**Analysis plan**

**Question 1: Is there an association between a person's protected characteristics and how many first/other author papers they have on submission of PhD?**

*Response:* No. of first/other author papers at PhD submission (separate models)

*Fixed explanatory variables:* gender, sexual orientation, ethnic background, socioeconomic status, disability status, age at submission

*Random explanatory variables:* year thesis submitted, country thesis submitted

*Interactions:* gender x [all other protected groups]

*Model:* Poisson GLMM

**Question 2: Is there an association between a person’s protected characteristics and how many postdoc applications they make following their PhD (proxy of time spent/success rate in getting a postdoc?)**

*Response*: No. of postdoc applications made

*Fixed*: gender, age, [all other protected groups], total no. papers

*Random*: year submitted thesis? (proxy for time)

*Interactions*: gender x age, gender x [all other protected groups], total no. papers x gender, total no. papers x [all other protected groups].

*Model:* Poisson GLMM

**Question 3:  Is there an association between a person’s protected characteristics and how many permanent position applications they make (proxy of time spent/success in getting a permanent position?)**

*Response:* No. of permanent position applications made

*Fixed:* gender, age, [all other protected groups], total no. papers, no. postdocs, no. of applications made for postdocs, years post-PhD

*Random:* year secured permeant contract/year submitted PhD (proxy for time), would also be good to have a random term for geographical location to account for any spatial autocorrelation.

*Interactions:* gender x age, gender x [all other protected groups], total no. papers x gender, total no. papers x [all other protected groups], no. postdocs x gender, no. postdocs x [all other protected groups].

*Model:* Poisson GLMM

**Question 4:  Is there an association between a person’s protected characteristics and whether or not they are successful in securing a permanent contract?**

*Response:* Whether not have a permanent contract

*Fixed:* gender, current age, [all other protected groups], no. postdocs, no. of first/other author papers, years post-PhD, grouped barrier (derived variable from qualitative data)

*Random:* year secured permanent contract/year submitted PhD (proxy for time)?

*Interactions:* gender x age, gender x [all other protected groups], no. postdocs x gender, no. postdocs x [all other protected groups], years post-PhD x gender, years post-PhD x [all other protected groups], no. of first/other author papers x gender, no. of first/other author papers x [all other protected groups]

*Model:* Binomial GLMM

**Question 5: Is there an association between a person’s protected characteristics and how many grants they have applied for?**

*Response:* How many grant applications made (will be zero-inflated)

*Fixed:* gender, current age, [all other protected groups], no. first/other author papers, no. postdocs, whether or not permanent, type of contract

*Random:* year submitted thesis (proxy for time)?

Interaction: gender x age, gender x [all other protected characteristics], no. postdocs x [gender and all other protected characteristics], whether or not permanent x [gender and all other protected characteristics], type of contract x [gender and all other protected characteristics], no. of first/other author papers x [gender and all other protected characteristics],

*Model:* Poisson or binomial GLMM

**Question 6: Is there an association between a person's protected characteristics and the type of contract they have?**

*Response:* Type of contract e.g. ‘Research' vs. ‘Teaching' or 'Teaching & Research' OR ‘Teaching & Research' & ‘Research' vs. ‘Teaching'

*Fixed:* gender, current age, [all other protected characteristics], whether or not permanent, no. postdocs, no. first/other author papers

*Random:* year handed in PhD (proxy for time again?)

Interaction: gender x age, gender x [all other protected characteristics], no. first/other author papers x [gender and all other protected characteristics]

*Model:* Binomial GLMM

**Question 7: Is there an association between a person's protected characteristics and whether or not they report having faced barriers to their identity?**

*Response:* Whether or not faced barriers

*Fixed:* gender, current age, [all other protected characteristics], whether or not permanent, no. postdocs, no. first/other author papers

*Random:* when handed in PhD (proxy of time)?

Interaction: gender x age, gender x [all other protected characteristics], whether or not permanent x [gender and all other protected characteristics], no. postdocs x  [gender and all other protected characteristics] , no. first/other author papers x  [gender and all other protected characteristics]

*Model:* Binomial GLMM

**Question 8: What are the most frequent words in peoples' responses to questions about the barriers they have faced?**

Using word cloud methods to generate frequencies for each word stem (excluding connecting words), then looking for strong associations between most frequent words

Consider a way of categorising by related keywords in order to give a very simplified summary?

**Question 9: What are the most frequent words in peoples' responses to questions about how they overcame these barriers?**

As above