**Assessing diversity and representativeness of big team science in psychology [Stage 1 RR Protocol]**

Weibiao Liu 1, Sakshi Ghai 2, Patrick Forscher 3, Hu Chuan-Peng 1\*

1. School of Psychology, Nanjing Normal University, Nanjing, China

2. Oxford Internet Institute, University of Oxford, Oxford, UK

3. Busara Center for Behavioral Economics, Nairobi, Kenya

**\*Corresponding author:** Hu Chuan-Peng (Email: [hu.chuan-peng@nnu.edu.cn](mailto:hu.chuan-peng@nnu.edu.cn), or, [hcp4715@hotmail.com](mailto:hcp4715@hotmail.com))

**Abstract**

Psychological science has a persistent diversity problem: samples, authors, and journal editors are predominantly from the Global North. Big team science, characterized by collaborative projects involving researchers worldwide, has emerged as a potential solution to these diversity and generalizability concerns. Despite its widespread adoption and claims of increased generalizability, the actual impact of big team science on improving diversity and representation in psychological studies remains uncertain. Here, we plan to systematically examine the diversity and representativeness of big team science by comparing big team science to (1) traditional psychological studies and (2) its target population (the world population or a specific population). Specifically, we will compare the demographics of participants and authors from big team science to the other two sources. In addition, we will examine how socioeconomic/cultural factors might contribute to the presence of researchers and participants in big team science and traditional psychological research in a country/region.

**Keywords:** Meta-science; Population psychology; Representativeness; Big team science; Generalizability

**1 Introduction**

Psychological science faces a generalizability crisis (Yarkoni, 2022). One critical reason is a well-known problem: Samples in psychological science are notoriously lacking in diversity and representativeness (Arnett, 2008; Henrich et al., 2010; Nielsen et al., 2017; Pollet & Saxton, 2019; Rad et al., 2018). For example, Arnett (2008) pointed out that most of the samples published in the six premier APA journals come from western countries, and this pattern has not changed in the past decades (Thalmayer et al., 2021). However, this issue is much broader than the WEIRD problem, i.e., participants are mainly from Western, Educated, Industrialized, Rich, and Democratic regions (Henrich et al., 2010), because heterogeneity also exists outside of WEIRD regions (Ghai, 2020; Ghai et al., 2023).

The lack of diverse and representative samples in psychological science leads to several severe consequences for the field. Theories in psychological science can not be generalized to people other than western undergraduate students (e.g., Tindle, 2021), findings relevant to clinical settings (e.g., Chen et al., 2023) and public policy (e.g., Ijzerman et al., 2020) are biased. Furthermore, ignoring people outside the western population created a situation in which we simply do not know what is unknown about the human mind and behavior (i.e., the "unknown unknown"; Adetula et al., 2022), which limited the scope of psychological science.

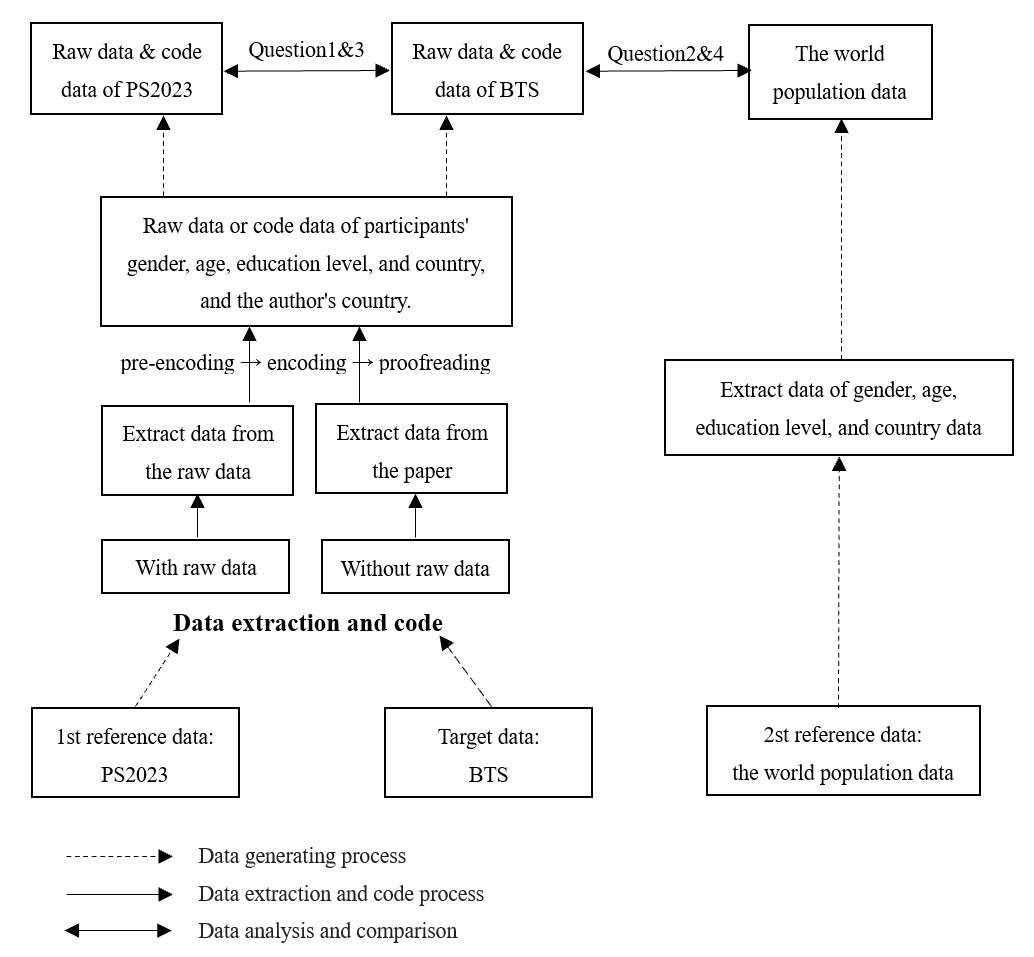
A closely related issue, which might be the reason for the lack of diverse and representative participants but less known to researchers, is the lack of diversity among authors publishing in psychology journals (Arnett, 2008; Lin & Li, 2022; Medin et al., 2017; Thalmayeret al., 2021). Arnett (2008) revealed that, on average, over 70% of first authors in six APA journals are from the United States So far, this pattern hasn't changed. Lin and Li (2022) found that 86% of authors in 68 top psychology journals were primarily from North American and West European countries. IJzerman et al (2021) also found that 90% of awards from the Society for Personality and Social Psychology go to researchers affiliated with the United States institutions. In the Open Science movement, majority of researchers and grassroots network are from North America and West Europe (Jin et al., 2023). The lack of author representation not only leads to biased sampling and methods (Yang et al., 2023), but also ignores the voices of underrepresented populations (Arnett, 2008; Lin & Li, 2022).

In recent years, big team science has emerged as a potential solution to the generalizability issue in psychological science (Tiokhin et al., 2019; Moshontz et al., 2018; Uhlmann et al., 2019). Big team science[[1]](#footnote-1) is loosely defined as projects which involve a relatively large number of collaborators spread across different labs, cultures, and regions, all of whom pool resources to complete larger, more ambitious projects than they could individually (Baumgartner et al., 2023; Forscher et al., 2022). Because researchers from different countries are involved in collecting data locally, big team science usually has a large number of participants from dozens of countries/regions, which generate much stronger impact than studies conducted by singles teams (Nogrady, 2023). Networks for big team science grew rapidly in psychology, such as the Psychological Science Accelerator (Moshontz et al., 2018), ManyLabs (Ebersole et al., 2016; Klein et al., 2018), ManyBabies (Byers-Heinlein et al., 2020), EEGManyLabs (Pavlov et al., 2021), and EEGManyPipelines (Trübutschek, 2022). Using these massive grassroots networks, the Open Science Collaboration (2015), for example, provided compelling evidence for the reproducibility problem in psychological science. Another example is the Psychological Science Accelerator, whose COVID-19 project collected a large dataset for testing the effect of emotional regulation, nudging, and the communication of health information (Dorison et al., 2020; Psychological Science Accelerator Self-Determination Theory Collaboration, 2022; Wang et al., 2021).

Focusing on the number of countries/regions without looking into the heterogeneity within each region (Ghai, 2021), findings from big team science might create an “illusion of generalizability” (Ghai et al., 2023). Indeed, published big team science often claim diversity of their data and generalizability or globalization of their findings. For example, the report from Many Labs 2 wrote in the abstract that "little heterogeneity between Western, educated, industrialized, rich, and democratic (WEIRD) cultures and less WEIRD cultures". Similarly, based on data collected in 61 countries/regions, Ruggeri et al. (2022) claimed that “temporal discounting is a robust, global, and generalizable effect”, Buchanan et al (2023) wrote that the Psychological Science Accelerator’s COVID-19 rapid-response dataset include “a diverse, global sample”. These claims about diversity and generalizability, however, may not be supported by the data. A re-analysis of the data from Many-Lab 2 showed that its samples were mostly from WEIRD people worldwide (Schimmelpfennig et al., 2023). Similarly, Ghai et al (2023) re-analyzed data from Ruggeri et al. (2022) and found that the samples from different countries were not representative of their own country's population, especially in terms of age and education level. A meta-science focused on Chinese psychological samples also found that Chinese samples in big team science are not representative of the Chinese adult population when using the census data as a reference (Yue et al., 2023). As for the authors’ diversity, although big team science includes authors from different countries, the leading authors (first authors and senior authors) are dominated by teams from "WEIRD'' countries, and researchers from non-WEIRD countries only help to collect data (Adetula et al., 2022; Coles et al., 2022; Forscher et al., 2022; Ghai et al., 2023). As researchers tend to investigate the topics that are familiar or of interest to them (Medin et al., 2017), and the topics that of interest to researcher in non-WEIRD countries may systematically differ from those in WEIRD countries (Adetula et al., 2022), big team science led by researchers from WEIRD countries, mainly focus on the topics that of interest to researchers from WEIRD countries. This bias in research topics is not only unfair but also leads to the "unknown unknown" problem mentioned above. These studies suggested that big team science, with its big team of authors and large sample size, may not have solved the diversity and representativeness problem yet as it promised.

More importantly, we should not expect the problem to be solved without much changes in research investment. As Forscher et al (2022) have pointed out, “insufficient investment of intellectual and nonintellectual resources into the typical psychology study” is the common cause for many challenges in psychology. Although big team science involved much more researchers in each project, the investment scheme was not changed: big team science is typically crowdsourced without much additional financial support for each participating team (e.g., Coles et al., 2022). In other words, the scarcity of non-intellectual resources faced by researchers from developing countries/regions has not changed (UNESCO, 2021a). The big team science studies offloaded financial costs onto the "data collectors" in exchange for authorship of the data collectors on the final paper. This arrangement may have filtered out many global south researchers who don't have established capacity for data collecting in the first place. For instance, capacity surveys conducted by the Psychological Science Accelerator showed that most labs are from the global north (Kijilian et al., 2022; Paris et al., 2020). Thus, socioeconomic factors, such as GDP per capita, research & development (R&D) Investment, number of universities, number of psychology researchers, urbanization, average years of schooling, may correlate with the participation of big team science. Also, because most big team science recruited collaborators and collected data using the internet, especially English social media for the former, it is possible that many cultural factors played a role in determining who are the collaborators and who are the participants. These factors may include English proficiency, cultural closeness to the WEIRD regions, internet penetration rate, and globalization of the economy of the local country/region. In short, even though big team science had mobilized much greater non-intellectual resources for those projects, the costs of making such investment vary by local socioeconomic and cultural factors, which are unchanged.

However, there is no comprehensive assessment of the diversity and representativeness of big team science samples and authors. The relationships between socioeconomic/cultural factors and engagement of big team science across regions are unknown either. To bridge this gap, we plan to first examine the samples and authors of big team science[[2]](#footnote-2) by comparing them to two different reference data sources: samples in traditional psychological studies and the target population of big team science (the world population or a specific population). The former will be non-big team science articles published in *Psychological Science* in 2023 (hereafter *PS2023*), and the latter will be census data for each country, United Nations Population Division, or World Bank. Then, we explored several socioeconomic/cultural factors that might be related to the lack of presence for some countries/regions. The results of the current study will provide insights for big team science to move toward a fairer science that contributes to all, regardless of region.



*Fig 1*. Overview of the current study. PS2023, the non-big team science articles published in *Psychological Science* in 2023; BTS, big team science; Question 1 & 3, are the samples and authors from BTS more diverse and representative than those from PS2023? Question 2 & 4, how representative are the samples and authors from BTS as compared to the world population?

**2 Method**

2.1 Data sources

Our target data will be all available big team science studies in psychology. The big team science is defined as “a method involving a relatively large number of collaborators who may be dispersed across labs, institutions, disciplines, cultures, and continents” (Forscher et al., 2023). However, there is no operationalization for how large is “a relatively large number”. Also, big team science may be used interchangeably with “crowdsourced studies'' and sometimes with “big-sample” studies. Here we distinguish these three types of studies from six aspects (Table 1). In most cases, crowdsourcing science can be used interchangeably as big team science, with one exception: when participating teams only involved in data analyses, such as NARPS (Botvinik-Nezer et al., 2020), EEGManyPipelines (Trübutschek, 2022). On the other hand, big team science is different from big sample studies, but they share two shared features: the large number of participants and participants’ countries/regions. Big-sample study might be led and finished by one or two teams, with help of their funding and resources, which is different from big team science (e.g., Coles et al., 2022).

Here we define big team science as studies with crowdsourcing data collection, a big team of authors (n >= 30), big N of authors’ affiliations (>=5), big N of participants (>= 1000). Typical big team science including Many Labs projects (Ebersole et al., 2016; Ebersole et al., 2020; Klein et al., 2014; Klein et al., 2018; Klein et al., 2022), the Human Penguin Project (Hu et al., 2019; IJzerman et al., 2018), and all completed projects from PSA (Chen et al., 2018, Jones et al., 2021; Wang et al., 2021). Other cross-cultural studies with big-team are also considered as big team science. We have compiled a list of big team science and big sample studies, please see [here](https://docs.google.com/spreadsheets/d/18lF1KohchjKR5nM2utvlnd9A-zc3iKyf4EXLJ_jMPgk/edit#gid=1741341307)[[3]](#footnote-3) for a complete list that we compiled. We also included some Big-sample studies finished in the traditional scientific approach, which will serve as a supplementary comparison to the big team science.

Table 1. Aspects of studies that may distinguish three types of “big” sciences.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Crowdsourcing data analysis | Crowdsourcing data collection | Big # of authors | Big # of authors’ countries | Big # of participants |
| Big Team Science | - | √ | √ | √ | √ |
| Crowdsourcing science | √ | √ | √ | - | - |
| Big-Sample/data science | - | - | - | - | √ |

The first reference data will be all non-big team science articles published in *Psychological Science* in 2023. We chose *Psychological Science* not only because of its prestigious position in the field, but also because it openly endorses the value of reproducibility and diversity (Bauer, 2022). In addition, *Psychological Science* encourages open data, and the percentage of open data is relatively high (e.g., 78% of authors disclosed their data in 2021) and gradually increasing over the years (Bauer, 2022). These open data may provide raw data that contains participants’ characteristics.

The second reference data, the world population census data, will come from several sources. We will primarily search the United Nations Population Division (its data is based on census data for each country; United Nations Population Division, 2022) and The World Bank (https://data.worldbank.org.cn/). In case the data is too old (before 2012), we will search for more recent data from other authentic sources, such as the government agencies for census or statistics of the countries. The source of data will be recorded in detail in our R scripts.

**2.2 Data extraction**

**2.2.1 Data extraction for PS2023**

We will extract data from PS2023. Since some papers from PS2023 shared raw data with demographic variables but some did not, we will distinguish these two types of studies when extracting data.

For studies with shared raw data (publicly available or available after emailing the authors) with demographic information of participants, we will extract data from the raw data.

For studies without raw data, we will extract data from the full text of the paper and supplementary materials. The extraction will follow these three stages: pre-encoding, encoding, and proofreading.

In the pre-coding stage, we will first develop the initial version of the coding manual (see https://github.com/Chuan-Peng-Lab/BTS\_Sample\_Stage1\_RR) based on previous research (Arnett, 2008; Nielsen et al., 2017; Pollet & Saxton, 2019; Rad et al., 2018; Yue et al., 2023). Then, at least two coders will code 5 randomly selected studies from the articles, compare the coding results, resolve differences, and revise the manual. After that, they will code another 5 randomly selected articles to improve the coding manual further. We will iterate this procedure until the differences between the two coders is negligible. The coding manual will be the formal one for coding the rest of articles.

Once the formal coding manual is established, we will re-code the previous articles based on the formal coding manual and code all the rest empirical studies in PS2023. All trained coders will code approximately 120 articles. Each article will be coded by at least two coders.

The sample information we extracted includes: article number, journal source, article title, number of studies, type of study, sample type, sample size, method of recruitment. More importantly, we will extract all the sample information if available, including gender, age, socioeconomic status, education level, race, occupation, religion, nationality, urban-rural status and the area where the subjects were recruited, if available. In addition, we will read abstracts and discussions (or conclusions) to extract the target population to which the studies intended to generalize and the statements on constraints of generalizability (Hoekstra & Vazire, 2021). Based on our previous experience (Yue et al., 2023), identifying the target population to which the original authors tried to generalize is difficult, we will pay special attention to this part of the manual and make it concrete.

To ensure the accuracy and objectivity of the coding process, two independent coders will conduct internal discussions after the first coding to resolve coding errors that were not coded according to the coding manual. The third coder will rate the coding consistency of each article from 0 to 1 (0 represents completely different, 1 represents the same). The consistency of the scores will be used to calculate the internal reliability of the scoring using the R package *irr* (Gamer et al., 2019). Finally, if there are still inconsistencies in the coding between the two coders, they will be resolved through collective discussion.

**2.2.2 Data extraction for big team science**

We will use the same strategy to code the big team science study as the empirical study published in *PS2023*. For big team science with demographic information in the shared raw data, we will directly use the information in the raw data; for big team science without demographic information in the shared raw data or without shared raw data, we will extract data from the article and supplementary using the same coding manual we used to extract data from the *PS2023* papers.

**2.3 Data analyses**

**2.3.1 Data preprocessing and visualization**

We will use R 4.3.1 (R Development Core Team 2023) for data preprocessing, visualization, and data analyses.

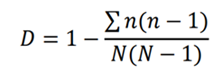
**2.3.2 Hypotheses testing**

We will use the Bayes factor to compare whether participants from two data sources are different on certain dimensions (e.g., sex ratio, age distribution) because Bayes factors can provide evidence for both null hypothesis and alternative hypothesis (Dienes, 2016; Dienes & Mclatchie, 2018; Hu et al., 2018; Wagenmakers et al., 2018). Specifically, we use the Bayesian multinomial test (equivalent to the frequentist goodness-of-fit test or *χ2*) to compare the sample characteristics from two data sources (sex ratio, age distribution, education level, SES, etc.). As in Yue et al (2023), the Bayesian multinomial test with non-informative prior of Dirichlet distribution and sample size *N* >= 1200 allows us to obtain evidence for the null hypothesis or the smallest effect size of interest with more than 80% of the chance.

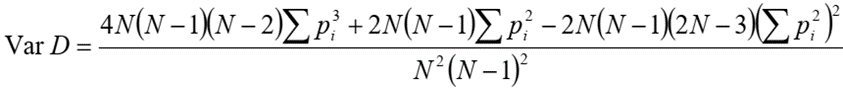
For our first question, is the sample characteristics in the big team science similar to the *PS2023*? We will visualize the gender, age, education, race, country/region, and other demographic information of participants from *PS2023* and big team science. Then, we will use Bayesian multinomial tests for each dimension of the samples.

For our second question, the extent to which big team science participants are representative of the world population, we compared big team science participants to population data. We visually compared big team science participants and the world population data in terms of gender, age, SES (if available), and geospatial distribution. Then, we will use Bayesian multinomial tests for each dimension of the samples as mentioned above.

For our third question, whether the national diversity of authors in big team science is similar to the empirical research in *PS2023*. More specifically, we will analyze the diversity for all and leading authors, the latter is defined by first author(s), corresponding author(s), and last author(s). Because big team science challenges traditional authorship by its large number of authors and potential different ways of contributions (Coles et al., 2023), we will also identify leading author(s) in the authors’ notes (e.g., project administration listed in CREDIT). For both types of authors, we will visualize and compare the national diversity of all or leading authors between the empirical research in *PS2023* and big team science. We use Simpson's diversity index (Simpson, 1949) to quantify the national diversity of authors. Simpson's diversity index estimates how likely two random samples drawn from a population belong to different groups or categories, given a fixed number of groups in the population. In our case, Simpson's diversity index, *D*, represents how likely two random authors drawn from all authors or leading authors are to be from two different countries. If the value of *D* is high, it means that the authors are more evenly distributed across different countries, otherwise it means that the authors are more likely to be from a few countries. The formula for Simpson's diversity index is as follows:

 (1)

in which *n* is the number of authors in each country and *N* is the total number of authors. The value of *D* is between 0 and 1, with higher values indicating greater diversity. After calculating the Simpson's diversity index for *PS2023* and big team science, they are compared using the diversity *t*-test as implemented in R. The diversity *t*-tests based on variance estimation (Brower et al., 1998), the variance estimation formula is as follows:

(2)

in which *pi* is the proportion of authors in each country, and *N* is the total number of authors. And we calculated the Bayesian factor based on the *t*-value and sample size (Ly et al., 2018). In addition, we will also combine visualization and Bayesian multinomial tests (Yue et al., 2023) to test the national diversity by comparing the proportion of developed countries/developing countries/least developed countries to which *PS2023* and big team science first and corresponding author institutions belong.

For the fourth question, whether the national diversity of authors in big team science is similar to the diversity of the world population. More specifically, we will visualize and compare the diversity for both all or leading authors of big team science and the world population. We also use Simpson's diversity index (Simpson, 1949) to quantify the world population diversity by country (in formula (1), *n* is the number of people in each country, and *N* is the total number of people). After calculating Simpson's indices for the world population, they are compared using the diversity *t*-test as implemented in R. We will calculate the Bayesian factor based on the t-value and sample size, and the prior is a two-tailed Cauchy distribution with scale *r* = 0.707 (Ly et al., 2018). In addition, we will also combine visualization and Bayesian multinomial tests (Yue et al., 2023) to test the national diversity by comparing the proportion of developed countries/developing countries/least developed countries to which the leading author institutions of big team science and the world population belong.

**2.3.3 Exploratory analyses**

To explore the impact of socioeconomic and cultural factors on the presence of a country in psychological research, we will further gather country-level socioeconomic and cultural factors and explore the relation between these factors and proportion of participants/authors in *PS2023* and big team science. Specifically, for socioeconomic factors, we will examine the following two factors: GDP per capita, R&D investment, number of universities per 100,000 people, number of psychology researchers per 100,000 people, urbanization, average years of schooling, globalization, and internet penetration rate; for cultural factors, we will examine the following cultural distance, linguistic distance, and language barriers (see Table 2). Because the dependent variable is proportional data based on count, overdispersed, and contains a large number of zeros, we will use Bayesian zero-inflated beta-binomial regression model for these analysis (Douma and Weedon, 2019), vias the R package *brms* (Bürkner, 2017). Since the *b* coefficient usually has a flat prior, we will set the *b* to a normal distribution with a mean of 0 and a standard deviation of 1 in prior distribution (Heiss, 2021), and intercept uses the default prior.

Table 2. Country level socioeconomic and cultural variables and corresponding data source

|  |  |
| --- | --- |
| **Country-level variable** | **Data** |
| GDP per capita | GDP per capita (from The World Bank, 2022a) |
| R&D investment | Research expenditure as a share of GDP (from UNESCO, 2021b) |
| Globalization level | Globalization index (from KOF Swiss Economic Institute, 2022) |
| Number of universities per capita | Number of universities per 100,000 people (from Webometrics Ranking of World Universities, 2023) |
| Number of psychology researchers per capita | Number of psychology researchers per 100,000 people (from International Union of Psychological consists, 2023) |
| Urbanization level | Proportion of urban population (from The World Bank, 2022b) |
| Average years of schooling | Average years of formal education for individuals aged 15-64 (from Our World in Data, 2023) |
| Internet penetration rate | Proportion of population using the internet (from The World Bank, 2022c) |
| Cultural distance from the United States | Cultural distance from the United States (from Muthkrishna et al., 2020) |
| Linguistic distance from the United States | Linguistic distance from the United States (from Melitz & Toubal, 2014) |
| English proficiency | English proficiency rank (from Education First, 2022) |

**3 Open Science Practices**

We will report all details of data extraction, data exclusion (if applicable), and data analytical choices in the manuscript or supplementary materials. All data extracted from paper will be open at Open Science Framework or other platforms. All scripts will be open at Github (https://github.com/Chuan-Peng-Lab/BTS\_Sample\_Stage1\_RR).

**4 Results**

**4.1 Overview of participants**

For XXX papers published in *Psychological Science* in 2023, there are XXX participants from XXX countries (Fig 2 A), with XXX authors from XXX institutes. There are *N* = XXX in XXX big team science, from XXX countries/regions (Fig 2 B), with XXX authors from XXX institutes.

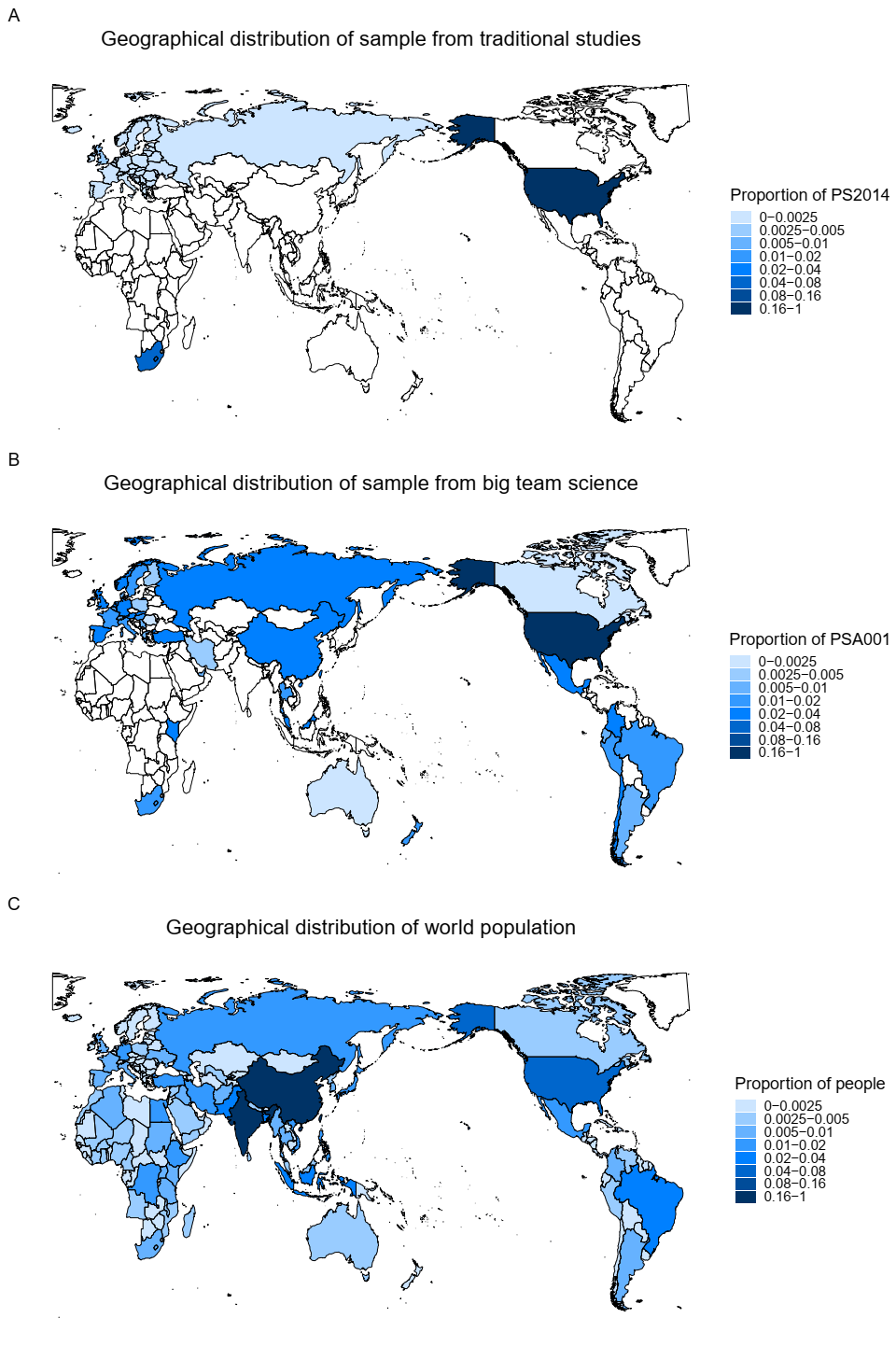


Figure 2. Geographical distribution of samples from traditional studies (A), BTS (B), and the world population (C). [*Note panel A will be replaced with data from PS2023 after collecting data; panel B will be replaced with data from all BTS after collecting data.*]

**4.2 Sample diversity and representativeness**

**4.2.1 Comparing with BTS and *PS2023***

4.2.1 Comparing with big team science and *PS2023*

[*We will compare the sample characteristics between big team science and PS2023 in the following dimension: sex ratio, age distribution, educational level, and geographical distribution. We will also explore the potential differences in racial/ethnicity and SES if possible. Below are the preliminary results (Figure C), we used data from Rad et al., (2018) for illustrating PS2023 and PSA001[[4]](#footnote-4) (Jones et al., 2021) for illustrating big team science.*

*More specifically, for data from Rad et al., (2018), we used data from 35 studies reported in 21 papers in the supplementary material. Data were coded by Yue et al. (2023) and reported in their replies to reviewers, including sample size, sex ratio, mean age, and standard deviation of age. We estimated the number of participants in each age bin based on mean and standard deviation of age (as in Yue et al., 2023) from each paper (In addition to a paper with original data on demographic variables).*

*First, we compare the sample characteristics from Rad et al (2018) and big team science in sex ratio and age distribution (Figure 3).*

*Second, we compare the sample characteristics from Rad et al (2018) and big team science in geographical distribution (Figure 2A & 2B) and population distribution (Figure 4A).*]

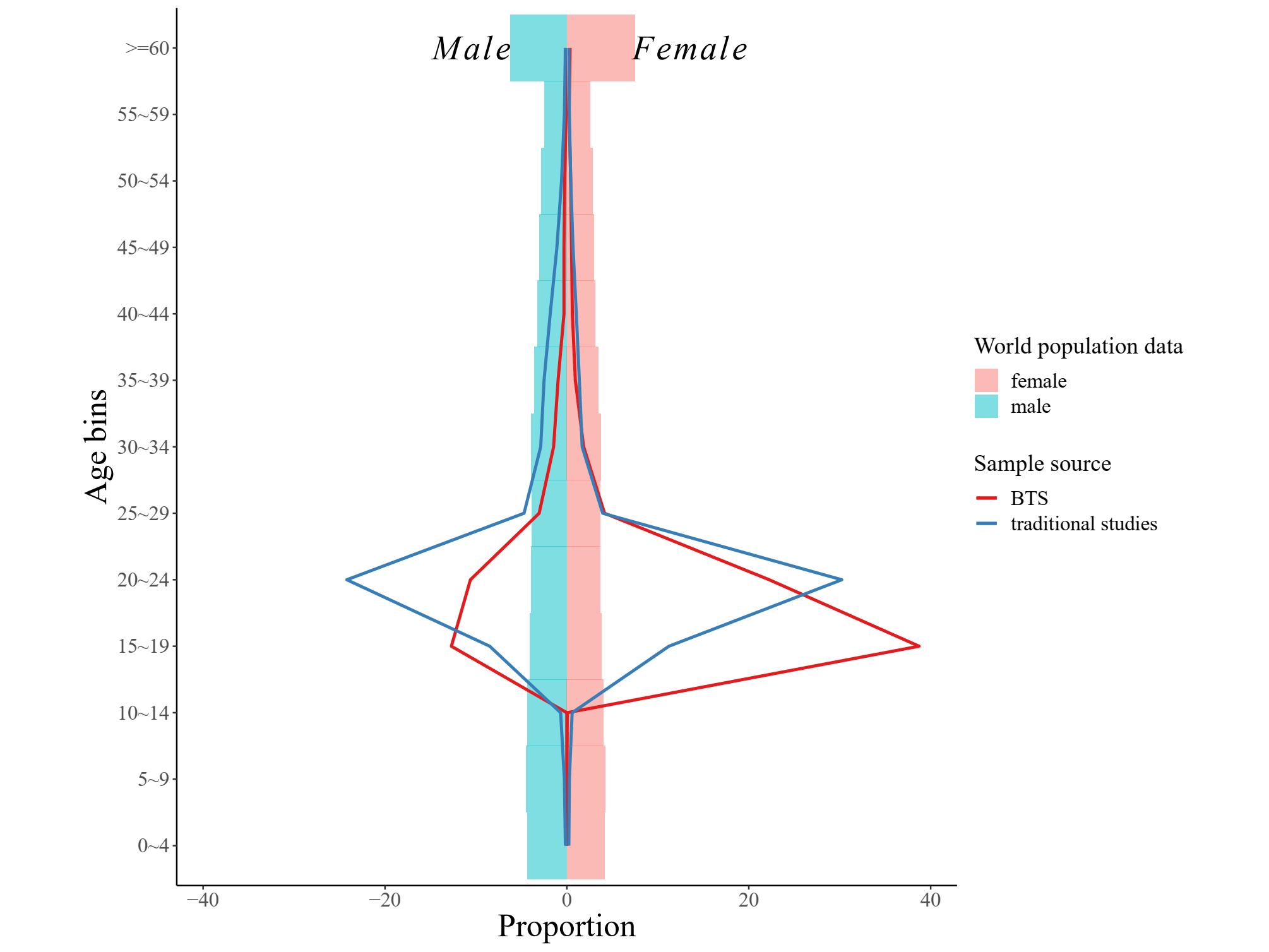


Figure 3. Preliminary results of sample’ sex ratio and age distributions from traditional studies (using Rad et al 2018’s data), and the big team science (using PSA001’s data), and the world population data. The Y-axis is the age bins, and the X-axis is the proportion. The blue lines represent the sample of traditional studies (using Rad et al 2018’s data); the red lines represent the big team science samples (using PSA001’s data), and the pyramid chart represents the world population data. The left side of the picture is the male and the right is the female.

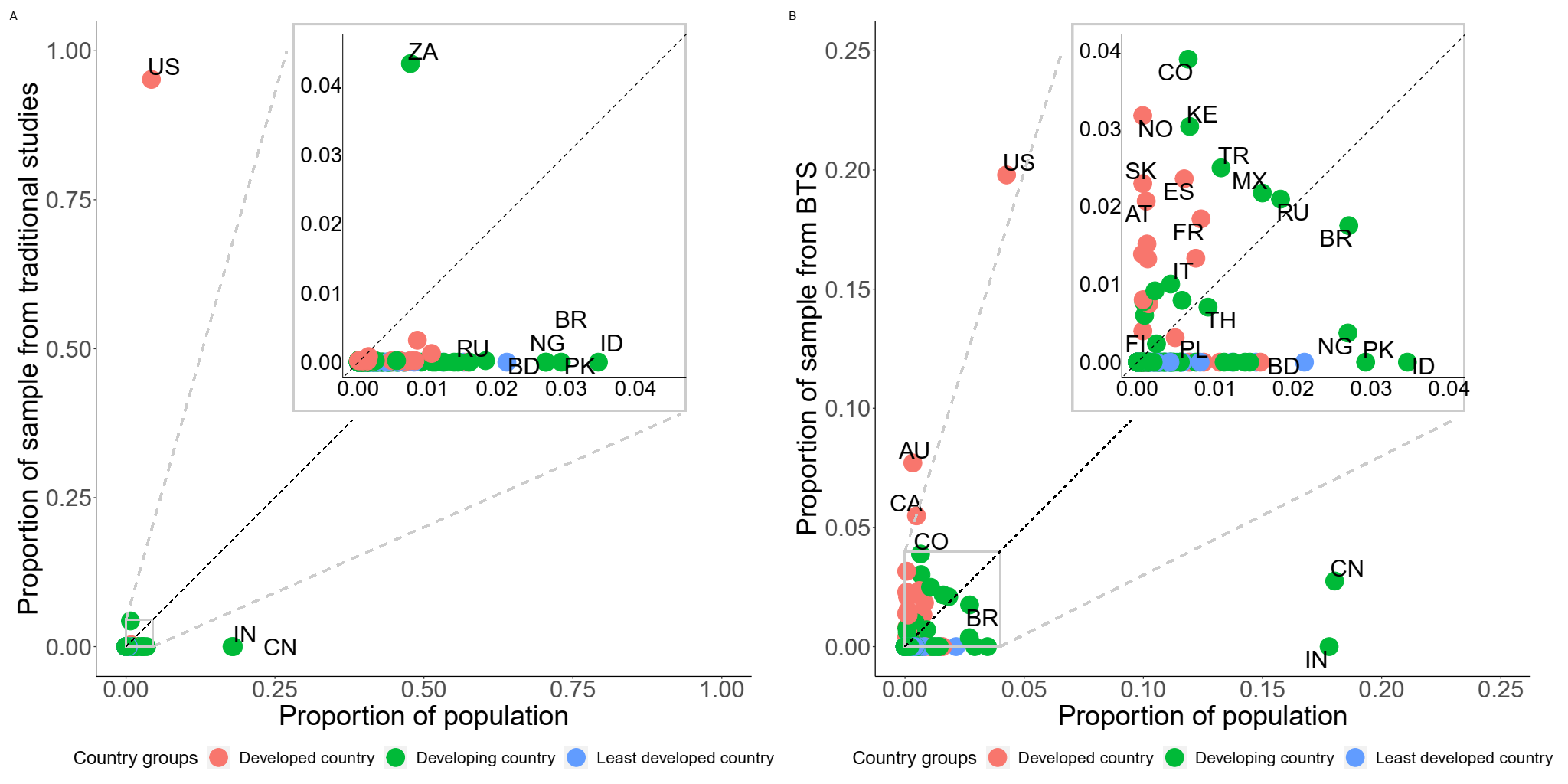


Figure 4. Preliminary results of sample’ country distributions from the traditional studies (using Rad et al 2018’s data), and the big team science (using PSA001’s data), and the world population data. The x-axis represents the proportion of each country's population in the world's population, and the y-axis represents the proportion of each country's sample. (A) Comparing the samples’ country distributions from the traditional studies (using Rad et al 2018’s data) to the population census data (year 2022). (B) Comparing the samples’ country distributions from the big team science (using PSA001’s data) to the population census data (year 2022).

**4.2.2 Comparing with big team science and the world population data**

[*We will compare the sample characteristics from big team science to the world population in the following dimension: sex ratio, age distribution, educational level, and geographical distribution/country distribution. We will also explore the potential differences in racial/ethnicity and SES if data is available. Below are the preliminary results, we used data from PSA001 for big team science and World Population Prospects for world population.*

*First, we compare the characteristics of big team science samples and the world population data in sex ratio and age distribution (Figure 3).*

*Second, we compare the characteristics of big team science samples and the world population data in geographical distribution (Figure 2B & 2C)/population distribution (Figure 4B) by illustrating the proportion of each country's participants in big team science and the corresponding proportion of each country’s population in the world population.*]

**4.3 Authors’ diversity and representativeness**

[*We will compare the geographical distribution and diversity of all authors, and first/corresponding authors' institutions from between big team science and PS2023. Below are the preliminary results (Figure C), we used non-big-team science articles published in Psychological Science in 2014(hereafter PS2014) data from Rad et al., (2018) for illustrating PS2023 and PSA001 for illustrating big team science.*]

4.2.1 Comparing with all or leading authors of big team science and *PS2023*

*First, we will compare the geographical distribution of all authors' institutions from PS2023 (Figure 5A) and big team science (Figure 5B), and compare the geographical distribution of leading authors' institutions from PS2023 (Figure 6A) and big team science (Figure 6B).*

*Second, we will compare the Simpson's diversity index of big team science all authors' institutions with all authors' institutions from traditional studies. The results revealed the diversity of all author institutions in big team science (D = 0.94) is significantly higher than PS2023(D = 0.52; t = 6.46, BF10 = 1768972).*

*Third, we will calculate and compare the Simpson's diversity index of leading authors for both big team science and traditional studies. The results revealed the diversity of leading author institutions in big team science (D = 0.92) is significantly higher than PS2023(D = 0.40; t = 4.37, BF10 = 163.33).*]

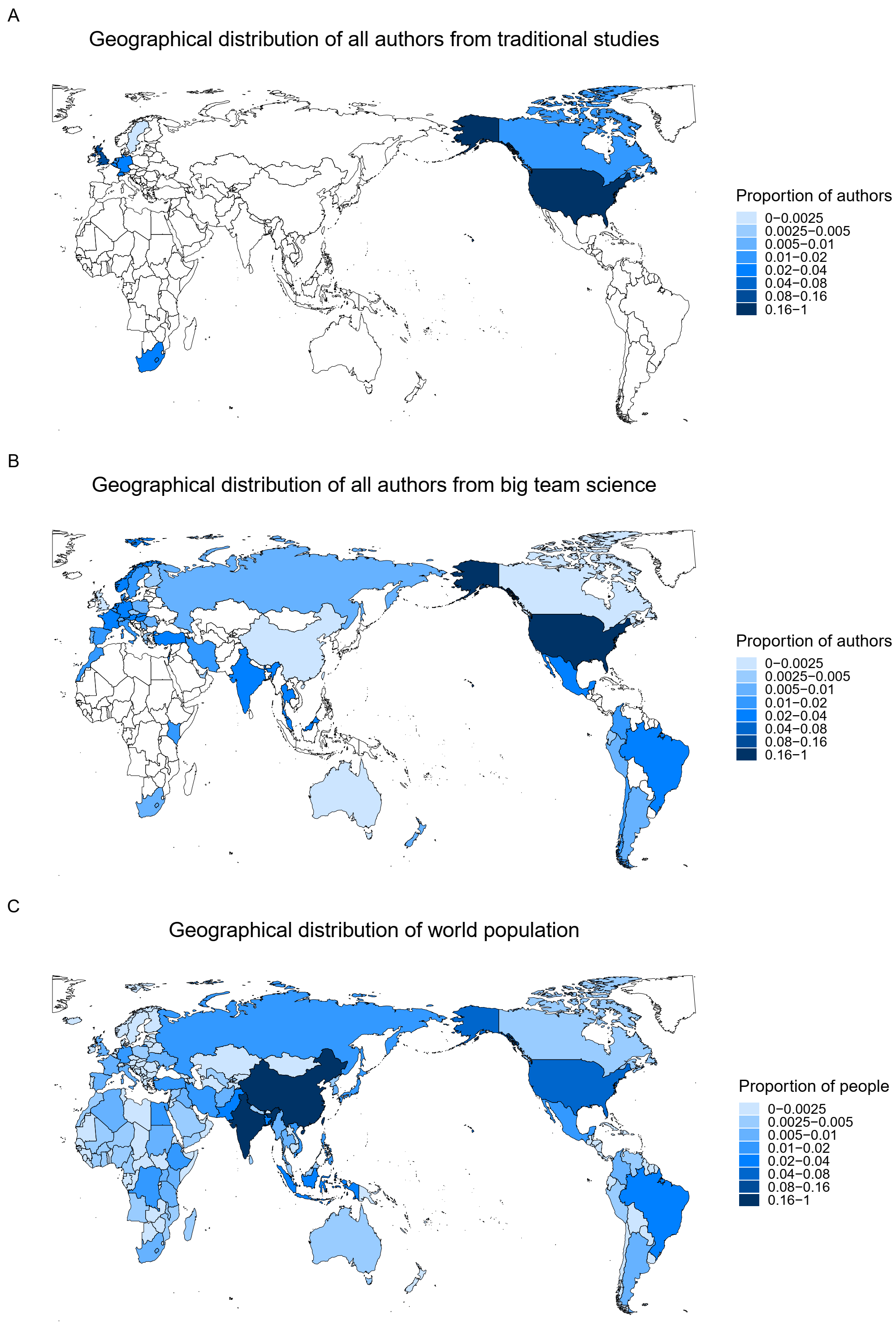


Figure 5.Geographical distribution of all author institutions from traditional studies (A), all author’ institutions from big team science (B), and the world population (C). [*Note panel A will be replaced with data from PS2023 after collecting data; panel B will be replaced with data from all big team science after collecting data.*]

4.2.2 Comparing all or leading authors of big team science with the world population data

[*First, we will compare the geographical distribution for both all authors' institutions from big team science (Figure 5B) and the world population data (Figure 5C), and compare the geographical distribution for both leading authors' institutions from and big team science (Figure 6B) and the world population data (Figure 6C).*

*Second, we will calculate and compare the Simpson's diversity index for both all authors of big team science and the world population. The results revealed the national diversity of the world population (D = 0.93) is significantly lower than the diversity of all author institutions in big team science (D = 0.94; t = 2.50, BF10 = 1.49).*

*Third, we will calculate and compare the Simpson's diversity index for both leading authors of big team science and the world population. The results revealed the difference between the national diversity of the world population (D = 0.93) and the diversity of author institutions in big team science (D = 0.92) has not reached statistical significance (t = -0.19, BF10 = 0.11).*]

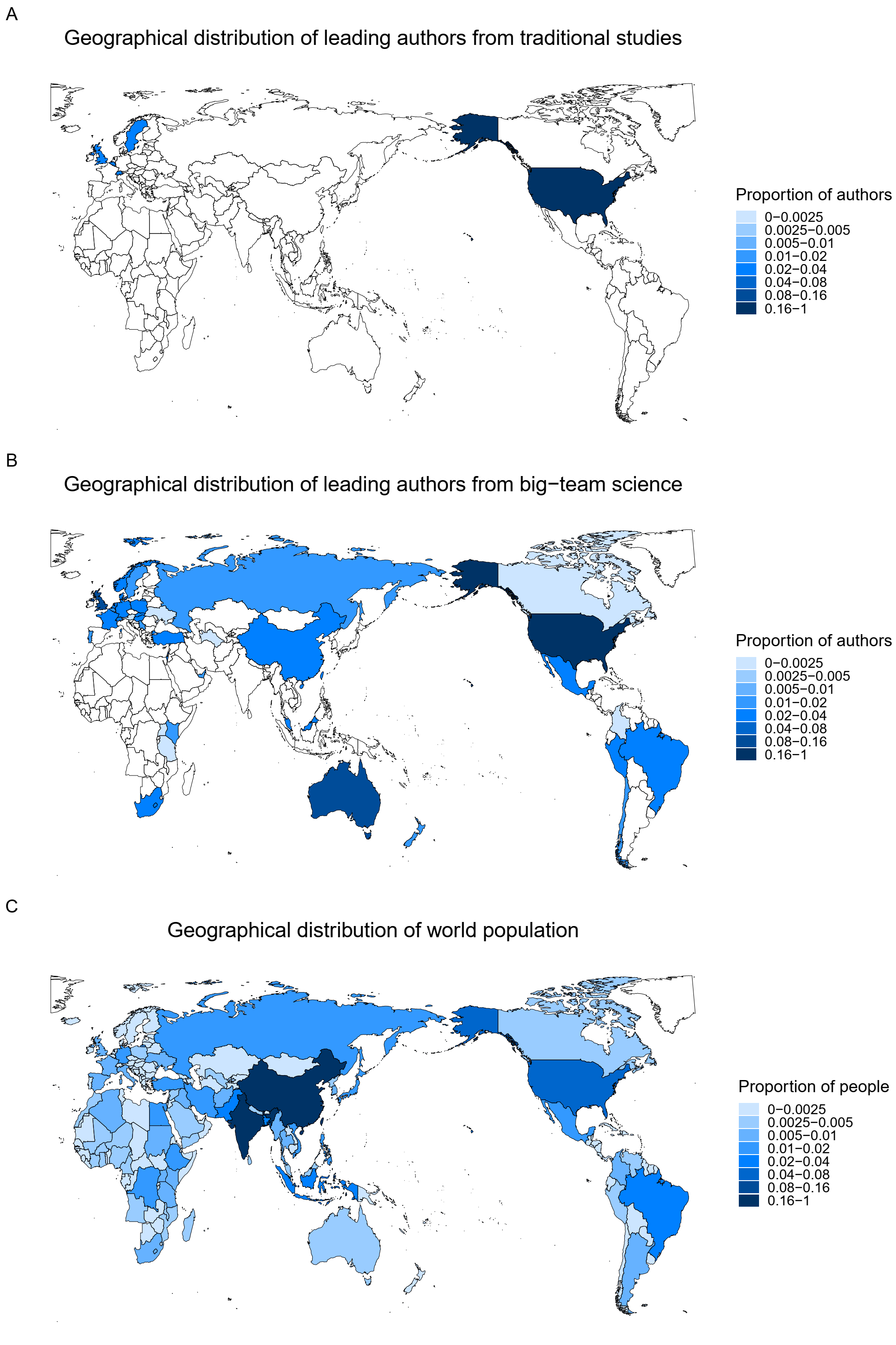


Figure 6.Geographical distribution of leading author institutions from traditional studies (A), leading author’ institutions from big team science (B), and the world population (C). [*Note panel A will be replaced with data from PS2023 after collecting data; panel B will be replaced with data from all big team science after collecting data.*]

**4.4 Exploratory Analysis**

[*We will report the results of exploratory analysis, including the relationship between the proportion of participants/authors and GDP per capita, R&D investment, number of universities per 100,000 people, number of psychology researchers per 100,000 people, Average years of schooling, Urbanization, Globalization, Internet penetration rate, cultural distance from the United States, linguistic distance from the United States, and English Proficiency of each country/region (Table3 & Figure 7). To explore the heterogeneity within countries, we also used three countries as examples and explored the relationship between the proportion of participants/authors in each state/province and the* *Urbanization, Internet penetration rate, and English Proficiency of each state/province in Nigeria, China, and the United States (Table4 & Figure 8). The purpose of presenting results here is to demonstrate how our analytical scripts work. The results will be replaced after collecting data.*]

Table 3. Regression analysis between country level socioeconomic/cultural factors and proportion of samples in traditional studies and big team science

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *PS2014* | | | Big team science | | |
|  | *Estimate* (*b*) | 95% *CI* | *BF10* | *Estimate* (*b*) | 95% *CI* | *BF10* |
| GDP per capita | 6.18 | [2.36, 12.40] | >3999 | 2.75 | [0.94, 46.80] | >3999 |
| R & D investment | 40.80 | [15.60, 90.90] | >3999 | 92.90 | [39.50, 209.00] | >3999 |
| Number of universities per capita | 34.20 | [5.87, 62.50] | >3999 | 7.09 | [3.94, 12.10] | >3999 |
| Number of psychology researchers per capita | -2.16 | [-13.20, 21.80] | 0.66 | 2.34 | [0.51, 43.60] | 1332.33 |
| Average years of schooling | 51.60 | [-0.01, 140.00] | 159 | 23.20 | [8.30, 44.20] | >3999 |
| Urbanization level | 5.45 | [-1.64, 44.90] | 21.73 | 2.39 | [0.52, 16.30] | 128.03 |
| Globalization level | 3.07 | [-5.21, 8.96] | 5.63 | 4.63 | [2.61, 7.62] | >3999 |
| Internet penetration rate | -0.59 | [-7.63, 5.68] | 0.68 | 2.87 | [1.15, 5.20] | >3999 |
| Cultural distance from the United States | -42.00 | [-96.20, -12.90] | >3999 | -327.00 | [-758.00, -105.00] | >3999 |
| Linguistic distance from the United Statesa | 5.35 | [-11.70, 15.50] | 0.27 | 18.50 | [-12.60, 46.20] | 0.13 |
| English Proficiency | -1.38 | [-37.08, 104.00] | 3 | 43.50 | [-3.47, 99.20] | 1.00 |

Note. Estimate (*b*) is the median of the posterior distribution of regression coefficients

a There are two methods for calculating language distance, one based on the Ethnology language trees (Fearon, 2003; Laitin, 2000) and the other based on the Automated Similarity Judgment Program (https://asjp.clld.org). The results shown in the table are based on the data from the former calculation method, while the data from the latter calculation method are similar.

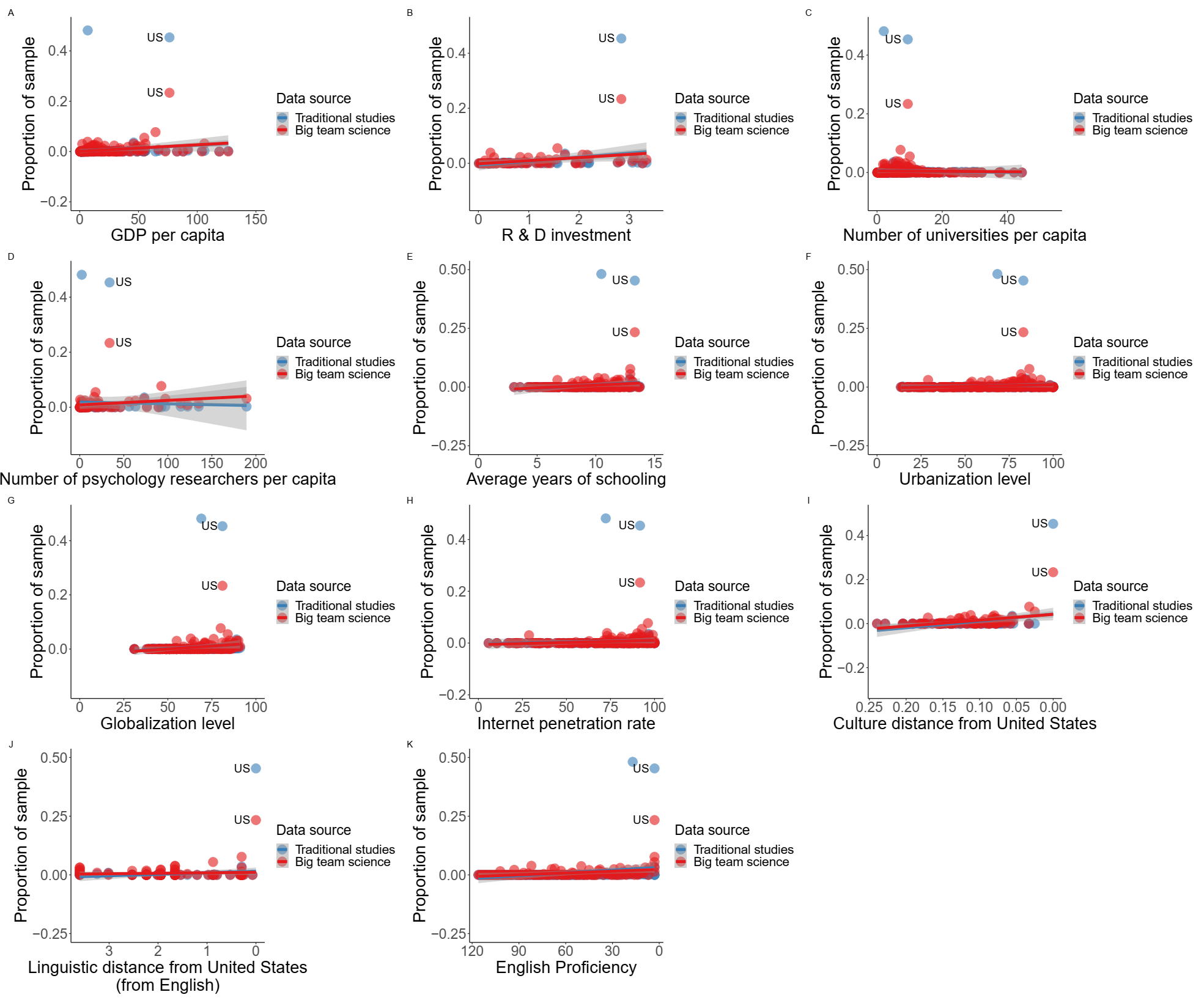


Figure 7. Country-level relationship between proportion of sample and socioeconomic/cultural factors for traditional studies (blue, using data from *Rad et al., (2018)* as an example) and big team science (red, using PSA 001 as an example). (A) Proportion of sample size and GDP per capita in traditional studies and big team science. (B) Proportion of sample size and R&D investment for traditional studies and big team science. (C) Proportion of sample size and number of universities per capita for traditional studies and big team science. (D) Proportion of sample size and number of psychology researchers per capita for traditional studies and big team science. (E) Proportion of sample size and average years of schooling for traditional studies and big team science. (F) Proportion of sample size and urbanization level for traditional studies and big team science. (G) Proportion of sample size and globalization level in traditional studies and big team science. (H) Proportion of sample size and internet penetration rate in traditional studies and big team science. I) Proportion of sample size and cultural distance from the United States in traditional studies and big team science. (J) Proportion of sample size and linguistic distance from the United States (from English) in traditional studies and big team science. (K) Proportion of sample size and English Proficiency in traditional studies and big team science.

Table 4. Correlation analysis between Urbanization level, Internet penetration rate and English Proficiency & proportion of participants across first-level administrative regions within participating countries (using Ruggeri et al. (2022) as an example).

|  |  |  |  |
| --- | --- | --- | --- |
|  | Big team science (using Ruggeri et al., (2022) as an example) | | |
|  | *Estimate* (*b*) | 95% *CI* | *BF*10 |
| **Urbanization level** |  |  |  |
| Kenyaa | 4.30 | [-20.40, 13.70] | 2.49 |
| China | 13.8 | [-11.30, 15.40] | 5.92 |
| United States | 4.91 | [-8.12, 22.30] | 2.82 |
| **Internet penetration rate** |  |  |  |
| Nigeria | -4.75 | [-32.40, 16.20] | 2.44 |
| China | 8.67 | [-28.00, 44.60] | 2.24 |
| United States | 4.96 | [-18.40, 21.60] | 1.55 |
| **English Proficiency** |  |  |  |
| Nigeria | -0.45 | [-1.71, 0.32] | 7.89 |
| China | -0.50 | [-1.46, 0.30] | 7.11 |
| Netherlandsb | 0.99 | [-1.42, 5.20] | 4.09 |

Note. Estimate (b) is the median of the posterior distribution of regression coefficients

a Here we use Kenya instead of Nigeria because we cannot find Nigeria's urbanization data by state.

b Here we use Netherlands instead of US because English is the native language of United States participants. BTS, big team science.

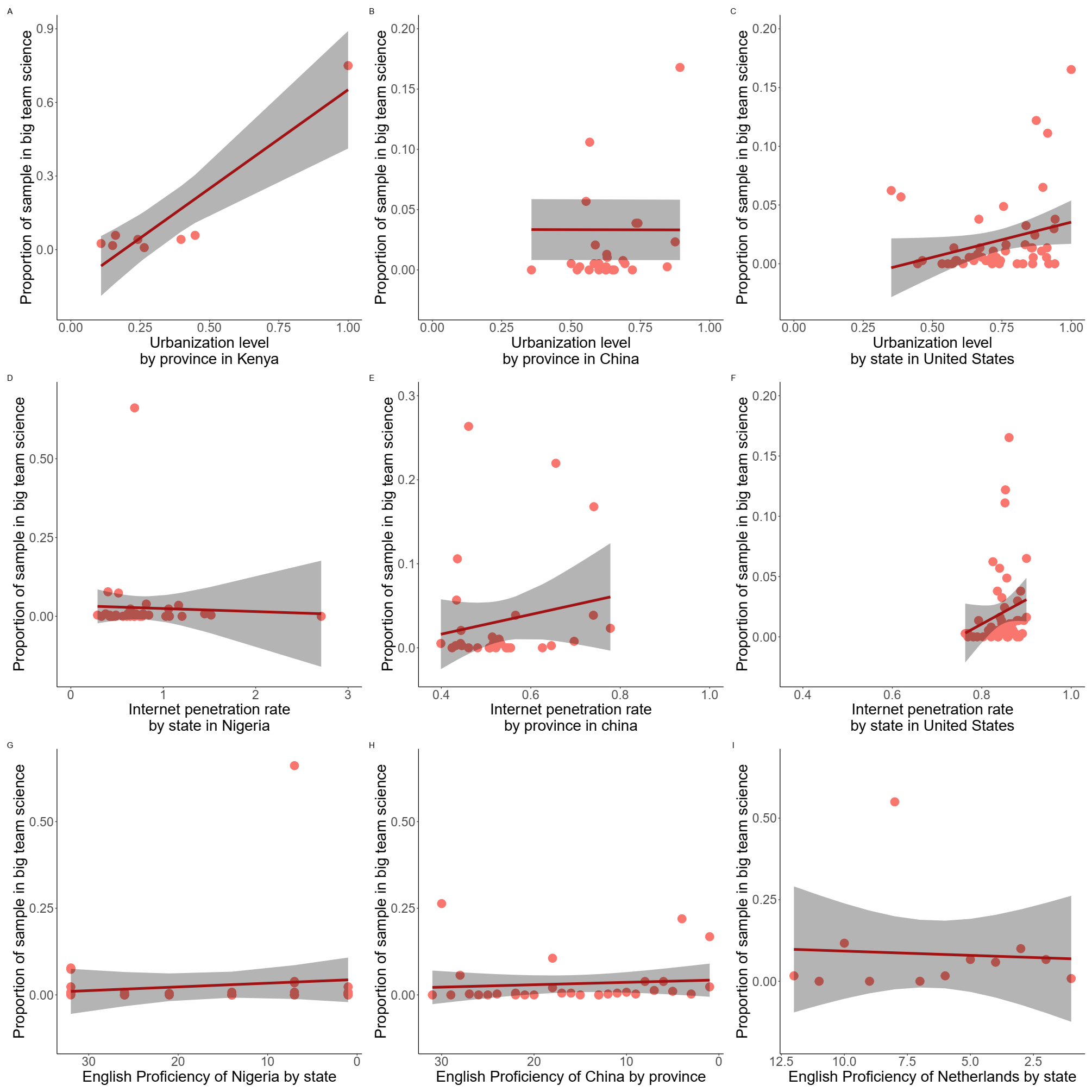


Figure 8. Relation between the proportion of each state’s/province’s sample in big team science (using Ruggeri et al. (2022) as an example) and Urbanization level in Kenya (A), China (B), and United States (C); Relation between the proportion of each state’s/province’s sample in big team science (using Ruggeri et al. (2022) as an example) and Internet penetration rate in Nigeria (D), China (E), and United States (F); Relation between the proportion of each state’s/province’s sample in big team science (using Ruggeri et al. (2022) as an example) and English Proficiency in Nigeria (G), China (H), and Netherlands (I).

**5 Discussion**

[Recap of the results]

[Limitations of the current study]

Table 1. Overview of research questions, hypotheses, analytical plan, and interpretations of the current study

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Analysis Plan** | **Interpretation given different outcomes** |
| 1. Is the sample from BTS more representative than samples from *PS2023*? | H1a1: The BTS have a more equal sex ratio than the *PS2023*.  H1a0: The BTS and the *PS2023* have the same sex ratio.  H1b1: The BTS has a more diverse age distribution than the *PS2023*.  H1b0: The BTS and the *PS2023* have the same age distribution.  H1c1: The BTS has a more diverse educational level than the *PS2023*.  H1c0: The BTS and the *PS2023* have the same Education level diversity.  H1d1: The BTS has a more diverse country distribution than *PS2023*.  H1d0: The BTS and the *PS2023* have the same country distribution. | We will combine visualization and Bayesian multinomial tests to test the hypothesis by comparing the sample characteristics in terms of sex ratio, age distribution, educational level, and country distribution of the *PS2023* and the BTS. We will also visualize the generalizability reports of the two sets of data. | H1a is supported if, compared to the *PS2023*, the BTS have a more equal sex ratio (visual inspection).  H1b is supported if, compared to the *PS2023*, the BTS have a more diverse age distribution (visual inspection). H1c is supported if, compared to the *PS2023*, the BTS have a more diverse educational level (visual inspection).H1d is supported if, compared to the *PS2023*, the BTS have a more diverse country distribution (visual inspection).    If BF10 > = 10, we infer there is relatively strong evidence that this hypothesis is supported. |
| 2. The extent to which participants in BTS are representative of the world population, compared with the world population data. | H2a1: The world population data has a more equal sex ratio than the BTS.  H2a0: The BTS and the world population data have the same sex ratio.  H2b1: The world population data has a more diverse age distribution compared to the BTS.  H2b0: Age distribution of the BTS and the world population data are the same.  H2c1: The world population data has a more diverse educational level than the BTS.  H2c0: Educational level of both the BTS and the world population data are the same.  H2d1: The world population data has a more diverse country distribution than the BTS.  H2d0: The country distribution of both the BTS and the world population data are the same. | We will combine visualization and Bayesian multinomial tests to test the hypothesis by comparing the sample characteristics in terms of sex ratio, age distribution, educational level, and country distribution of the BTS and the world population data. | H2a is supported if, compared to the BTS, the world population data has a more equal sex ratio (visual inspection).  H2b is supported if, compared to the BTS, the world population data has a more diverse age distribution (visual inspection).  H2c is supported if, compared to the BTS, the world population data has a more diverse educational level (visual inspection).  H2d is supported if, compared to the BTS, the world population data has a more diverse country distribution (visual inspection).    If BF10 > = 10, we infer there is relatively strong evidence that this hypothesis is supported. |

Table 1. (Continued)

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Analysis Plan** | **Interpretation given different outcomes** |
| 3. Whether the national diversity of all authors/first and corresponding authors in BTS are similar to *PS2023*? | H3a1: Compared to all authors of the *PS2023*, the BTS’s all authors have more national diversity.  H3a0: The country diversity of all authors of the *PS2023* and the BTS are the same.  H3b1: Compared to the first and corresponding authors of the *PS2023*, the BTS’s first and corresponding authors have more country diversity.  H3b0: The country diversity of the first authors and corresponding authors of *PS2023* and the authors of BTS are the same.  H3c1: Compared with *PS2023*, the first author and corresponding author of BTS have a larger proportion from developing countries or least developed countries.  H3C0: *PS2023* and BTS' first author and corresponding author have the same proportion from developing countries or least developed countries. | We will visualize the national diversity of authors between *PS2023* and BTS for all authors/first and corresponding authors.  We will also use the diversity *t*-test to compare the *PS2023* and the BTS.  We will combine visualization and Bayesian polynomial tests to test the hypothesis by comparing the proportion of developed countries/developing countries/least developed countries to which *PS2023* and BTS first and corresponding author institutions belong. | H3a is supported if, compared to all authors of the *PS2023*, the BTS’s authors have more national diversity (visual inspection).  H3b is supported if, compared to first and corresponding authors of the *PS2023*, the BTS’s first and corresponding authors have more national diversity. (visual inspection).  H3c is supported if, compared with *PS2023*, the first author and corresponding author of BTS have a larger proportion from developing countries or least developed countries.  If BF10 > = 10, we infer there is relatively strong evidence that this hypothesis is supported. |
| 4. Whether the national diversity of all authors/first and corresponding authors in BTS are similar to world population diversity by country? | H4a1: Compared to the world population diversity by country, the BTS’s all authors have less national diversity.  H4a0: All authors' national diversity of BTS and the national diversity of world population are the same.  H4b1: Compared to the world population diversity by country, the BTS’s first author and corresponding author have less country diversity.  H4b0: All first and corresponding authors' national diversity of BTS and the national diversity of world population are the same.  H4c1: Compared to the world population, the first author and corresponding author of BTS have a smaller proportion from developing countries or least developed countries.  H4c0: The proportion of the first author and corresponding author of BTS from developing countries or least developed countries is the same as that of the world population. | We will visualize the national diversity of all authors/first and corresponding authors in BTS and the world population diversity by country.  We will also use the diversity *t*-test to compare the national diversity of authors in BTS and the world population diversity by country.  We will combine visualization and Bayesian polynomial tests to test the hypothesis by comparing the proportion of developed countries/developing countries/least developed countries to which BTS’ the first and corresponding author institutions and the world population belong. | H4a is supported if, compared to the world population diversity by country, the BTS’s all authors have more national diversity (visual inspection).  H4b is supported if, compared to the world population diversity by country, the BTS’s first and corresponding authors have more national diversity (visual inspection).  H4c is supported if, compared to the world population, the first author and corresponding author of BTS have a smaller proportion from developing countries or least developed countries.  If BF10 > = 10, we infer there is relatively strong evidence that this hypothesis is supported. |

Note. *PS2023*, non-big team science articles published in *Psychological Science* in 2023; BTS, big team science.

**Acknowledgements**

This research was funded by

**Author contributions**

W.L., G.S., P.F., and H.C-P. conceived and designed the study. W.L., and H.C-P. extracted and analysis the pilot data. W.L., G.S., P.F., and H.C-P. wrote the paper.

**Competing interests**

The authors declare no competing interests.

**Reference**

Adetula, A., Forscher, P. S., Basnight-Brown, D., Azouaghe, S., & IJzerman, H. (2022). Psychology should generalize from—not just to—Africa. Nature Reviews Psychology, 1(7), 370-371. https://doi.org/10.1038/s44159-022-00070-y

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist, 63*(7), 602–614. https://doi.org/10.1037/0003- 066X.63.7.602

Bauer, P. J. (2022). Psychological science stepping up a level. *Psychological science, 33*(2), 179-183. https://doi.org/10.1177/09567976221078527

Baumgartner, H. A., Alessandroni, N., Byers-Heinlein, K., Frank, M. C., Hamlin, J. K., Soderstrom, M., Voelkel, J. G., Willer, R., Yuen, F., & Coles, N. A. (2023). How to build up big team science: a practical guide for large-scale collaborations. *Royal Society Open Science, 10*(6). https://doi.org/10.1098/rsos.230235

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., . . . Kong, X. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature, 582*(7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Brower, J. E., Zar, J. H., & Ende, C. v. (1998). Field and laboratory methods for general ecology (4th ed.). WCB McGraw-Hill.<http://www.loc.gov/catdir/description/mh023/97015263.html>

Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, J. K., Kline, M., Kominsky, J. F., Kosie, J. E., Lew-Williams, C., Liu, L., Mastroberardino, M., Singh, L., Waddell, C. P. G., Zettersten, M., & Soderstrom, M.(2020). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Canadian Psychology/Psychologie Canadienne, 61*(4), 349.<https://doi.org/10.1037/cap0000216>

Chen, Z., Hu, B., Liu, X., Becker, B., Eickhoff, S. B., Miao, K., Gu, X., Tang, Y., Dai, X., Bhattarai, R., Leonov, A., Xiao, Z., Feng, Z., Chen, J., & Hu, C.-P. (2023). Sampling inequalities affect generalization of neuroimaging-based diagnostic classifiers in psychiatry. *BMC Medicine, 21*(1). https://doi.org/10.1186/s12916-023-02941-4

Chen, S., Buchanan, E. M., Kekecs, Z., Miller, J. K., Szabelska, A., Aczel, B., Bernabeu, P., Forscher, P., Szuts, A., Vally, Z., Al-Hoorie, A. H., Helmy, M., da Silva, C. S. A., da Silva, L. O., de Moraes, Y. L., Hsu, R. M. C. S., Mafra, A. L., Dixson, B., Peters, K., … Zickfeld, J. (2023). Investigating object orientation effects across 18 languages. *PsyArXiv*. https://doi.org/10.31219/osf.io/2qf6w

Coles, N. A., Hamlin, J. K., Sullivan, L. L., Parker, T. H., & Altschul, D. (2022). Build up big-team science. *Nature, 601*(7894), 505–507. https://doi.org/10.1038/d41586-022-00150-2

Coles, N. A., DeBruine, L. M., Azevedo, F., Baumgartner, H. A., & Frank, M. C. (2023). ‘Big team’ science challenges us to reconsider authorship. *Nature Human Behaviour, 7*(5), 665-667. https://doi.org/10.1038/s41562-023-01572-2

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 7*2, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review, 25*(1), 207–218. https://doi.org/10.3758/s13423-017-1266-z

Dorison, C., Lerner, J. S., Heller, B., Rothman, A. J., Kawachi, I., Wang, K., Rees, V. W., Gill, B., Gibbs, N., Ebersole, C. R., Vally, Z., Tajchman, Z., Zsido, A. N., Zrimsek, M., Chen, Z., Ziano, I., Gialitaki, Z., Ceary, C. D., Lin, Y., . . . Ren, D. (2022). In COVID-19 Health Messaging, Loss Framing Increases Anxiety with Little-to-No Concomitant Benefits: Experimental Evidence from 84 Countries. *Affective Science, 3*(3), 577–602. https://doi.org/10.1007/s42761-022-00128-3

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J., Banks, J., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J., . . . Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology, 67*, 68–82. https://doi.org/10.1016/j.jesp.2015.10.012

Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D., Buttrick, N. R., Chartier, C. R., Corker, K. S., Corley, M., Hartshorne, J. K., IJzerman, H., Lazarević, L. B., Rabagliati, H., Ropovik, I., Aczel, B., Aeschbach, L. F., Andrighetto, L., Arnal, J., Arrow, H., Babincak, P., . . . Montealegre, A. (2020). Many labs 5: Testing Pre-Data-Collection Peer review as an intervention to increase replicability. *Advances in Methods and Practices in Psychological Science, 3*(3), 309–331. https://doi.org/10.1177/2515245920958687

Education First (2022). EF EPI English Proficiency. Retrieved from *https://www.ef.com/wwen/epi/*.

Espinosa, J., Stevens, J.R., Alberghina, D., Alway, H.E.E., Barela, J.D., Bogese, M., Bray, E.E., Buchsbaum, D. Byosiere, S.-E., Byrne, M., Cavalli, C. M., Chaudoir, L. M., Collins-Pisano, C., DeBoer, H.J., Douglas L.E.L.C., Dror, S., Dzik, M.V., Ferguson, B., Fisher, L., . . . Zylberfuden, S.G. (2023). ManyDogs 1: A multi-lab replication study of dogs’ pointing comprehension. *Animal Behavior & Cognition, 10*(3), 232-28. https://doi.org/10.26451/abc.10.03.03.2023

Fearon, J. D. (2003). Ethnic and cultural diversity by country. Journal of economic growth, 8, 195-222. https://doi.org/10.1023/A:1024419522867

Forscher, P. S., Basnight-Brown, D. M., Dutra, N., Adetula, A., Silan, M., & IJzerman, H. (2021). Psychological science needs the entire globe, Part 3. *APS Observer*, *35*.

Forscher, P. S., Wagenmakers, E. J., Coles, N. A., Silan, M. A., Dutra, N., Basnight-Brown, D., & IJzerman, H. (2022). The Benefits, Barriers, and Risks of Big team science. *Perspectives on Psychological Science*.<https://doi.org/10.1177/17456916221082970>

Gamer, M., Lemon, J., & Singh, I. F. P. (2019). irr: Various coefficients of interrater reliability and agreement (R package version 0.84.1) [Computer software]. [https://CRAN.R-project.org/package=irr](https://cran.r-project.org/package=irr)

Ghai, S., Forscher, P. S., & Hu, C.-P. (2023, May 16). The illusion of generalizability in one big team science study. https://doi.org/10.31234/osf.io/avcsp

Gutchess, A., & Rajaram, S. (2022). Consideration of culture in cognition: How we can enrich methodology and theory. *Psychonomic bulletin & review*, 1-18. https://doi.org/10.3758/s13423-022-02227-5

Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica, 4*(1), 1-9.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences, 33*(2-3), 61-83.<https://doi.org/10.1017/S0140525X0999152X>

Heiss, A. (2021). A Guide to Modeling Proportions with Bayesian Beta and Zero-Inflated Beta Regression Models.  <https://doi.org/10.59350/7p1a4-0tw75>.

Hoekstra, R., & Vazire, S. (2021). Aspiring to greater intellectual humility in science. *Nature Human Behaviour, 5*(12), 1602-1607. https://doi.org/10.1038/s41562-021-01203-8

Hu, C.-P., Kong, X.-Z., Wagenmakers, E.-J., Ly, A., & Peng, K. (2018). Bayes Factor and Its Implementation in JASP: A Practical Primer (in Chinese). *Advances in Psychological Science, 26*(6), 951–965. https://doi.org/10.3724/SP.J.1042.2018.00951

Hu, C.-P., Yin, J., Lindenberg, S., Dalgar, I., Weissgerber, S. C., Vergara, R. C., Cairo, A. H., Čolić, M., Dursun, P., Frankowska, N., Hadi, R., Hall, C. J., Hong, Y., Joy-Gaba, J. A., Lazarević, D., Lazarević, L. B., Parzuchowski, M., Ratner, K. G., Rothman, D. J., . . . IJzerman, H. (2019). Data from the Human Penguin Project, a cross-national dataset testing social thermoregulation principles. Scientific Data, 6(1). https://doi.org/10.1038/s41597-019-0029-2

IJzerman, H., Dutra, N., Silan, M., Adetula, A., Brown, D. M. B., & Forscher, P. (2021). Psychological science needs the entire globe, Part 1. *APS Observer, 34*.

IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L. M., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour, 4*(11), 1092–1094. https://doi.org/10.1038/s41562-020-00990-w

IJzerman, H., Lindenberg, S., Dalgar, I., Weissgerber, S. C., Vergara, R. C., Cairo, A. H., Čolić, M., Dursun, P., Frankowska, N., Hadi, R., Hall, C. J., Hong, Y., Chuan-Peng, H., Joy-Gaba, J. A., Lazarević, D., Lazarević, L. B., Parzuchowski, M., Ratner, K. G., Rothman, D. J., . . . Zickfeld, J. (2018). The Human Penguin Project: Climate, social integration, and core Body temperature. *Collabra: Psychology, 4*(1), 37. https://doi.org/10.1525/collabra.165

International Union of Psychological consists. (2023). *National Members*. https://www.iupsys.net/about/members/national-members/

Jamil, T., Ly, A., Morey, R. D., Love, J., Marsman, M., & Wagenmakers, E.-J. (2017). Default “Gunel and Dickey” Bayes factors for contingency tables. *Behavior Research Methods, 49*(2), 638–652. https://doi.org/10.3758/s13428-016-0739-8

Jin, H., Wang, Q., Yang, Y., Zhang, H., Gao, M., Jin, S. J., Chen, Y., Xu, T., Zheng, Y., Chen, J., Xiao, Q., Yang, J., Wang, X., Geng, H., Ge, J., Wang, W., Chen, X., Zhang, L., Zuo, X., & Chuan-Peng, H. (2023). The Chinese Open Science Network (COSN): Building an open science community from scratch. *Advances in Methods and Practices in Psychological Science*, *6*(1), 251524592211449. https://doi.org/10.1177/25152459221144986

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxsom, N. G., Lewis, S., Foroni, F., Willis, M., Cubillas, C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N., Jang, C., . . . Baskin, E. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour, 5*(1), 159–169. https://doi.org/10.1038/s41562-020-01007-2

Kijilian, C., Sharifian, M., Paris, B., Iyer, A., & Silan, M. A. (2022, September 2). PSA Study Capacity Report 2022-2023. https://doi.org/10.31234/osf.io/7krhn

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C., Nosek, B. A., Hilgard, J., Ahn, P. H., Brady, A. J., Chartier, C. R., Christopherson, C. D., Clay, S. L., Collisson, B., Crawford, J. T., Cromar, R., Gardiner, G., Gosnell, C. L., Grahe, J., Hall, C. J., Howard, I., . . . Ratliff, K. A (2022). Many Labs 4: Failure to replicate mortality salience effect with and without original author involvement. *Collabra: Psychology*, *8*(1), 35271.<https://doi.org/10.1525/collabra.35271>

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., . . . Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology, 45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., . . . Lazarević, L. B. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science, 1*(4), 443–490. https://doi.org/10.1177/2515245918810225

KOF Swiss Economic Institute. (2022) *KOF Globalisation Index.* https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalisation-index.html

Laitin, D. D. (2000). What is a language community? American Journal of Political Science, 44(1), 142. https://doi.org/10.2307/2669300

Lin, Z., & Li, N. (2022). Global diversity of authors, editors, and journal ownership across subdisciplines of psychology: Current state and policy implications. *Perspectives on Psychological Science*, 17456916221091831.<https://doi.org/10.1177/17456916221091831>

Lukács, G. (2021). neatStats: An R package for a neat pipeline from raw data to reportable statistics in psychological science. *The Quantitative Methods for Psychology, 17*(1), 7–23. <https://doi.org/10.20982/tqmp.17.1.p007>

Ly, A., Raj, A., Etz, A., Marsman, M., Gronau, Q. F., & Wagenmakers, E. J. (2018). Bayesian reanalyses from summary statistics: a guide for academic consumers. Advances in Methods and Practices in Psychological Science, 1(3), 367-374. https://doi.org/10.1177/2515245918779348

Medin, D., Ojalehto, B., Marin, A., & Bang, M. (2017). Systems of (non-) diversity. *Nature Human Behaviour, 1*(5), 0088. https://doi.org/10.1038/s41562-017-0088

Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics, 93*(2), 351-363. https://doi.org/10.1016/j.jinteco.2014.04.004

Miller, R., Lambert, M. L., Frohnwieser, A., Brecht, K. F., Bugnyar, T., Crampton, I., ... & Clayton, N. S. (2022). Socio-ecological correlates of neophobia in corvids. *Current Biology, 32*(1), 74-85. https://doi.org/10.1016/j.cub.2021.10.045

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., ... & Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science, 1*(4), 501-515.<https://doi.org/10.1177/2515245918797607>

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. Psychological science, 31(6), 678-701. https://doi.org/10.1177/0956797620916782

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology, 162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Nogrady, B. (2023). Hyperauthorship: the publishing challenges for ‘big team’ science. *Nature, 615*, 175-177. https://doi.org/10.1038/d41586-023-00575-3

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Our World in Data. (2023 July 17). *Average years of schooling.* https://ourworldindata.org/grapher/mean-years-of-schooling-long-run

Paris, B., IJzerman, H., & Forscher, P. S. (2020, December 2). PSA 2020-2021 study capacity report. https://doi.org/10.31234/osf.io/v9zma

Pavlov, Y. G., Adamian, N., Appelhoff, S., Arvaneh, M., Benwell, C. S. Y., Beste, C., Bland, A. R., Bradford, D. E., Bublatzky, F., Busch, N. A., Clayson, P. E., Cruse, D., Czeszumski, A., Dreber, A., Dumas, G., Ehinger, B., Ganis, G., He, X., Hinojosa, J. A., . . . Mushtaq, F. (2021). #EEGManyLabs: Investigating the replicability of influential EEG experiments. *Cortex, 144*, 213–229. https://doi.org/10/gjx8qt

Pollet, T. V., & Saxton, T. K. (2019). How diverse are the samples used in the journals ‘Evolution & Human Behavior’ and ‘Evolutionary Psychology’? *Evolutionary Psychological Science, 5*(3), 357–368. https://doi.org/10.1007/s40806-019-00192-2

Psychological Science Accelerator Self-Determination Theory Collaboration. (2022). A global experiment on motivating social distancing during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences, 119*(22), e2111091119. https://doi.org/10.1073/pnas.2111091119

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences, 115*(45), 11401–11405.<https://doi.org/10.1073/pnas.1721165115>

Rahal, R.-M., Fiedler, S., Adetula, A., Berntsson, R. P.-A., Dirnagl, U., Feld, G. B., Fiebach, C. J., Himi, S. A., Horner, A. J., Lonsdorf, T. B., Schönbrodt, F., Silan, M. A. A., Wenzler, M., & Azevedo, F. (2023). Quality research needs good working conditions. *Nature Human Behaviour*, 1–4. https://doi.org/10.1038/s41562-022-01508-2

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [https://www.R-project.org/](https://www.r-project.org/)

Ruggeri, K., Panin, A., Vdovic, M., Većkalov, B., Abdul-Salaam, N., Achterberg, J., Akil, C., Amatya, J., Amatya, K., Andersen, T. L., Aquino, S., Arunasalam, A., Ashcroft-Jones, S., Askelund, A. D., Ayacaxli, N., Sheshdeh, A. B., Bailey, A., Arroyo, P. B., Mejía, G. B., . . . Meyer, S. A. (2022). The globalizability of temporal discounting. *Nature Human Behaviour, 6*(10), 1386–1397. https://doi.org/10.1038/s41562-022-01392-w

Schimmelpfennig, R., Spicer, R., White, C. J. M., Gervais, W., Norenzayan, A., Heine, S. J., Henrich, J., & Muthukrishna, M. (2023). A problem in theory and more: Measuring the moderating role of culture in Many Labs 2. *PsyArXiv*. https://doi.org/10.31234/osf.io/hmnrx

Simpson, E. H. (1949). Measurement of diversity. *Nature, 163*(4148), 688-688.

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist, 76*(1), 116–129. https://doi.org/10.1037/amp0000622

The World Bank. (2022a). *World development indicators: GDP per capita.* <https://data.worldbank.org.cn/indicator/NY.GDP.PCAP.CN>

The World Bank. (2022b). *World development indicators: Urban population (proportion in total population).* <https://data.worldbank.org.cn/indicator/SP.URB.TOTL.IN.ZS?end=2022&start=1960&view=chart>

The World Bank. (2022c). *World development indicators: Individuals using the Internet (% of population).* <https://data.worldbank.org.cn/indicator/IT.NET.USER.ZS>

Tindle, R. (2021). Improving the global reach of psychological research. *Discover Psychology, 1*, 1-4. https://doi.org/10.1007/s44202-021-00004-4

Tiokhin, L., Hackman, J., Munira, S., Jesmin, K., & Hruschka, D. (2019). Generalizability is not optional: Insights from a cross-cultural study of social discounting. *Royal Society open science, 6*(2), 181386.<https://doi.org/10.1098/rsos.181386>

Trübutschek, D., Yang, Y., Gianelli, C., Cesnaite, E., Fischer, N. L., Vinding, M. C., Marshall, T.R., Algermissen, J., Pascarella, A., Puoliväli, T., Vitale, A., Busch, N. A., Nilsonne, G., Nilsonne, G. (2022). EEGManyPipelines: A large-scale, grass-root multi-analyst study of EEG analysis practices in the wild. *MetaArXiv*.<https://doi.org/10.31222/osf.io/jq342>

Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T., Kidwell, M. C., Lai, C. K., McCarthy, R. J., Riegelman, A., Silberzahn, R., & Nosek, B. A. (2019). Scientific Utopia III: Crowdsourcing Science. *Perspectives on Psychological Science, 14*(5), 711–733. https://doi.org/10.1177/1745691619850561

UNESCO. (2021a). *UNESCO Science Report: The race against time for smarter development; executive summary* (SC-2021/WS/7).<https://unesdoc.unesco.org/ark:/48223/pf0000377250>

UNESCO. (2021b). *UNESCO Science Report: The Race Against Time for Smarter Development. S. Schneegans, T. Straza and J. Lewis (eds).* UNESCO Publishing: Paris. https://www.unesco.org/reports/science/2021/en/statistics

United Nations Population Division. (2022). *World population prospects*. https://population.un.org/wpp

van Doorn, J., Ly, A., Marsman, M., & Wagenmakers, E.-J. (2020). Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman’s *ρ*. *Journal of Applied Statistics, 47*(16), 2984–3006. https://doi.org/10.1080/02664763.2019.1709053

Wagenmakers, E., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., Van Kesteren, E., Van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wang, K., Goldenberg, A., Dorison, C. A., Miller, J. K., Uusberg, A., Lerner, J. S., ... & Štrukelj, E. (2021). A multi-country test of brief reappraisal interventions on emotions during the COVID-19 pandemic. *Nature Human Behaviour, 5*(8), 1089-1110.<https://doi.org/10.1038/s41562-021-01173-x>

Webometrics Ranking of World Universities. (2023). *Countries arranged by Number of Universities in Top Ranks*. https://www.webometrics.info/en/distribution\_by\_country

World Population Prospects. (2022). *United Nations population estimates and projections.* https://population.un.org/wpp/

Yang, Q., Zhang, W., Liu, S., Gong, W., Han, Y., Lu, J., Jiang, D., Nie, J., Lyu, X., Liu, R., Jiao, M., Qu, C., Zhang, M., Sun, Y., Zhou, X., & Zhang, Q. (2023). Unraveling controversies over civic honesty measurement: An extended field replication in China. Proceedings of the National Academy of Sciences of the United States of America, 120(29). https://doi.org/10.1073/pnas.2213824120

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences, 45*. https://doi.org/10.1017/S0140525X20001685

Yue, L., Zuo, X.-N., & Hu, C.-P. (2023) The WEIRD problem in a “non-WEIRD” context: A meta-research on the representativeness of human subjects in Chinese psychological research, in principle acceptance of Version 7 by Peer Community in Registered Reports.<https://osf.io/ehw54>

1. It should be noted that big team science exists not only in psychology, but also in other disciplines such as ManyBirds (Miller et al., 2022) and ManyDogs (Espinosa et al., 2023). In this study, we focus only on big team science in the field of psychology. [↑](#footnote-ref-1)
2. We limited our analysis to big-team science studies in psychology, though we realized that there are big-team science projects that study non-human subjects such as (Miller et al., 2022) and ManyDogs (Espinosa et al., 2023) [↑](#footnote-ref-2)
3. This list was compiled independently for our purpose, the list by Dwayne Lieck and Daniel Lakens <https://osf.io/wx4zd/> was used from cross-checking. [↑](#footnote-ref-3)
4. To avoid the double dipping issue, **we will exclude the data of PSA001 and Ruggeri et al. (2022)** **in stage 2**, but will additionally report the data that include these two datasets in the supplementary materials. [↑](#footnote-ref-4)