Mapping Inequities in Global Vaccine Sentiment Research

Duilio Balsamo [1,2], Vittoria Offeddu [1], Zhina Aghamohammadi [1], Chiara Chiavenna [1,2], Laura P. Leone [1,3,4], Elena D'Agnese [1,5,6], Deepak Sharma [7], Aleksandra Torbica [1,8], Soheil Shayegh [9,10], Javier Andreu-Perez [11], Alessia Melegaro [1,2*]

1. DONDENA Centre for Research on Social Dynamics and Public Policy, Bocconi University, Milan, Italy

2. BIDSA Bocconi Institute for Data Science and Analytics, Bocconi University, Milan, Italy

3. Nuffield College, University of Oxford, Oxford, UK

4. Department of Sociology, University of Oxford, Oxford, UK

5. Department of Economics and Management, University of Pisa, Pisa, Italy

6. Department of Statistics, Computer Science, Applications, University of Florence, Florence, Italy

7. Department of Computer Science, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

8. CERGAS Centre for Research on Health and Social Care Management, Bocconi University, Milan, Italy

9. RFF-CMCC European Institute on Economics and the Environment, Milan, Italy

10. CMCC Foundation - Euro-Mediterranean Center on Climate Change, Milan, Italy

11. Centre for Computational Intelligence, School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

* Corresponding author: Alessia Melegaro (alessia.melegaro@unibocconi.it, +39 347 6724239)

Abstract

Introduction: Negative public sentiment towards vaccination (PSV) poses significant challenges to the effectiveness of immunization programs, with dramatic effects on morbidity and mortality for vaccine-preventable diseases. Yet, health research is often shaped by economic and geopolitical factors rather than countries' epidemiological or healthcare needs. This study examines global patterns and drivers of PSV research on five vaccines -polio, measles, human papillomavirus, influenza, and SARS-CoV-2- and evaluates the COVID-19 pandemic's impact on research volume, focus, and distribution.

Methods: We conducted a machine-learning-assisted literature search on PSV without geographical, language or time constraints. Using natural language processing, network and statistical analyses, we examined the global PSV research landscape and identified geographical, epidemiological, and economic drivers.

Results: We analysed 13,287 articles and detected consistent literature growth from 1980 onwards, with vaccine-specific peaks following key licensing events. After 2020, publication volumes rose above projections for influenza (32%; 95%CI = 20%, 46%) but declined for polio (-56%; 95%CI = -68%, -26%) and measles (-17%; 95%CI = -33%, 9%). Although PSV research had global coverage, its distribution was markedly imbalanced, largely shaped by country-specific economic factors. A few high-income countries (HIC) produced 72% of publications and the likelihood of a country being studied varied by income and vaccine. Foreign authorship also increased as the income of the studied country decreased (over 75% in LMIC vs. below 50% in HIC).

Conclusion: PSV research reveals persistent inequities, with a misalignment between countries leading research and populations most in need of its outcomes. These inequities, further exacerbated by COVID-19 disruptions, reflect systematic imbalances in global health. Our

46 findings underscore the need to decolonise research by fostering leadership, agenda-setting, and
47 accountability that centres on the necessities of affected communities. Achieving this will require
48 funding and publication reforms that promote equitable collaborations and elevate local priorities
49 alongside long-standing global health objectives.

50 **Key messages**

51 ***What is already known on this topic***

52 Public sentiment towards vaccination (PSV) significantly influences vaccination campaign
53 outcomes. While bibliometric studies have explored PSV research patterns, economic and
54 contextual drivers in this research field remain underexplored, and international collaboration
55 dynamics lack in-depth analysis.

56 ***What this study adds***

57 This study identifies significant economic disparities shaping PSV research, with high-income
58 countries dominating the field and directing studies toward low-income settings with limited local
59 author representation. It also highlights the COVID-19 pandemic's role in shifting research focus
60 from critical childhood vaccines to SARS-CoV-2, exacerbating these imbalances.

61 ***How this study might affect research, practice or policy***

62 By revealing inequities in PSV research, this work underscores the need for more equitable and
63 inclusive research initiatives, prioritizing public health needs while addressing systemic disparities
64 in global health research and policy.

Introduction

The primary goal of global health research is to promote health equity and strengthen healthcare systems by bridging global knowledge with local interventions.[1] However, global health initiatives are heavily influenced by economic factors.[2] Researchers from low- and middle -income countries (LMIC) face limited access to global health research grants,[3] and are underrepresented in peer-reviewed publications.[4] Furthermore, 90% of all health research funding targets health issues of industrialized countries,[5] marginalizing lower-income regions, where health burdens are typically greater. This perpetuates inequities, with researchers from high-income countries (HIC) dominating research activities also in LMIC, resulting in inadequate research agendas that fail to meet local public health needs.[2] Vaccine uptake and childhood immunization patterns across the globe represent a significant example of this dynamic.

Despite vaccines' efficacy, global uptake remains suboptimal in most countries, pockets of zero-dose individuals persist,[6] and public sentiment towards vaccination (PSV) is imbued with widespread hesitancy.[7] Over time, considerable research efforts have been directed towards investigating the intricate mechanisms shaping PSV, resulting in a heterogeneous, multi-disciplinary, and fast-growing body of literature.[8] Numerous reviews attempted to systematize and summarize the existing evidence on PSV.[9] However, due to the sheer amount and complexity of the published data, this task is proving increasingly difficult, especially without the aid of machine-assisted methods. Identifying the main drivers of PSV research might facilitate the definition of future research priorities and provide an evidence-based rationale to support a more appropriate allocation of resources.

To what extent PSV research is shaped by the interplay between economic factors, local disease endemicity, and the effective need to study PSV locally is not fully understood,[8,10–12] especially in light of disruptive events, such as the COVID-19 pandemic. The exceptional circumstances in

89 which the SARS-CoV-2 vaccine was licensed, together with the determining role it has played in
90 curbing the pandemic, significantly intensified research efforts and public discussions on PSV.[13]
91 However, the pandemic also diverted a significant portion of global academic efforts and funding
92 from other research to COVID-19-related contributions, which benefitted from fast publication
93 tracks and great academic attention, putting research on other diseases and vaccines at considerable
94 disadvantage.[14–16]

95 In this work, we investigated the global patterns and drivers of scientific research on PSV,
96 including studies on public attitudes, beliefs, emotions, trust, hesitancy and uptake. To enable a
97 fine-grained yet broad perspective on the literature, we used machine-learning-assisted article
98 selection to perform a large-scale literature search of peer-reviewed publications. We screened for
99 a selected set of vaccines *i.e.*, those against polio, measles, Human PapillomaVirus (HPV),
100 influenza, and SARS-CoV-2, chosen to be representative of different target populations, licensing
101 and distribution histories, and viral pathogens with distinct transmission routes and epidemiology.
102 Using a combination of Natural Language Processing and Large Language Model techniques we
103 enriched each record with standardised geographic metadata about authorship and place of study,
104 and linked each record with epidemiological and economic country-specific metadata. We
105 analysed publication trends for all vaccines, and by benchmarking publication volumes before and
106 after 2020, we quantified the extent to which the COVID-19 pandemic has altered pre-existing
107 publication dynamics. Then, leveraging network and statistical tools, we modelled and analysed
108 the geographical, epidemiological, and economic factors driving scientific focus and academic
109 contribution in the PSV research field. The integrated perspective introduced in this work charts
110 the overall structure of PSV literature at large scale, allowing for direct comparisons across a set
111 of vaccines with distinct characteristics and public attention patterns.

112 Our paper represents a significant contribution to the literature both in terms of methodology as
113 well as implications for academic research and global health policy. By combining novel
114 techniques with traditional statistical approaches, we offer a strong example of how diverse
115 methods can be jointly applied to address complex research questions. By outlining the structural
116 inequities that influence PSV research agendas, our findings provide important insights to guide
117 future research priorities, with the ultimate goal of a more equitable, balanced, and resilient global
118 health research agenda.

119

120 **Methods**

121 *Database construction*

122 To retrieve scientific publications on PSV, we developed a comprehensive search string according
123 to validated guidelines,[17] incorporating keywords regarding public knowledge, attitudes,
124 awareness, emotions and beliefs related to vaccination, vaccine hesitancy and refusal, trust, and
125 uptake (Table S1). To optimize the specificity of our search, we limited vaccine-related keywords
126 to the article title, and other search terms to the title and abstract. We refined our search string
127 through subsequent search cycles and qualitative screening of the first n=100 delivered papers,
128 iteratively removing 1) keywords retrieving irrelevant literature, *e.g.* "vaccinia", and 2) individual
129 terms that were likely to occur together with a synonymous search term in true positive cases and
130 retrieving irrelevant literature otherwise. The literature search was conducted in the MEDLINE
131 and Web of Science Core Collection databases until December 2023, without geographical,
132 language or other time constraints. We chose this database combination to cover a broad spectrum
133 of scientific disciplines while obtaining the best possible balance between sensitivity and

specificity.[18] We removed duplicates using the article DOIs and titles and excluded papers not mentioning any of the vaccines of interest in their titles or abstracts.

Given the large number of retrieved articles, we employed supervised Machine Learning (ML) techniques to identify studies directly related to PSV. We began by manually annotating a sample of 1,000 articles by title and abstract, following a predefined set of inclusion criteria. Articles were retained if they addressed at least one dimension of vaccine sentiment, such as studies on knowledge, attitudes, and practices, or assessed determinants of vaccine acceptance, uptake, or timeliness. Eligible study types included both quantitative and qualitative primary research, secondary data analyses, theoretical frameworks, mathematical models exploring vaccination determinants as well as reviews and editorials. After the first round of screening, performed by two independent reviewers, we obtained an inter-rater agreement *Cohen's k = 0.8*, indicating substantial agreement. A third reviewer resolved the disagreements, resulting in 376 articles labelled as directly related to PSV and 624 as unrelated.

Next, using this annotated sample, we trained multiple ML classifiers, fine-tuning the pipelines' hyperparameters in a 5-fold cross-validation setting to maximize classification performance. Model evaluation was conducted on a held-out test set (30% of the annotated data), achieving high performance across models, with all classifiers reaching an F1-score > 0.86 and a Matthews Correlation Coefficient (MCC) > 0.78.[19] The best-performing model - *Random Forest Classifier* with *bag-of-words* and *tf-idf* preprocessing- achieved an *F1-score of 0.89*, and *MCC of 0.82* on the test set. This model was then applied to classify the full dataset. For each retained article, we identified the *primary vaccine* studied, defined as the most frequently mentioned vaccine in the title and/or abstract, as well as the other mentioned vaccines. If multiple vaccines were referenced equally, the first vaccine appearing in the text was assigned as the primary one.

157 *Data pre-processing and enrichment*

158 Using OpenAI's GPT (Supplement-Methodological details) we enriched our corpus by extracting
159 from titles and abstracts of collected articles the *authors' countries*, *i.e.* the countries of the
160 authors' institutions, and the *studied countries*, *i.e.,* the countries of the analysis. Each country was
161 then assigned a list of categorical attributes, including income level as of 2021,[20] time-dependent
162 vaccine-specific coverage, [21,22] and time-dependent disease incidence for each related disease.[23,24]
163 Country income was chosen as a proxy for a country's research capacity since research institutions,
164 such as universities and public health organisations, largely rely on national funding and
165 infrastructure to conduct their projects. Disease incidence dynamically captures the
166 epidemiological pressure that a country is facing. Together with varying levels of vaccination
167 coverage, these factors are likely to influence the scientific -and public- concern about a health
168 threat, potentially fuelling or slowing down research. All attributes were categorized into four
169 groups (L=low; LM=lower-middle; UM=upper-middle; H=high) based on the empirical quantiles
170 (Supplement-Methodological details). Polio incidence was classified as either "No cases" or
171 "Reported cases". Due to incomplete data on disease incidence and vaccination coverage for HPV
172 and influenza, we limited the analysis including all attributes to measles and polio.

173 *Time trend analysis*

174 To characterize the publication trends over time, we modelled the volume of published papers for
175 each vaccine from 2010 to 2019 fitting an exponential function. We conducted this analysis
176 starting from 2010 to ensure inclusion of the period following the licensure of both the HPV and
177 2009 H1N1 pandemic influenza vaccines. The impact of the SARS-CoV-2 pandemic on PSV-
178 related research was assessed by comparing the predicted publication volumes for 2020-2023,
179 based on the fitted exponential models, with the actual volume produced after 2020, calculating
180 the *percentage difference* (Supplement-Methodological details).

Geographic network construction

We developed a custom network analysis pipeline to map the geographic direction of global research on PSV. In particular, for each 5-year time interval, we created vaccine-specific networks where each node represents a country, and each directed weighted edge, a connection between countries *i* and *j*, was formed proportionally to the number of papers authored or co-authored by institutions from country *i* focusing on country *j.* Articles with multiple authors' countries and/or studied countries contributed to multiple edges in the network. We further characterized each node *i* by counting the number of 1) *Domestic* studies, papers on country *i* authored by local researchers alone (the weighted self-loops of node *i*); 2) *Inbound* studies, papers on country *i* authored or co-authored by foreign institutions (the weighted in-degree minus the self-loops); 3) *Outbound* studies, papers on any foreign country authored or co-authored by institutions from country *i* (the out-degree minus the self-loops).

Statistical analysis on research drivers

We employed a regression framework to investigate the drivers of global PSV research. First, to investigate the probability of a country being *studied*, we fitted logistic regression models using a binary dependent variable indicating whether a country had been the object of at least one research paper on a given vaccine (1 if studied, 0 if not) (Supplement-Methodological details). Second, for countries that were *studied*, we analysed which countries' factors influenced the proportion of papers authored or co-authored by foreign institutions (*i.e., Inbound* / (*Inbound* + *Domestic*)). We achieved that through a Beta regression model, suitable for modelling proportions (Supplement-Methodological details). Independent variables in both models included 1) time interval, classified as pre-pandemic (2010-2014; 2015-2019) and post-pandemic (2020-2023) periods, 2) country income level, 3) country time-varying vaccination coverage level, and 4) country time-varying disease incidence level. All models except the ones for polio (for which not enough data was

205 available) included an interaction term between income and time to test whether the probability of
206 being studied or the proportion of foreign authorship varied over time across income groups
207 (Supplement-Methodological details).

208 *Ethics approval*

209 This study did not include human participants and is thus exempt from Institutional Review Board
210 approval.

211 **Results**

212 *The unfolding of PSV research over time*

213 Our query across two databases yielded 154,552 articles from 1902 to 2023 (Figure 1). We
214 removed duplicates (*43%*) and screened for references to any of the five selected vaccines,
215 retaining 17,587 articles. Of these, 97.5% were written in English, 2.2% in other European
216 languages, and 0.2% in non-European ones. We then used ML-assisted techniques to exclude non-
217 relevant papers (*24%* of articles on the selected vaccines). We retained a total of *13,287* articles
218 (*18%* of the initial yield) for analysis. Of these, 44% focused on SARS-CoV-2 as the *primary*
219 *vaccine* (n= 5,818), 25% on influenza (n=3,328), 23% on HPV (n=3,078), 7% on measles (n=883),
220 and 1% on polio (n=180).

221 The volume of PSV literature did not rise significantly until the late 1980s (Figure 1). We observed
222 exponential growth across all vaccines until 2020, with vaccine-specific surges corresponding to
223 the licensing of the HPV and SARS-CoV-2 vaccines, and a volume spike in influenza-related
224 publications following the development of the novel influenza vaccine against the 2009/H1N1
225 pandemic strain. Using the fitted exponential models (Figure S2), we predicted the volume of
226 publications post-SARS-CoV-2 and compared them against the measured volumes, finding a

227 higher-than-expected increase in the volume of influenza-related research (Perc. Diff. = *32%;*
228 *95%CI = 20%, 46%)*. In contrast, the volume of PSV research on the two routine childhood
229 vaccinations, polio and measles, fell below expectation after the pandemic, with decreases of *Perc.*
230 *Diff. = -56% (95%CI = -68%, -26%)* and *Perc. Diff. = -17% (95%CI = -33%, 9%)* respectively.

231 Notably, an increasing proportion of papers mentioned more than one vaccine in the text (7·3%
232 before 2020 and 12·9% after 2020). In particular (Figure 2), vaccines sharing similar target
233 populations, such as polio and measles, or those protecting against pathogens with similar
234 characteristics, such as influenza and SARS-CoV-2, two respiratory viruses with pandemic
235 potential, were more frequently studied together. Additionally, several studies on all *primary*
236 *vaccines* began co-mentioning SARS-CoV-2 during the pandemic. This applied even to vaccines
237 with relatively different characteristics, such as polio, measles, or HPV, suggesting a potential
238 impact of the SARS-CoV-2 pandemic on PSV research. Despite historical trends, the share of
239 articles studying influenza alone also declined after the pandemic, with a large proportion (19·7%)
240 of recent influenza-related articles also referencing the new coronavirus.

241 *Geographic patterns in global health research*

242 We mapped the geographic distribution of global PSV research from 2010 onwards (Figure 3). To
243 reconstruct the direction of research we evaluated country-specific proportions of publications
244 authored by local institutions alone (*Domestic),* authored by or with institutions from a foreign
245 country (*Inbound*), or focused exclusively on foreign countries (*Outbound)*.

246 Overall, published PSV literature covered most countries worldwide, with HIC in North America
247 and Europe authoring the majority of studies (72·2 %). However, notable differences emerged in
248 the spatial distribution of published articles across the examined vaccines. PSV research on the
249 polio vaccine was more geographically clustered compared to other vaccines, with hotspots of

250 research in Sub-Saharan Africa, the Middle East, and South Asia. Conversely, research efforts on
251 the influenza vaccine were relatively frequent in Europe and North America but scarce on the
252 African continent, with over 70%, over 40%, and less than 30% of countries in the respective
253 continents involved, *i.e.* authoring or being studied in at least one paper (Figure S3, top row).

254 We also observed income-related global imbalances in terms of research direction (Figure S3,
255 bottom row). Despite the high number of studies on polio and measles vaccines authored by
256 institutions in North America, Europe, and Australia, the direction of these studies was
257 predominantly *Outbound* for HIC *(Median proportion Outbound studies, High-income; Polio:*
258 *82%; Measles: 61%)*, and mostly *Inbound* for low-income countries *(Median proportion Inbound*
259 *studies, Low-income: Polio: 100%; Measles: 78%)*. In the same time frame, little *Domestic* and
260 even less *Outbound* research emerged from low- and lower-middle-income countries. Similar but
261 reversed income-dependent trends were observed in the distribution of PSV research on the HPV
262 and influenza vaccines, although in HIC the shares of *Domestic*, *Inbound*, and *Outbound* research
263 efforts were more equally distributed for these two vaccines *(Median proportion Domestic,*
264 *Inbound, Outbound studies, High-income; Influenza: 24%, 37%, 26%; HPV: 33%, 32%, 33%)*.
265 Lastly, we measured comparable shares of *Outbound* research on SARS-CoV-2 across income
266 levels *(Median proportion Outbound studies SARS-CoV-2: Low-income: 21%; Low-middle-*
267 *income: 31%; Upper-middle-income: 33%; High-income: 40%)*, with the amount of *Inbound*
268 research decreasing with increasing country income, mirroring patterns observed for other
269 vaccines.

270 *Drivers of global research flow*

271 Through vaccine-specific logistic regression models, we assessed the potential factors influencing
272 the probability of countries being the focus of global PSV research. The reconstructed marginal
273 probabilities are shown in Figure 4 and Figure S4, upper panels, coefficients are reported in Table

274 S2 and Table S3. We found that a significantly lower proportion of countries has been studied in
275 polio research compared to all the other vaccines. For all vaccines, the probability of being studied
276 was significantly associated with the country's income level, although this dependence followed
277 opposite directions for different vaccines. In the reference period (2010-2014), lower-income
278 countries had a significantly higher probability of being the object of polio or measles PSV
279 research compared to HIC. The disparity was greatest for the measles vaccine in the reference
280 period, with HIC having 98% lower odds of being studied compared to the low-income group *(H*
281 *vs L: OR=0·02; 95%CI = 0·00, 0·11)* (Table S2). This income disparity remained constant over
282 time for polio and appeared to level up for measles, with increased interest in higher-income
283 settings but a decrease in lower-income ones. In contrast, HIC had a significantly higher
284 probability of serving as study locations for the HPV and influenza vaccines, perpetuating existing
285 income disparities without significant time-and-income variation. Additionally, the probability of
286 being studied for the influenza vaccine appeared to be significantly higher in the years following
287 the COVID-19 pandemic *(2020-23 vs 2010-14: OR=7·57; 95%CI = 1·16, 149·71)* (Table S3).

288 For both measles and polio, local disease incidence affected the probability of countries to be
289 studied (Figure S4, Table S2). Countries where clinical polio cases were reported had 2·7 times
290 higher odds of being studied compared to countries where the disease was eliminated *(Reported*
291 *cases vs No cases: OR=2·67; 95%CI= 0·95, 7·37)*. Analogously, the probability of being studied
292 was significantly higher among countries experiencing comparatively high measles incidence *(H*
293 *vs L: OR=10·11; 95% CI= 4·45, 24·76)*. Conversely, we did not detect a consistent effect of
294 national vaccination coverage on the probability of being studied for either vaccine.

295 *Determinants of foreign authorship*

296 For countries that have been the object of PSV research, we modelled the proportion of articles
297 (co-)authored by foreign countries using Beta regression models. The fitted values are shown in

298 Figure 4 and Figure S4, bottom panels, the coefficients are reported in Table S2 and Table S3.
299 Across all vaccines, during the reference period (2010-2014), the proportion of papers involving
300 foreign research efforts increased as the income level of the *studied* country decreased. This
301 income-related disparity appeared to diminish over time for the measles, HPV, and influenza
302 vaccines. Conversely, we observed no significant time-dependent change for polio research.

303 For both polio and measles, countries reporting clinical disease cases had a smaller proportion of
304 articles (co-)authored by foreign researchers. We found no consistent effect of national vaccination
305 coverage on the proportion of foreign authorship.

306 **Discussion**

307 This work examines the drivers of global health research using innovative methodologies, offering
308 novel insights and analysis that, to our knowledge, are unexplored in the existing literature. We
309 integrated an extensive literature search, machine learning-assisted article screening and data
310 enrichment, network analysis, and statistical methods to develop a novel database of research
311 papers on PSV, which offers a valuable resource for addressing broad global health research
312 questions. Going beyond traditional bibliometric methods,[8,10–12] and merging information on
313 country-specific epidemiological and economic factors as well as direction of research, we
314 investigated patterns and drivers shaping the global research efforts on PSV.

315 Our findings portrayed PSV research as a rather recent research field, partially shaped by vaccine-
316 specific epidemiological conditions and unforeseen global health challenges, such as pandemic
317 threats, but strongly influenced by country-specific economic determinants that steer academic
318 focus and the opportunity for scientific collaborations.

319 By modelling the volumes of historical research data, we revealed that overall PSV publications
320 have grown exponentially until 2020. Despite the eradication of polio and measled being within

321 reach[25], the research output on these vaccines remained relatively modest throughout the study
322 period. Moreover, notwithstanding the rise in public vaccine hesitancy in recent years,[26,27] it did
323 not intensify as much as expected after COVID-19 pandemic, proving vulnerable to sudden
324 disruptions. In contrast, we observed higher-than-expected post-pandemic publication volumes
325 regarding the influenza vaccine, likely due to its analogies with the SARS-CoV-2 vaccines. These
326 findings align with abundant evidence of the negative impact of the SARS-CoV-2 pandemic on
327 routine childhood vaccinations, both in terms of vaccine uptake[28,29] and public trust,[30] and may be
328 attributed to the temporary redirection of academic interest and research funding towards COVID-
329 19-related research. The influence of SARS-CoV-2 on PSV research was further highlighted by a
330 post-pandemic increase in articles on other *primary vaccines* mentioning the SARS-CoV-2
331 vaccine. This pattern, especially striking for vaccines that share few or no similarities with SARS-
332 CoV-2 vaccine, likely reflects a momentary shift of the global health research agenda due to the
333 very peculiar and worrisome situation of the pandemic threat. However, it also suggests that PSV
334 research is highly sensitive to time- and context-dependent funding and publication opportunities,
335 which in turn influence researchers' actions.[16] This result underscores the importance to maintain
336 sustained academic attention on long-standing research objectives, such as vaccine-preventable
337 diseases elimination and eradication efforts.

338 Our statistical analyses of geographical, economic, and epidemiological factors consistently
339 showed that country income levels critically shape the landscape of PSV research, inducing an
340 economically driven imbalance in both the global distribution and direction of research efforts.

341 In agreement with previous literature,[8,10–12] we found that HIC produced the majority of global
342 PSV research, typically redirecting their research on polio and measles towards LMIC and, on the
343 other hand, attracting research efforts on HPV and influenza vaccines.

344 Our regression analysis also suggested a gradual shift in academic attention from LMIC to HIC,
345 particularly for measles, but also for HPV and influenza. We showed that such a trend may be
346 partially explained by changes in the epidemiology of the related diseases, although other factors
347 likely played a significant role as well. These may include a country's historical experience with
348 the disease, the presence of local outbreaks, local vaccination recommendations and policies,
349 political context, the unequal distribution of research funding, global publication bias, and
350 countries' involvement in international public health endeavours, such as disease elimination
351 initiatives. These results also suggest that the gap left by shifting attention to HIC is not currently
352 filled with *Domestic* research in LMIC.

353 Income-dependent disparities were also evident through the disproportionate amount of foreign
354 authorship in studies concerning LMIC, especially in PSV research on polio and measles. This
355 suggests a structural involvement of external researchers in LMIC, even in cases concerning very
356 local health challenges. Except for polio research, where the historical imbalances persist, foreign
357 research contributions became more evenly distributed across income groups for all vaccines,
358 indicating an encouraging increase in academic cross-income-group collaborations. Combining
359 this result with an overall increased attention to higher-income countries, however, our analyses
360 also suggest that the levelling of foreign efforts might be driven by increased collaborations among
361 HIC, likely at the expense of lower-income ones, where such research is often most needed.

362 Overall, our regression analyses provided strong evidence that a country's lower income
363 significantly limits its capacity to perform *Domestic* research, even in areas with high disease
364 incidence, resulting in an inequitable and likely inefficient reliance on external research activities
365 to address critical local health challenges.

366 We acknowledge some limitations. First, our literature search strategy primarily employed
367 English-language keywords and a substantial portion of non-English scientific literature may not

368 be indexed in the two databases used. As a result, we may have missed relevant studies published
369 in other languages lacking an English abstract. Because English-language publications are not
370 fully representative of the global scientific landscape, this may have introduced a language bias
371 into our literature corpus. However, the vast majority of global scientific output is currently
372 published in English.[31] Hence, any language bias is likely to be limited in scope and not expected
373 to substantially affect the main patterns identified in the analysis.

374 Second, although our ML-assisted article selection achieved strong performance metrics, some
375 misclassification may have occurred, potentially leading to the inclusion of a few irrelevant articles
376 or the exclusion of some pertinent ones. Yet, given the scale of the dataset and the inherent
377 subjectivity in relevance judgements, even a fully manual screening process would likely have
378 introduced comparable inconsistencies. Importantly, the volume and the consistency of the
379 retained articles still provide a reliable basis for mapping broader scientific trends.

380 Lastly, inclusion of additional contextual factors that may shape global attention on PSV research
381 could offer deeper explanatory power. However, in order to ensure complete, consistent and
382 reliable cross-country comparison on a global scale, we chose to leverage only indicators provided
383 by internationally recognised organizations. Collating national datasets may have introduced
384 uncertainty and unknown sources of potential bias, ultimately compromising the robustness and
385 comparability of our analysis.

386

387 **Conclusion**

388 The persistent misalignment between countries with the financial capacity to conduct PSV
389 research and those most in need of its outcomes highlights entrenched inequities within this field.
390 These disparities are not unique to PSV research but are rather a symptomatic of broader structural

391 imbalances that pervade global health research. Research priorities -particularly in HIC- are often
392 shaped by internal agendas and funding opportunities rather than global need. This hampers the
393 potential impact of research efforts in achieving broader health outcomes, such as increased
394 vaccine uptake in underserved settings.[32]

395 Our findings suggest that such imbalances are likely to be exacerbated during periods of global
396 disruptions, as demonstrated by the COVID-19 pandemic, with potentially detrimental
397 consequences for both ongoing public health interventions and future research directions. By
398 providing a clear view of the main drivers of PSV research and revealing the risks of shifting
399 academic focus, this study highlights the need to decolonise PSV research and, broadly, global
400 health research. Addressing historical imbalances requires promoting equitable, inclusive research
401 partnerships, and redistributing leadership, ownership, and priority-setting power across the global
402 research community. Key principles for establishing equitable research dynamics include
403 sustainability, policy-oriented objectives, transparency, equitable data access, and meaningful
404 community engagement.[33–37] Existing ethical guidance can help researchers design projects in
405 alignment with these values,[38] and funding mechanisms should embed strong incentives to
406 promote such practices.[39] Future funding allocation criteria should prioritise research with the
407 highest potential health impact for vulnerable populations - such as improving vaccine coverage
408 among under-immunised children - over purely academic or innovation-driven considerations.[40]
409 Grants must be accessible to institutions in LMIC directly, fostering application diversity and
410 enhancing local research capacity.[3,41] Achieving key global health targets, such as the eradication
411 of polio and measles, requires sustainable investments guided by clear long-term strategies and
412 implemented within well-defined timeframes, particularly in low-resource settings.[42] Currently,
413 decision-making power remains highly concentrated in a small group of actors, mostly from HIC,
414 yet continued support may be subject to political or economic fluctuations. This can lead to funding

415 disruptions with potentially serious implications for the populations affected. A sustainable
416 transition toward increased domestic investment in health research is therefore critical and must
417 include robust exit strategies, early planning, stakeholder engagement, and dedicated pre-transition
418 resources. [3,43,44] Scientific journals should also play a crucial role in reshaping inequitable research
419 practices by establishing systems to identify and discourage submissions that violate equity
420 principles related to content and authorship.[45,46] Our work highlights the need for a systemic
421 change across funding structures, authorship norms, and publication practices in order to redress
422 long-standing inequities in global health research, and to ensure that research efforts ultimately
423 benefit the most affected.[47]

References

1 Koplan JP, Bond TC, Merson MH, Reddy KS, Rodriguez MH, Sewankambo NK *et al.* Towards a common definition of global health. *The Lancet* 2009; **373**: 1993–1995.

2 Smeeth L, Kyobutungi C. Reclaiming global health. *The Lancet* 2023; **401**: 625–627.

3 Charani E, Abimbola S, Pai M, Adeyi O, Mendelson M, Laxminarayan R *et al.* Funders: The missing link in equitable global health research? *PLOS Global Public Health* 2022; **2**: e0000583.

4 Cash-Gibson L, Rojas-Gualdrón DF, Pericàs JM, Benach J. Inequalities in global health inequalities research: A 50-year bibliometric analysis (1966-2015). *PLoS One* 2018; **13**: e0191901.

5 Global Forum for Health Research & World Health Organization. The 10/90 (ten ninety) report on health research 2003-2004. 2004.https://iris.who.int/handle/10665/44386 (accessed 16 Jul2024).

6 The WHO Global Health Observatory. IA2030 IG 2.1: Number of zero dose children. 2024.https://www.who.int/data/gho/indicator-metadata-registry/imr-details/7792 (accessed 19 Sep2024).

7 de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. *The Lancet* 2020; **0**. doi:10.1016/S0140-6736(20)31558-0.

8 Sweileh WM. Bibliometric analysis of global scientific literature on vaccine hesitancy in peer-reviewed journals (1990-2019). *BMC Public Health* 2020; **20**: 1–15.

9 Fernandes S, Jit M, Bozzani F, Griffiths UK, Anthony JG. A bibliometric analysis of systematic reviews on vaccines and immunisation. 2020; **36**: 2254–2261.

10 Chen C, Yang Q, Tian H, Wu J, Chen L, Ji Z *et al.* Bibliometric and visual analysis of vaccination hesitancy research from 2013 to 2022. *Hum Vaccin Immunother* 2023; **19**. doi:10.1080/21645515.2023.2226584.

11 Sweileh WM. Bibliometric analysis of scientific research publications on COVID-19 vaccine hesitancy (2020 – 2021). *Electronic Journal of General Medicine* 2023; **20**: em500.

12 Jaca A, Iwu-Jaja CJ, Balakrishna Y, Pienaar E, Wiysonge CS. A global bibliometric analysis of research productivity on vaccine hesitancy from 1974 to 2019. *Hum Vaccin Immunother* 2021; **17**: 3016–3022.

13 Emery N, Dugerdil A, Flahault A. Vaccine Hesitancies Across the World in the Era of COVID-19. *Am J Public Health* 2022; **112**: 1579–1581.

14 He J, Liu X, Lu X, Zhong M, Jia C, Eliseo D *et al.* The impact of COVID-19 on global health journals: an analysis of impact factor and publication trends. *BMJ Glob Health* 2023; **8**: 11514.

15 Riccaboni M, Verginer L. The impact of the COVID-19 pandemic on scientific research in the life sciences. *PLoS One* 2022; **17**: e0263001.

16 Aviv-Reuven S, Rosenfeld A. Publication patterns' changes due to the COVID-19 pandemic: a longitudinal and short-term scientometric analysis. *Scientometrics* 2021; **126**: 6761–6784.

17 Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. *Journal of the Medical Library Association* 2018; **106**. doi:10.5195/jmla.2018.283.

18 Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Syst Rev* 2017; **6**: 1–12.

19 Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 2020; **21**: 1–13.

20 World Bank Group. World Bank Country and Lending Groups. 2024.

21 World Health Organization. Polio (Pol3) Immunization coverage estimates by country. 2024.https://apps.who.int/gho/data/node.main.A831 (accessed 15 Jul2024).

22 World Health Organization. Measles-containing-vaccine second-dose (MCV2) immunization coverage by the locally recommended age (%). 2024.https://www.who.int/data/gho/data/indicators/indicator-details/GHO/measles-containing-vaccine-second-dose-(mcv2)-immunization-coverage-by-the-nationally-recommended-age-(-) (accessed 15 Jul2024).

23 World Health Organization. Poliomyelitis - number of reported cases. 2024.https://www.who.int/data/gho/data/indicators/indicator-details/GHO/poliomyelitis---number-of-reported-cases (accessed 15 Jul2024).

24 WOrld Health Organization. Measles - number of reported cases. 2024.https://www.who.int/data/gho/data/indicators/indicator-details/GHO/measles---number-of-reported-cases (accessed 15 Jul2024).

25 Higgins DM, O'Leary ST. A World without Measles and Rubella: Addressing the Challenge of Vaccine Hesitancy. *Vaccines 2024, Vol 12, Page 694* 2024; **12**: 694.

26 Pandey A, Galvani AP. Exacerbation of measles mortality by vaccine hesitancy worldwide. 2023. doi:10.1016/S2214-109X(23)00063-3.

27 Harvard International Review. Polio in a post vaccine world the consequences of anti-vaccination sentiments. 2023.https://hir.harvard.edu/polio-in-a-post-vaccine-world-the-consequences-of-anti-vaccination-sentiments/.

28 Shet A, Carr K, Danovaro-Holliday MC, Sodha S V., Prosperi C, Wunderlich J *et al.* Impact of the SARS-CoV-2 pandemic on routine immunisation services: evidence of disruption and recovery from 170 countries and territories. *Lancet Glob Health* 2022; **10**: e186–e194.

29 Lassi ZS, Naseem R, Salam RA, Siddiqui F, Das JK. The Impact of the COVID-19 Pandemic on Immunization Campaigns and Programs: A Systematic Review. *Int J Environ Res Public Health* 2021; **18**: 1–19.

30 Lazarus J V., White TM, Wyka K, Ratzan SC, Rabin K, Larson HJ *et al.* Influence of COVID-19 on trust in routine immunization, health information sources and pandemic preparedness in 23 countries in 2023. *Nature Medicine 2024 30:6* 2024; **30**: 1559–1563.

31 Scientific publishing has a language problem. *Nat Hum Behav* 2023; **7**: 1019–1020.

502 32 Agnandji ST, Tsassa V, Conzelmann C, Köhler C, Ehni HJ. Patterns of biomedical science
503 production in a sub-Saharan research center. *BMC Med Ethics* 2012; **13**. doi:10.1186/1472-
504 6939-13-3.

505 33 Monette EM, McHugh D, Smith MJ, Canas E, Jabo N, Henley P *et al.* Informing 'good' global
506 health research partnerships: A scoping review of guiding principles. *Glob Health Action* 2021;
507 **14**. doi:10.1080/16549716.2021.1892308.

508 34 Evertsz N, Bull S, Pratt B. What constitutes equitable data sharing in global health research?
509 A scoping review of the literature on low-income and middle-income country stakeholders'
510 perspectives. *BMJ Glob Health* 2023; **8**: 1–11.

511 35 Achana F, Boah M, Chirawurah D, Owusu-Marfo J, Ndago JA, Santah C *et al.* Equitable global
512 health research collaborations with a mind of human dignity. *Front Educ (Lausanne)* 2023; **8**:
513 1–5.

514 36 Hodson DZ, Etoundi YM, Parikh S, Boum Y. Striving towards true equity in global health: A
515 checklist for bilateral research partnerships. *PLOS Global Public Health* 2023; **3**: 1–15.

516 37 Saleh S, Masekela R, Heinz E, Abimbola S, Morton B, Vercueil A *et al.* Equity in global health
517 research: A proposal to adopt author reflexivity statements. *PLOS Global Public Health* 2022;
518 **2**: 10–12.

519 38 Pratt B. Research for Health Justice Framework.
520 2025.https://researchforhealthjustice.squarespace.com/general-2 (accessed 7 May2025).

521 39 Pratt B, Hyder AA. Designing research funding schemes to promote global health equity: An
522 exploration of current practice in health systems research. *Dev World Bioeth* 2018; **Jun;18**:
523 76–90.

524 40 Murray M, Mubiligi J. An approach to building equitable global health research collaborations.
525 *Ann Glob Health* 2020; **86**: 1–3.

526 41 Chikwari CD, Tadesse AW, Shanaube K, Shepherd A, McQuaid CF, Togun TO. Achieving
527 equitable leadership in Global Health partnerships: barriers experienced and strategies to
528 improve grant funding for early- and mid-career researchers. *BMC Global and Public Health*
529 2024; **2**. doi:10.1186/s44263-024-00047-4.

530 42 Stenberg K, Hanssen O, Edejer TTT, Bertram M, Brindley C, Meshreky A *et al.* Financing
531 transformative health systems towards achievement of the health Sustainable Development
532 Goals: a model for projected resource needs in 67 low-income and middle-income countries.
533 *Lancet Glob Health* 2017; **5**: e875–e887.

534 43 Olufadewa I, Adesina M, Ayorinde T. Global health in low-income and middle-income
535 countries: a framework for action. *Lancet Glob Health* 2021; **9**: e899–e900.

536 44 Huffstetler HE, Bandara S, Bharali I, Kennedy Mcdade K, Mao W, Guo F *et al.* The impacts
537 of donor transitions on health systems in middle-income countries: a scoping review. *Health*
538 *Policy Plan* 2022; **37**: 1188–1202.

539 45 Khan SA. Decolonising global health by decolonising academic publishing. *BMJ Glob Health*
540 2022; **7**: 1–6.

541 46 Olusanya BO, Mallewa M, Ogbo FA. Beyond pledges: Academic journals in high-income
542 countries can do more to decolonise global health. *BMJ Glob Health* 2021; **6**: 1–3.

543 47 Büyüm AM, Kenney C, Koris A, Mkumba L, Raveendran Y. Decolonising global health: If
544 not now, when? *BMJ Glob Health* 2020; **5**: 1–4.

545

546

547 **Figure legends**

548 **Figure 1** Annual number of published scientific articles on public sentiment towards 5 selected
549 vaccines from 1980 onwards. Dashed lines indicate the total number of articles mentioning each
550 vaccine, reported in different colours, in their title and/or abstract. Solid lines indicate the total
551 number of articles focusing on a specific vaccine as the *primary vaccine*. Vertical lines mark
552 vaccine licensing dates. The licensing date for influenza corresponds to the vaccine against the
553 2009/H1N1 pandemic strain, while the licensing of the 1945 inactivated seasonal influenza
554 vaccine is omitted. Polio (1955) and measles (1963) vaccine licensing dates precede the emergence
555 of PSV research. For visualization purposes, the number of papers published on SARS-CoV-2 is
556 evaluated and reported quarterly.

557

558 **Figure 2** Co-occurrence of *primary vaccines* and *mentioned vaccines* from 2010 onwards. For
559 each 5-year interval, the color-coded heatmap shows the percentage of articles focusing on a
560 primary vaccine (y-axis) and mentioning another vaccine (x-axis). Percentages on the diagonal
561 represent the proportion of papers that focus exclusively on a *primary vaccine*. Articles mentioning
562 more than one additional vaccine are included more than once in the matrices. The total number
563 of articles focusing on each *primary vaccine* for a given period is reported under the y-axis labels.

564

565 **Figure 3** Geographic distribution of *author countries* and *studied countries* by vaccine from 2010
566 onwards. Pie slices represent country-specific shares of *Domestic*, *Inbound,* or *Outbound* studies.
567 *Domestic*: studies on country *i* authored by local researchers alone. *Inbound*: studies on country *i*
568 authored or co-authored by foreign institutions. *Outbound:* studies on any foreign country
569 authored or co-authored by institutions from country *i*. The size of each pie chart is proportional

570 to the total number of studies involving a country. Background colour corresponds to countries'
571 income levels. Panels *(f – j)* focus on European countries.

572

573 **Figure 4** (Top row) Estimated probabilities, with 95% confidence intervals, of being the subject
574 of PSV research on a given vaccine across time points and income groups. (Bottom row) Estimated
575 proportion, with 95% confidence intervals, of PSV research articles on a given vaccine involving
576 foreign researchers (*Inbound/(Inbound+Domestic)),* by income level of studied countries. L=low
577 income; LM=lower-middle income; UM= upper-middle income; H= high income.

Data availability statement

The database containing all scientific publications included in this analysis will be made available on reasonable request for scientific research purposes. For further information or to request access, please contact the corresponding author.

Competing interests statement

None declared

Funding statement

Supported by the ERC Grant "IMMUNE" (ID: 101003183), PNRR Grant BEHAVE-MOD (ID: SA-02. P0001), and the Bocconi Covid Crisis Lab, supported by the "Fondazione Romeo ed Enrica Invernizzi".

Patient and Public Involvement statement

It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research

Author contributions

DB, VO, ZA, CC, LPL, AT, SS, JAP, and AM contributed to the conceptualization by formulating the research idea, objectives, and research questions. VO, LPL, ED, and AM developed the search

string, screening strategy, and annotated the dataset for relevance. DS and JAP collected and curated the dataset, while DB, VO, ZA, and CC enriched it with relevant attributes and performed data cleaning. DB, VO, ZA, CC, and AM designed the study methods and performed technical analyses, including machine-assisted literature searches, enrichment via natural language processing, network analysis, and other formal analyses. CC and ZA conducted the regression analysis. DB, VO, ZA, CC, and AM produced the results and drafted the initial manuscript. All authors interpreted the results, reviewed, and edited the drafts. DB and VO led and coordinated the project, AM supervised and secured funding.