Universiteiten
van Nederland

Casus Wetenschappelijke Integriteit

2024

Datamanipulatie proefschrift– ongegrond

Universiteit Leiden

Onderwerp

Klagers hebben de klacht gericht tegen een voormalig medewerker. Klagers hebben getracht een bepaald onderzoek te reproduceren, maar kwamen daarbij tot afwijkende resultaten dan Beklaagde heeft gerapporteerd. Ook wijzen klagers op duplicaties in een excel-bestand, terwijl dat niet voor de hand ligt.

Samenvatting advies CWI d.d. :

Het verloop van de procedure

Complainants submitted a notice of complaint regarding a suspected violation of academic integrity by letter of [date].

A statement of defence was filed by [Accused] on [date].

On Monday [date] the Committee held a hearing with both parties present. [Accused] was present online due to his stay abroad. Complainants both attended the hearing in Leiden in person.

Klacht

Complainants state, in short, that [Accused] was awarded his PhD on [date], for his thesis titled [Title]. Their main concern lies with Chapter [number] of the thesis, where the development of a method to [description] is outlined. Complainants explained that the manuscript based on this chapter was under revision at [Journal], but they withdrew it when it was found that some of [Accused]'s data could not be reproduced. The data described in this manuscript have been collected using three model systems: [Subject 1], [Subject 2] and [Subject 3].

According to Complainants, the first problems arose when the new PhD student on this project was unable to reproduce crucial aspect of the data in the [subject 1]. Different batches were used for the replication of the experiment. Additionally, the original samples from these cells that were discovered in the freezer and had been collected by [Accused] were used. The results could not be reproduced. Complainants refer to Figure 1 in their notice of complaint.
Complainants state that analysis of raw data from a MSc student who had been supervised by [Accused] also did not show the effect as it had been presented by [Accused]. The [subject 1] could not be reproduced by another research group at the [Institute] either.

Complainants claim that for the [Subject 2] experiments, different dynamics and magnitude of the [description] were found. They refer to Figure 2. After repeating the experiment, the results that [Accused] had produced were generally reproduced, apart from a large peak in [description] concentration, which was an important result in the manuscript. This is set out in Figure 3 of the notice.

Vereniging Universiteiten van Nederland

Lange Houtstraat 2
Postbus 13739
2501 ES Den Haag

Tel +31 70 302 14 00
E-mail post@unl.nl
Web universiteitenvannederland.nl

KvK 40480226
IBAN NL61 INGB 0001 5964 15
BTW NL007088784B01

Complainants bring forward that also the samples from the [Subject 3] experiment were reanalyzed. The same samples previously analyzed by [Accused] were used for that. The [description] analysis for [description] samples from [description] were done again. In all four cases, the data for the [description] groups were reproduced, but the data for the [description] were found to be significantly higher than originally found by [Accused]. This is shown in Figure 3 of the notice.

The raw data from this experiment that [Accused] supplied upon request raised further concern according to Complainants. First, he provided an excel file that showed an [description], even though the experiment was done [description]. Second, when asked for the complete data set, a new Excel folder was provided several days later (Appendix 2 of the notice), in which entire blocks of data seemed to be duplicated.

Finally, Complainants state that an [description] was run, and similarly to the data from the [description], the results of the [description] were reproduced, but the [description] reported by [Accused] (Figure 3).

Verweer
[Accused] substantiated that there could be several reasons why the data are not reproducible, and has referred to literature.

He indicated, among other things, that some samples may have become contaminated as students conducted some experiments. [Accused] also pointed out that some samples had been [description] for an extended period of time, possibly leading to different results. Also, some samples may have been contaminated due to the storage time or the fact that students conducted some of the experiments.

He further remarks that the discrepancies noted in the replication of the [subject 1] data could be attributed to the number of passages of the [subject 1] used and the usage of [description].
Also collaboration and division of labor in data generation and different batches of [description] can lead to different results.

[Accused] also pointed out that some samples had been frozen for an extended period of time, possibly leading to different results.
In the opinion of [Accused] the new data brought forward by Complainants illustrates conflicting results and a lack of accuracy.

[Accused] points out the results were reviewed and discussed every week and systemically over [number] years with [Complainant]. Moreover, assays were repeated and reported by different students.

Overwegingen
The Committee considers that the core of the complaint is that the data used by [Accused] in his thesis cannot be reproduced. Complainants therefore suspect that the data were fabricated or manipulated, or that the data were created in a sloppy way. Several standards of the Code of Conduct apply to this complaint, in particular those mentioned in Chapter 3.3, *Conduct*.

If it turns out that the complainants' suspicions are well-founded, several standards will have been violated. The Committee's investigation focuses first on the factual merit of the alleged circumstances of the complaint. In this sense, the standards set out in the Code of Conduct were not taken as the starting point of the investigation.

In addition, the Committee took into account in its assessment data management, which is named in Chapter 4.4 of the Code of Conduct. The annex contains the standards of the mentioned chapters.

The Committee further considers as follows.

The Committee thanks [Accused], [Complainants] for an open discussion regarding certain results performed and analyzed by [Accused] (and his master students) that could not be reproduced in the laboratories of [Complainants], and an excel folder of qPCR data of [Accused] in which certain data results were duplicated.

The experiments that were discussed concern a thesis chapter entitled “[title]”. This was also submitted to Nature Communications and upon receival of reviewer comments and initiation of follow-up research by PhD students in the group of [Complainant] and [Complainant], certain experiments thereof that were previously conducted by [Accused] (and his master students) could not be repeated. In addition, a request by [Complainant] to [Accused] for primary data of [description] results showed duplication errors in the excel report. The Nature Communications paper was redrawn during the review process.

The experiments center around 4 Figures.

Figure 1.
This concerns an experiment with [Subject 1] in which the response to [description] was measured. Whereas original results by [Accused] show a pronounced/significant [description], this was not observed by PhD students and a post-doc in [Complainant]’s lab upon data reanalysis of [Accused]’s samples and analysis of new samples (with [Subject 1] used by [Accused] or with other [Subject 1] batches from other lab sources). As detailed below, it remains difficult for the Committee to explain these differential findings by [Accused] and by PhD students/post-doc in the [Complainant] lab.

The [description] data included in the manuscript are an integration of three independent experiments. Examination of the individual experiments performed by [Accused] (and master students) shows little variation, and hence the statistical significance. There are only minor differences in [description] measurements; therefore large variation of control samples does not explain the [description]. [Complainant] found data entries in the lab-journal of a master student of [Accused] which revealed an absence of [description] (in line with [Complainant]’s repeat experiments). It is unclear to the Committee why these data were excluded from the analyses reported in the manuscript. It is not clear whether [Accused] was aware of these results or whether they were excluded because this experiment was not correctly performed?
With the information presented and discussed, the Committee cannot exclude the possibility that variation between [Accused]’s results and those from [Complainant]’s lab is caused by technical issues or differences in the experimental set up. [Accused] has cultured the (presumed) [Subject 1] for a long time, which may have changed the behavior of cells. [description] may have deteriorated upon storage, and thereby affected the reanalysis of the same [description] samples. The [subject 1] used by [Accused] and [Complainant]/[Complainant] were not authenticated to prove that the cell lines used were indeed [Subject 1], and importantly to rule out cross contamination with other cell lines. Different cells or contamination with other cells may affect the outcome of experiments. Moreover, experiments are sensitive to [description] conditions (prior handling before [description], particular serum batch and cell confluency).

Figure 2.
This figure concerns a kinetic expression analysis of [description] in whole [Subject 2] after [description]. The data from [Accused] and data reanalysis of new samples show similar kinetic trends, but the results after 5 hours differ. Both sets of data are the result of one experiment in which many (nearly hundred) [Subject 2] were combined and in which [description]. These results may well fall within the experimental error of these sets of animal experiment; with only one biologically independent experiment performed by

[Accused] and [Complainant]'s lab, we do not know the experimental variation within and between different investigators. Different batches of [Subject 2], exact age of [Subject 2], and the method of wounding may have varied and affected the experimental outcome.

Figure 3.
This Figure relates to [description] analysis of [Subject 2] samples after [description]. Whereas the results from [Accused] show a strong [description] upon wounding (with is [description] treatment), this is not observed by reanalysis of new samples. The experiment by [Accused] and by [Complainant]'s lab (two independent experiments) were performed only once by [Accused] and [Complainant]'s lab each. Experimental variation within different independent experiments (reproducibility) of the data is therefore not clear. The [description] can be challenging and the analysis was performed externally in the [name] lab (Leiden University). There is a notable difference in [description]in the results by [Accused] and by [Complainant]'s lab for which the cause is unclear. With the information presented and discussed, the Committee cannot exclude the possibility that differential results of [Accused] and [Complainant]'s lab are caused by technical issues and experimental set up.

Figure 4.
This experiment relates to [description] (and from [description] measurements of [description] [Subject 3] plasma samples). Whereas the [description]could be repeated, the results with [description]could not be repeated by [name] upon rerunning the same samples. The Committee cannot explain why the different [description] were obtained (as part of the results, i.e. [description] effect was repeated, and therefore showed [description] samples).

Excel folder [description]data
This refers to an excel folder of [description] data of [Accused] in which certain data were duplicated. This data folder was sent to [Complainant] by [Accused], then living abroad, who instructed his [description] to find it on his computer. [Accused] did not have remote access to his computer. [Accused] stated that he had not checked the file that was sent to [Complainant] by his [description] and notes that [Complainant] had access to the original excel data on a USB stick that he left upon his departure from [Complainant]'s lab. It remains unclear to the Committee how the duplication errors occurred and whether this was done intentionally by [Accused].

Conclusie
The results from [Accused] vary considerable from reanalysis of samples of [Accused] and independent [description] repeat experiments performed by PhD students/post-doc from [Complainant]'s lab. [Accused] has been instrumental in making (primary) data available and storing original [description] samples in such a way that they could be found by [Complainant] and used for reanalysis. However, why certain [description] results could not be repeated, why data from a master student were not included in the report, and how erroneous duplications in excel folder occurred, is not clear to the Committee. Better central data management of experimental results of [Accused] and his master students could have prevented some of these problems. It would be beneficial to implement a centralized, safeguarded data management system for original raw data.

The Committee cannot exclude the possibility that the irreproducibility is due (at least in part) to technical issues or variation in experimental setup. In addition, it cannot be proven that [Accused] deliberately falsified or fabricated data.

Advies van de Commissie
In view of the above, the Committee advises the Executive Board to declare the complaint unfounded.

Universiteiten
van Nederland

Gelet op het vorenstaande adviseert de Commissie de klachten ongegrond te verklaren.

Aanvankelijk oordeel College van bestuur d.d. 29 augustus 2024:
Advies van de CWI wordt overgenomen.

LOWI:
De kwestie is niet voorgelegd aan het LOWI.

Definitief oordeel van het College van Bestuur d.d. 10 oktober 2024:
Het College stelt als definitief oordeel vast dat de klacht inzake wetenschappelijke integriteit ongegrond is.