Gender Conformity and Wellbeing:

An Experiment in Ordering Effects

Aliya Saperstein

Stanford University

Tagart Sobotka

University of California-Santa Barbara

Abstract

Gender disparities in health, between women and men, are widespread and well known. However, growing use of more inclusive gender measures allows for expanding our understanding not only to additional categories such as transgender or nonbinary people but also to variation within categories, by perceived femininity and masculinity. In this study, we use nonbinary categorical and gradational gender measures and a survey experiment to tease apart whether perceptions of gender influence reported wellbeing and vice versa. We find evidence of a bidirectional relationship, with both reported gender perceptions and wellbeing affected by question order. We also find that people who experience larger gaps between their gender self-conception and how most people see them report worse wellbeing, especially when they were first asked to reflect on their gender. Our results are consistent across two samples of U.S. adults: a nationally representative sample and a targeted sample of gender and sexual minorities.

This abstract was prepared for submission to the 2025 conference of the International Union for The Scientific Study of Population. The LGBT sample was supported through a special competition for targeted samples sponsored by Time-sharing Experiments for the Social Sciences, NSF Grant 0818839, Jeremy Freese and James Druckman, Principal Investigators. We are also grateful to Elizabeth Deneen for research assistance and to the staff at NORC/AmeriSpeak for managing data collection.

Background

Interest in using nonbinary gender measures to study contemporary gender inequality has increased in recent years with a growing number of surveys around the world now including both nonbinary categorical and more gradational femininity and masculinity measures (Alexander et al. 2021). Although the topical focus of such analyses is expanding, most past studies have used these measures to assess gender disparities in health (see Hart et al. 2019 for a review). In line with theories that link masculinity with risk-taking and poor health behaviors, early studies tended to find that self-reported femininity was associated with better health outcomes. At the same time, we might expect people who have poor health to be seen as less masculine by others, if they are unable to perform stereotypical masculine acts of agency and strength. More recent work has also pointed to a role for gender conformity more generally, with both cisgender and transgender people who are seen to be nonconforming by others reporting worse wellbeing.

Collectively, this body of work raises the question of whether the relationship between gender and health might run in both directions. We designed an original survey experiment to test this possibility, which has both substantive and methodological implications for the study of gender and wellbeing. With growing use of more inclusive gender measurements in research, often with the express purpose of better understanding disparities in wellbeing (NASEM 2022), an explicit study of these potential ordering effects is necessary to further advance the field.

Data and methods

We draw on two sets of original data that both involved a survey experiment in which the order of health and gender related questions were randomized across participants. The overall design of the study was similar between samples and both were in the field in April-May 2023. Our first sample (n=1,495) includes adults recruited through the online crowdsourcing platform Prolific using U.S. nationally representative quotas for binary sex, age and race. Our second sample was a targeted study of previously identified gender- and sexual-minority adults sponsored by Time-sharing Experiments for the Social Sciences (TESS) and conducted by NORC. Given the challenges of reaching this population, participants were recruited through the platforms AmeriSpeak (n=1,057) and Lucid (n=1,346). Reflecting this targeted sample, the percent of participants identified as cisgender or straight/heterosexual was significantly lower in the TESS sample (89% and 5%, respectively) than the Prolific sample (98% and 85%, respectively). The purpose of the two samples was to be able to examine whether measured effects would be similar in both general population and LGBT-focused studies. We show pooled estimates here for the sake of simplicity, as the analyses below were generally consistent across samples.

Gender measures

We followed recent recommendations from the U.S. National Academies of Sciences, Engineering and Medicine (NASEM 2022) and use a two-step measure that asks about sex assigned and birth and current gender identity. The categorical gender question offered four possible responses: man, woman, transgender, or “I use a different term.” Those who reported using a different term were also asked to specify, with “nonbinary” being the most common write-in response in both TESS (43%) and Prolific (80%) samples.

We also assessed first and third order perceptions of participants gender using paired 7-point feminine and masculine scales that ranged from Not at all=0 to Very=6 (see Magliozzi et al. 2016). Respondents were first asked "How do you see yourself?" and asked to respond regarding both femininity and masculinity followed by "How do most people see you?" The order in which the feminine and masculine scales were presented was randomized between conditions but kept consistent between first- and third-order assessments to reduce confusion. From these four measures, we constructed four variables that reflect levels of polarization and difference in how participants evaluated themselves on the scales. First, we constructed polarization scores by calculating the absolute difference between first order masculinity and femininity scores (*Polar-Self*) and third order masculinity and femininity scores (*Polar-See)*. Higher values indicate greater polarization on participants' first or third order evaluations on the gender scales but not the direction of difference (see Magliozzi et al. 2016). We also created *Feminine Diff* and *Masculine Diff* variables by calculating the absolute value of the difference between the respective third and first order scales. Higher values indicate larger gaps in self and other perceptions but not the direction of the difference.

Health Measures

To measure health, we drew on items from the General Social Survey that ask respondents to self-rate three dimensions of their health—quality of life, physical health, and mental health—as either Excellent, Very good, Good, Fair, or Poor. First, participants were asked: "In general, would you say your quality of life is:". Then, "In general, how would you rate your physical health?" Finally, participants were asked "In general, how would you rate your mental health, including your mood and ability to think?" Responses were recoded so higher scores indicate better reported health and wellbeing. Mean scores on each of the health measures between the samples ranged from 2.89-3.36, indicating that participants, on average, saw their health as "good." However, pairwise ttests showed mean scores were significantly lower for the gender and sexual minority sample, indicating lower reported wellbeing on all three measures.

Survey experiment

We utilized a between-subject design and randomized the order in which participants saw the gender and health measures (*Health Scales First* and *Gender Scales First*). Doing so allows us to test whether question ordering affects how participants evaluate their gender and wellbeing.

Results

We first consider whether the gender scale responses were influenced by the question order and then whether the wellbeing responses were also affected. We go on to examine a potential mechanism for the wellbeing results: whether thinking about their gender first was especially influential for respondents who had greater differences in their first and third order scale scores.

Gender polarization and question order

We find that U.S. adults report more polarized gender perceptions when they were asked to consider their femininity and masculinity before they were asked to report on their health and quality of life. The relationship was strongest for our third-order gender measure (how most

people see you), compared their (first-order) gender self-conception, and strongest when the feminine scale appeared before the masculine scale in each gender scale pair.

Table 1 shows differences in average polarization scores for the full pooled sample (Model 1) and for the pooled sample by scale presentation order. Model 2a reports the average effect of seeing gender questions first when also seeing the feminine scale before the masculine scale while Model 3a reflects the opposite scale presentation condition. All estimates for the effect of question order are positive, indicating more polarized gender responses when the gender questions came first, but only the estimate for seeing gender question first and the feminine scales being presented first is measurably different from zero at conventional levels of statistical significance.

The final two columns of Table 1 show the question ordering difference in reported gender perceptions also holds in the presence of basic controls for characteristics that differ between our two samples. As might be expected, cisgender respondents, older respondents, and straight respondents report most other people perceive their gender in more traditional polarized ways, while gay and lesbian respondents reported most people perceived smaller differences between their femininity and their masculinity. Table A1 provides descriptive results for all gender scales, separately for self-identified women, men, transgender people and people who prefer another gender term. The patterns of greater polarization when the gender scales appeared first holds for women and men for both first and third-order gender scales; there are less consistent patterns by question order for gender self-conceptions among transgender people and those who prefer another gender term.

Reported wellbeing and question order

Importantly, the question order also appears to influence respondents' reports of their health and quality of life. As Table 2 shows, people who saw the gender scales first consistently reported lower quality of life and worse physical and mental health. These results hold in the presence of controls for age, categorical gender (women, men, transgender, another gender) and sexual orientation, which represent the major demographic differences across the two samples. The magnitude of the reported wellbeing differences by question order are smaller than those observed for the gender scales and are only statistically significant at conventional levels for quality of life in the presence of controls. However, supplementary analyses suggest the question order effects on reported wellbeing also vary more across subgroups in the sample, so the pooled estimates shown here likely underestimate how seeing the gender scales first shaped the wellbeing reports of some respondents.

A possible mechanism: facing gender nonconformity

When the gender scales appeared before the health scales, respondents first answered how they saw themselves in terms of their femininity and masculinity, then how they thought most people see them, and only then did they report their quality of life and physical and mental health. It is possible, then, that thinking about how other people perceive their femininity and masculinity prompted more negative assessments of wellbeing, perhaps especially if it reminded respondents of how much others' view differs from their self-conception. If so, we would expect this effect to

be strongest among people with the largest differences between their first- and third-order gender perceptions.

This is precisely what we find. Larger gaps between how people see themselves and how they think most other people see them is associated with reporting lower quality of life and worse physical and mental health. Crucially, though, this relationship is strongest, and in some cases only holds, when respondents saw the gender scales first. We illustrate this with the difference in masculinity scale scores for the pooled sample, though results were generally consistent across samples and also similar for the difference in femininity scores.

Table 3 shows regression results separately for self-identified women and men, across the three health measures. Estimates shown in the first column correspond to when respondents saw the health scales first and the second column corresponds to when respondents saw the gender scales first. All but one of the estimates are negative, indicating that larger differences in masculinity perceptions are generally associated with worse health, but only in the "gender first" columns are the estimates consistently statistically significant. There is one estimate in the "health first" columns that is also statistically significant and that is for men and mental health, suggesting that larger gaps in masculinity perceptions may take a toll on men's mental health in general, in addition to the effects of the experimental manipulation that primed gender before health.

Conclusion and implications

Substantively, these results suggest there is a bidirectional relationship between gender and health: people's perceptions of their gender affect their health, or at least how they report it in surveys, and vice versa. This supports perspectives that "doing gender" and "doing health" are intertwined (see, e.g., Courtenay 2000). Our results also echo the potential health consequences for perceived gender nonconformity when it is not part of one's self-conception (Hart et al. 2019), in this case in relation to men and their perceived masculinity.

Methodologically, our findings also have implications for research practice. Both sets of scales appear to be sensitive to context. Starting off our brief survey with the gender scales produced more polarized responses, especially when respondents saw the feminine scales presented before the masculine scales in each pair of gender questions. Being presented with gradational (nonbinary) gender measures that also upended the traditional gender order may have produced a threat response that was particularly reflected in how respondents reported most other people would perceive their gender. When these respondents then went on to report their health, those with larger differences between how they saw their own gender and how they reported most other people saw them also reported worse health and wellbeing. Thus, surveys that incorporate nonbinary gender measures will need to carefully consider their placement to avoid question order affects. In the U.S. context, they likely will work better surrounded by more gender-neutral demographic questions such as race/ethnicity (i.e. items for which responses do not generally exhibit or cue gender disparities).

References

Alexander, Amy C., Catherine Bolzendahl, and Lena Wängnerud. 2021. "Beyond the binary: new approaches to measuring gender in political science research." *European Journal of Politics and Gender* 4 (1): 7-9.

Courtenay, Will H. 2000. "Constructions of masculinity and their influence on men's well-being: a theory of gender and health." *Social science & medicine* 50 (10): 1385-1401.

Hart, Chloe Grace, Aliya Saperstein, Devon Magliozzi, and Laurel Westbrook. 2019. “Gender and Health: Beyond Binary Categorical Measurement.” *Journal of Health and Social Behavior* 60(1):101–18. doi: 10.1177/0022146519825749.

Magliozzi, Devon, Aliya Saperstein, and Laurel Westbrook. 2016. "Scaling up: Representing gender diversity in survey research." *Socius* 2. doi: 10.1177/2378023116664352

National Academies of Sciences, Engineering, and Medicine. 2022. *Measuring Sex, Gender Identity, and Sexual Orientation*. Washington, DC: The National Academies Press. doi: 10.17226/26424.

Table 1. OLS regressions of polarized gender perceptions on randomized question order

	1	2a	3a	2b	3b
Third-order gender polarization (polarsee)	Pooled sample	Feminine scale first	Masculine scale first	with controls	
Gender scale first	.194***	.301**	0.089	.273**	.113
	(.063)	(.088)	(.090)	(.084)	(.087)
Cisgender respondents				.232	.432*
				(.169)	(.170)
Age (ref. 18-29)					
Age 30-44				-.067	.101
				(.105)	(.111)
Age 45-59				.338**	.419**
				(.125)	(.013)
Age 60 and up				.882***	.716***
				(.136)	(.134)
Sexual orientation (ref. Bisexual)					
Gay or lesbian				-.367**	-.433***
				(.117)	(.120)
Straight				.687***	.543***
				(.104)	(.108)
Another sexual orientation				-.432*	-.278
				(.188)	(.210)
Constant	3.18***	3.19***	3.17***	2.66***	2.43***
	(.045)	(.062)	(.064)	(.177)	(.177)
N	3881	1952	1929	1950	1929

Note: Gender polarization is the absolute value of the difference between respondents feminine and masculine scale responses. It runs from 0 to 6, with 6 representing a dichotomous sense of gender (i.e., being "very" on one scale and "not at all" on the other). Cisgender respondents reported the same binary gender identity as their sex assigned at birth. Standard errors in parentheses. ***p<.001 **p<.01 *p<.05

Table 2. OLS regressions of reported wellbeing on question order

	1	2
Quality of life	Pooled sample	with controls
Gender scales first	-.054+	-.064.*
	(.032)	(.032)
Physical health		
Gender scales first	-.049	-.062+
	(.063)	(.032)
Mental health		
Gender scales first	-.052	-.064+
	(.038)	(.035)

Note: Models estimated separately by wellbeing measure, which are on the same 5-point scale with lower scores indicating worse wellbeing. Controls not shown are: self-identified gender (woman, man, transgender, another gender), age, and sexual orientation. All models also control for whether the feminine scales appeared first. Standard errors in parentheses. *p<.05 +p<.10

Table 3. OLS regressions of reported wellbeing on masculinity perception gaps, by question order

	Women		**Men**	
Quality of life	Health first	Gender first	Health first	Gender first
Masculinity difference	-0.0157	-.115***	-0.04	-.173***
	(.033)	(.033)	(.043)	(.045)
Physical health	Health first	Gender first	Health first	Gender first
Masculinity difference	0.007	-.109**	-0.006	-.178***
	(.033)	(.034)	(.045)	(.044)
Mental health	Health first	Gender first	Health first	Gender first
Masculinity difference	-0.042	-.180***	-.101*	-.213***
	(.039)	(.039)	(.050)	(.049)
N	1020	1013	750	826

Note: Models estimated separately by self-identified gender, question order and health measure. Masculinity difference is the absolute value of the difference between how respondents said most people saw their masculinity and how they saw themselves in terms of masculinity. All models control for whether the feminine scales appeared first (not shown). Standard errors in parentheses. ***p<.001 **p<.01 *p<.05

Table A1. Average gender scale scores, by randomized question order and self-identified categorical gender

Women

	Feminine scale first			*Masculine scale first*		
	Health first	Gender first	Difference	Health first	Gender first	Difference
Gender scales						
Polarsee	3.15	3.46	-0.31	3.2	3.21	-0.01
Femsee	4.43	4.55	-0.12	4.37	4.37	0
Mascsee	1.75	1.54	0.21	1.67	1.67	0
Polarself	2.86	3.02	-0.16	2.89	2.93	-0.04
Femself	4.4	4.48	-0.08	4.37	4.43	-0.06
Mascself	1.93	1.78	0.15	1.85	1.86	-0.01

Men

	Feminine scale first			*Masculine scale first*		
	Health first	Gender first	Difference	Health first	Gender first	Difference
Polarsee	3.39	3.62	-0.23	3.32	3.42	-0.1
Femsee	1.51	1.33	0.18	1.47	1.4	0.07
Mascsee	4.38	4.58	-0.2	4.45	4.45	0
Polarself	3.1	3.34	-0.24	3.16	3.2	-0.04
Femself	1.68	1.48	0.2	1.58	1.55	0.03
Mascself	4.41	4.54	-0.13	4.37	4.45	-0.08

Transgender people

	Feminine scale first			*Masculine scale first*		
	Health first	Gender first	Difference	Health first	Gender first	Difference
Polarsee	2.41	2.79	-0.38	2.28	2.35	-0.07
Femsee	3.12	2.97	0.15	3.89	3.25	0.64
Mascsee	3.59	3.48	0.11	2.5	3.9	-1.4
Polarself	2.59	2.07	0.52	2.44	2.15	0.29
Femself	3.35	3.33	0.02	3.75	3	0.75
Mascself	3.29	3.38	0.15	2.69	4.05	-1.36

Another gender

	Feminine scale first			*Masculine scale first*		
	Health first	Gender first	Difference	Health first	Gender first	Difference
Polarsee	2.42	2.85	-0.43	2.2	2.69	-0.49
Femsee	3.68	2.75	0.93	3.57	3.6	-0.03
Mascsee	2.32	3	-0.68	2.8	2.48	0.32
Polarself	1.05	2	-0.95	1.66	1.5	0.16
Femself	2.76	2.65	0.11	2.71	3.12	-0.41
Mascself	2.92	3.35	-0.13	3.63	2.9	0.73

Note: The "see" scales asked about third-order gender perceptions (how do most people see you?). The "self" scales asked about first-order gender perceptions (how do you see yourself?). The polarization scores are the absolute value of the difference between the feminine and masculine responses. N=3,881; 2,045 women; 1,581 men; 119 transgender people; 135 another gender.