|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Question | Hypothesis | Sampling plan | Analysis Plan | Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis | Interpretation given different outcomes | Theory that could be shown wrong by the outcomes |
| Research Question 1: What is the prevalence of Direct Behavioral Measures in the field of psychology from 2009-2023? | Psychology has not regularly collected the necessary measurements to treat behavior as its explanandum from 2009-2023, and the field is therefore not accurately described as the ‘science of human behavior’. | 450 articles (15 articles per journal, 5 journals per subfield, 6 subfields) were rated for use of at least one Direct Behavioral Measure was. | Bayesian hierarchical model from which Marginal Global Means (aka Average Marginal Effects) for prevalence were extracted along with their 95% Credible Intervals using the percentile method.  If Direct Behavioral Measures are shown to have an MGM prevalence of less than 10%, we would conclude that psychology did not, within the observed time period, regularly collect the necessary measurements to treat behavior as its explanandum, and the field therefore *is not* accurately described as the ‘science of human behavior’. | Sample size was primarily determined by the availability of resources, based on the amount of time it takes to code the articles. | If Direct Behavioral Measures are shown to have an MGM prevalence of greater or equal to 10%, we would conclude that psychology did, within the observed time period, regularly collect the necessary measurements to treat behavior as its explanandum, and the field therefore *is* accurately described as the ‘science of human behavior’. | Not a psychological theory, but rather a description of our collective research practices during the time period and their alignment with our field’s collective self-image and definition, according to our review of 22 common introductory textbooks and professional dictionaries. |
| Research Question 2: Has psychology capitalized on the technological changes of the last two decades to collect greatly more data on overt behavior (e.g., via smartphones and smart devices)? | We hypothesize that psychology has not capitalized on the technology available to collect these data. | Same rationale as above. | Same model as the above, this time extracting the MGM trend for centred year and its 95% CRs.  If the data show a detectable positive trend trend in the use of Direct Behavioral Measures (i.e., the lower bound of the CR is greater than zer0), we will conclude that psychological research has increased its use of behavioural measures during the time period (e.g., by capitalizing on the opportunity that technology has afforded for direct observations of behavior). | Same rationale as above. | A detectable negative trend or no detectable trend, based on the CR’s exclusion of 0, will be interpreted as evidence against the claim that the use of behavioral measures has increased over the time period or a lack of evidence to support the claim that they have increased over the time period, respectively. | Not a psychological theory, but rather a description of our collective research practices during the time period and their potential future alignment with our field’s collective self-image and definition, according to our review of 22 common introductory textbooks and professional dictionaries. |
| Research Question 3 defines a more descriptive and exploratory approach to analysis (see the manuscript) |  |  |  |  |  |  |
| Research Question 4 defines a more descriptive and exploratory approach to analysis (see the manuscript) |  |  |  |  |  |  |

**Guidance Notes**

* **Question**: articulate each research question being addressed in one sentence.
* **Hypothesis**: where applicable, a prediction arising from the research question, stated in terms of specific variables rather than concepts. Where the testability of one or more hypotheses depends on the verification of auxiliary assumptions (such as positive controls, tests of intervention fidelity, manipulation checks, or any other quality checks), any tests of such assumptions should be listed as hypotheses. Stage 1 proposals that do not seek to test hypotheses can ignore or delete this column.
* **Sampling plan**: For proposals using inferential statistics, the details of the statistical sampling plan for the specific hypothesis (e.g power analysis, Bayes Factor Design Analysis, ROPE etc). For proposals that do not use inferential statistics, include a description and justification of the sample size.
* **Analysis plan**: For hypothesis-driven studies, the specific test(s) that will confirm or disconfirm the hypothesis. For non-hypothesis-driven studies, the test(s) that will answer the research question.
* **Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis**: For hypothesis-driven studies that employ inferential statistics, an explanation of how the authors determined a relevant effect size for statistical power analysis, equivalence testing, Bayes factors, or other approach.
* **Interpretation given different outcomes**: A prospective interpretation of different potential outcomes, making clear which outcomes would confirm or disconfirm the hypothesis.
* **Theory that could be shown wrong by the outcomes**: Where the proposal is testing a theory, make clear what theory could be shown to be wrong, incomplete, or otherwise inadequate by the outcomes of the research.