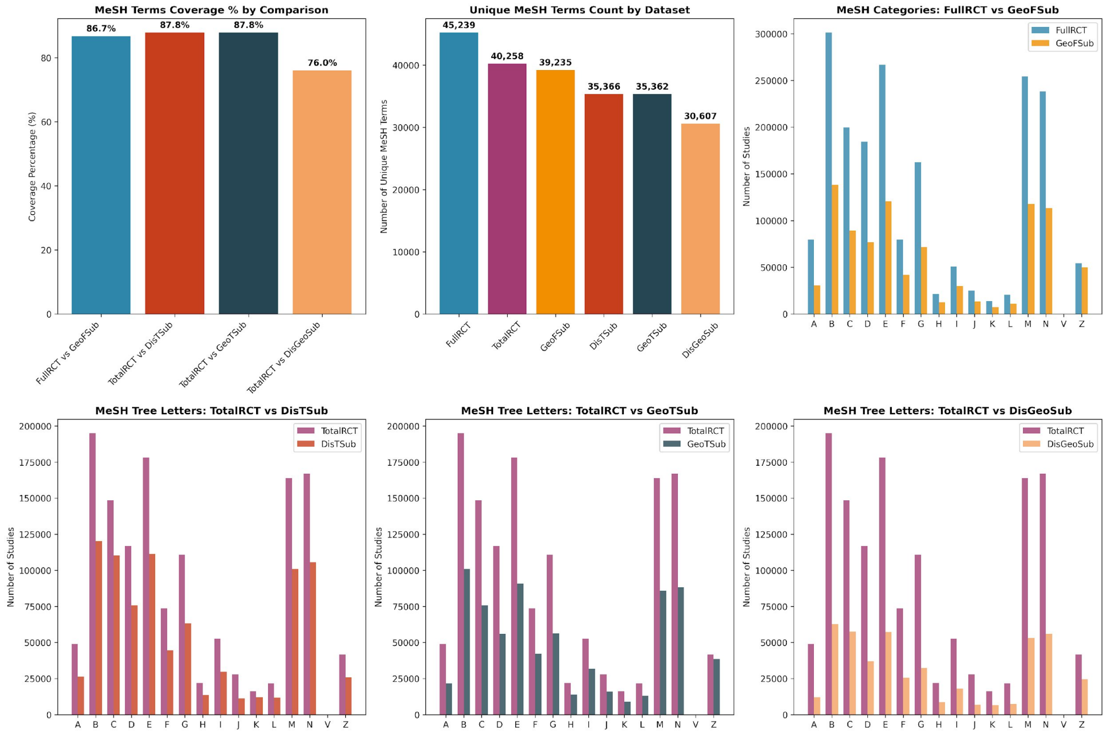
6.4 Journal analysis



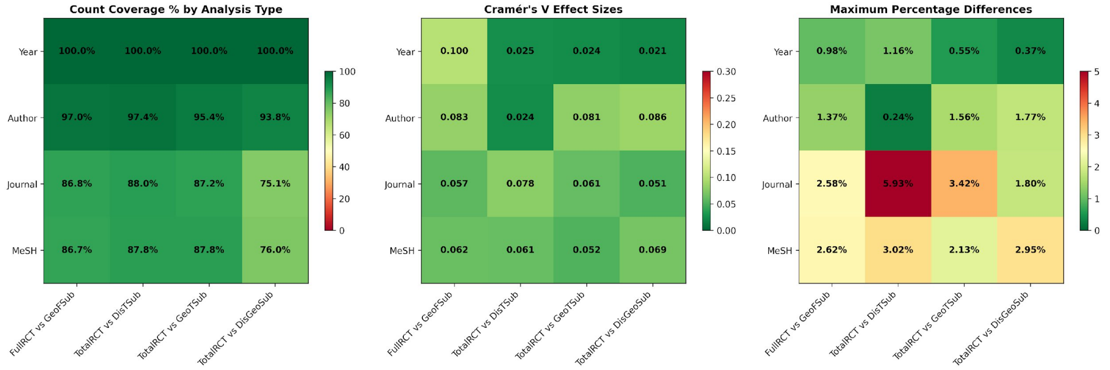
Supplementary Methods Fig.4. Journal and disciplinary representativeness

following dataset filtering. Journal coverage showed the greatest variation, ranging from 78.7% to 91.6%, yet all comparisons maintained very small effect sizes. Journal category coverage remained perfect (100%) across all datasets. Effect sizes were: FullRCT vs GeoFSub (V=0.0565), TotalRCT vs DisTSub (V=0.0778), TotalRCT vs GeoTSub (V=0.0608), and TotalRCT vs DisGeoSub (V=0.0506). Despite larger absolute coverage differences, maximum percentage differences in category distributions remained manageable (1.80%-5.93%).

6.5 MeSH term analysis



Supplementary Methods Fig.5. Research content representativeness through MeSH term analysis. MeSH term coverage ranged from 76.0% to 87.8%, with perfect retention of high-level research categories (100% coverage). Effect sizes were consistently very small: FullRCT vs GeoFSub (V=0.0610), TotalRCT vs DisTSub (V=0.0511), TotalRCT vs GeoTSub (V=0.0523), and TotalRCT vs DisGeoSub (V=0.0689). Maximum percentage differences in topical distributions ranged from 2.13% to 3.12%.



Count Coverage % by Analysis Type

	FullRCT vs GeoFSub	TotalRCT vs DisTSub	TotalRCT vs GeoTSub	TotalRCT vs DisGeoSub
Year	100.0%	100.0%	100.0%	100.0%
Author	97.0%	97.4%	95.4%	93.8%
Journal	86.8%	88.0%	87.2%	75.1%
MeSH	86.7%	87.8%	87.8%	76.0%

Cramér's V Effect Sizes

	FullRCT vs GeoFSub	TotalRCT vs DisTSub	TotalRCT vs GeoTSub	TotalRCT vs DisGeoSub
Year	0.100	0.025	0.024	0.021
Author	0.083	0.024	0.081	0.086
Journal	0.057	0.078	0.061	0.051
MeSH	0.062	0.061	0.052	0.069

Maximum Percentage Differences

	FullRCT vs GeoFSub	TotalRCT vs DisTSub	TotalRCT vs GeoTSub	TotalRCT vs DisGeoSub
Year	0.98%	1.16%	0.55%	0.37%
Author	1.37%	0.24%	1.56%	1.77%
Journal	2.58%	5.93%	3.42%	1.80%
MeSH	2.62%	3.02%	2.13%	2.95%

Supplementary Methods Fig.6. Summary of representativeness across analytical domains. Domain-level analysis revealed consistent patterns: Temporal: Mean

Cramér's V = 0.0422, average count coverage = 100%. Geographic: Mean Cramér's V = 0.0683, average count coverage = 95.9%. Journal: Mean Cramér's V = 0.0614, average count coverage = 84.3%. Content: Mean Cramér's V = 0.0611, average count coverage = 84.6%. All domains demonstrated average count coverage above 84% and mean effect sizes well below the 0.1 threshold for representativeness concerns. Sequential filtering showed minimal cumulative bias. The most restrictive filtering (TotalRCT to DisGeoSub, representing 32.63% total coverage) maintained very small effect sizes across all domains, with the largest effect size being V=0.0566 for geographic distribution—still well within the highly representative range.

This comprehensive representativeness analysis provides robust evidence that progressive filtering and subsetting procedures did not introduce systematic bias across temporal, geographic, disciplinary, or content dimensions. All subset datasets demonstrated high representativeness of their parent populations, with effect sizes consistently in the "very small" category according to established interpretive guidelines. These findings support the validity of using these filtered datasets for epidemiological and systematic research purposes without concerns about selection bias affecting generalizability.

The consistent pattern of very small effect sizes across diverse analytical domains suggests that the filtering criteria were effectively random with respect to the measured characteristics, preserving the essential distributional properties of the original datasets. This methodological validation strengthens confidence in subsequent analyses conducted using these refined datasets.