WORK AND WELL-BEING IN SCIENCE

An International Study

Work and Well-Being in Science:

An International Study

Methodology Report

November 2021

Authors

Julie Pacer

Renee Chen

Jodi Walton

Abt Associates | 6130 Executive Boulevard | Rockville, MD 20852

CONTENTS

1. Study Overview

The **Work and Well-Being in Science: An International Study** aims to understand factors that affect the well-being of physicists and biologists in different national contexts. The pre-test and the main study were conducted by Abt Associates (Abt) for the Catholic University of America (CUA) in four countries: India, Italy, the United Kingdom (UK) and the United States (USA). Biologists and physicists at PhD-granting academic institutions and national labs or research institutes were invited to participate in a web survey. The survey was administered in English for all countries with the addition of Italian for Italy. The median survey length was 23 minutes.

The pre-test was conducted between March and April 2021 and invited 100 physicists and 100 biologists to participate in the study through an advance letter mailing and email invitation. A total of 20 physicists and 16 biologists completed the survey. All respondents were offered an e-gift card for completing the survey in a country-specific amount equivalent to $25 (USD). For more details on the pre-test, please see the Pre-Test Report (***WWB_Pre-Test_Report.docx***).

The main study consisted of two waves. Wave 1 was conducted between May and June 2021 and invited 12,246 physicists and 11,398 biologists to participate in the study through an advance email and email invitation. Wave 2 was conducted between August and October 2021 and re-invited respondents who did not complete the survey in wave 1 to participate through an email invitation. A total of 1,837 physicists, 1,381 biologists and 224 others completed the survey for the main study. All respondents were offered an e-gift card in a country-specific amount equivalent to $20 (USD), a chance to enter a raffle for an Apple iPad, a personalized report, and a general summary report for completing the survey.

2. Sampling Preparation

Abt constructed the sampling frame for this study by a) identifying or developing appropriate lists of PhD-granting institutions and research organizations in the USA, UK, India, and Italy, then b) sampling institutions within the USA and UK that have extensive research output and correspondingly large numbers of researchers, and within India and Italy where institutional information was available, and finally by c) selecting individual eligible researchers from the sampled institutions to be included on the sampling frame.

In the first stage of sampling the applicable departments (biology/physics), where the lists of institutions of higher education with additional metadata were available, departments were stratified by size and discipline. In the second stage of sampling individuals, to the extent the data could be found on the frame of individuals constructed by web scraping,[1] individual researchers were stratified in terms of career stage and gender. We describe the process for inclusion of departments within these institutions relevant to biology/physics as well as the specifics of the sampling procedure for each country in further detail in the sub-sections below.

2.1 Overview

2.1.1 Identifying Lists of Institutions
The processes differ by country as the organization of science disciplines and availability of standardized national data differs markedly as described in the sub-sections below.

2.1.2 Selection of Departments
For biology, Abt included the following subfields: cell biology, developmental biology, structural biology, molecular biology, biochemistry, neuroscience, immunology, microbiology, genetics, plant science/ botany, animal-related research/ zoology, physiology, nutrition, ecology, environmental biology, evolution, infectious disease, and other very specific medical research (i.e., cancer, diabetes, etc.). Abt excluded any subfield outside of these listed categories, including areas of research that are sometimes included within biology but overlap with other disciplines, as well as the following subdisciplines: biomaterials, biotechnology, bionanotechnology, pharmacology/pharmaceutical research, clinical trials, health sciences, horticulture, kinesiology, translational biomedicine, agriculture and environment, cognitive science, and departments with a broadly medical focus. Given the interdisciplinary nature of these subfields, such departments would have yielded very few biologists, not justifying the effort of screening them. Likewise, mathematical biology and computational biology were only included if they were explicitly housed within a biology department. Biology departments within medical schools were not considered eligible.

Determining eligible organizations from which to draw physicists was more straightforward. In general, physics/physical science departments are clearly named and easily identified. Abt excluded ambiguous cases (e.g., materials science) and overtly interdisciplinary departments from the sampling frame (e.g.,

---

[1] See Appendix B for detail on web scraping.

departments named "Department of Science" or "Department of Mathematics, Physics and Computer Science" in very small schools). Across our regions of interest, we encountered departments that combined physics with other disciplines, such as astronomy, engineering, and earth sciences. When possible, we excluded researchers from these other related disciplines from the sampling frame and sample—especially if they were housed in distinct subunits within a mixed department—using information we gleaned from their doctoral degrees, courses taught, and publications. We did include astronomers from mixed physics and astronomy departments if they were not in a distinct "astronomy" division or academic cluster. If a department did not contain a majority of evident physicists, the department was excluded. Departments of biophysics were considered physics departments (unless explicitly housed within a biology department, in which case they were considered biology.)

In some cases, a department explicitly had a combined focus on multiple disciplines including some discipline that would normally be excluded. These combined departments were excluded if the excludable discipline was clearly unrelated to biology or physics (e.g., a Department of Biology and Chemistry) but included if the excludable discipline was related to biology or physics (e.g., a Department of Biology and Biotechnology).

2.1.3 Stratification by Institution Size and by Individual Career Stage and Gender

Abt collected systematic government data (if any was available) on each of the institutions we included in the frame. This included the size of the university or department, measured in the number of qualifying faculty or graduating doctorates. We used text processing tools (including the R package "gender"[2] and the Name Parser API[3]) to classify the researchers' gender, so that we could eventually oversample women, especially in physics. It is important to note the challenge with gender identification in the sampling frame because computer algorithms cannot always correctly identify individuals' gender from their names, and human coders can only infer gender from a profile picture on the department website or a personal webpage if such a picture is available.

2.2 Specific Details by Country

2.2.1 USA

The initial list of the USA universities that grant PhDs in physical and biological sciences was obtained from the Integrated Postsecondary Education Data (IPEDS) at the National Center for Educational Statistics website (http://nces.ed.gov/ipeds/). The system does not provide the count of all researchers. Instead, using their data extraction query system https://nces.ed.gov/ipeds/datacenter/Data.aspx), Abt obtained the counts of degrees awarded in 2018–2019 as a correlate for the total number of researchers. The data are broken down by institution (UnitID), award/degree level (doctor's degree – research/scholarship (new degree classification); doctorate degree – professional or doctorate degree; other degrees were omitted), and field of study. For the latter, NCES uses 6-digit CIP codes. The relevant codes included are listed in the

---

[2] Mullen, L.; Blevins, C; & Schmidt, B. (2020). gender. R package version 0.5.4. https://CRAN.R-project.org/package=gender
[3] https://parser.name/

Appendix A. Exhibit 2-1 shows the breakdown of degrees granted by the Carnegie Foundation classification and institution size, with several leading institutions highlighted.

Exhibit 2-1. Doctoral degrees granted in physics and biology, USA, 2018-2019.
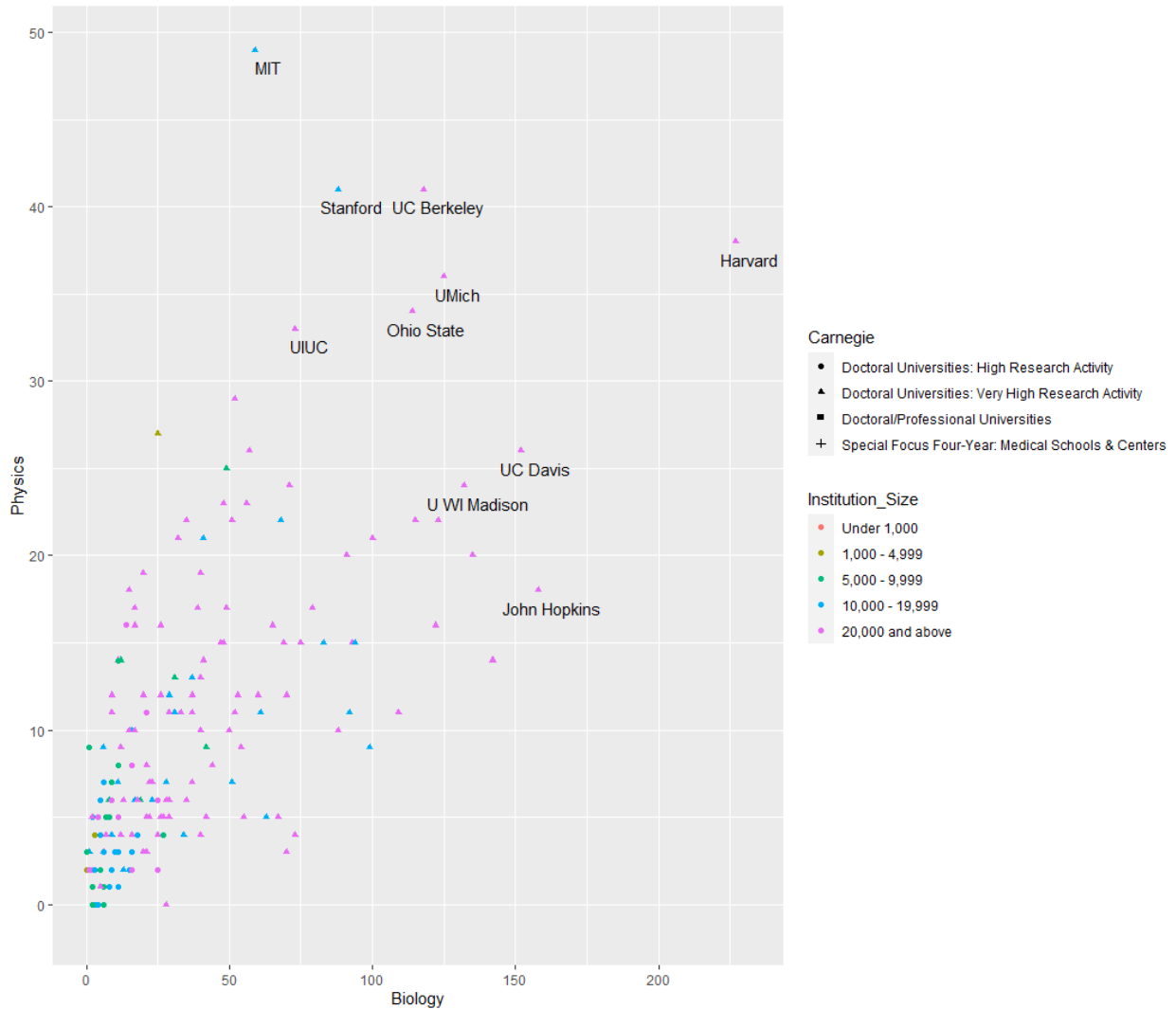
The sampling approach for the USA institutions

To build the frame of physics and biology researchers in the USA, Abt began by identifying institutions of higher learning that grant PhDs in these two areas and their sub-areas of interest as indicated by the CUA team. Based on the IPEDS data on 2018-19 doctoral degrees conferred, there were 178 institutions that had granted doctoral degrees in physics and 257 institutions that had granted doctoral degrees in biology, for a total of 271 institutions. There was a total of 1,875 doctoral degrees in physics granted and 6,910 doctoral degrees in biology and biomedical sciences granted (the latter includes degrees in bioinformatics, biomathematics, and biotechnology).

The institutions were broken down into three strata:

1. Top decile (28 institutions), to be sampled at the highest rate (initially proposed 100% resulting in 28 institutions).
2. Middle range (108 institutions), to be sampled at a lower rate (the initially proposed 20% rate would result in 22 institutions.)
3. Bottom half of the distribution (135 institutions), to be sampled at the lowest rate (the initially proposed 10% rate would result in 15 institutions.)

Institutions in the first stratum graduate between 15 and 50 doctorates in physics a year; in the middle range, 3 to 25 doctorates, and in the bottom half, at most 10 physics doctorates. In biology, the top institutions would graduate 50 to 200 doctorates; the middle range, 10 to 70 doctorates; and in the bottom half, up to 25 doctorates. Nearly all the institutions classified by the Carnegie Foundation as "Very High Research Activity" (former Research I) are found in the top two strata; the bottom half is predominantly "High Research Activity" (Research II) universities. The universities that graduate fewer doctorates were expected to be smaller, i.e., also have fewer faculty and post-docs (on which we do not have any administrative data information). Comparisons of the sampling frame measure of size with the scraped number of contacts per university/department largely confirmed that, although correlations were less than perfect.

In summary, institution size was determined using the following steps:

- We obtained IPEDS data on number of doctoral degrees confirmed in physics and biology in 2018-19 for USA institutions.
- We divided that data (the whole frame, not just the sample) into deciles. Therefore, when the institutions were listed in order of number of degrees conferred (collapsing across biology and physics), the ones that formed the top 10% of that list were Decile #1. The ones that formed the second-highest 10% were Decile # 2, etc., through the bottom 10% of the list that formed Decile #1.
- Decile # 1 institutions (top 10%) were classified as "Large." Decile 2-5 institutions (40% of the frame) were classified as "Medium." Decile 6-10 institutions (bottom 50% of the frame) were classified as "Small."

Because size was determined by number of biology doctoral degrees conferred plus number of physics doctoral degrees conferred, it may not perfectly correspond to overall enrollment. It's also worth noting that size designations may not seem perfect at the department level: for example, a physics dept from a "medium" institution might be more similar in size to physics departments from "small" institutions, because the "medium" designation was based on physics and biology in combination.

Abt initially scraped and fully processed researchers' contact information for 11 physics departments in the USA to assess how the scraping results (number of researchers identified) and the scraping process (professional staff time) correlated with the measures of size (number of doctorate graduates). The number of researchers identified in these institutions ranged from 15 to 337; the amount of time that our staff spent scraping one department ranged from 1.5 hours to 6.5 hours, and critically depended on how well-structured the given university webpages were. Abt consistently found that websites of the smaller universities were more difficult to scrape, probably because these universities lack unified faculty management systems that would have produced well-structured pages typical for larger universities. We found correlations of the observed counts to the available frame data to be weaker than expected, and

universities within the same expected size group to vary widely both in terms of the size of the ultimate list and the time to verify the contact information.

Websites were ultimately scraped using the list of universities that had been generated using the sampling process described above. The USA institutions were scraped until a minimum of 2,778 individuals from each discipline was obtained.[4] Due to limited resources, however, only a subset of the sampled universities was scraped to reach that number. The subset prioritized institutions across a range of sizes and whose websites could easily be scraped with the web scraping tool. The final USA sample included physics departments from 19 institutions (3 small, 5 medium, and 11 large) and biology departments from 20 institutions (3 small, 6 medium, and 11 large). These biology departments included a partial scrape of one medium-sized biology department at Kent University, which was not in the original sample. This was scraped in error, the goal being University of Kent which is a UK institution. Kent University was ultimately included because it met eligibility criteria for USA.

2.2.2   UK

The frame construction process for the UK institutions was facilitated by the data from Research Excellence Framework (REF), a periodic evaluation of research activities of all UK higher education institutions. The previous evaluation was conducted in 2014; the next one is planned for 2021. Each department is ranked based on the research output and research impact of their research staff. The REF submission data are given at the level of research groups (e.g., departments and research centers), and are further detailed at the level of individual researchers. Thus, the data for the REF Unit of Assessment 5 – Biological Sciences contain information on 2,492 researchers affiliated with 191 research groups in 43 universities, and the data for the REF Unit of Assessment 9 – Physics contain information on 1,773 researchers affiliated with 175 research groups in 40 universities. For each university, the assessment rubrics include outputs, impact and environment; additional variables include number of staff, research degrees awarded, research income, and the overall ranking.

Exhibit 2-2 displays the institutions in the REF 2014 data by the effective number of research staff rated at the highest REF level, 4*, which is the product of the 4* percentage rating and the total number of FTE Category A staff submitted in the overall rating of the institution.

---

[4] A minimum of 2,778 individuals from each country-discipline combination were scraped for a total of 23,634 sample cases which were intended to achieve 6,000 completed surveys at a 27% response rate.

Exhibit 2-2. Research staff in physics and biology, UK, 2014 REF data.
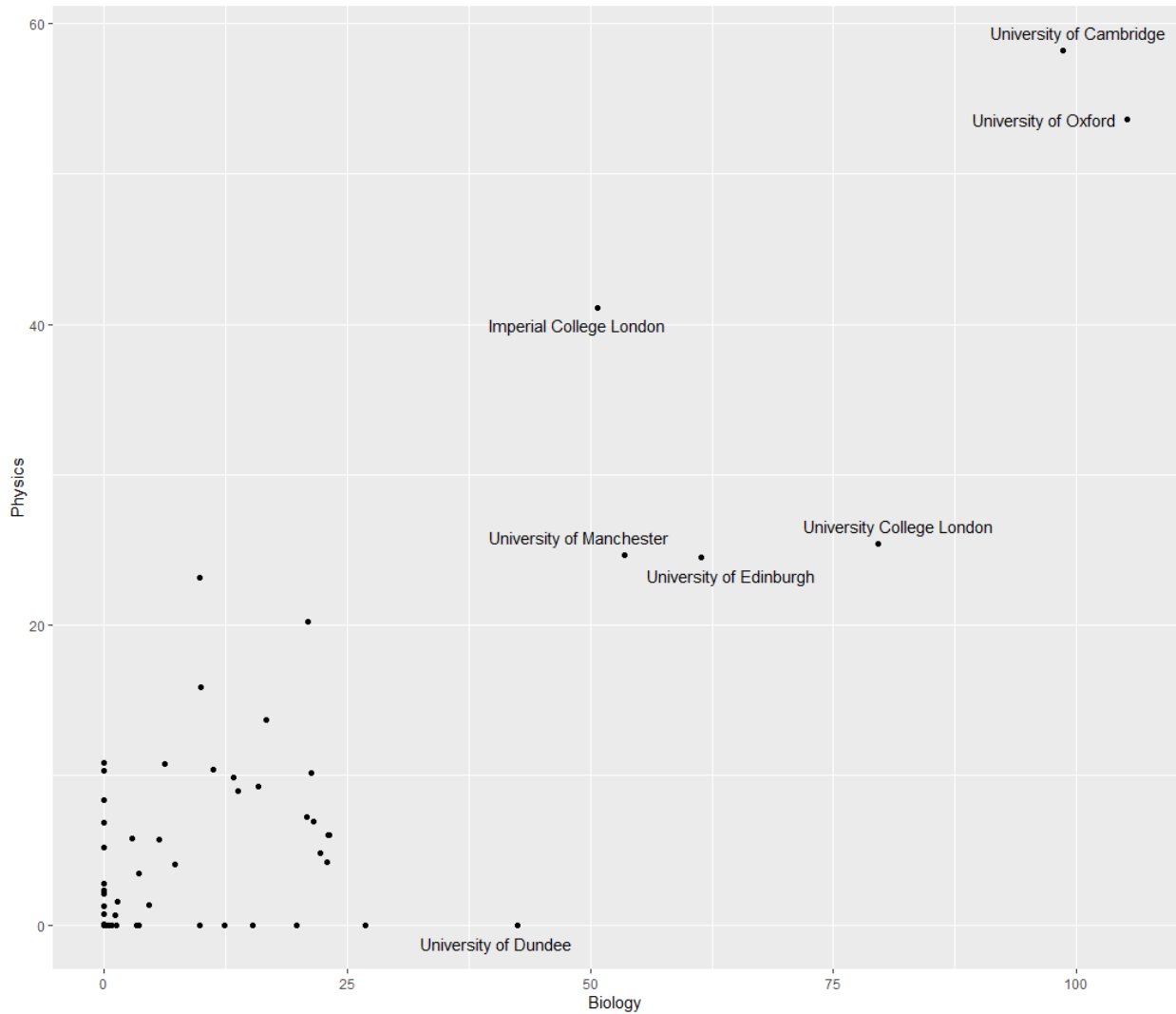
We determined institution size for UK using the following steps:

- In the absence of a comprehensive resource indicating number of doctoral degrees granted by UK universities, we used the clusters defined by Boliver (2015)[5] that divided UK universities into one of four clusters: academic selectivity, research activity, teaching quality, socio-economic exclusivity and economic resources.
  - We expect these criteria to correlate positively with university size. The universities in each cluster are listed in an appendix in Boliver's publication.

---

[5] https://dro.dur.ac.uk/14978/

- We only used three of the clusters, with cluster 1 universities classified as "Large," cluster 2 classified as "Medium," and cluster 3 classified as "Small." Cluster 4 university departments were generally not deemed eligible for the sampling frame.
- Each institution that we sampled was matched to Boliver's list to get the size designation. Only one sampled institution (the Institute of Cancer Research, which was a non-university research institution) was not in Boliver's list; in the absence of size information, we did not categorize that institution.

Because Boliver's criteria were not simply based on university size, again, the designations may not perfectly correspond to overall enrollment. For example, Boliver's algorithm classified University College London as Cluster 2, therefore it appears as "medium."

In total, we scraped two of the cluster 1 universities, 22 of the cluster 2 universities, five of the cluster 3 universities, and one institution that was classified as "specialist" (Institute of Cancer Research). These 30 UK institutions were scraped until a minimum of 2,778 records were obtained for biology and physics each. This resulted in physics departments from 21 different institutions (3 small, 17 medium, and 1 large) and biology departments from 11 different institutions (2 small, 6 medium, 2 large, and 1 specialist institution for which size data was unavailable).[6]

2.2.3 India

For India, Abt used a combination of the list of universities provided by the University Grants Commission of India and institutions ranked by Times Higher Education (THE) and the National Institute Ranking Framework (NIRF) of the Ministry of Education of India.[7]

Although the tally of total faculty and PhDs awarded is not systematically listed information, it is factored into the ranking system used by NIRF in ranking universities: for the 2020 rankings, the final overall ranking framework was applied to institutions that had at least 1,000 students since 2017 onwards and used 16-18 parameters organized in five major groups. Several of these parameters are common to those employed globally and serve as pointers to ambience for teaching, learning and research. However, there are a few India-centric parameters, reflecting aspirations of the rising numbers of young people enrolled in higher education institutions.

Vale and Dell (2009) listed 19 institutions with 802 faculty, 266 of which were junior faculty, 218 women faculty, 2,595 PhD students, and 308 post-docs in major institutes and universities conducting life science research in India (all these institutions have departments relevant to the sample).[8] The authors also listed

---

[6] Two universities were scraped for both their physics and their biology department.

[7] https://www.nirfindia.org/nirfpdfcdn/2020/pdf/Report/IR2020_Report.pdf

[8] Vale, R. D., & Dell, K. (2009). The biological sciences in India: aiming high for the future. The Journal of cell biology, 184(3), 342–353. https://doi.org/10.1083/jcb.200812123

a total of 350 institutions and universities in biology that conferred PhD degrees. Gupta and Dhawan (2009) count 1,307 universities and colleges (as of 1993–2001) with significant physics research output.[9]

We know that these tallies have changed significantly since then; we have used the top 100 universities as ranked by NIRF and the top 100 universities ranked in THE to obtain 100 universities for the sampling frame in physics and 40 universities/institutes in biology. Further, the original RASIC frame provided us additional physics departments.

Indian institutions were scraped until a minimum of 2,778 records were obtained for biology and physics each. This resulted in inclusion of physics departments from 82 different institutions and biology departments from 63 different institutions. While the original list of universities obtained from combining the NIRF, the THE and the RASIC lists had more universities, upon our review we concluded that they were not eligible for this study per CUA eligibility definitions or did not have information on faculty and staff available on their websites.

2.2.4   Italy

Although we were unable to find specific information on either the number of faculty or the number of doctoral degrees granted in Italy, some relevant information was obtained from the Italian Institute of Statistics.[10] In 2017, there were 2,170 professors of physics and 4,556 professors of biology in the country. Of these, 976 were female in physics and 2,440 were female in biology. Although numbers on PhDs granted were unavailable, approximately 15% of the population aged 15 and older held post-undergraduate degrees in 2019. In 2016, 13,809 students earned bachelor's or master's degrees in Life and Natural Sciences from Italian colleges or universities, and 8,818 students earned bachelor's or master's degrees in mathematics and physical science.

In developing the Italian sampling frame, Abt began by compiling a list of universities from the website of the Italian Ministry of University Instruction and Research (MUIR)[11] and a list of research institutions from the Italian National Research Council[12]. This resulted in an original list of 28 research institutes and 106 universities. After obtaining further information about which departments should be excluded (i.e., interdisciplinary departments), as well as eliminating some that did not list faculty emails, Abt narrowed down this list to 27 research institutes and 43 universities. Obtaining access to the RASIC frame allowed us to identify several additional research institutes, bringing the number of research institutes up to 54. URLs for institutions' websites were most often obtained manually through a web search, except for those research institutes listed by the National Resource Council and the MUIR, which often provided websites.

Italian institutions were scraped until a minimum of 2,778 records were obtained for biology and physics each. This resulted in physics departments from 24 different institutions and biology departments from 31 different institutions.

---

[9] Gupta, B.M., & Dhawan, S.M. (2009). Status of physics research in India: An analysis of research output during 1993–2001. Scientometrics 78, 295–316. https://doi.org/10.1007/s11192-007-1926-2

[10] http://www.miur.it/guida/guide.htm

[11] https://www.cnr.it/en

[12] https://www.istat.it/en/education-and-training

2.3 Cases for Exclusion

A minimum of 2,778 individuals from each country-discipline combination were scraped for a total of 23,634 sample cases which were intended to achieve 6,000 completed surveys at a 27% response rate. This total was later reduced to exclude individuals that were incorrectly included in the sampling file. Most of the exclusions consisted of individuals with job titles that indicate they were administrative and technical staff (i.e., Administrative Officer, various Engineer and Junior Engineer titles, Lab Assistant/Lab Technician, Personal Technical Administrator, etc.). Other exclusions were duplicate sample entries.

2.4 Sample Opt-in

There were 10 individuals from an institution in India that opted into the sample as a result of hearing about the study from a colleague upon survey launch (wave 1). These individuals were determined to be eligible to participate because they were from an institution and department that was included in the sample and had been omitted for some reason. It is possible they were omitted in error or due to personnel changes within a department between the time the sampling frame was generated and the survey launched. These respondents reached out to Abt via email requesting to be included in the study.

2.5 Summary

The distribution of cases in sampling frame by country, field, and institution size is summarized in Exhibit 2-3 below.

Exhibit 2-3. Sampling Frame Summary

Country, Field, Size	Total Original		Excluded		Final Sample	
	N	%	N	%	N	%
USA	**5,667**	**24%**	**210**	**26%**	**5,457**	**24%**
Biology	**2,874**	**12%**	**156**	**19%**	**2,718**	**12%**
Large	2,055	9%	82	10%	1,973	9%
Medium	708	3%	65	8%	643	3%
Small	111	0%	9	1%	102	0%
Physics	**2,793**	**12%**	**54**	**7%**	**2,739**	**12%**
Large	1,966	8%	34	4%	1,932	8%
Medium	763	3%	19	2%	744	3%
Small	64	0%	1	0%	63	0%
UK	**6,635**	**28%**	**167**	**21%**	**6,468**	**28%**
Biology	**2,798**	**12%**	**58**	**7%**	**2,740**	**12%**
Large	1,609	7%	24	3%	1,585	7%
Medium	886	4%	26	3%	860	4%
Small	176	1%	4	0%	172	1%
Unavailable	127	1%	4	0%	123	1%
Physics	**3,837**	**16%**	**109**	**14%**	**3,728**	**16%**

Large	531	2%	6	1%	525	2%
Medium	2,819	12%	93	12%	2,726	12%
Small	487	2%	10	1%	477	2%
India	**5,660**	**24%**	**242**	**30%**	**5,418**	**24%**
Biology	2,856	12%	162	20%	2,694	12%
Physics	2,804	12%	80	10%	2,724	12%
Italy	**5,682**	**24%**	**185**	**23%**	**5,497**	**24%**
Biology	2,870	12%	115	14%	2,755	12%
Physics	2,812	12%	70	9%	2,742	12%
Grand Total	**23,644**	**100%**	**804**	**100%**	**22,840**	**100%**

3.   Questionnaire Development and Overview

The survey questionnaire was developed by CUA and reviewed and refined by Abt. As much as possible, the survey implemented questions from other surveys including the RASIC study. Abt's Institutional Review Board and legal team reviewed the consent language to ensure compliance with the General Data Protection Regulation (GDPR). The GDPR gives citizens of the EU and the UK the right to know what personal data will be collected about them, how those data will be used, how long the data will be stored, the reason for the length of storage time, and the right to ask organizations to delete their personal data.

The survey questionnaire began with an introduction to the study. This was followed by the informed consent section to ensure respondents could freely provide their choice to participate in the survey based on full knowledge and understanding of the project and its implications. The consent process informed respondents of the nature of the research, sample selection, survey administration and length, question themes, how the data collected would be used, that their participation is voluntary, risks and benefits of participation, that the data are confidential, and contact information for study representatives. In addition, respondents in the UK and Italy were assured of GDPR compliance through a downloadable addendum. Only respondents that consented to participate could complete the survey.

In addition to consent, eligible respondents confirmed affiliation with the sampled institution and their primary discipline. The survey asked questions about general health and well-being, the meaningfulness of scientific work, the role of aesthetics (awe, wonder, and beauty) in scientific work, workplace culture, the ways in which their work and life were affected by the COVID-19 pandemic, as well as demographic information. Personal information was collected only if they opted to provide it to participate in follow-up interviews.

The survey was programmed in English and Italian. It was offered in English to all respondents. The Italian version was additionally offered to Italian respondents. Of the 1,145 respondents from Italy who completed the survey, 990 respondents (86%) took the survey in Italian. The survey questionnaire and GDPR addendum from the consent process are in Appendices C-D.

3.1   Pre-test Survey

Two questions were included in the pre-test survey to help the project team assess the pre-test design and any necessary changes. At the beginning of the survey, respondents were asked if they received the advance letter. At the end of the survey, a pre-test debrief question was included. Respondents were asked if they had any comments, suggestions or issues encountered to report based on completing the survey. The feedback provided by the respondents was considered when updating the questionnaire for the main study.

3.2   Wave 1 Changes

After the pre-test, the following changes were implemented for wave 1 data collection:

- The advance letter mailing was replaced with an advance email to sample members.
- The survey questionnaire was updated to change existing question text and add and remove questions. The screener was updated to remove certain screening criteria.

- The e-gift card incentive was decreased from a country-specific amount equivalent of $25 (USD) to $20 (USD).

- New incentives for an Apple iPad raffle and a personalized report of findings were added.

- The study communications were updated.

3.3   Wave 2 Changes

After wave 1, the following changes were implemented for wave 2 data collection:

- The survey landing page text was updated and included positive feedback and quotes from wave 1 respondents that completed the survey.

- A progress bar was added to each screen of the web survey.

- Formatting of the consent section was modified to improve readability.

- An automated thank-you email was sent to respondents upon completing the survey, thanking respondents for completing and asking them to encourage colleagues to complete.

- The survey invitation email was updated to include the number of scientists that had already completed the survey, the updated survey length, and a link to a one-minute video of the principal investigator encouraging participation.

- The number of reminder emails and the frequency of contact was reduced.

- Targeted reminders for those who partially completed the survey were added.

- CUA sent a reminder email to encourage respondents to participate and confirm the legitimacy of the study.

- A postcard mailing was added to a sub-sample of Italy, UK and USA sample members reminding them to look for the survey communications in their email to access the survey.

- Promotion via Abt's social media channels were implemented.

4. Study Website

A study website was developed to legitimize the study and provide a centralized location to access study information. This was important to accommodate sample members across the different countries and time zones. The website included the study background, a downloadable copy of the support letter (described below), related research, frequently asked questions which included a downloadable copy of the GDPR addendum from the consent portion of the survey questionnaire, and contact information for study representatives. Portions of the website were available in Italian. Abt provided content for the study website while the design was provided by Grey Matter Group (GMG), a marketing and design agency with whom CUA partnered. GMG also provided the study branding including name, logo, and colors in collaboration with CUA and Abt.

5. Study Communication and Schedule

5.1 Advance Email

Abt sent an advance email to all sample members on May 21, 2021, to introduce the study, request participation in the web survey, and notify them that the invitation to participate would be sent soon via email. It included study details such as purpose of the research, information on sample selection, study design, incentive, confidentiality, study website, and contact information for study representatives. The sample member could contact the study via email to update their email address. Three sample members provided an updated email address.

The advance email included a link to a letter of support from external study advisors. This letter endorsed the study and provided additional encouragement to participate by reiterating the importance of the research. This one-page letter was signed by fourteen prominent physicists and biologists affiliated with institutions across all four countries for wave 1. For wave 2, four additional physicists and biologists signed the letter. The advance email also included links to two articles on the challenges to scientists' mental health and the role of beauty in science in the science journal, *Nature*.

5.2 Survey Invitation and Reminder Emails

The email invitation to the survey was sent to all sample members on May 26, 2021, for wave 1. The email invitation included most of the content of the advance email and provided a web link with direct access to the survey. Further, it included a hyperlink to the support letter and study website which gave access to the frequently asked questions and additional materials. Survey non-respondents received up to eight email reminders to complete the survey. For Italian sample members, emails were sent in English and included a link to view the email in Italian for the survey invitation and reminders 1-5. In efforts to boost response rates in Italy, reminders 6-8 were sent in Italian and included a link to view the email in English.

For wave 2, all sample members who did not complete the survey or opted out of emails were re-invited to participate in the survey on August 25, 2021. The survey invitation for wave 2 included most of the content from the wave 1 invitation and a new link to a one-minute video from the principal investigator. A total of five reminders were sent for wave 2 with reminders 3 and 4 tailored for non-response and partial completes. For Italian sample members, emails were sent in Italian with a link to view the email in English for the survey invitation, reminder 1, reminder 3 and reminder 5. Reminders 2 and 4 were sent in English with a link to view the email in Italian.

5.3 Department Head Email

Abt also sent an email to the department heads to verify the legitimacy of the study and request they promote completion to those in their department who had been invited to the survey. This email was sent during wave 1 on June 17, 2021.

5.4   Postcard Mailing

For wave 2, Abt mailed a postcard to a sub-sample of 6,150 sample members on September 10, 2021. The postcard was mailed to a random 40% of the wave 2 sample in the USA, UK, and Italy. The reduced sample was identified to receive the postcard to add a mail mode of contact where resources would allow. India was excluded because of the cost of mailing to India and because they had a high completion rate in wave 1. The postcard described the study and incentives for completion and asked recipients to check their email for the personalized invitation link to the survey, including checking their spam folder. It did not include direct access information to the survey so it could be posted as a flyer within the department or otherwise shared with colleagues that may have been invited to the survey. An Italian language version of the postcard was sent to the sample members in Italy.

5.5   Social Media Posts

For wave 2, Abt promoted the study on their social media pages. Abt posted about the study on their Facebook and Twitter pages on September 10, 2021. The posts were intended to promote the legitimacy of the study without inviting all researchers to participate since the sample was scientifically selected. The posts briefly described the study and directed readers to the project page on the Abt Associates website.

5.6   Thank You Email

Abt sent an email to all 2,214 sample members who completed the survey in wave 1 on June 29, 2021, to thank them once again for participating in the study. The email also informed respondents that the iPad raffle winners had been selected and to expect their personal and general reports by the end of the study. In addition, this email solicited suggestions from respondents for ways to encourage other invited scientists at their institution to complete the survey. Fifty-nine (59) responses were received with a wide range of comments, including praise, criticism, notes of thanks, as well as suggestions that were both tenable ("underline that this survey doesn't need to much time to be completed") to untenable (suggesting specific names and email addresses of others who might be interested in participating).

For wave 2, Abt set up an automated thank you email that was sent at survey completion. In addition to thanking the respondent for participating, the email encouraged respondents, who were comfortable doing so, to promote the survey to their colleagues in case they had received an invitation.

5.7   Abt Study Communication Schedule

The full schedule of the study communication is in Exhibit 5-1 below. All emails were sent at the local time unless otherwise specified. The materials that were emailed, mailed, or posted for the sample members are provided in Appendices E-L.

Exhibit 5-1.   Abt Study Communication Schedule

	Email	Day	Date	Local Time*
Wave 1				
AE	Advance Email	Friday	May 21, 2021	10:00 AM
1-SI	Wave 1 Survey Invitation	Wednesday	May 26, 2021	10:00 AM
1-R1	Wave 1 Reminder 1	Friday	May 28, 2021	9:00 AM
1-R1.5	Wave 1 Reminder 1.5	Wednesday	June 2, 2021	8:00 AM
1-R2	Wave 1 Reminder 2	Friday	June 4, 2021	9:00 AM
1-R3	Wave 1 Reminder 3	Tuesday	June 8, 2021	11:00 AM
1-R4	Wave 1 Reminder 4	Monday	June 14, 2021	10:00 AM
1-R5	Wave 1 Reminder 5	Thursday	June 17, 2021	4:00 PM
1-DH	Wave 1 Department Head	Thursday	June 17, 2021	4:00 PM
1-R6	Wave 1 Reminder 6	Wednesday	June 23, 2021	8:00 AM
1-R7	Wave 1 Reminder 7	Friday	June 25, 2021	11:00 AM
1-R8	Wave 1 Reminder 8	Monday	June 28, 2021	8:00 AM
TY	Thank You Email	Tuesday	June 29, 2021	4:00 PM
Wave 2				
2-SI	Wave 2 Survey Invitation	Wednesday	August 25, 2021	9:00 AM
2-R1	Wave 2 Reminder 1	Friday	August 27, 2021	10:00 AM
2-R2	Wave 2 Reminder 2	Wednesday	September 1, 2021	4:00 PM
2-R3	Wave 2 Reminder 3 (a/b)	Thursday	September 9, 2021	4:00 PM
PM	Postcard Mailing	Friday	September 10, 2021	N/A
SMP	Social Media Postings	Friday	September 10, 2021	N/A
2-R4	Wave 2 Reminder 4 (a/b)	Monday	September 20, 2021	10:00 AM
2-R5	Wave 2 Reminder 5	Friday	September 24, 2021	8:00 AM

*All emails were sent at the local time. Mountain Time was used for USA based on the majority time zone of sample members.

5.8   Other Study Communication

CUA coordinated with the science journal *Nature* to promote the study. A letter to the Editor from the study team was published in the May 27, 2021, printed issue of *Nature*. An ad for the study was also included in an email blast *Nature* sent to their subscribers that same day. A second email was sent by *Nature* on September 21, 2021, to promote the study and encourage their subscribers to check their emails to see if they've received a survey invitation from Abt.

For wave 2, CUA sent a reminder email to sample members who had not completed the survey to verify the legitimacy of the study and request they complete the survey. This email was sent directly by CUA on September 13, 2021.

6.   Data Collection

Data collection for wave 1 launched on May 26, 2021, at 10:00 AM locally for India, Italy, the UK and the USA (Mountain Time). The survey fielded for approximately four and a half weeks and respondents were asked to complete the survey by June 28, 2021, at 11:59 PM locally. The survey remained open until data collection for wave 2 was launched on August 25, 2021, at 9:00 AM locally. Wave 2 fielded for approximately five and a half weeks and closed on October 4, 2021, at 9:30 AM ET. Exhibits 6-1 and 6-2 chart the email communications listed in Exhibit 5-1 and show the resulting completes, by wave. This illustrates that survey completion strongly aligned with the email communications. Note the complete dates are based on the complete date and time in the data collection platform which is in Eastern Time, not the local date and times. Exhibits 6-3 and 6-4 show the cumulative completes over the field period. The data collection progress is shown by wave for comparison of wave 1 and wave 2. The wave 1 survey remained available to respondents during the eight-week interim between closing wave 1 closing and launching wave 2. However, respondents were not actively pursued via email reminders nor other form of outreach. Thirty-four (34) completes were obtained in the interim. Overall, the charts below show an increasing trend throughout both waves since they are tracking the number of completes collected. While wave 2 does not show substantial gains compared to wave 1, the gain itself is notable considering the sample consisted of the same sample members as wave 1, or the wave 1 non-respondents. An additional 1,228 completed interviews were collected in wave 2, or 36% of the total completed interviews.

Exhibit 6-1.   Wave 1 Email Communications Sent and Completes by Day
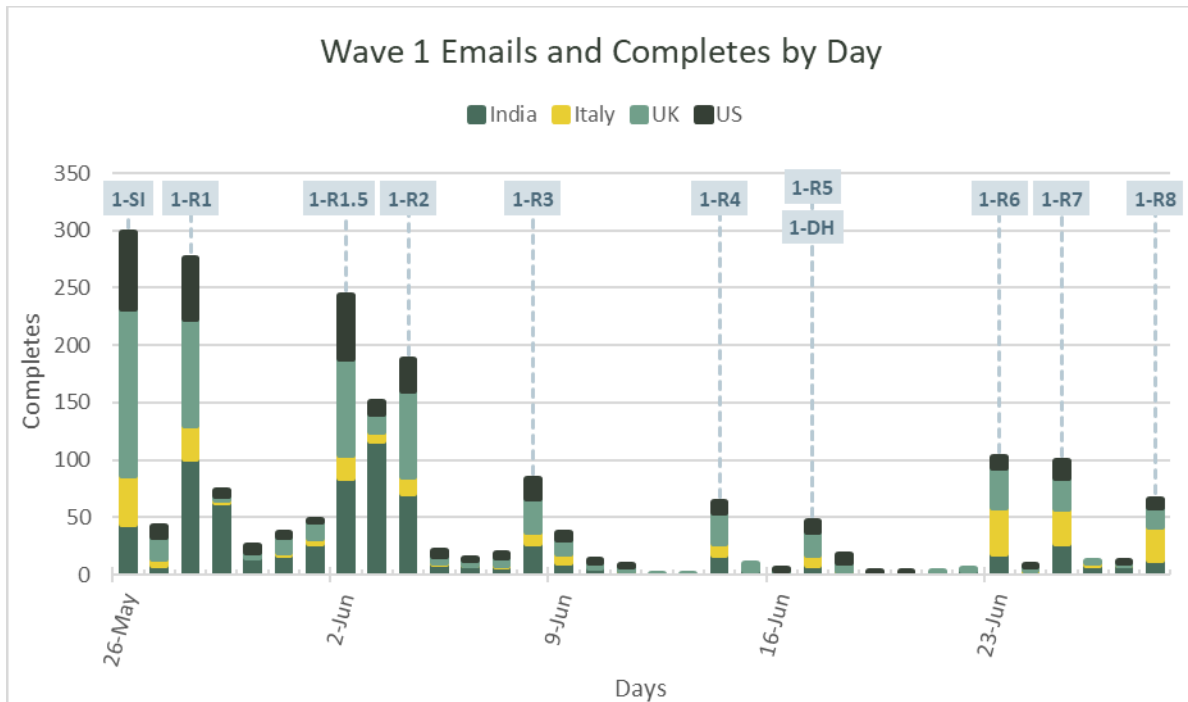
Exhibit 6-2.   Wave 2 Email Communications Sent and Completes by Day
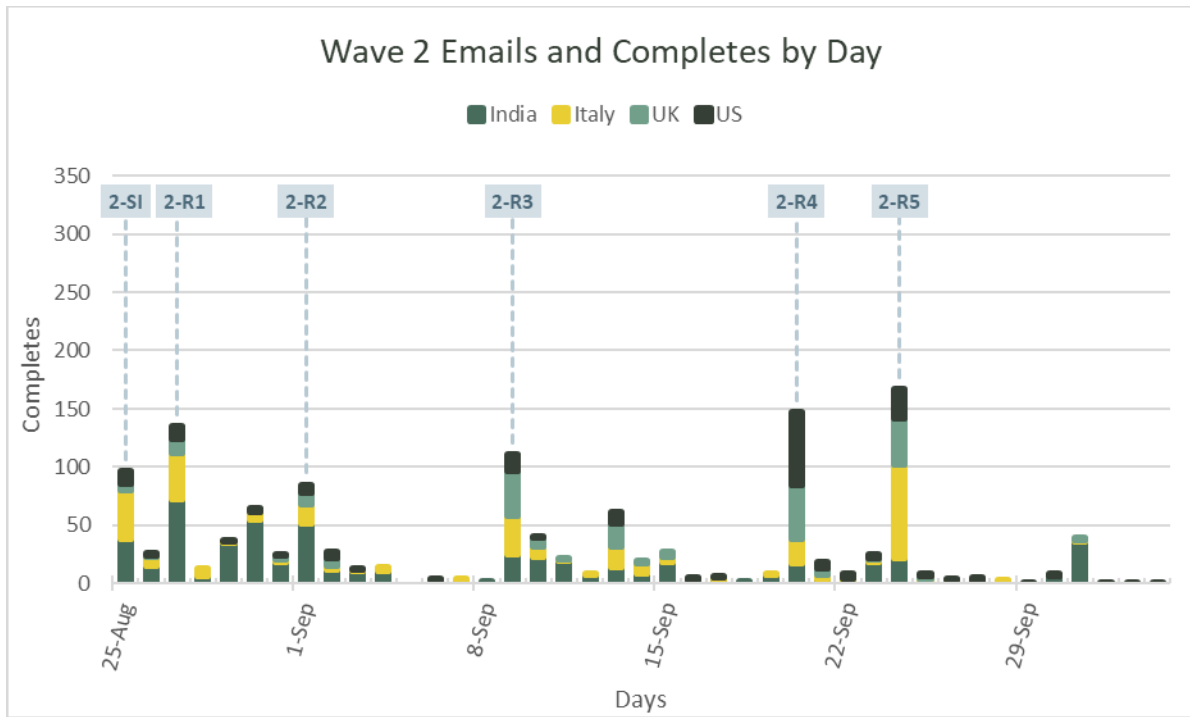
Exhibit 6-3.   Wave 1 Cumulative Completes
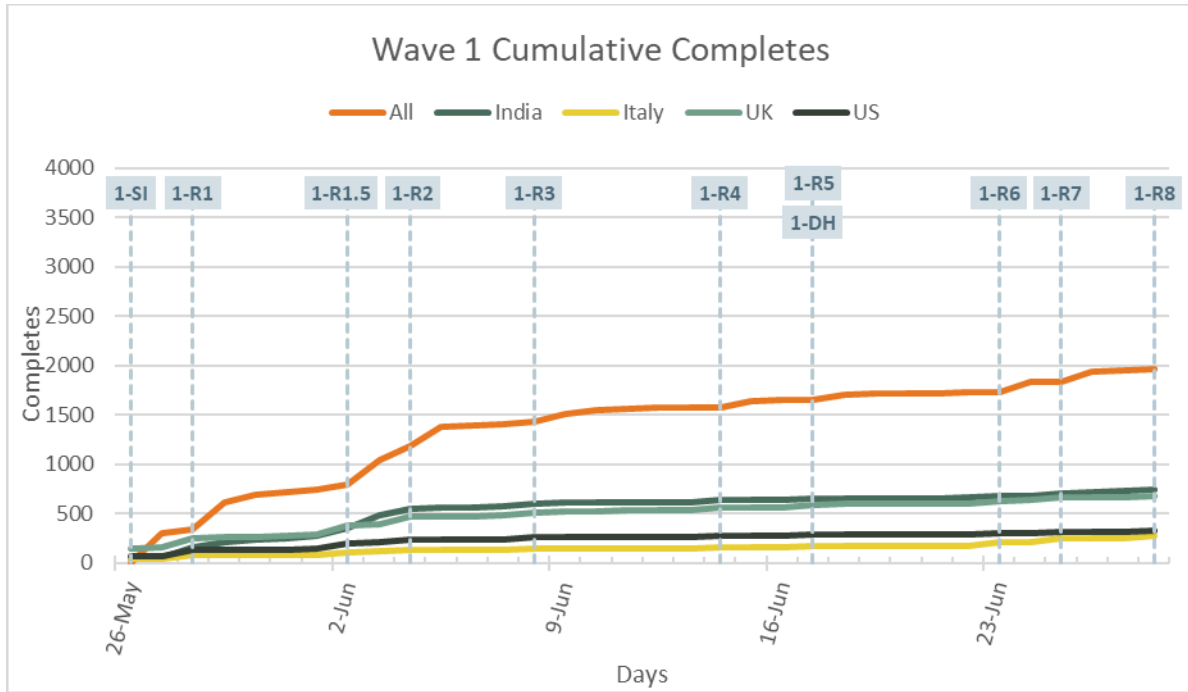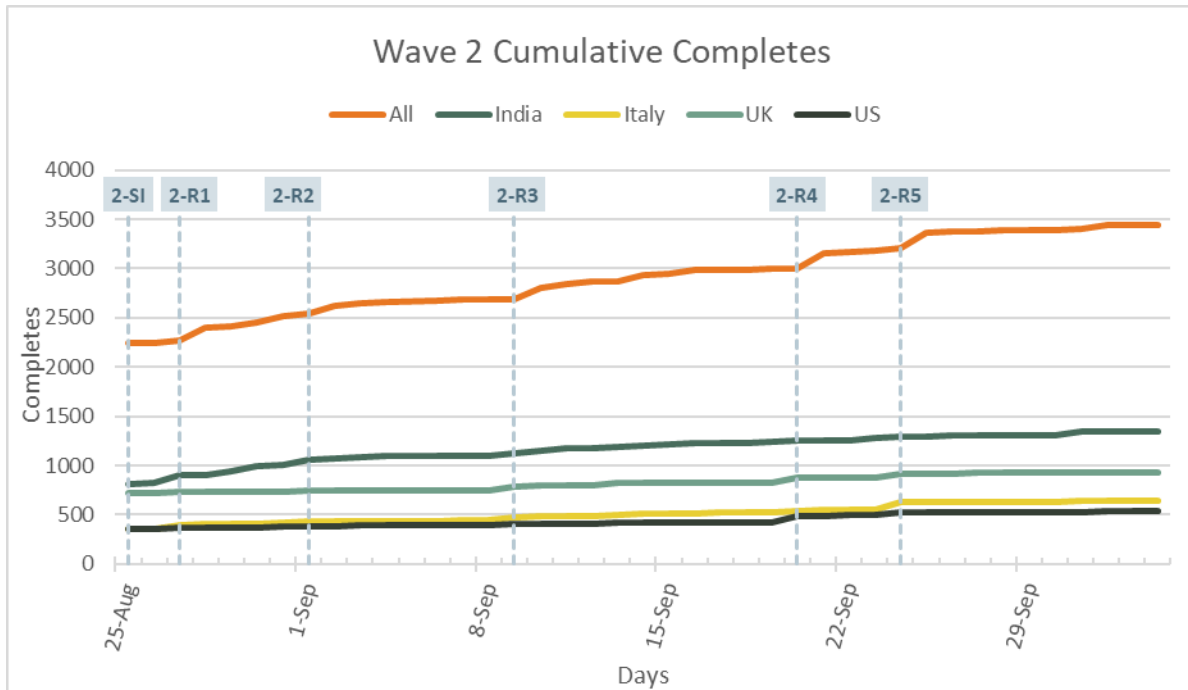

Exhibit 6-4.   Wave 2 Cumulative Completes

The web survey was optimized for mobile completes. Thirteen percent (13%) of respondents completed the survey on a smartphone. Exhibit 6-5 shows the distribution of the sample by completion status as well

as country, discipline, and gender. Of the respondents that completed in the "Other" discipline category, 13 percent reported their discipline as Chemistry or Biochemistry and seven percent as Psychology or Psychiatry. Additional other mentions included a variety of science disciplines such as Astrophysics, Biophysics, Neuroscience, and Microbiology, among others. The "Other" category did not include administrators nor support staff.

Exhibit 6-5.   Cumulative Completes

All Countries		Completes		Incompletes		Screened Out		Total	
Country		**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
US		535	16	452	15	21	11	1,008	15
UK		925	27	1,729*	57	42	23	2,696	41
Italy		637	19	444	15	64	35	1,145	17
India		1,345	39	386	13	58	31	1,789	27
Total		**3,442**	**100**	**3,011**	**100**	**185**	**100**	**6,638**	**100**
Discipline	**Gender**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Physics	Male	1,302	38	82	3	-	-	1,384	21
Physics	Female	519	15	46	2	-	-	565	9
Physics	Other	16	0	0	0	-	-	16	0
Physics	Not asked	-	-	2	0	-	-	2	0
Physics	**Total**	**1,837**	**53**	**130**	**4**	**-**	**-**	**1,967**	**30**
Biology	Male	642	19	54	2	-	-	696	10
Biology	Female	732	21	45	1	-	-	777	12
Biology	Other	7	0	0	0	-	-	7	0
Biology	Not asked	-	-	0	0	-	-	0	0
Biology	**Total**	**1,381**	**40**	**99**	**3**	**-**	**-**	**1,480**	**22**
Other		**224**	**7**	**15**	**0**	**-**	**-**	**239**	**4**
Did not get to these questions		-	-	2,767	92	185	100	2,952	44
Total		**3,442**	**100**	**3,011**	**100**	**185**	**100**	**6,638**	**100**

* Note that UK Incompletes includes 962 cases from University of Oxford that gave the false impression of a partial interview. They appeared to have accessed the first screen of the survey which flagged them as "Incomplete" however no additional data were collected. It is likely that the university email system includes a spam filter that tests links within emails such as the survey invitation email. This seems isolated to the University of Oxford because the 962 cases are 56 percent of all UK Incompletes. All other UK institutions range from eight to less than one percent.

6.1   Incentives

6.1.1   E-gift Card

All respondents were offered an e-gift card incentive as a gesture of appreciation for their time and contribution to the survey. Abt partnered with a gift card vendor that utilized an Application Programming Interface (API) which allowed respondents to redeem their e-gift card on a self-initiated portal at the end of the web survey. Incentives were issued in local currency in an amount approximately equivalent to $20 (USD): ₹1500 for India, €15 for Italy and £15 for the UK. Of the 3,442 respondents who completed the survey, 403 respondents (12%) declined the incentive.

6.1.2   Apple iPad Raffle

All respondents who completed the survey by a specified date were offered a chance to enter a raffle for an Apple iPad. The raffle winners were randomly selected by Abt at the end of each wave. The iPads were ordered by CUA and shipped directly to the winners from their local Apple store. For wave 1, respondents must have completed the survey by June 4, 2021, to be eligible for the raffle. A total of four winners were selected, one per country. Of the 1,546 respondents eligible for the raffle, 317 respondents (21%) declined to participate. For wave 2, all respondents who completed the survey during wave 2 were offered an opportunity to enter a raffle for an Apple iPad, regardless of when they completed. Only one winner was selected across all countries. Of the 1,192 respondents eligible for the raffle in wave 2, 387 respondents (32%) declined to participate.

6.1.3   Personalized Report

All respondents who completed the survey by a specified date received a personalized report at the end of the study. The report compared the respondent's key aggregate scores to others in their country and discipline. CUA produced the personalized reports and Abt sent them via email to the respondents. For wave 1, respondents had to have completed the survey by June 23, 2021. For wave 2, all respondents who completed the survey received the report, regardless of when they completed. A total of 3,164 respondents received the report.

6.1.4   General Summary Report

All respondents who completed the survey received a general summary report of findings at the end of the study as a way of thanking them for their response and to promote information sharing and collaboration among scientists. This report was same for all respondents and provided overall key aggregated findings from all countries. to promote information sharing and collaboration.

CUA produced the report and Abt sent the report to the respondents via email. A total of 3,442 respondents received the report.

6.2   Project Email

Abt set up a project email for respondents to contact if they had any questions or issues. There was little communication with respondents via email. Communication consisted of inquiries related to the e-gift card, few requests to opt out of the survey (refusals), and a few requests to opt into the survey. The requests to opt in were facilitated for 10 individuals who were determined to be eligible to participate. They were eligible if they were from an institution and department that was included in the sample but they had not been sent an email invitation. This may have been due to personnel changes within a department between the time the sampling frame was generated, and the survey launched. These respondents reached out to Abt to opt in based on hearing about the study from their colleagues. Other requests to opt in were declined if the individual was not deemed eligible, i.e., if they had heard about the study from a colleague but did not meet the criteria for inclusion the sample (geography, institution, department, or career level).

7.   Response Rate

The final response rate for the Work and Well-Being in Science survey was 15.2%. Exhibit 7-1 below displays the final dispositions and response rate calculation using Response Rate 3 (RR3) computed according to current American Association for Public Opinion Research (AAPOR) Standard Definitions of case codes and outcome rates.[13] The response rate calculation is defined as the sum of the completed interviews divided by the sum of the completed, partially completed, refusals, non-contacts, other, and a portion of cases with unknown eligibility.[14] In this survey there were no cases with unknown eligibility. Sample members that did not respond were counted as eligible because they were drawn from the sampling frame of researchers that would qualify for the survey. Sample members that were counted Not Eligible (No eligible respondent) were those that screened out of the survey. They are excluded from the RR3 denominator.

Exhibit 7-1.   Response Rate

Disposition	AAPOR Code	N	%
Interview (Category 1)			
Complete	1.100	3,442	15.1%
Partial	1.200	1,093	4.8%
Eligible, Non-Interview (Category 2)			
Refusal	2.110	956	4.2%
Other, non-refusals	2.900	17,164	75.1%
Unknown Eligibility, Non-Interview (Category 3)			
N/A			
Not Eligible (Category 4)			
No eligible respondent	4.900	185	0.8%
Total		**22,840**	**100.0%**
Response Rate Formula			
AAPOR Response Rate 3* Formula: I/((I+P) + (R+NC+O) + e(UH+UO))			15.2%

Key: I = Completed interview (1.1); P = Partial interview (1.2); R = Refusal (2.1); NC = Non-contact (2.2); O = Other non-response (2.3); UH = Unknown if household, UO = Other unknown eligibility; NE = Not eligible (4)
* The e coefficient in AAPOR Response Rate 3 was computed as (I+R+NC+O)/((I+R+NC+O)+NE).

---

[13] American Association for Public Opinion Research (AAPOR). (2016). Standard definitions. Retrieved from https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.

[14] Because the total for partially completed interviews is not part of the denominator, the UK universities that gave the false impression of a partial interview did not affect the response rate. These cases counted as either a partial interview or non-contact.

8. Final Weights

This section describes the process involved in creating the analysis steps. The sampling frame of scraped individuals contained 22,840 entries, described in Exhibit 8-1.

Exhibit 8-1. Sampling Frame Summary

Country	Sample	Respondents	Institutions
IN	5,418	1,345	130
IT	5,497	637	40
UK	6,468	925	30
US	5,457	535	33
Total	**22,840**	**3,442**	**233**

For USA and UK, the sample of institutions was stratified. For the USA, stratification was by the institution size (three groups: largest and medium, mostly comprised of the very high research intensity doctoral universities; and small, comprised of all other universities). For UK, the sample was stratified according to the existing clusters (Oxford + Cambridge; other pre-1992 universities divided into higher and lower tiers). Additional institutions were also sampled in the UK whose function is pure research without undergraduate instruction (e.g., Institute of Cancer Research.) In India and Italy, all eligible institutions were sampled. The institution weights were defined as inverse probabilities of selection from the respective countries' frames of institutions, shown in Exhibit 8-2.

Exhibit 8-2. Weights by Country

Country	Institution Stratum	Institution Weight
US	1	1.555556
US	2	13.888889
US	3	18.000000
UK	0	1.000000
UK	1	1.000000
UK	2	1.772727
UK	3	8.800000
IT	0	1.000000
IN	0	1.000000

To define person-level weights, response propensities were modeled as interactions of:

- discipline (biology and physics),
- country,
- gender (male, female, unknown),
- seniority.

The latter was coded into 10 groups including "uncoded/other" that would contain groups of size 40 or less in the original data as shown in Exhibit 8-3.

Exhibit 8-3. Groups for Seniority

Seniority	N
AssistProfLecturer	1,408
AssocProf	1,920
GradStudent	5,684
LeadResearcher	440
Missing	314
Postdoc	1,425
Professor	2,863
Researcher	4,447
TeachingProf	1,083
Uncoded	3,256
Total	**22,840**

The specific response propensity model included all pairwise interactions except country-by-seniority, and random effects of institution. This model was selected by the best Akaike Information Criterion (AIC).[15] Person-level weights were obtained as inverse predicted response propensities.

```
## Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
##  Family: binomial  ( logit )
## Formula: svy_response ~ S_CountryCode * S_Gender + S_CountryCode * S_Field +
##    S_Gender * S_Field + S_Gender * seniority + S_Field * seniority +    (1 | S_Institution)
##    Data: WWB_frame
##
##     AIC     BIC   logLik deviance df.resid
## 18055.2 18497.2 -8972.6 17945.2   22785
##
## Scaled residuals:
##    Min    1Q Median    3Q    Max
## -1.3557 -0.4389 -0.3358 -0.2515 5.9069
##
## Random effects:
##  Groups      Name      Variance Std.Dev.
##  S_Institution (Intercept) 0.2356   0.4854
## Number of obs: 22840, groups:  S_Institution, 247
##
## Fixed effects:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -1.3491202  0.1829883  -7.373 1.67e-13 ***
## S_CountryCodeIT            -0.7336109  0.1719757  -4.266 1.99e-05 ***
## S_CountryCodeUK             0.0446932  0.1772814   0.252 0.800961
## S_CountryCodeUS            -0.6327524  0.1747412  -3.621 0.000293 ***
## S_GenderM                  -0.1508200  0.1859026  -0.811 0.417202
```

---

[15] Akaike 1974, https://ieeexplore.ieee.org/document/1100705

```
## S_GenderUnknown                              -0.2003477 0.3426933 -0.585 0.558799
## S_FieldPhysics                               -0.0920746 0.1939103 -0.475 0.634907
## seniorityAssocProf                           -0.0215009 0.2186754 -0.098 0.921676
## seniorityGradStudent                          0.0967980 0.1912852  0.506 0.612828
## seniorityLeadResearcher                       0.0701821 0.3070709  0.229 0.819216
## seniorityMissing                              0.0513620 0.2808650  0.183 0.854899
## seniorityPostdoc                              0.2025431 0.2317440  0.874 0.382121
## seniorityProfessor                           -0.3769061 0.2292291 -1.644 0.100128
## seniorityResearcher                           0.0238790 0.1971566  0.121 0.903599
## seniorityTeachingProf                        -0.3424987 0.2929920 -1.169 0.242416
## seniorityUncoded                             -0.1452080 0.1968054 -0.738 0.460621
## S_CountryCodeIT:S_GenderM                    -0.0560819 0.1441156 -0.389 0.697169
## S_CountryCodeUK:S_GenderM                    -0.5537900 0.1207703 -4.585 4.53e-06 ***
## S_CountryCodeUS:S_GenderM                    -0.5175867 0.1379559 -3.752 0.000176 ***
## S_CountryCodeIT:S_GenderUnknown              -0.0484451 0.1935117 -0.250 0.802319
## S_CountryCodeUK:S_GenderUnknown              -0.6265227 0.2180778 -2.873 0.004067 **
## S_CountryCodeUS:S_GenderUnknown              -0.3422837 0.2230501 -1.535 0.124892
## S_CountryCodeIT:S_FieldPhysics               -0.0498012 0.1876345 -0.265 0.790689
## S_CountryCodeUK:S_FieldPhysics               -0.2264865 0.1798110 -1.260 0.207821
## S_CountryCodeUS:S_FieldPhysics                0.0361136 0.1855872  0.195 0.845713
## S_GenderM:S_FieldPhysics                      0.0003605 0.0973482  0.004 0.997045
## S_GenderUnknown:S_FieldPhysics               -0.2429062 0.1589780 -1.528 0.126532
## S_GenderM:seniorityAssocProf                 -0.0019899 0.2442570 -0.008 0.993500
## S_GenderUnknown:seniorityAssocProf            0.3843027 0.4070448  0.944 0.345104
## S_GenderM:seniorityGradStudent                0.2547166 0.1958461  1.301 0.193397
## S_GenderUnknown:seniorityGradStudent          0.6408327 0.3513491  1.824 0.068164 .
## S_GenderM:seniorityLeadResearcher            -0.2483255 0.3454924 -0.719 0.472290
## S_GenderUnknown:seniorityLeadResearcher -1.1348316 0.7316679 -1.551 0.120897
## S_GenderM:seniorityMissing                    0.4340146 0.3227461  1.345 0.178704
## S_GenderUnknown:seniorityMissing              0.4722924 1.3354274  0.354 0.723591
## S_GenderM:seniorityPostdoc                    0.2440993 0.2529277  0.965 0.334497
## S_GenderUnknown:seniorityPostdoc             -0.0221979 0.4082288 -0.054 0.956636
## S_GenderM:seniorityProfessor                  0.3459368 0.2422284  1.428 0.153251
## S_GenderUnknown:seniorityProfessor            0.4471781 0.4088999  1.094 0.274125
## S_GenderM:seniorityResearcher                 0.1686048 0.2072548  0.814 0.415923
## S_GenderUnknown:seniorityResearcher           0.0091961 0.3590213  0.026 0.979565
## S_GenderM:seniorityTeachingProf               0.2780646 0.3183058  0.874 0.382349
## S_GenderUnknown:seniorityTeachingProf   0.0191274 0.5029568  0.038 0.969664
## S_GenderM:seniorityUncoded                   -0.0356429 0.2157078 -0.165 0.868757
## S_GenderUnknown:seniorityUncoded              0.0018073 0.3765459  0.005 0.996170
## S_FieldPhysics:seniorityAssocProf             0.1778563 0.2250849  0.790 0.429426
## S_FieldPhysics:seniorityGradStudent           0.4589127 0.1974595  2.324 0.020121 *
## S_FieldPhysics:seniorityLeadResearcher -0.0417892 0.4885465 -0.086 0.931834
## S_FieldPhysics:seniorityMissing               0.7386066 0.4264148  1.732 0.083250 .
## S_FieldPhysics:seniorityPostdoc               0.1471987 0.2442775  0.603 0.546783
## S_FieldPhysics:seniorityProfessor             0.1987310 0.2108565  0.942 0.345940
## S_FieldPhysics:seniorityResearcher            0.1105877 0.2043614  0.541 0.588412
## S_FieldPhysics:seniorityTeachingProf   -0.1740522 0.2909321 -0.598 0.549668
## S_FieldPhysics:seniorityUncoded               0.3996383 0.2081998  1.919 0.054922 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
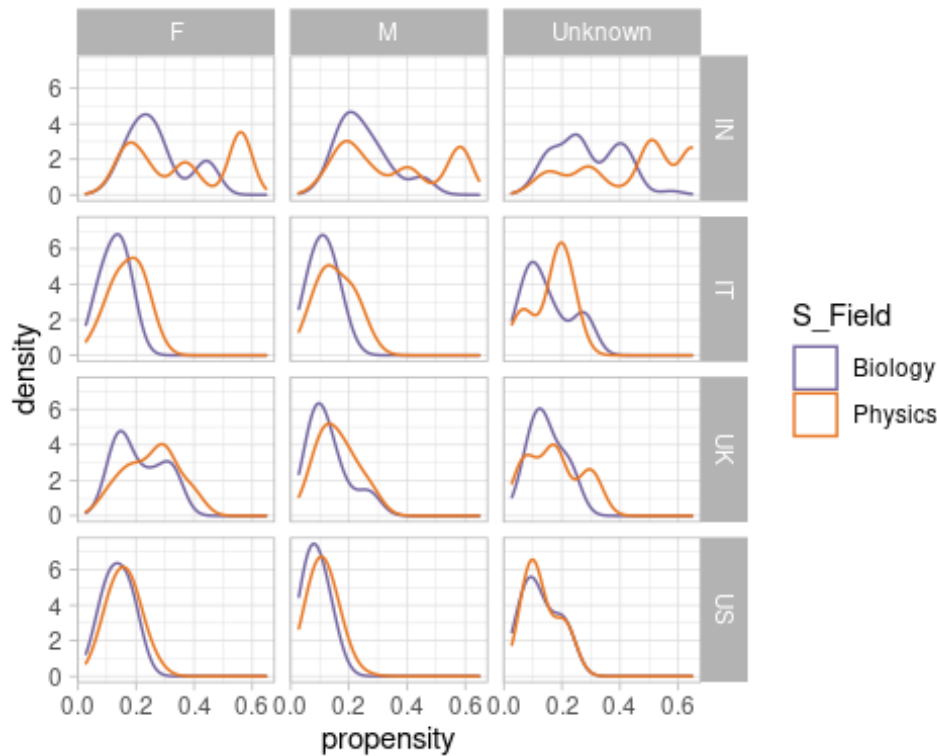
```
##
## Correlation matrix not shown by default, as p = 54 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## optimizer (Nelder_Mead) convergence code: 4 (failure to converge in 10000 evaluations)
## Model failed to converge with max|grad| = 0.00930532 (tol = 0.002, component 1)
## failure to converge in 10000 evaluations
```


The final weights were obtained as the product of institution (sampling) and person (response propensity) weights. See Exhibit 8-4 for the summary of the final weights.

Exhibit 8-4. Final Weights Summary

Country	Institution Stratum	N	Institution Weight	min pweight	mean pweight	max pweight
IN	0	1,345	1.000000	1.544088	3.885177	12.63498
IT	0	637	1.000000	3.622796	8.410995	28.48838
UK	0	17	1.000000	4.748151	7.869569	12.94376
UK	1	197	1.000000	6.070837	10.294587	28.87994
UK	2	622	1.772727	4.338537	10.249722	38.05959
UK	3	89	8.800000	24.104175	67.040095	138.95516
US	1	345	1.555556	6.028744	17.050926	55.83088
US	2	175	13.888889	53.143326	118.045634	253.82015
US	3	15	18.000000	134.359519	222.169970	364.14555
Total		**3,442**				

To conduct analysis in Stata with appropriate corrections for the complex survey design, the following specification should be used:

```
svyset n_Institution [pweight=pweight], strata(inst_size_stratum)
fpc(pop_institutions)
```

The variables used in this specification are:

- n_Institution: institution, the primary sampling unit (numbered 1 to 253, with gaps; a scrambled sequential numbering of universities on the frame)

- pweight: analysis weight correcting for unequal probabilities of selection and non-response

- inst_size_stratum: indicator of the sampling stratum (single stratum for Italy and India; three sampling strata for the USA and UK)

- pop_institutions: total number of institutions on the sampling frame.

There is a non-trivial danger of wiping out some of the lesser populated strata and PSU when analyzing subpopulations and variables with missing data, which runs the risk of producing missing standard errors. To avoid that, the filters for non-missing values of the analysis variables should be entered into the subpop() option. E.g., instead of

```
svy, subpop(if S_Gender == "F"): tab hflifesat_1, se
```

(which works OK, but is used here for illustration), one may want to run

```
svy , subpop( if S_Gender == "F" & !mi(hflifesat_1 ) ): tab hflifesat_1, se
```

where the analysis variable(s) are restated as non-missing in the subpop() option.