Academic Research Climate Amsterdam
Data analysis plan

Academic Research Climate Amsterdam
Data analysis plan

1. **General description ARCA survey project**

This study investigates the Academic Research Climate in Amsterdam (ARCA). To get insight in this climate, we explore three factors known to play a role in this climate. Firstly, we look at the organisational research climate. Organizational climate here is defined as "the shared meaning organizational members attach to the events, policies, practices, and procedures they experience and the behaviors they see being rewarded, supported, and expected." (Ehrhart, Schneider & Macey, 2013, p. 115). Secondly, we look into the publication pressure, focusing on the extent to which researchers report publication pressure as this could be an indicator of a hyper-competitive research climate (Tijdink, 2014). Lastly, we examine the frequency and impact of both major and minor misbehaviors that academics perceive in their daily research practice at these 4 institutions.

2. **Relation protocol, data analysis plan, survey**

This data analysis plan builds on the ARCA protocol, that can be found here or via http://amsterdamresearchclimate.nl/full-protocol/

This data analysis plan is intended to guide the analysis of all data gathered with the ARCA survey that can be found here or via https://surfdrive.surf.nl/files/index.php/s/rhEAWrUap69jQ6a. Please note that numbering in this document can be somewhat confusing, as Qualtrics assigns a new number (e.g. 71) when you adjust the wording of what was initially item number 7. Below it will be specified what items belong to which instrument, which then again forms an outcome variable. It will also be explained which items are determinants and each variable can be found (see sections 12 and 13 of this document).

3. **Programs used for analysis and data collection**

All data were collected using Qualtrics with which the VU Amsterdam has a specific license agreement.
All data will be analyzed using SPSS.

4. **Type of study**

Cross-sectional, descriptive/explorative

5. **Sample**

Our population consists of all 7548 academic researchers in Amsterdam (total population sample). We included researchers that engage in academic research for more than one day per week (>0.2 fte).

6. **Missing values**

Respondents had to fill in all items of an instrument before going to the next section ('forced response',see survey). If respondents only filled in one instrument and then stopped, we will use their responses for response calculations of the completed instrument(s), but not for any of the instruments which they didn't fill in completely. We will not use imputation techniques for the other items the respondent failed to fill in.

Responses for major and minor misbehaviors were collected using missingness by design. This means that respondents only were presented a random selection of 20 items out of 60 only. We will only analyze an item if there are at least 15 responses to that item.

7. **Outlier treatment**

As the SOuRCe has specific criteria for a response to be valid, we will not look for outliers in the SOuRCe responses and only use the valid scale scores in accordance with the SOuRCe user guide (Thrush et al. 2014).

For the PPQ sum score, we will visually inspect each item's score using a scatter plot and delete extreme outliers (more than 3SD) if all team members reach consensus about this (for team members, see Protocol, section 2, p. 4).

As for the major and minor misbehaviors, we will visually inspect for outliers using a scatterplot for overall frequency and impact scores. If scores strongly deviate from average scores (more than three standard deviations) and if all team members believe that a respondent's responses are nonsense (e.g. a respondent only clicked the most right or most left button), then this score can be excluded from analyses, but only if all team members reach consensus about this. A sensitivity analysis will be conducted to see the difference with and without outliers, if the difference alters the effect, both outcomes will be reported.

8. **Expectations about normality**

We will assess normality of dependent variables and residuals from regression analyses visually using QQ-plots and histograms. If strongly skewed to the left, we will attempt a log transformation to normalize the data or use Mann Whitney or Kruskal Wallis tests for not-normal data. Otherwise we will use independent samples t-test, one-way ANOVA or linear regression whichever appropriate.

9. **Response rate calculations**

We define the following response rates:
1.Total response rate: number of respondents that completed full questionnaire / total invitees
2. Corrected response rate: number of respondents that completed full questionnaire / total invitees whose e-mail addresses were correct (that means subtracting those denoted as "bounced" in Qualtrics from the total) and who were at the time of the survey actually employed at an academic institution in Amsterdam (subtracting those people that notified the research team via e-mail that they were no longer employed at an Academic institution in Amsterdam from the total).
3. Survey response rate: number of respondents that completed full questionnaire / number of invitees that started the questionnaire
4. Corrected SOuRCe response rate: number of respondents that completed the SOURCE / total invitees whose e-mail addresses were correct and who were at the time of the survey actually employed at an academic institution in Amsterdam
5. SOuRCe survey response rate: number of respondents that completed the SOURCE / number of invitees that started the questionnaire
6. Corrected PPQ respondents: number of respondents that completed the PPQ / total invitees whose e-mail addresses were correct and who were at the time of the survey actually employed at an academic institution in Amsterdam
7. PPQ survey response rate: number of respondents that completed the PPQ / number of invitees that started the questionnaire
8. Corrected 60 items respondents: number of respondents that completed the random 20 items / total invitees whose e-mail addresses were correct and who were at the time of the survey actually employed at an academic institution in Amsterdam
9. 60 items survey response rate: number of respondents that completed the random 20 items / number of invitees that started the questionnaire

10. Research questions

This data analysis plan focuses on the following research questions (Protocol, section 5.1, p. 5):
a. What are the perceptions of academic researchers in Amsterdam regarding the organizational climate for responsible research practices both in one's general organizational setting and in one's specific department, stratified for academic rank and disciplinary field?
b. What is the level of publication pressure in the four academic institutions in Amsterdam, stratified for academic rank and disciplinary field?
c. What do scientists in the four academic institutions in Amsterdam consider to be the most detrimental research misbehaviors, stratified for academic rank and disciplinary field?
d. What are the effects of publication pressure and the organizational climate on what scientists in the four academic institutions in Amsterdam perceive to be the most frequent research misbehaviors?

11. Research questions operationalized

The research questions and their related subquestions were operationalized as follows:

a. The research climate will be assessed with the Survey of Organizational Research Climate (henceforth: SOuRCe), consisting of 7 subscales.
a. 1 How do academic researchers in Amsterdam experience their organizational climate?
a. 2 Do academic researchers in Amsterdam working in different disciplinary fields experience their organizational research climate differently?
a. 3 Do academic researchers in Amsterdam from different academic ranks experience their organizational research climate differently?

b. The level of publication pressure will be measured by getting the sum score of the 18-item revised version of the Publication Pressure Questionnaire (henceforth: PPQ) and of three subscales: stress resulting from publication pressure, attitude regarding the publication climate, and lastly resources to cope with stress.
b. 1 What is the level of publication pressure of researchers in Amsterdam?
b. 2 Does the pressure to publish differ between academic disciplinary fields?
b. 3. Does the pressure to publish differ between academic ranks?

c. We will answer this question using frequency adjusted impact by ranking the major and minor misbehaviors according to how detrimental they were reported to be (the higher the average product of frequency and impact; the more important).
c. 1 What is the top 5 of most detrimental misbehaviors (defined as above)?
c. 2 What is the top 5 of most detrimental misbehaviors in different disciplinary fields?
c. 3 What is the top 5 of most detrimental misbehaviors in different academic ranks?

d. We look into this question from the idea that in a setting where the organizational climate is bad and the publication pressure is high, more major and minor misbehaviors will be observed. Specifically, we will employ the SOuRCe subscales and the PPQ subscales as independent variables (determinants) and the most frequently perceived misbehaviors as dependent variables (outcome variable).
d. 1 What is the effect of the SOuRCe subscale scores on the 5 most frequently perceived major and minor misbehaviors?
d. 2 What is the effect of the PPQ subscales scores on the 5 most frequently perceived major and minor misbehaviors?
d. 3 What is the combined effect of the SOuRCe subscales and the PPQ sub scales on the 5 most frequently perceived major and minor misbehaviors?

Based on the questions stipulated above, we have also listed what we specifically expect to see in our data in Table 1*.

Nr.	Explorative hypotheses*
a. 2	SOuRCe subscale scores differ per discipline.
a. 3	SOuRCe subscale scores differ per academic rank.
b. 2	PPQ subscale scores and overall sum scores will differ between disciplinary fields.
b. 3	PPQ subscale scores and overall sum scores will differ between academic ranks.
c. 2	Top 5 rankings will be different for different disciplines.
c. 3	Top 5 rankings will be different for different academic ranks.
d. 1	The more negative the organizational climate, the stronger the association with the most frequently perceived misbehaviors
d. 2	The higher the publication pressure, the stronger the association with the most frequently perceived misbehaviors
d. 3	The higher the publication pressure and the more negative the organizational climate, the stronger the association with the most frequently perceived misbehaviors *Note: as publication pressure and the organizational climate may correlate, we expect some overlap but the effect of the organizational climate and the publication pressure to be greater combined than either the effect of publication pressure alone, or the effect of the organizational climate alone*

Table 1. * *As these questions are explorative, we will not hypothesise the direction of the difference we are interested in; these are all nondirectional.*

12. **Outcome variables**

Below the instrument is specified, followed by the created variables.

a. Survey of Organizational Research Climate (Thrush et al. 2007; Crain, Martinson & Thrush, 2012; Martinson, Thrush & Crain, 2013; Wells et al. 2014)

   The SOuRCe consists 28 items comprising seven subscales. For each subscale, a variable will be created when the responses to items of that subscale are valid. This is calculated in the following way: NITEM = Number of items in the subscale
   NVAL = Number of items with valid (non-missing) data for person i
   SUMVAL = Sum of values of items with valid data for person i
   IF NITEM/NVAL >= 2,
   Scale Score = SUMVAL/NVAL
   These seven subscale scores result in in seven new variables:
   (full name, code, number of items)
   ● Institutional RCR Resources (IRCRRR 6 items)
   ● Institutional Regulatory Quality (IRQ 3 items)
   ● Subunit Integrity Norms (SIN 4 items)
   ● Subunit Integrity Socialization (SIS 4 items)
   ● Subunit Advisor /Advisee Relations (SAR 3 items)
   ● Subunit Integrity Inhibitors (SII 6 items)
   ● Subunit Expectations (SE 2 items)
   See survey p. 6 – 19, Q6 – Q29.

Publication Pressure Questionnaire (Tijdink et al. 2014; Tijdink, Verbeke & Smulders, 2014; Tijdink, Vergouwen & Smulders, 2014)
The PPQ consists of a sumscore, as such one variable will be created: PPQ_SS
The revised PPQ consists of three subscale scores, namely the sumscore of subscale publication stress: PPQ_PS, the sumscore of subscale Publication Attitude: PPQ_PA and lastly the sumscore of subscale Resources: PPQ_RE
See survey p. 52 – 55, PPQ Q1-Q18.

Major and minor misbehaviors questionnaire (Bouter et al. 2016)
Note: Participants were administered a random selection of 20 items out of the total pool of 60 items to decrease the length of the survey. However, we expect with our total sample that we will have enough responses per item to get an indication how frequent this behavior is judged to have happened over the last three years and what its impact is on the validity of the study's findings. We will thus pool all responses together (*analysed only when a cell has at least 15 responses)*

The list of major and minor misbehaviors consists of 60 items. For each of these 60 item, the perceived frequency (Nr_F) and perceived impact (Nr_I) is measured. Also, a product score of reported frequency times reported impact will be computed. This results in 60 new items with the following format: Number of item_Frequency*Impact (Nr_FxI)
See survey p. 21 – 51, Q37 – Q101.

13. **Determinants**

- Academic rank: (a) PhD-student, (b) Postdoc/Assistant professor, (c) Associate/Full Professor
See survey p. 50, Rank/Q36.
Note: within the subgroup PhD students, we differentiate between internal PhD students and external PhD students
See survey p. 50, Q111.
- Disciplinary field: (a) Life Sciences and Medical Sciences, (b) Natural Sciences and Engineering Sciences, (c)Social Sciences and Behavioral Sciences, (d) Humanities, Language, Communication, Law and Arts. We based this definition of fields on http://www.cwts.nl/pdf/nowt_-classification_sc.pdf
See survey p. 50, Discipline/Q35.
~~- Gender: (a) Male, (b) Female~~
~~See survey p.50, Gender/Q34~~

14. **Confounders and effect modifiers**

1. Academic rank if discipline is main variable of interest
2. Disciplinary field when rank is main variable of interest
3. Gender when rank or discipline are predictor of interest

Note: we will not use the two other outcome measures (e.g. PPQ and 60 major and minor misbehaviors) when one outcome variable is of interest for the model (e.g. SOURCE). We will only inspect the correlations between the different outcome variables, but not inspect them as confounders or effect modifiers as we have only measured cross sectionally at one time.

a. Treatment of confounders
We will check for confounders when the model with the relevant determinant is in fact significant. We will first also check if the confounding variable by itself has a significant effect (i.e. effect gender on outcome variables).
   i) How to check for confounder
   We will check for confounders using regression analysis, whereby we start the

model with only the main determinant and the outcome variable. Then we insert the confounder to the model as a second variable. We will then check the change in the main determinant.

ii) What counts as a confounder
A confounder is a variable that induces a change of more than 10% in the value of the determinant, when added to the model. We will add potential confounders one by one and check if it exceeds the 10% rule per confounder, it will be added to the multivariable model.

iii) Confounder with dummies
When our main determinant takes multiple values, e.g. academic rank (3 options; 2 dummies) and a potentially confounding variable is added to the model, we will treat this variable as confounding when the beta coefficient of at least one of the dummies changes relevantly by >10%.

ii) Treatment of effect modifiers
After having checked for confounders, we will check for effect modifiers. We will test for effect modifiers by starting with a model that has only the outcome variable and determinant in it. We will then create a new variable, that is the product of the determinant and the effect modifier (two-way interaction). We will then add both the effect modifier and this new variable to the model and check whether the newly created variable is significant ($p<0.05$). If so, we will take into account this variable as an effect modifier.

All analyses will also be done for the SOuRCe and PPQ subscales.

15. Univariable analyses

a. SOuRCe

Descriptives:

If not normally distributed, we will use median and interquartile range.

a.1 Mean scale scores for total sample, per subgroup rank and per subgroup discipline
Standard deviations scale scores for total sample, per subgroup rank and per subgroup discipline

Analyses:

a. 2 Regression analysis with SOuRCe subscales as outcome variable and rank as predictor

a. 3 Regression analysis with SOuRCe subscales as outcome variable and discipline as predictor

a. 2. 1 Pairwise post-hoc regression analyses with SOuRCe subscale to see which academic ranks score differently, applying the Bonferroni correction where justified (only in case overall effect in a.2. is significant).

a. 3. 1 Pairwise post-hoc regression analyses with SOuRCe subscale to see which disciplinary fields score differently, applying the Bonferroni correction where justified (only in case overall effect in a.3. is significant).

Statistics

Mean, SD, Regression coefficients, 95% Confidence Intervals, F value, df, p-value for two-sided 5% significance level, Standardized Mean Differences

b. PPQ

Descriptives:

If not normally distributed, we will use median and interquartile range.

b.1 Mean PPQ_SS score for total sample, per subgroup rank and per subgroup discipline
Standard deviations PPQ_SS for total sample, per subgroup rank and per subgroup discipline

Analyses:

b. 2 Regression analysis with PPQ_SS as outcome variable and rank as predictor to see whether CI includes 0
b. 3 Regression analysis with PPQ subscales as outcome variable and discipline as predictor to see whether CI includes 0
b. 2. 1 Pairwise post-hoc regression analyses with PPQ_SS to see which academic ranks score differently, applying the Bonferroni correction (only in case overall effect in b.2. is significant).
b. 3. 1 Pairwise post-hoc regression analyses with PPQ subscales to see which disciplinary fields score differently, applying the Bonferroni correction (only in case overall effect in b.3. is significant).
Statistics
Mean, SD, Regression coefficients, 95% Confidence Intervals, F value, df, p-value for two-sided 5% significance level, Standardized Mean Differences

c. Major and minor misbehaviors
Descriptives:
- Product scores per item of frequency and impact
- Mean product scores Frequency*Impact per item for total sample as well as per subgroup rank and per subgroup discipline
c.1 Top five of most detrimental items with the highest Frequency*Impact score total sample
c. 2 Top five of most detrimental items with the highest Frequency*Impact score per subgroup rank and
c. 3 Top five of most detrimental items with the highest Frequency*Impact score per subgroup per subgroup discipline

d. Influence of PPQ & SOuRCe scores on perceived frequency misbehaviors
Descriptives:
- Mean frequency scores per item for total sample as well as per subgroup rank and per subgroup discipline
d. 1 Regression analysis with mean perceived frequency scores as outcome variables and SOuRCe subscales as predictors.
d. 2 Regression analysis with mean perceived frequency scores as outcome variables and PPQ subscales as predictors.
d. 3 Regression analysis with mean perceived frequency scores as outcome variables and both PPQ subscales and SOuRCe subscales as predictors.
Statistics
Mean, SD, regression coefficients, F value, df, p-value corrected for multiple testing; using p value of 0.001 instead of the regular 0.05 [calculated as follows: initial p value 0.05 / (10 subscales in total * 5 most frequent items = 50) = 0.001]

16. Multivariable analyses

Here we will list our multivariate analyses, using the same numbering per analysis as in section 15. For example: a.2 refers to "Regression analysis with SOuRCe subscales as outcome variable and rank as predictor". Here we correct for potential confounders and effect modifiers. If it says "corrected for X and or Y", we will separately investigate X and Y to see whether they function as confounders/effect modifiers. We can then add the next confounder/effect modifier to the model and test which model works best.

a. SOuRCe
a. 2 corrected for discipline and or gender as confounder or if discipline and or gender is an effect modifier; with discipline and or gender added to the model as effect modifier (hence, discipline and two-way interaction between discipline and rank).
a. 3 corrected for rank and or gender as confounder or if rank and or gender was an effect

modifier; with rank and or gender added to the model as an effect modifier
a. 2.1 corrected for discipline and or gender as confounder or if discipline and or gender is an effect modifier; with discipline and or gender added to the model as effect modifier
a. 3.1 corrected for rank and or gender as confounder or if rank and or gender was an effect modifier; with rank and or gender added to the model as an effect modifier

b. PPQ
b. 2 corrected for discipline and or gender as confounder or if discipline and or gender is an effect modifier; with discipline and or gender added to the model as effect modifier
b. 3 corrected for rank and or gender as confounder or if rank and or gender was an effect modifier; with rank and or gender added to the model as an effect modifier
b. 2.1 corrected for discipline and or gender as confounder or if discipline and or gender is an effect modifier; with discipline and or gender added to the model as effect modifier
b. 3.1 corrected for rank and or gender as confounder or if rank and or gender was an effect modifier; with rank and or gender added to the model as an effect modifier

c. Not applicable.

d. d. 1. While correcting for SOuRCe subscales as confounders or if effect modifier; with SOuRCe subscales added to the model as effect modifier.
d. 2. While correcting for PPQ subscales as confounders or if effect modifier; with PPQ subscales added to the model as effect modifier.
d. 3. Combining the effect while correcting for gender as confounder of if gender is an effect modifier; with gender added to the model as effect modifier.

17. Factsheet analyses

a. Faculties
We will provide faculties with the absolute average values from outcome variables a, b, c from their employees, which we contrast with a benchmark of overall average scores from our total sample. The faculties we refer to can be found in the Protocol (p. 7-8).

We will only provide such a report if at least 25 employees filled the total survey.

Factsheets for faculties will thus be based on the following analyses:
- Mean SOuRCe subscale scores (with SD)
- Benchmark overall sample Mean SOuRCe subscale scores (with SD)
- Mean PPQ sum score (with SD)
- Benchmark overall sample Mean PPQ sum score (with SD)
- Top 5 most detrimental misbehaviors according to their faculty staff (Frequency*Impact according to 25+ employees)
- Benchmark overall Top 5 most detrimental misbehaviors
- Top 60 of reported frequencies per behavior
- Benchmark Top 60 overall sample
- Top 60 of reported impact per behavior
- Benchmark Top 60 overall sample

b. Research institutes
We will provide faculties with their employees' absolute average values of outcome variables a, b, c, which we contrast with a benchmark of overall average scores from our total sample (minus that subset). The research institutes we refer to can be found in the Protocol (p. 7-8).

We will only provide such a report if at least 25 employees filled the total survey.

Factsheets will list the following analyses:
- Mean SOuRCe subscale scores (with SD)
- Benchmark overall sample Mean SOuRCe subscale scores (with SD)
- Mean PPQ sum score (with SD)
- Benchmark overall sample Mean PPQ sum score (with SD)
- Top 5 most detrimental misbehaviors according to their faculty staff (Frequency*Impact according to 25+ employees)
- Benchmark overall Top 5 most detrimental misbehaviors
- Top 60 of reported frequencies per behavior
- Benchmark Top 60 overall sample
- Top 60 of reported impact per behavior
- Benchmark Top 60 overall sample

18. Potential comparative analyses

Below we have outlined a few potential comparative hypotheses. These analyses are inspired by what was previously found using the SOURCE, PPQ and 60 major and minor misbehaviors. We aim here to replicate with our own data whether we find similar findings as in the literature. To test for this, we formulated specific hypotheses we are going to test with our own data. This hypothesis is followed by the reference to the paper that found the finding we expect to replicate.

d. The top 5 rankings from academic researchers working in Biomedical- and Social Sciences will be largely similar. The top 5 of natural sciences will differ from Biomedical- and Social Sciences. The top 5 most detrimental misbehaviors from the humanities will be starkly different from all other disciplinary fields (Bouter et al., 2016). This will be inspected descriptively.
e. Postdoc/Assistant Professors will report the highest mean PPQ sum scores compared to mean scores assigned by PhD students or Associate/Full Professors (Tijdink, Smulders & Vergouwen, 2014).
f. There will be a correlation between PPQ subscale stress scores and top 5 most detrimental misbehaviors items of at least 0.3 (Tijdink, Smulders & Vergouwen, 2014).
g. Researchers will assign higher frequency scores to behaviors defined as minor misbehavior (all behaviors minus the items that specifically regard plagiarism, fabrication and falsification). Overall, at least a third of our sample should assign the score “once or twice” or “3 times or more” to these items (Martinson et al., 2005; Fanelli, 2009).
h. PhD-students and Postdoc-Assistant Professors will assign significantly higher most detrimental misbehavior scores compared to Associate/Full professors (e.g. frequency*impact scores) (Tijdink, Verbeke, & Smulders, 2014).

19. References

Bouter, L. M., Tijdink, J., Axelsen, N., Martinson, B. C., & ter Riet, G. (2016). Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. *Research Integrity and Peer Review, 1*(1). doi:10.1186/s41073-016-0024-5

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One, 4*(5), e5738. doi:10.1371/journal.pone.0005738

Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. *Nature, 435*.

Martinson, B. C., Crain, A. L., De Vries, R., & Anderson, M. S. (2010). The importance of organizational justice in ensuring research integrity. *J Empir Res Hum Res Ethics, 5*(3), 67-83. doi:10.1525/jer.2010.5.3.67

Martinson, B. C., Nelson, D., Hagel-Campbell, E., Mohr, D., Charns, M. P., Bangerter, A., . . . Wells, J. A. (2016). Initial Results from the Survey of Organizational Research Climates (SOuRCe) in the U.S. Department of Veterans Affairs Healthcare System. *PLoS One, 11*(3), e0151571. doi:10.1371/journal.pone.0151571

Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational climate and culture. *Annu Rev Psychol, 64*, 361-388. doi:10.1146/annurev-psych-113011-143809

Tijdink, J. K., Verbeke, R., & Smulders, Y. M. (2014). Publication pressure and scientific misconduct in medical scientists. *J Empir Res Hum Res Ethics, 9*(5), 64-71. doi: 10.1177/1556264614552421

Tijdink, J. K., Smulders, Y. M., Vergouwen, A. C. M., de Vet, H. C. W., & Knol, D. L. (2014). The assessment of publication pressure in medical science; validity and reliability of a Publication Pressure Questionnaire (PPQ). Quality of Life Research, 23(7), 2055-2062. DOI: 10.1007/s11136-014-0643-6

Tijdink, J. K. (2016). Publish & Perish (Doctoral dissertation). Retrieved from https://vu.on.worldcat.org/

Thrush C. R, Martinson B. C, Crain A.L., Wells J.A. (2014). User's Manual for the *Survey of Organizational Research Climate (SOuRCe),* version 2. June, 2014.

Wells, J. A., Thrush, C. R., Martinson, B. C., May, T. A., Stickler, M., Callahan, E. C., & Klomparens, K. L. (2014). Survey of organizational research climates in three research intensive, doctoral granting universities. *J Empir Res Hum Res Ethics, 9*(5), 72-88. doi: 10.1177/1556264614552798