

IRB #: HRPP-2023-239
Title: Where postdoctoral journeys lead
Creation Date: 12-22-2023
End Date:
Status: Approved
Principal Investigator: Bedoor AlShebli
Review Board: NYUAD Institutional Review Board
Sponsor:

Study History

Submission Type	Initial	Review Type	Exempt	Decision	No Human Subjects Research
-----------------	---------	-------------	--------	----------	----------------------------

Key Study Contacts

Member	Talal Rahwan	Role	Co-Principal Investigator	Contact	tr72@nyu.edu
Member	Bedoor AlShebli	Role	Principal Investigator	Contact	bka3@nyu.edu
Member	Bedoor AlShebli	Role	Primary Contact	Contact	bka3@nyu.edu

Initial Submission

Screening Questions

*required

Does the proposed activity involve RESEARCH?

Research means a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.

☒ Yes

☐ No

*required

Does proposed activity involve HUMAN SUBJECTS?

Human subject means a living individual about whom an investigator (whether professional or student) conducting research:

- Obtains information or bio-specimens through intervention or interaction with the individual and uses, studies, or analyzes the information or biospecimens; or
- Obtains , uses, studies, analyzes or generates identifiable private information or identifiable biospecimens.

☒ Yes

☐ No

*required

Are you engaging in the following activities WITHIN THE EMIRATE OF ABU DHABI?

Please select all that apply

Biomedical research (e.g. involving biospecimen collection and/or storage, genetic analysis, etc.)

Health related research (e.g. Public health, mental health related)

Any research involving EEG, MEG and/or MRI scanning

Social behavioral research (i.e. not biomedical) that involves collection and/or storage of human tissue samples (e.g. hair, saliva, blood, etc)

☒ None of the above

The '**Principal Investigator**' (PI), '**Co-Principal Investigator**' (Co-PI) and '**Primary Contact**' (PC) are special roles with certain certification abilities throughout the protocol routing process.

PI: This person has to be NYUAD standing faculty. For student- or postdoc-led projects, the faculty adviser can be listed as PI on the protocol. Only the PI has the ability to initially certify a new submission. Once this is done, the IRB can begin review.

Co-PI: Next leading research personnel after the PI.

Primary Contact: The system automatically selects the individual completing this form. The PC can be the PI or Co-PI, or another member of the research team, e.g. a student.

*****Can't find a member of the research team under FIND PEOPLE? Please contact the IRB Office at irbnyuad@nyu.edu.*****

*required

Primary Contact

If you are responsible for the completion of this form, then you will automatically be selected as 'Primary Contact' here to avoid being locked out of the application.

Name: Bedoor AlShebli

Organization: Division of Social Science

Address: PO Box 129188 , Abu Dhabi,

Phone:

Email: bka3@nyu.edu

*required

Is this a student-led project?

Yes

☒ No

*required

NYUAD Principal Investigator/Faculty Adviser

Name: Bedoor AlShebli

Organization: Division of Social Science

Address: PO Box 129188 , Abu Dhabi,

Phone:

Email: bka3@nyu.edu

(The following three questions are specific to NYUAD personnel, there will be an opportunity to input information on external research personnel further down)

*required

Is there a NYUAD 'Co-Principal Investigator' on this protocol?

☒ Yes

No

*required

NYUAD Co-PI:

Name: Talal Rahwan

Organization: Division of Science

Address: PO Box 129188 , Abu Dhabi,

Phone:

Email: tr72@nyu.edu

*required

Are there 'NYUAD Student Investigators' on this protocol?

Yes

✓ No

*required

Are there 'NYUAD Additional Investigators' on this protocol?

Yes

✓ No

*required

Please describe the research activities your NYUAD Co-PI, Student Investigators and/or Additional Investigators will be involved in on this project.

In this project, we investigate the career trajectories of postdoctoral researchers. Postdoctoral training is often characterized as a challenging and stressful period during which many promising PhDs see their academic aspirations thwarted by factors beyond their control. Using a unique dataset of academic publications and career information, we analyze the varying levels of success among postdoctoral researchers. Our measure of academic success is based on citation patterns observed two to five years into a faculty career. We examine how factors such as relocation, changes in research topics, and early well-cited papers during postdoctoral positions influence early-career success.

To achieve our objectives, we utilize two datasets:

1. **Microsoft Academic Graph:** A publicly accessible repository of over 250 million research publications, available for free download online. It includes details such as author names, paper titles, author affiliations, publication years, and more. Additionally, it provides a citation network to track citation activity over time.
2. **Dataset of 300 Million CVs:** Collected and parsed from an online professional network and purchased from DataHut, this dataset includes publicly available CVs with links to their online profiles.

All the data we use is publicly available. We leverage these datasets to identify academics who have undertaken postdoctoral positions at some point in their careers and compare them to those who have not.

*required

Are any external (non-NYUAD) investigators or personnel engaged in this research?

Generally, "engaged" individuals are those who intervene or interact with participants in the context of the research or who will obtain individually identifiable private information for research purposes. See [OHRP Engagement Guidance](#) for information.

External (non-NYUAD) personnel are subject to their own institutional review and/or local oversight requirements. Investigators are responsible for determining if other requirements apply and are encouraged to maintain documentation of any additional approvals/determinations for this study.

☒ Yes

☐ No

Please list all Non-NYUAD investigators and provide the following information about them:

- *First and Last Name*
- *Institutional Affiliation Name (Country, where relevant)*
- *Project responsibilities (e.g., research design, recruitment, consent, data collection, analysis of identifiable or de-identified data)*

1- Petter Holme, Department of Computer Science, Aalto University, Espoo, Finland

Petter is responsible for research design, analyzing the data, and writing the paper. Petter does not have access to the original datasets.

2- Yueran Duan, School of Economics and Management, China University of Geosciences, Beijing, China

Yueran is responsible for analyzing the data. Yueran does not have access to the original datasets.

3- Shahan Ali Memon, Information School, University of Washington, Seattle, USA

Shahan is responsible for data parsing and filtering the data. He merged the two datasets and identified our sample of postdocs and other academics.

4- Qing Guan, School of Economics and Management, China University of Geosciences, Beijing, China

Qing Guan is responsible for writing the paper, does not handle the data.

Training in human subjects research ethics is required for all Investigators. If you've completed CITI training through an NYU or NYUAD subscription then your training status should be linked through Cayuse.

However if you've completed CITI training within the last 3 years **under another institution's subscription**, please attach a PDF of the completion certificate/report here.

Depending on the nature of your research, the acceptable training options are:

1. *CITI Social & Behavioral Basic/Refresher Course;*
2. *CITI Biomedical Basic/Refresher course;*
3. *CITI Course for Investigators Conducting Exempt Research.*

Make sure to affiliate with 'New York University Abu Dhabi' in the [CITI Program](#).

[citiCompletionReport.pdf](#)

[citiCompletionCertificate.pdf](#)

[Talat - CITI training certificate \(Conflict of Interest\) Expires 15-Apr-2026.pdf](#)

[Talal - CITI training report \(Conflict of Interest\) Expires 15-Apr-2026.pdf](#)

[Talal - Social-Behavioral-Educational Research - Certificate \[Expires 28-Feb-2027\].pdf](#)

[Talal - Investigators conducting Exempt research - Certificate \[Expires 28-Feb-2027\].pdf](#)

[Shahan - citiCompletionReport.pdf](#)

*required

Purpose of Study

Please describe the purpose of your proposed research, making clear the question the research is attempting to address. Avoid the use of technical terms or discipline specific language. Your explanation must be clear to those unfamiliar with your field. References are not necessary.

In this project, we investigate the career trajectories of postdoctoral researchers. Postdoctoral training is often characterized as a challenging and stressful period during which many promising PhDs see their academic aspirations thwarted by factors beyond their control. Using a unique dataset of academic publications and career information, we analyze the varying levels of success among postdoctoral researchers. Our measure of academic success is based on citation patterns observed two to five years into a faculty career. We examine how factors such as relocation, changes in research topics, and early well-cited papers during postdoctoral positions influence early-career success.

To achieve our objectives, we utilize two datasets:

1. **Microsoft Academic Graph:** A publicly accessible repository of over 250 million research publications, available for free download online. It includes details such as author names, paper titles, author affiliations, publication years, and more. Additionally, it provides a citation network to track citation activity over time.
2. **Dataset of 300 Million CVs:** Collected and parsed from an online professional network and purchased from DataHut, this dataset includes publicly available CVs with links to their online profiles.

*required

Proposed Project **Start Date**

01/01/2022

*required

Proposed Project **End Date**

31/12/2025

*required

Has this study been previously approved by any IRB or Research Ethics Committee?

Yes

☒ No

*required

Research Funding

Indicate all possible sources of funding (submitted, pending and/or awarded) that may/will support the proposed project:

☒ Investigators' internal research funds

NYUAD competitive or other research funding program (e.g. REF, etc.)

U.S. Federal Funding (NIH, NSF, etc.)

Other External (non-NYUAD) Funding

No funding associated with this project

*required

Cooperating Institutions

Will any part of the research be conducted with the **cooperation of another institution or involve a collaboration?**

Select 'Yes' if you will be coordinating with another organization/institution on any research activities. Survey firms DO NOT apply here.

Example institutions include:

- *Hospitals or health care providers*
- *Educational Institutions such as universities and schools (Including NYU WSQ and Shanghai)*
- *Governmental Organizations*
- *Non-Governmental Organizations, Agencies, Associations*

☒ Yes

☐ No

☐ Unsure, to be determined

*required

International Research

Will any research be conducted **on the ground** in an **international setting outside the U.A.E or the U.S.?**

Online research and interviews conducted remotely DO NOT apply here.

☐ Yes

☒ No

*required

If you are collaborating with an Institution with an **FWA**, you may request an **IRB reliance agreement**

A Federal Wide Assurance (FWA) is a registration with the U.S. Department of Health and Human Services (HHS) that an institution must have in order to receive HHS support for research involving human subjects. NYU and NYUAD both have FWAs.

Regardless of funding source for your research, in order to enter into a reliance agreement with another IRB in which we accept their determination for your role in the research, **the collaborating institution must also have an FWA.**

Would you like to request an **IRB reliance agreement** for your collaboration with any institution?

(Please contact the IRB office at irbnyuad@nyu.edu if you'd like to discuss this option)

Yes

☒ No

*required

Have you determined exactly which institutions/organizations you will be collaborating or cooperating with?

☒ Yes, I know which institutions I will be working with.

No, this is still being determined or I know some but not all.

*required

How many institutions are you collaborating with?

1

2

✓ 3

4

5

6+

*required

Provide the following information for **collaborating institution 1**.

Institution name (do not abbreviate):

Department of Computer Science, Aalto University, Espoo, Finland

*required

Name of contact person:

Petter Holme

*required

Role of the institution in the research:

Petter is responsible for research design, analyzing the data, and writing the paper

*required

Provide the following information for **collaborating institution 2**.

Institution name (do not abbreviate):

School of Economics and Management, China University of Geosciences, Beijing, China

*required

Name of contact person:

Yueran Duan and Qing Guan

*required

Role of the institution in the research:

Yueran is responsible for analyzing the data

Qing Guan is responsible for writing the paper

*required

Provide the following information for **collaborating institution 3**.

Information School, University of Washington, Seattle, USA

*required

Name of contact person:

Shahan Ali Memon

*required

Role of the institution in the research:

Shahan is responsible for data parsing and filtering. He merged the two datasets and identified our sample of postdocs and other academics.

Upload any **letters of cooperation** or **IRB/REC approvals or equivalent** from the institutions listed above.

*required

Will you be using any Human Subjects data/information/biospecimens for this research that were or will be collected for another purpose?

This includes the use of human subjects information that is originally created (or will be created) for purposes other than the proposed research, some examples include:

- *educational records/grades,*
- *medical records,*
- *census data,*
- *posts on Internet message boards or social media,*
- *existing individually identifiable research data from another study,*
- *Data sets collected by governmental or non-governmental agencies that contain identifiable information*

DO NOT select 'Yes' if your research only draws on published literature, news articles, historical/contemporary records, etc. that do not include individually identifiable information about people.

☒ Yes

☐ No

*required

In addition to this secondary use of identifiable data/information/biospecimens:

Will you conduct any primary data collection (e.g., surveys, interviews, observations) or research activity (e.g., treatment or intervention)?

☐ Yes

☒ No

*required

Select the type(s) of secondary data you will be using/accessing:

Publicly available identifiable information

- ✓ *This means anyone can access this information or data online or from a particular source (i.e. libraries, government offices, etc.) without any special permissions or data use agreements.*

Private identifiable Information

This is information that an individual can reasonably expect will not be made public. Generally, data sets that require specific permission from the data owner or are restricted access are considered "private."

- *Examples: Medical records, data sets that require a data use agreement, educational or HR records, etc.*

*required

Will the data include information about residents of the EU/EEA?

✓ Yes

No

*required

Will the data be processed by an organization established in the EU/EEA?

Yes

✓ No

*required

For publicly available information, please describe:

1. The nature of the data, i.e., type of information.
 2. The source of the data and why you believe it is "publicly" available.
-

If possible, provide the URL for the data set

Microsoft Academic Graph (MAG) can be found here:

<https://www.microsoft.com/en-us/research/project/m...> For the Dataset purchased from DataHut, each CV has an link to an online user profile. These links are for publicly available profiles. I would like to note that ALL the analysis is done using MAG. DataHut's data is only used to identify if a scientist was a postdoc at some point in their career or not (based on their CV) and when that period was. No other data on the CV was used. All other data used in the analysis was extracted from MAG.

Also, I would like to note that the majority of the dataset is US-based. That data might include postdocs who were in European institutes at some point in their career, which is why I selected yes above. Furthermore, all of our analysis is in aggregate form and not information on individuals can be found anywhere in the analysis.

This study involves data covered by the European Union General Data Protection Regulation (EU GDPR). For more information about the regulation, see [EU GDPR](#), and [NYU EU GDPR](#).

*required

How will your data be stored?

✓ Internet-based storage

*required

Select all that apply:

NYU requires that all Internet-based data be stored on platforms run by a service provider/vendor with a formal agreement with NYU to ensure compliance with the EU GDPR.

NYU Qualtrics

NYU Box

✓ NYU Google Suite -- e.g., Gmail, Google Drive, Google Forms (*must use NYU access to Google Suite, not personal access*)

✓ Other

*required

Identify the name(s) and URL(s) for each platform and provide a justification for why they must be used:

Note that the IRB may need to obtain documentation from the University and/or platform(s) to ensure compliance with GDPR.

NYU's HPC cluster.

Upload documentation of the security of the other platform(s) and their compliance with GDPR.

Physical storage devices or locations, e.g., external hard drive, USB stick, local hard drive, cellphone, audio recording device, filing cabinets

Other Internet-based storage

*required

Will the data from EU/EEA participants or processed by an EU/EEA organization include any of the following "special categories" of data?

- Data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership
- Genetic or biometric data
- Data concerning health, sex life, or sexual orientation

Yes

☒ No

Please confirm that you will comply with NYU's policy regarding breach of GDPR data.

See policy under 'Appendix A: Breach of GDPR Data' [here](#). Note that NYU has 72 hours to report breaches to the regulatory authorities; thus, you must report breaches to the NYU Data Protection Officer IMMEDIATELY.

☒ Confirmed

Unconfirmed

Investigator CITI Training Certificate/Reports

[citiCompletionReport.pdf](#)

[citiCompletionCertificate.pdf](#)

[Talal - CITI training certificate \(Conflict of Interest\) Expires 15-Apr-2026.pdf](#)

[Talal - CITI training report \(Conflict of Interest\) Expires 15-Apr-2026.pdf](#)

[Talal - Social-Behavioral-Educational Research - Certificate \[Expires 28-Feb-2027\].pdf](#)

[Talal - Investigators conducting Exempt research - Certificate \[Expires 28-Feb-2027\].pdf](#)

[Shahan - citiCompletionReport.pdf](#)

Research Proposal

Collaborations: Reliance Agreement

Collaborations: Institutional Letters and Reviews

Collaborations: Institutional Letters and Reviews 2

Data Usage Agreement

[DataHub Dataset No Share Agreement.pdf](#)

Recruitment: Screening form

Recruitment: Physical posters, flyers, pamphlets, etc.

Recruitment: Emails to Participants

Recruitment: video/digital recordings

Recruitment: Letters to institutions

Recruitment: All other materials

Instrument: Interview questions and/or surveys

Instrument: PDF of online survey

Instrument: Experimenter instructions and tasks

In-person Research Exception Request (COVID-19)

International: Translator Statement of Confidentiality

International: Survey firm training materials

International: Local IRB/REC

Adult consent: consent forms

Adult consent: consent scripts/texts

Adult consent: altered consent forms, scripts, texts

Child Research: Verbal assent script(s)

Child Research: Signed assent form(s)

Child Research: Parental Permission

Child Research: Altered Parental Permission
