Pre-Analysis Plan: Nonbinary Hiring Discrimination and the Politicization of Pronouns

Date: 4/27/2023
Investigators: Taryn Eadie (University of Toronto)

Background

A significant amount of research has evaluated labour market discrimination faced by women, racial minorities, and sexual minorities. However, limited research exists evaluating labour market discrimination faced by transgender people. This research is warranted: from the 2015 US Transgender Survey, 46% of respondents report being verbally harassed and 9% physically attacked in the last year for being transgender. Further, from the same survey 30% of respondents report being fired, denied a promotion, or otherwise mistreated in the work place in the last year (James et al., 2016). At the same time, the transgender population is growing significantly among younger generations. Data from the Pew Research Centre shows that while only 0.3% of Americans 50 or over identify as transgender, this percentage is 1.6% for those 30-49 and 5.1% for those 18-29. Further, under the transgender umbrella nonbinary people are the majority, making up around two thirds of the transgender population (Brown, 2022).

In recent years, use of the gender-neutral pronouns like “they/them” and the convention of asking for and declaring preferred pronouns (at the start of meetings, in email signatures, etc.) has become politicized. Whether to share pronouns is divisive in and of itself, with opinion split along political lines. A YouGov poll conducted in the United States shows that while 40% of Republicans think that “people should generally not say / display their pronouns unless asked,” that holds for only 10% for Democrats (Ballard, 2022). As a result, sharing any pronouns (whether gender neutral or binary “he/him,” “she/her”) may act as a (left-leaning) political signal, regardless of an individual’s apparent gender. As a result, when evaluating response to “they/them” pronoun disclosure it is important to parse out the additional political signals at play.

A resume audit study design will be leveraged to estimate hiring discrimination against nonbinary applicants (signaled on resumes via “they/them” pronouns listed below the name) and cisgender applicants who disclose pronouns (signaled on resumes via binary pronouns congruent with name-implied sex—for example, “she/her” for Emily and “he/him” for Jacob). Outcomes among these two treatment groups will be compared to a control applicant who does not disclose pronouns on their resume. Data collection will include resume characteristics, job posting text, employer information, and employer response to application (this will be done via phone and email monitoring). This will allow for statistical testing of differences in means across groups and estimation of how treatment (pronoun disclosure) and its interactions influence employer response.

Study Timeline

Tasks	Start Date	Duration
Send fictitious resumes to job postings	May 15, 2023	22 weeks
Preliminary analytics, power analysis update	July 10	2 weeks
Collect employer responses	May 15, 2023	30 weeks
Final analysis and write up	October 16, 2023	6 months

Experimental Design

A. Geographies

Fictitious resumes will be sent in the following geographies:

CBSA	State	Population		2020 Presidential Votes		Category
		Count	Density	Democratic	Republican	
Denver-Aurora-Lakewood	CO	2,967 K	356	61%	36%	Democratic
Colorado Springs	CO	746 K	278	42%	54%	Republican
Salt Lake City	UT	1,233 K	160	52%	43%	Democratic
Provo-Orem	UT	648 K	120	26%	68%	Republican
Seattle-Tacoma-Bellevue	WA	3,980 K	678	67%	30%	Democratic
Spokane-Spokane Valley	WA	582 K	103	44%	52%	Republican

Pairs of CBSAs were selected that are (1) in states which have legislation prohibiting labour market discrimination on the basis of both sexuality and gender identity, (2) have a population of at least 500K, and (3) where one can be categorized as Democratic and the other Republican. CBSA population data is sourced from the United States Census Bureau (2021a), land square footage from TIGERweb (United States Census Bureau, 2020), and 2020 Presidential voting records from the MIT Election Data and Science Lab (2018).

B. Occupations

Fictitious resumes will be sent to the following occupations:

Occupation	Worker Count	Worker Composition			Customer Interaction	
		% Male	% Female	Category	Score	Category
Receptionist	638 K	9%	91%	Female-Dominated	87	High
Cashier	739 K	28%	72%		86	
Housekeeper	722 K	15%	85%		58	Medium
Nursing Assistant	804 K	11%	89%		47	Low
Administrative Assistant	1,499 K	6%	94%		47	
Retail Salesperson	1,332 K	62%	38%	Non-Dominated	93	High
Server	527 K	36%	64%		75	
Cook	1,041 K	59%	41%		52	Medium
Baker	122 K	44%	56%		37	Low
Assembler / Fabricator	701 K	62%	38%		17	
Construction Laborer	1,161 K	97%	3%	Male-Dominated	59	Medium
Truck Driver	2,601 K	95%	5%		53	
Warehouse Worker	1,237 K	80%	20%		46	Low
Janitor / Building Cleaner	1,378 K	70%	30%		44	
Landscaper	630 K	94%	6%		32	

An equal number of occupations were selected in each Worker Composition category, where those with high worker counts and job postings were prioritized. In addition, occupations in a mix of Customer Interaction categories were included. Note that there are very few male-dominated occupations with high customer interaction, hence there are no occupations fitting this description.

Worker count and composition data is from the American Community Survey (United States Census Bureau, 2022). Data on Customer Interaction is taken from O*NET scores for the importance of "performing for people or working directly with the public. This includes serving customers in restaurants and stores, and receiving clients or guests." Association between ACS occupation codes and O*Net occupation codes was sourced from O*NET OnLine (2023).

C. Names

The goal when selecting first and last names was to generate names that are generic and racially ambiguous. The following first names (where some imply the applicant is female and others male) will be used in this study. Each name has been grouped with other names that are similar in terms of warmth and competence associations.

Implied Sex	First Name	Baby Name Popularity (1990s)		Name Association Scores		
		Rank	Count	Warmth	Competence	Category
Female	Madeline	92	37 K	3.31	3.51	High
Female	Olivia	38	76 K	3.33	3.51	
Female	Jennifer	16	148 K	3.30	3.50	
Female	Katelyn	62	53 K	3.15	3.13	Medium
Female	Leah	97	34 K	3.13	3.11	
Female	Nicole	17	136 K	3.15	3.11	
Female	Dominique	136	24 K	2.58	2.66	Low
Female	Mercedes	189	15 K	2.50	2.58	
Female	Felicia	187	16 K	2.37	2.57	
Male	William	18	218 K	3.48	3.66	High
Male	Nicholas	6	275 K	3.59	3.59	
Male	Michael	1	462 K	3.54	3.52	
Male	Dennis	178	18 K	3.10	3.06	Medium
Male	Jeremy	47	78 K	3.12	3.05	
Male	Seth	90	42 K	3.11	3.05	
Male	Devon	124	28 K	2.85	2.75	Low
Male	Marco	185	17 K	2.88	2.75	
Male	Dominic	138	26 K	2.89	2.73	

First names were chosen that were (1) in the top 200 popular names given to babies born in the 1990s, and (2) were in a set of 3 names with similar warmth and competence associations. Data on 1990s baby name popularity is from United States Social Security (Social Security, 2022) and data on name association scores is from Newman et al. (2018).

The following last names will be used in this study:

Last Name	Rank	Count	Racial Breakdown			
			White	African American	Hispanic	Other
Green	41	430 K	57%	37%	2.6%	3.6%
Bryant	128	193 K	58%	36%	2.2%	3.8%
Franklin	251	121 K	54%	39%	2.7%	4.2%
Solomon	745	47 K	58%	32%	3.2%	6.9%

Last names were chosen from the most common 1,000 last names in the United States where (1) percent White ranges from 50-60%, and (2) percent African American is over 30%. From this list of 51 potential last names, the above names were randomly chosen. Data on last name frequency and racial breakdown is from US Social Security United States Census Bureau (2021b).

D. Resume Design

A process for generating occupation-specific resumes has been developed using a program by Lahey and Beasley (2009). The characteristics over which resumes are randomized are equivalent across geographies, with the exception of Work Experience where company names are city specific (position titles and descriptions are independent of geography). For all occupations and geographies, fictitious resumes are generated for an applicant born in 1999 (i.e., fictitious applicants are 24 in 2023).

Resumes are generated one pair at a time, and within a characteristic, resumes can be matched same (i.e., if the first resume is randomly assigned characteristic A, then the matched pair will also be given characteristic A) or matched different (i.e., if the first resume is randomly assigned characteristic A, then the matched pair will be randomly assigned a characteristic aside from A). Resumes are randomized across the following characteristics:

- Name: randomly drawn from a grouped list of first names, and ungrouped last names
  - Probability: equal chance of a name in each first name group (implied sex and association category) then equal chance of each first name within the group. Finally, equal probability of each last name regardless of first name
  - Matched: same first name group, different first and last name
- Pronouns: one of they/them, binary pronouns congruent with implied sex, or no pronouns
  - Probability: equal chance of either disclosing pronouns or not, then $\frac{2}{3}$ chance of they/them and $\frac{1}{3}$ chance of binary pronouns given disclosure
  - Matched: different—at least one resume in a matched pair has no pronouns
- Contact Information: one of two local phone numbers, email related to first and last name
  - Probability: equal probability of receiving each phone number
  - Matched: different
- Objective: randomly drawn from a list of objectives
  - Probability: equal probability of receiving each objective
  - Matched: different
- Highest Education: one of GED, high school, Associate's degree, Bachelor's degree
  - Probability: informed by prevalence within the occupation
  - Matched: same level of education, different specialization (if applicable)
- Work Experience, 2015-2017: in the last two years of high school, applicants either did not work or may have held one of two positions
  - Probability: equal probability of receiving each position (including no position)
  - Matched: same (working / not working), different (position / description)
- Work Experience, after 2017: applicants have 2-3 jobs spanning this period
  - Probability: jobs are selected without replacement from 6 possible positions, where 3 are the occupation of interest (i.e., relevant experience) and 3 are other positions commonly held by workers before becoming the occupation of interest
  - Matched: same (years of relevant work experience), different (non-relevant positions, job order)

- Skills: 6 skills are randomly drawn for each applicant where 4 are generic (drawn from the same list across all occupations) and 2 are occupation specific
    - Probability: equal probability across all options
    - Matched: different, across both generic and occupation specific skills

Resumes are then randomly assigned one of two resume formats, which are designed to look as different as possible (different font, resume categories are ordered differently, etc.).

E. Job Application Targets

To improve power of secondary analyses, applications will be balanced across geography and occupation type. The target sample size is 3,240 matched resume pairs (where each pair includes one of two treatments along with a control) sent to job postings—or a target of 6,480 total resumes distributed. Because the quantity of job postings varies with occupation, job application targets vary by occupation:

Occupation	Application Target		
	Percentage	Total Count	Per City Count
Receptionist	8.3%	270	45
Cashier	5.0%	162	27
Housekeeper	6.7%	216	36
Nursing Assistant	8.3%	270	45
Administrative Assistant	5.0%	162	27
Retail Salesperson	11.7%	378	63
Server	6.7%	216	36
Cook	8.3%	270	45
Baker	3.3%	108	18
Assembler / Fabricator	3.3%	108	18
Construction Laborer	11.7%	378	63
Truck Driver	5.0%	162	27
Warehouse Worker	8.3%	270	45
Janitor / Building Cleaner	5.0%	162	27
Landscaper	3.3%	108	18

Summing across occupation categories yields total targets:

Occupation Category	Application Target		
	Percentage	Total Count	Per City Count
Female-Dominated	33.3%	1080	180
Non-Dominated	33.3%	1080	180
Male-Dominated	33.3%	1080	180
High Customer Interaction	31.7%	1026	171
Medium Customer Interaction	31.7%	1026	171
Low Customer Interaction	36.7%	1188	198

Summing across the intersection of occupation categories yields total targets:

Worker Composition	Customer Interaction	Application Target		
		Percentage	Total Count	Per City Count
Female-Dominated	High	13.3%	432	72
Female-Dominated	Medium	6.7%	216	36
Female-Dominated	Low	13.3%	432	72
Non-Dominated	High	18.3%	594	99
Non-Dominated	Medium	8.3%	270	45
Non-Dominated	Low	6.7%	216	36
Male-Dominated	High	-	-	-
Male-Dominated	Medium	16.7%	540	90
Male-Dominated	Low	16.7%	540	90

Actual application counts will be constrained by job posting availability; while the above targets were based on preliminary investigations of job postings within the CBSAs of interest, actual counts may differ.

F. Data Collection Process

A team of research assistants will search major job posting websites (Indeed, Monster, Craigslist, etc.) for occupation vacancies in the CBSAs of interest. When an appropriate job posting is found, a pair of fictitious, randomized, matched, formatted resumes will be generated and sent in in response. To reduce cost, job postings will only be applied to if the application process involves uploading a resume PDF and answering simple, standardized questions that can be easily determined from the randomized resume (e.g., how many years of relevant experience do you have?) or that can have a standard general response (e.g., can you reliably commute to work at this location?—yes). When applying, information on job posting, employer, and resume characteristics will be recorded in an encrypted database.

Employer response will be carefully tracked via phone and email. For each geography, two phone lines will be set up using an area code local to the area. For each first and last name, an email will be set up—with 18 first names and 4 last names, this implies 72 emails will be set up. Phone voicemails and emails will be monitored on an ongoing basis to identify applications which receive a positive employer response. If an employer reaches out at least twice, they will be contacted and told that the applicant has already accepted another position.

Hypotheses Tested

For simplicity, I denote applicants who send resumes with nonbinary pronouns “T1” matched to control “C1,” and resumes with binary pronouns congruent with name-implied sex “T2” matched to control resumes with no pronouns “C2.”

A. Primary Hypotheses

P1. Determine whether T1 achieve lower response rates compared to C1

P2. Determine whether T2 achieve lower response rates compared to C2

P3. Determine whether T1 achieve lower response rates compared to T2

P4. Determine whether differences in response rates between T1, C1 are different for males versus females

P5. Determine whether the difference in response rates between T1, C1 are different between Republican and Democratic geographies

P6. Determine whether the difference in response rates between T1, C1 are different in occupations with high, medium, and low customer interaction

P7. Determine whether the difference in response rates between T1, C1 are different in female-dominated, male-dominated, and non-dominated occupations

B. Secondary Hypotheses

S1. Determine whether differences in response rates between T2, C2 are different for males versus females

S2. Determine whether the difference in response rates between T2, C2 are different between Republican and Democratic geographies

S3. Determine whether the difference in response rates between T2, C2 are different in occupations with high, medium, and low customer interaction

S4. Determine whether the difference in response rates between T2, C2 are different in female-dominated, male-dominated, and non-dominated occupations

S5. Determine whether the difference in response rates between T1, C1 change as relevant experience increases

S6. Determine whether the difference in response rates between T1, C1 change with employer characteristics (employer size, for Equal Opportunity Employers, etc.)

S7. Determine whether the difference in response rates between T1, C1 change with job posting characteristics (existence of key text like "diversity," etc.)

Econometric Specifications

A. Notation

Logistic regression (logit) models will be leveraged, using notation:

$$\mathrm{P}\left(y_{ij} = 1\right) = \frac{1}{1 + e^{-z}}$$

where y_{ij} is an indicator variable which equals 1 if applicant i received a positive response from firm j and z is the model specification specific to analyses described below. For all regressions, standard errors will be clustered at the firm level.

B. Primary Hypotheses

To test P1 and P2, a logistic regression will be run to estimate $\hat{\delta}$ given:

(1) $z = \alpha_j + D_i\delta + X_i'\beta_1 + Z_j'\beta_2$

where α_j are firm fixed effects, D_i is an indicator variable which equals 1 if the resume has treatment pronouns ("they/them" for T1 vs C1, binary pronouns for T2 vs C2), X_i is a vector of resume characteristics that may influence baseline employer response rates (years of relevant work experience, education level, etc.), and Z_j is a vector of occupation and firm characteristics that may influence baseline employer response rates (occupation, firm size, etc.). Multiple specifications will be run, where some will include (α_j, X_i, Z_j) and some will exclude them (when "excluding" α_j it is replaced with α). Note that the specification excluding (α_j, X_i, Z_j) is equivalent to a proportion test; the specification excluding (X_i, Z_j) but including α_j is equivalent to a McNemar (1947) test of differences between matched pairs.

To test P3, a logistic regression will be run to estimate $\hat{\delta}$ given:

(2) $z = \alpha + D_i\delta + X_i'\beta_1 + Z_j'\beta_2$

where D_i is an indicator variable which equals 1 if the resume has "they/them" pronouns. Multiple specifications will be run, where some will include and some exclude (X_i, Z_j).

To test P4, logit (1) will be run separately for females and males. In addition, a logistic regression will be run to estimate $\hat{\delta}$ given:

(3) $z = \alpha_j + D_i\delta_1 + [D_i \cdot S_i]\delta_2 + X_i'\beta_1 + Z_j'\beta_2$

where S_i is an indicator variable which equals 1 if sex implied by name is male. Multiple specifications will be run, where some will include and some exclude (α_j, X_i, Z_j).

To test P5, logit (1) will be run separately for firms in Democratic versus Republican geographies. In addition, a logistic regression will be run to estimate $\hat{\delta}$ given:

(4) $z = \alpha_j + D_i\delta_1 + \left[D_i \cdot G_j\right]\delta_2 + X_i'\beta_1 + Z_j'\beta_2$

(5) $z = \alpha_j + D_i\delta_1 + \left[D_i \cdot V_j\right]\delta_2 + X_i'\beta_1 + Z_j'\beta_2$

where R_j is an indicator variable which equals 1 if the geography is Republican and V_j is the Republican vote share in geography within which the firm is located. Multiple specifications will be run, where some will include and some exclude (α_j, X_i, Z_j).

To test P6, logit (1) will be run separately for occupations with high, medium, and low customer interaction scores. In addition, a logistic regression will be run to estimate $\hat{\delta}$ given:

(6) $z = \alpha_j + D_i\delta_1 + [D_i \cdot HI_j]\delta_2 + [D_i \cdot LI_j]\delta_3 + X_i'\beta_1 + Z_j'\beta_2$

(7) $z = \alpha_j + D_i\delta_1 + [D_i \cdot CIS_j]\delta_2 + [D_i \cdot LI_j]\delta_3 + X_i'\beta_1 + Z_j'\beta_2$

where HI_i is an indicator variable which equals 1 if the occupation is high customer interaction, LI_j is an indicator variable which equals 1 if the occupation is low customer interaction, and CIS_j is the O*NET customer interaction score associated with the occupation. Multiple specifications will be run, where some will include and some exclude (α_j, X_i, Z_j).

To test P7, logit (1) will be run separately for female-dominated, non-dominated, and male-dominated occupations. In addition, a logistic regression will be run to estimate $\hat{\delta}$ given:

(8) $z = \alpha_j + D_i\delta_1 + [D_i \cdot FD_j]\delta_2 + [D_i \cdot MD_j]\delta_3 + X_i'\beta_1 + Z_j'\beta_2$

(9) $z = \alpha_j + D_i\delta_1 + [D_i \cdot SD_j]\delta_2 + [D_i \cdot MD_j]\delta_3 + X_i'\beta_1 + Z_j'\beta_2$

where FD_j is an indicator variable which equals 1 if the occupation is female-dominated, MD_j is an indicator variable which equals 1 if the occupation is male-dominated, and SD_j is the difference in proportion of female- to male- workers in the occupation. Multiple specifications will be run, where some will include and some exclude (α_j, X_i, Z_j).

B. Secondary Hypotheses

To test S1, S2, S3, S4 I will follow the same process as P4, P5, P6, P7 but focus on T2, C2 rather than T1, C1.

To test S5, logit (1) will be run separately for applicants with low (2 years or less) or high (3 years or more) relevant work experience. In addition, a logistic regression will be run to estimate $\hat{\delta}$ given:

(10) $z = \alpha_j + D_i\delta_1 + [D_i \cdot RLE_i]\delta_2 + X_i'\beta_1 + Z_j'\beta_2$

where RLE_i is years of relevant work experience. Multiple specifications will be run, where some will include and some exclude (α_j, X_i, Z_j).

To test S6, logit (1) will be run separately for different firm groups. In addition, a logistic regression will be run to estimate $\hat{\delta}$ given:

(11) $z = \alpha_j + D_i\delta_1 + [D_i \cdot FG_j]\delta_2 + X_i'\beta_1 + Z_j'\beta_2$

(12) $z = D_i\delta_1 + [D_i \cdot F_j]\delta_2 + X_i'\beta_1 + Z_j'\beta_2$

where EG_j is an indicator variable denoting firm group (e.g., it may equal 1 if firms are large or if a firm is an Equal Opportunity Employer) and F_j is a firm value (e.g., number of employees). Multiple specifications will be run, where some will include and some exclude (α_j, X_i, Z_j).

To test S7, logit (1) will be run separately for different job posting types. In addition, a logistic regression will be run to estimate $\hat{\delta}$ given:

$$(13) \qquad z = \alpha_j + D_i\delta_1 + [D_i \cdot JP_j]\delta_2 + X_i'\beta_1 + Z_j'\beta_2$$

where JP_j is an indicator variable denoting job posting group (e.g., it may equal 1 if the job posting contains the word "diversity"). Multiple specifications will be run, where some will include and some exclude (α_j, X_i, Z_j).

C. All Hypotheses

Finally, logistic regressions will be run to estimate $\hat{\delta}$ given:

$$(14) \qquad z = \alpha_j + D_i\delta_1 + [D_i \cdot S_i]\delta_2 + [D_i \cdot G_j]\delta_3 + [D_i \cdot HI_j]\delta_4 + [D_i \cdot LI_j]\delta_5 + [D_i \cdot FD_j]\delta_6 + [D_i \cdot MD_j]\delta_7 + [D_i \cdot RLE_j]\delta_8 + [D_i \cdot FG_j]\delta_9 + [D_i \cdot JP_j]\delta_{10} + X_i'\beta_1 + Z_j'\beta_2$$

$$(15) \qquad z = \alpha_j + D_i\delta_1 + [D_i \cdot S_i]\delta_2 + [D_i \cdot V_j]\delta_3 + [D_i \cdot CIS_j]\delta_4 + [D_i \cdot SD_j]\delta_6 + [D_i \cdot RLE_j]\delta_8 + [D_i \cdot F_j]\delta_9 + [D_i \cdot JP_j]\delta_{10} + X_i'\beta_1 + Z_j'\beta_2$$

Multiple specifications will be run, where some will include and some exclude (α_j, X_i, Z_j). This analysis contributes to the validity of (most) hypotheses.

Robustness Checks

Heckman and Siegelman (1993) critique audit studies by showing that if there are differences in the variance of unobservable variables between treatment and control groups, this can bias discrimination estimates both upwards and downwards. The Neumark (2012) method (modified to be based on logistic rather than probit regression) will be used to identify unbiased discrimination estimate $\hat{\delta}$. Note that this approach requires resumes to randomly vary in quality, which is achieved by this research design since years of relevant experience and education is randomized per the above-described process. This approach requires an identifying assumption: that β_1, β_2 are equal across treatment and control groups (i.e., the extent to which resume, occupation, and employer characteristics influence probability of positive employer response is equal across T1, T2, C1, C2). This assumption will also be tested via the approach described in Neumark (2012).

Power Analysis

Consider proportion test $H_0: p_T - p_C = 0$, $H_1: p_T - p_C < 0$, where p_T is positive employer response for the treatment group, and p_C is positive employer response for the control group. Note that in all tables, $p_T - p_C$ is expressed as percentage points as is Minimum Detectable Effect (MDE). Note that to be conservative, all tables (including T1, C1 or T2, C2 comparisons) calculate power for a proportion test rather than a McNemar test (McNemar tests have higher power especially when there is more concordance in positive employer response; when positive employer response is totally discordant, power is lowest and similar to proportion test power).

Comparing T1 to C1 or T2 to C2, T1 (T2) occurs just as often as C1 (C2). Assuming $p_C = 10\%$, the power of this test at different $p_T - p_C$ is:

Sample Size	MDE (80% Power)	Test Power given $p_T - p_C$						
		-2.0%	-2.5%	-3.0%	-3.5%	-4.0%	-4.5%	-5.0%
Full sample: 2,160 pairs	-2.20%	72%	89%	97%	99%	100%	100%	100%
$\frac{1}{2}$ sample: 1,080 pairs	-3.05%	47%	64%	79%	89%	96%	99%	100%
$\frac{1}{3}$ sample: 720 pairs	-3.70%	34%	49%	63%	75%	86%	94%	98%

Assuming $p_C = 15\%$, the power of this test at different $p_T - p_C$ is:

Sample Size	MDE (80% Power)	Test Power given $p_T - p_C$						
		-2.0%	-2.5%	-3.0%	-3.5%	-4.0%	-4.5%	-5.0%
Full sample: 2,160 pairs	-3.08%	47%	64%	78%	88%	95%	99%	100%
$\frac{1}{2}$ sample: 1,080 pairs	-3.67%	37%	50%	63%	76%	86%	92%	96%
$\frac{1}{3}$ sample: 720 pairs	-4.57%	25%	36%	48%	59%	71%	79%	87%

Comparing T1 to T2, T1 occurs twice as often as T2 (i.e., in a sample of 3,240 there will be 2,160 T1 and 1,080 T2). Assuming $p_{T2} = 10\%$, the power of this test at different $p_{T1} - p_{T2}$ is:

Sample Size	MDE (80% Power)	Test Power given $p_{T1} - p_{T2}$						
		-2.0%	-2.5%	-3.0%	-3.5%	-4.0%	-4.5%	-5.0%
Full sample: 3,240 total	-2.65%	58%	76%	89%	96%	99%	100%	100%
$\frac{1}{2}$ sample: 1,620 total	-3.71%	34%	47%	60%	75%	86%	93%	97%
$\frac{1}{3}$ sample: 1,080 total	-4.47%	25%	35%	46%	58%	70%	81%	89%

Assuming $p_{T2} = 15\%$, the power of this test at different $p_{T1} - p_{T2}$ is:

Sample Size	MDE (80% Power)	Test Power given $p_{T1} - p_{T2}$						
		-2.0%	-2.5%	-3.0%	-3.5%	-4.0%	-4.5%	-5.0%
Full sample: 3,240 total	-3.21%	44%	60%	74%	87%	93%	98%	99%
$\frac{1}{2}$ sample: 1,620 total	-4.47%	26%	37%	49%	60%	71%	80%	87%
$\frac{1}{3}$ sample: 1,080 total	-5.38%	20%	26%	35%	44%	54%	64%	74%

When comparing T1 to T1 in group A and B, T1(A) occurs just as often as T1(B). Assuming $p_{T1_A} = 10\%$, the power of this test at different $p_{T1_B} - p_{T2_A}$ is:

Sample Size	MDE (80% Power)	Test Power given $p_{T1_B} - p_{T1_A}$						
		-2.0%	-2.5%	-3.0%	-3.5%	-4.0%	-4.5%	-5.0%
$\frac{1}{2}$ sample: 2,160 total	-3.05%	47%	64%	79%	89%	96%	99%	100%
$\frac{1}{3}$ sample: 1,440 total	-3.70%	34%	49%	63%	75%	86%	94%	98%

Assuming $p_{T1_A} = 15\%$, the power of this test at different $p_{T1_B} - p_{T1_A}$ is:

Sample Size	MDE (80% Power)	Test Power given $p_{T1_B} - p_{T1_A}$						
		-2.0%	-2.5%	-3.0%	-3.5%	-4.0%	-4.5%	-5.0%
$\frac{1}{2}$ sample: 2,160 total	3.67%	37%	50%	63%	76%	86%	92%	96%
$\frac{1}{3}$ sample: 1,440 total	4.57%	25%	36%	48%	59%	71%	79%	87%

Target sample may be modified in July, after seeing preliminary results. Little is known about hiring discrimination based on pronoun disclosure, so expected effect size is unclear at this time.

References

Brown, A. (June 7, 2022). About 5% of young adults in the U.S. say their gender is different from their sex assigned at birth. Pew Research Centre. https://www.pewresearch.org/fact-tank/2022/06/07/about-5-of-young-adults-in-the-u-s-say-their-gender-is-different-from-their-sex-assigned-at-birth/

Ballard, J. (August 2, 2022). How Americans feel about gender-neutral pronouns in 2022. YouGov America. https://today.yougov.com/topics/politics/articles-reports/2022/08/02/how-americans-gender-neutral-pronouns-2022-poll

Heckman, J., Siegelman, P. (1993). The Urban Institute audit studies: Their methods and findings. *Clear and convincing evidence: Measurement of discrimination in America, Fix and Struyk*: 187-258. Washington, D.C.: The Urban Institute Press. https://works.bepress.com/peter_siegelman/33/

James, S. E., Herman, J. L., Rankin, S., Keisling, M., Mottet, L., & Anafi, M. (2016). The Report of the 2015 U.S. Transgender Survey. Washington, DC: National Center for Transgender Equality. https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf

Lahey, J., Beasley, R. A. (2009). Computerizing audit studies. *Journal of Economic Behavior & Organization*, 70(3): 508-514. https://dx.doi.org/10.2139/ssrn.1001038

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12: 153-157. https://psycnet.apa.org/doi/10.1007/BF02295996

MIT Election Data and Science Lab (2018). County Presidential Election Returns 2000-2020. Harvard Dataverse, V11. https://doi.org/10.7910/DVN/VOQCHQ

Neumark, D. (2012). Detecting discrimination in audit and correspondence studies. *The Journal of Human Resources*, 47(4): 1128-1157. https://www.jstor.org/stable/23798528

Newman, L. S., Mingxuan, T., Caldwell, T. L., Duff, K. J., Winer, S. E. (2018). Name norms: A guide to casting your next experiment. *Personality and Psychology Bulletin*, 44(10): 1435-1448. https://doi.org/10.1177/0146167218769858

O*NET OnLine (April 18, 2023). Performing for or working directly with the public. https://www.onetonline.org/find/descriptor/result/4.A.4.a.8

Social Security (March, 2022). Top names of the 1990s. United States government. https://www.ssa.gov/oact/babynames/decades/names1990s.html

United States Census Bureau (May, 2022). Detailed occupation by sex education age earnings: ACS 2019. American Community Survey. https://www.census.gov/data/tables/2022/demo/acs-2019.html

United States Census Bureau (October 8, 2021a). County population by characteristics: 2010-2019. https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html

United States Census Bureau (October 8, 2021b). Frequently occurring surnames from the 2010 Census. https://www.census.gov/topics/population/genealogy/data/2010_surnames.html

United States Census Bureau (January 1, 2020). TIGERweb decennial state-based data files: county. TIGERweb. https://tigerweb.geo.census.gov/tigerwebmain/TIGERweb2020_counties_census2020.html