Gendered candidate turnover in 14 established parliamentary democracies. Biases in the secret garden of politics.

A Data Management Plan created using DMPonline.be

Creator: Gertjan Muyters

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Gertjan Muyters

Grant number / URL: 1228524N

ID: 206728

Start date: 01-10-2024

End date: 30-09-2027

Project abstract:

Unequal candidate turnover might be one of the reasons why full equality in gender representation is often not accomplished, even after the introduction of gender quotas. This research project builds further on my doctoral research on candidate turnover. The key theoretical frame is the demand and supply model of political recruitment. First, this project digs deeper into the questions: Whether candidate turnover is higher for women than for men? How quota impact this renewal rate across gender? If this gendered turnover is contingent on established turnover drivers at the systemic or party level? Hypotheses will be tested by means of a RIGLS multilevel fractional logit model and a MCMC augmented beta regression on a general and a weighted turnover measure. We will make use of comparative data, containing candidates of represented political parties within 14 established parliamentary democracies, both with list-PR and majority electoral systems. Second, to better understand selectorate preferences, we will complement the statistical analysis on electoral data with a discrete-choice-experimental design to investigate so-called stated preferences of national and regional party leaders. Third, this project investigates the supply of candidates by combining the PPDB with the ESS data to investigate the link between gendered candidate selection and citizens' political engagement.

Last modified: 24-04-2024

Gendered candidate turnover in 14 established parliamentary democracies. Biases in the secret garden of politics.

DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Gendered candidate turnover in 14 established parliamentary democracies. Biases in the secret garden of politics.

GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Gendered candidate turnover in 14 established parliamentary democracies. Biases in the secret garden of politics.

Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

For the first research line, we expand the existing self-created database on candidate turnover in established democracies. Data sources: government websites, respective ministries of interior. For the covariates we turn to online databases (IDEA, ParlGov, PPDB). Candidate gender is gathered from electoral bulletins; in second instance we turn to name analysis (with AI-'Gender Guesser' classification algorithm).

For the second research line (discrete choice experiment), we use Qualtrics to create an online questionnaire that is send via e-mail, together with research information and consent, to (regional) party leaders of represented political parties. Sampling method: snowball sampling. Party selectors are presented with about 15 choices, each time between three sets of candidate attributes (with different levels). Moreover, party selectors are asked to give some personal information about themselves: gender, age, educational attainment and previous actions within the party selection process regarding gender representation.

For the third research line, we make use of the European Social Survey data on campaign engagement. More specifically these three items will be used: 'wrkprty', 'badge', and 'pstplonl'.

These three research lines result in three data tables each consisting of numerical data. They will be stored in .xls format (which can be easily imported in PowerBI and/or R for ETL and analysis).

Total storage space needed will be less than 1 TB (around 100 GB).

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Designation of responsible person
2. Storage capacity/repository
    - during the research
    - after the research

During the research, the data will be stored on the KU Leuven PGI data storage drive. The shared network drive is based on the standard file storage solution of ICTS KU Leuven. This storage is accessible outside the KU Leuven network via VPN and there is self-service version control. This environment offers a secure option to safe the two final data matrices, including personal information provided by party selectors. In addition, this secure environment will be used to store the auxiliary data table containing invidvidual electoral candidate renewal and gender statistics, i.e. before aggregation (and in casu anonimisation) by calculating the (weighted) turnover at the electoral list level. This auxiliary data table is pseudonymised by means of a candidate-specific ID.

Moreover, bearing in mind flexiblity concerning data accessibility and data management, data will be saved as well on Teams-site, during the research. Taking into account additional precautionary measures when placing data on non-European cloud solutions, this option will only be used for the anonimesed aggregated candidate renewal data table from the frist research branch and the publicly available and pseudonymised ESS data (used in the third research branch of this project).

Taking into account the relatively small data volume (< 1TB) of our numerical data tables, it will be convenient to ensure data preservation for at least 5 years after the end of the research using the shared network drive (for which Inge Vermeulen is the contact person). Data in the teams environment will not be stored after the end of the research.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

Within the second research branch, we collect sensitive personal information (gender and political views) of party selectors. This is crucial information for the research goal in the second branch of this project: to search for and, if found, understand possible biased patterns in the demand-side of the candidate selection process. Therefore, we will not only ask for informed consent on the use of this data within the scope of this research project, but also to preserve this data in the secured KUL environment for 5 years after the end of the research (with a view to possible replication studies). Without this second informed consent, these sensitive personal data will not be stored after the end of the research.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Taking into account additional precautionary measures when placing data on non-European cloud solutions, the Teams-site will only be used for the anonimesed aggregated candidate renewal data table from the frist research branch and the publicly available and pseudonymised ESS data (used in the third research branch of this project).

Within the second research branch, we collect sensitive personal information (gender and political views) of party selectors. This is crucial information for the research goal in the second branch of this project: to search for and, if found, understand possible biased patterns in the demand-side of the candidate selection process. Therefore, we will not only ask for informed consent on the use of this data within the scope of this research project, but also to preserve this data in the secured KUL environment for 5 years after the end of the research (with a view to possible replication studies). Without this second informed consent, these sensitive personal data will not be stored after the end of the research.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Concerning data on electoral candidates, we gathered info via electoral bulletins on which the candidates agreed beforehand to run for public office. Yet, unavailable gender data will be generated using the AI-'Gender Guesser' classification algorithm. Data will only be shared, upon request, after aggregating renewal rates and gender shares at the list level. Possible questions regarding the AI algorithm can be addressed to the postdoctoral researcher (Gertjan Muyters).

Data sanity has been/will be assured in the following ways: official government data are collected; complex procedural tasks were performed both using the MS Access and R environments; a battery of sanity checks were/will be performed by re-arranging and aggregating the data – in addition to random manual checks.

Gendered candidate turnover in 14 established parliamentary democracies. Biases in the secret garden of politics.

FWO DMP (Flemish Standard DMP)

---

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Individual candidate renewal data table (auxiliary) - branch 1	Data table at the indidividual candidate level (hierarchical structure: Election year > Country > Political party > Electoral district > Candidate). This table contains candidate gender, a dummy indicating if generated with gender guesser algorithm or provided in electoral bulletins, candidate age and occupation (if provided in the electoral bulletins) and an indicator variable for renewal status.	Reuse existing data and add new data (additional country samples + gender data)	Digital	Compiled/aggregated data	.txt, .R, .pbix	< 100 GB	NA

aggregated candidate renewal data table - branch 1	Data table at the electoral list (i.e. party-in-district) level. Two aggregated dependent variables: unweighted and weighted candidate renewal. Independent variables at party level and election level. Data sources: government websites, respective ministries of interior + online databases (IDEA, ParlGov, PPDB).	Reuse existing data and add new data (additional country samples + gender data & party leader gender data)	Digital	Compiled/aggregated data	.txt, .R, .pbix	< 1 GB	NA
DCE - branch 2	Choice sets (varying levels of attributes) of party selectors + personal info on party selectors	Generate new data	Digital	Experimental	.txt, .R	< 1 GB	NA
ESS - branch 3	3 ESS items on campain engagement combined with party level and election level data from second date table	Reuse existing data	Digital	Observational	.csv, .R	< 1 GB	NA

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

I will expend my own dataset on candidate renewal (project number: 1228524N).
Moreover I will make use of the ESS: https://www.europeansocialsurvey.org/.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data

We perform a DCE experiment on party selectors to study gendered candidate demand (i.e. DCE dataset). Therefore we will first inform possbile resondents on the goals and methods of this research, the necessity of their responses for answering the research questions at hand, and the way we will process and store the (personal) data. This will be done in short in the corresponding e-mail and at the start of informed consent form.
Next we will ask for an informed consent, using a form based on the KUL template. This will presented to the respondents before they answer the questionnaire (a box for informed consent will need to be highlighted before the respondents can proceed to the actual questionnaire).
In addition we will explicitly ask for the respondent's permission for the use and storage during the research + the possibility for the data storage for 5 years after the end of the research. These will be done at the end of the Qualtrics questionnaire.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

We take into account additional data security measures to store and handle sensitive personal data (gender & political views of party selectors), more specifically in the second research branch of this project (i.e. DCE dataset).

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

We create a codebook, listing all variable names together with their short discription and their operationalisation. Moreover, we will provide the relevent code and/or data management procedure upon request (in line with publication policy of journals where our work is published).

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

We will use the github metadata format.

3. Data storage & back-up during the research project

Where will the data be stored?

During the research, the data will be stored on the KU Leuven PGI data storage drive. The shared network drive is based on the standard file storage solution of ICTS KU Leuven. This storage is accessible outside the KU Leuven network via VPN and there is self-service version control. This environment offers a secure option to safe the final data tables, including personal information provided by party selectors. In addition, this secure environment will be used to store the auxiliary data table containing invidvidual electoral candidate renewal and gender statistics, i.e. before aggregation (and in casu anonimisation) by calculating the (weighted) turnover at the electoral list level. This auxiliary data table is pseudonymised by means of a candidate-specific ID.

How will the data be backed up?

Moreover, bearing in mind flexiblity concerning data accessibility and data management, data will be saved as well on Teams-site, during the research. Taking into account additional precautionary measures when placing data on non-European cloud solutions, this option will only be used for the anonimesed aggregated candidate renewal data table from the frist research branch and the publicly available and pseudonymised ESS data (used in the third research branch of this project).

Given the relatively small data volume of the numeric data table, the DCE data table will be backed up on an external hard drive, only used for this purpose.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

The options explained above (shared drive + Teams-site) are guarenteed up to 5 TB, while our data volume will remain significantly under 1 TB.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The shared network drive is based on the standard file storage solution of ICTS KU Leuven. This environment offers a secure option to safe the final data tables.

Moreover, bearing in mind flexiblity concerning data accessibility and data management, data will be saved as well on Teams-site, during the research. Taking into account additional precautionary measures when placing data on non-European cloud solutions, this option will only be used for the anonimesed aggregated candidate renewal data table from the frist research branch and the publicly available and pseudonymised ESS data (used in the third research branch of this project).

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

No costs for data storage and backup are expected for the choosen options.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

Within the second research branch, we collect sensitive personal information (gender and political views) of party selectors. This is crucial information for the research goal in the second branch of this project: to search for and, if found, understand possible biased patterns in the demand-side of the candidate selection process. Therefore, we will not only ask for informed consent on the use of this data within the scope of this research project, but also to preserve this data in the secured KUL environment for 5 years after the end of the research (with a view to possible replication studies). Without this second consent, these sensitive personal data will not be stored after the end of the research. Otherwise they will be stort for 5 years.

Where will these data be archived (stored and curated for the long-term)?

Data will remain on the PGI shared drive for 5 years. In addition non-sensitive personal data (thus data from branch 1 and branch 3) will be placed in a GitHub secure repository that permits to control the access users have to assets, asset properties, and asset types.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

No costs for data preservation are expected for the choosen options.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in a restricted access repository (after approval, institutional access only, …)

Sensitive personal information (branch 2) will not be made available.
Aggregated data from branch 1 and branch 3 wlll be made available, in line with the publication policy of the journals where we publish the results based on the underlying data.

If access is restricted, please specify who will be able to access the data and under what conditions.

Sensitive personal information (branch 2) will not be made available. If given informed consent by the respondents, these data will be stored for 5 year after the end of the research to enable possible replication studies conducted by our research group (PGI). Without this consent, the data will not be stored afther the end of the research project.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Privacy aspects

Choice set information (i.e. choosen levels of attributes by party selectors in branch 2) and Sensitive personal information (branch 2: gender & political views) will not be made available. They will only be shared to reviewers in the publication process if necessary and if it can be guaranteed that data remain intern.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Data can be shared upon request via shared drive. Moreover, the unrestricted data will be placed on a repository (candidate repository: github or ManGo).

When will the data be made available?

The unrestricted data will be made available upon publication of research results or upon request (depending on the publication policy of the journal where results based on the underlying data are published).

Which data usage licenses are you going to provide? If none, please explain why.

In case no journal licence would apply, we will grant a licence based on the conditions specified in the informed consent form that respondents fill in.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

Not specific to the datasets, but to the journal articles that are based on the data.

What are the expected costs for data sharing? How will these costs be covered?

There are no expected costs related to these data sharing plans (other than the time invested by the postdoctoral researcher).

6. Responsibilities

Who will manage data documentation and metadata during the research project?

postdoctoral researcher

Who will manage data storage and backup during the research project?

postdoctoral researcher

Who will manage data preservation and sharing?

PGI data contact in cooperation with the postdoctoral researcher (as corresponding author of publications based on these data)

Who will update and implement this DMP?

postdoctoral researcher