Pre-Analysis Plan:

Hard garbling methods for monitoring harassment at scale

Laura Boudreau
Columbia Business School
l.boudreau@columbia.edu

Sylvain Chassang
Princeton University
chassang@princeton.edu

Ada González-Torres
Ben-Gurion University
adagt@bgu.ac.il

Rachel Heath
University of Washington
rmheath@uw.edu

December 16, 2025

Contents

1 Introduction

1.1 Abstract

This research studies how survey design affects transmission of sensitive information within organizations. We partner with a large, international apparel buyer and conduct a phone-based survey experiment with workers across several of its Bangladeshi supplier factories to study mechanisms behind garbling methods shown to increase workers' willingness to report misbehavior by managers (Boudreau et al., 2023). Boudreau et al. (2023) was the first test of hard garbling (HG) outside of a lab, building on Chassang and Padró i Miquel (2018); Chassang and Zehnder (2024). We extend our prior work to test for differences in reporting between higher and low HG rates, to assess workers' preferences over different HG rates, and to assess worker welfare consequences of the ability to report harassment in a way that allows possible deniability.

1.2 Motivation

The garment industry in Bangladesh and other export manufacturing in low-income countries raise female labor force participation (Standing, 1999; Heath and Mobarak, 2015), frequently pay better than workers' alternate options (Verhoogen, 2008; Frías et al., 2012; Amiti and Davis, 2012), and increase investment in children's human capital (Qian, 2008; Atkin, 2009; Heath and Mobarak, 2015). But in light of the poor conditions that characterize many low-skill manufacturing sectors, some researchers have questioned whether manufacturing jobs are actually better for workers than their alternatives (Blattman and Dercon, 2018; Blattman et al., 2019). Indeed, garment workers in Bangladesh face both safety violations (culminating in the tragic collapse of the Rana Plaza in 2013, which killed over 1100 workers), and high rates of physical and sexual abuse (Khosla, 2009; Begum et al., 2010; Subramanian, 2020; Gibbs et al., 2019; Boudreau et al., 2023).

However, organizations' ability to address harassment is limited by their ability to elicit information from relevant parties. Reporting harassment is a difficult step

3

for individuals concerned with possible retaliation and reputational costs. To assess whether survey techniques that guarantee workers plausible deniability increase reporting, we conducted an experiment in coordination with a large Bangladeshi apparel manufacturer (Boudreau et al., 2023). Within a phone-based survey of 2,245 workers at two plants of this manufacturer, we implemented hard garbling (HG) techniques that automatically randomly record a share of reported "no's" as "yeses,", thus ensuring plausible deniability to workers who report harassment. HG increased reporting of physical harassment by 290%, sexual harassment by 271%, and threatening behavior by 45%, from reporting rates of 1.5%, 1.8%, and 9.9%, respectively, under the status quo of direct elicitation (DE).

In addition, we replicated our HG results in a separate community-based survey (Boudreau et al., 2021), and a pilot experiment that precedes this extension.[1] We replicated our main HG treatment effect, namely finding that HG significantly raises reports of threats, physical harassment, and sexual harassment.

1.3 Goals

This survey experiment extends the results in Boudreau, Chassang, González-Torres, and Heath (2023) in important ways: Our main goals are (1) to understand the mechanisms underlying HG by conducting a test for differences in reporting between HG rates and introducing a new test focused on whether HG leads to reporting of false positives, (2) assess the workers' reported preferences over different HG rates, and (3) assess the consequences of the ability to report harassment in a way that allows possible deniability on mental health, locus of control and job satisfaction. In addition, we aim to achieve a sample size to detect gender differences if there are any.

---

[1]To achieve cost-effectiveness, the pilot experiment drew from the community-based survey sample. Extension survey 1 assigned different garbling rates across individuals, with our initial trouble-shooting phase using i.i.d. garbling (sample size of 614 respondents, from August-October, 2024), while our second phase applied blocked garbling (1192 respondents, from October-November, 2024). Extension survey 2 switched garbling rates within individuals across different outcomes (dropping DE for power) and added follow-up questions on preferences for DE, HG and different garbling rates (830 respondents, from December 2024-January 2025).

Varying the level of garbling tests the mechanism through which HG works and is an important input into the design of reporting systems to assess harassment. We test garbling rates of 1% and 20%. The greater the share of automatically-recorded yeses, the more protection afforded to reporters of harassment, which theory predicts will lead to more reports. Higher reporting in the high garbling treatment is thus a test of workers' understanding of the mechanism behind HG. But this protection comes at two costs. First, more automatic yeses (i.e., more false positives) limits the usefulness of the data for targeting harassment reduction interventions. As such, assessing the elasticity of reporting to the garbling rate helps firms and policy-makers understand the tradeoff between minimizing underreporting and being able to target governance interventions using garbled data. Second, higher garbling rates introduce more noise in the statistical data analysis, which limits what can be learned about the true values of population parameters.

If we find that reporting rates are similar under the low and high garbling rates, we will explore the leading mechanisms for why this may be the case. One possibility is that participants understand garbing and fundamentally value *any* plausible deniability, especially if they believe (potentially accurately) that anyone who accesses leaked information will know that garbling was involved but not the rate of garbling used. If so, this is a highly policy-relevant finding, because higher garbling rates come at potential costs to both the worker and the policy-maker. For the worker, higher garbling rates mean a greater chance that their name is associated with a report of harassment, which, if leaked, could lead to the type of retaliation that discourages reporting in the first place. The worker may also have preferences for their report to be more informative to policy-makers, which is more so the case with the lower garbling rate.

The survey assesses workers' comprehension of the HG mechanism. First, it includes questions about workers' comprehension of probabilities, which in the low socioeconomic status population we study, might limit respondents' responsiveness to variations in the garbling rate. Second, it includes multiple choice questions about the HG mechanism. Third, we will ask workers about fatal accidents in their firm under HG and DE. Because we have a proxy for the true occurrence of this event

5

from newspaper articles, this question will allow to assess whether HG leads to false positive reports, which we hypothesize would most likely be due to higher rates of confusion but in principle could also be due to strategic misreporting.

Finally, the survey assesses workers' preferences for different reporting methods. This will help us understand workers' perception of the welfare consequences for workers like themselves. In particular, we ask respondents to state their preferences for garbling vs direct elicitation, and among garbling, for high versus low garbling rates, for future surveys conducted by our buyer partner to allow workers to voice direct input into the design of garbling systems. These questions allow us to assess whether workers' preferences over garbling rates are such that they prefer fewer false positives.

Beyond HG's effects on reporting, we assess the welfare consequences of the ability to report harassment in a way that allows plausible deniability on workers' mental health and job satisfaction. We focus on the intent to treat effect of assignment to the HG system, as opposed to the causal effects of reporting for workers induced to report under HG, because of our interest in the welfare consequences of the HG system for all workers exposed to it. Such effects present a potential pathway to scale for the project; workers with higher job satisfaction have lower turnover rates, improving productivity for firms and potentially prompting them to adopt garbling systems.

Finally, compared to Boudreau et al. (2023), in this survey, we aim for a more even gender split of respondents. In our prior survey, we surveyed all eligible employees from participating plants at our partner firm, which generated a sample that is 81% women. Results showed that HG increased reporting for both sexes but was more important in increasing men's reporting. In this implementation, we will aim for a more equal gender split to increase power to detect gender differences.

1.4   Research Questions

Our main research question is:

1. How does survey design affect transmission of sensitive information in a real-

world organizational setting? More precisely:

(a) Does providing respondents with plausible deniability through hard garbling increase reporting of sensitive information in an organizational setting?

(b) Is respondents' willingness to report sensitive to the garbling rate at which 'nos' are flipped to 'yeses' (1% vs. 20%?)? If so, why?

As noted above, if we do not find a difference in reporting between the 1% and 20% garbling rates, this does not mean that participants don't understand garbling; it could be that respondents fundamentally value *any* plausible deniability, especially if they believe (potentially accurately) that anyone who accesses leaked information will know that garbling was involved but not the rate of garbling used.

Our secondary research questions are:

1. Which reporting system, and under HG, which garbling rate, do respondents believe benefit garment workers the most?

2. Does hard garbling affect men and women differently?

3. Does the ability to report harassment in a way that allows plausible deniability improve workers' well-being in the short-run?

1.5   Research Design

We randomly assign workers to one of three treatment conditions. At a minimum, we stratify the assignment by factory × production section × gender. When permitted by the data, we stratify by production team instead of production section. In large production teams or sections, we additionally stratify by workers' position type.

1) Direct elicitation (DE): directly asking the survey respondent about sensitive information. DE is the status quo survey method and the control condition.

7

– With approximately 33% of DE respondents, after the reporting experiment, we explain HG and elicit their preferences for DE vs. HG.

2) Hard garbling (HG): for a yes or no question, where "yes" is the more sensitive answer, we exogenously flip "no" answers to "yes" with a fixed probability. We fix the flipping rate at the factory leve (HG01) or the stratum-level (HG20), and we examine two variations in this flipping rate. In this way, if a respondent's answer is saved as "yes", it is impossible to know whether they actually responded yes or no, however, we can still learn about overall patterns from workers' responses.

   2a) HG with 20 out of 100 responses automatically recorded as "Yes."
   2b) HG with 1 out of 100 responses automatically recorded as "Yes."

We expect to find a difference between the HG arms and DE, but we do not have a strong prior on whether there is a difference between the HG arms. This is because it may be the case that higher garbling does not necessarily lead to more reporting, if workers primarily value any plausible deniability and/or have preferences for their reports to be more informative for decision-makers. Our survey will also allow to us explore other leading explanations for why we may not find reporting differences under the two rates, such as limited comprehension of probabilities.

In the HG arm, we ask all individuals what their preferred elicitation system is. We also ask this question to a random 33% of the DE-arm, to whom we explain the HG-mechanism after they have been asked about experienced harassment. We do this to help understand the extent to which experience with the HG mechanism affects preferences for it. We only apply this to a subset of the DE arm so that we can study impacts of HG relative to DE on worker well-being.

We estimate the effects of these treatments on two primary outcome variables: physical and sexual harassment. Details of variable construction are provided in Section 3.1.1, Table 2.

The scripts for the DE and HG conditions are displayed in Appendix A.1. Relative to our prior implementations, in this survey, we explained the mechanisms using the

most comparable security assurances possible. This aims to address the concern that our prior results were driven by stronger general security assurances under the HG mechanism, as opposed to the plausible deniability that HG affords to respondents.

2   Research Strategy

2.1   Setting and Context

Context: Our empirical setting is Bangladesh's apparel sector, which is critical to the country's economy. While factories vary, 50-60% of workers in the sector are women, while more than 90% of managers are men. The sector has long struggled with an international reputation for poor working conditions and limited labor rights. Workers who experience harassment or other forms of mistreatment by management often have few to no options to obtain resolution: Internal reporting systems are often corrupt or ineffective and legal institutions are weak.

Collaboration with an international apparel buyer: The project will be conducted in collaboration with a large, international apparel buyer that is interested in addressing harassment in their supplier factories, especially of women. The firm is allowing us to sample its supplier factories in Bangladesh, coordinating supplier factories' participation, and ensuring timely transfer of supplier factories' worker lists for the survey implementation. We main the buyer's anonymity due to the sensitivity of the study's focus on harassment.

2.2   Sampling and treatment assignment

2.2.1   Sampling, sample size and power

Sampling size and method: We will draw a sample of 5,377 workers from employee lists in 3-5 factories that supply to our apparel partner, stratifying using the variables discussed above. We use data from a related survey by Boudreau,

Chassang, and González-Torres (2021) to inform power calculations for two binary outcome variables (reports of physical and sexual harassment).

Power calculation assumptions:

- Power: $\beta = 0.9$.

- Significance level: We set $\alpha = 0.10$ and conservatively apply the Bonferroni adjustment with 2 primary outcomes of $\frac{0.10}{2} = 0.050$.

- Within-factory component of the variance (residual variance of outcome estimated from a regression with factory FE and demographic controls) $= 0.008$ for both physical and sexual harassment.

- Share of sample in treatment group: 2/3.

- Hypothesis test: 1-sided. This is justified based on the theoretical prediction that HG increases willingness to report compared to DE.

Minimum detectable effect (MDE): With these assumptions, we can detect a small to moderate minimum detectable effect (MDE) size of 0.15 standard deviations (SDs) for physical and sexual harassment. This effect is very realistic, given that the effects of HG in Boudreau et al (2022) for physical and sexual harassment were each 0.37 of a SD.

Attrition from the sample: We are able to replace workers who decline to participate or who are unreachable; for each stratum, we have a randomly-ordered list of back-up workers that we will use to replace workers who decline to participate or who are unreachable. In Boudreau et al. (2023), the response rate was 63%, which was primarily due to outdated phone numbers. As such, we anticipate that we may need to approach around 8,535 workers in order to achieve our target sample size of 5,377.

There may be attrition between our main survey experiment and our follow-up survey. While this attrition does not threaten our ability to measure the primary

treatment effects of interest, it would impact our ability to detect one of the secondary treatment effects of interest, which is the impacts of increased reporting on workers' well-being. In Boudreau et al. (2023) we had an attrition rate from baseline to follow-up survey of 7% and it was uncorrelated with treatment condition. We expect attrition to be low also here, because we are conducting the follow-up survey 7-10 days after our main survey. We will report the overall attrition rate in the paper as well any differences in the attrition rates across treatment conditions.

Table 1: Main Experiment - Target sample sizes by treatment arm

		Sample size	TOTAL
Direct elicitation	Standard	1201	1793
	Explain HG	592	
Hard garbling	20 out of 100	1792	3584
	1 out of 100	1792	
			5,377

Notes: The DE condition is the benchmark for the main research questions. For secondary research questions, we will test if the HG explanation affects respondents' mental health and job satisfaction outcomes in the follow-up survey. If not, we will pool the "Explain HG" and standard DE groups when analyzing our secondary research questions. The HG arm with the 20% garbling rate is the most protective.

2.2.2 Treatment assignment

The **unit of randomization** is a worker, **stratified** by factory, production team or section, and sex. The randomization is done in Stata. We first sample 15 or 30 workers per stratum depending on stratum size. We then randomly assign workers to the DE, HG20, or HG01 conditions. For the DE arm, we then randomly assign approximately 33% of DE respondents to the HG explanation condition, maintaining

the global share of DE workers in this condition at 33%. Table 1 provides our planned sample sizes, for each treatment arm, with HG20, HG01, DE being assigned 33% of the sample. For each participating factory, we will conduct 10 randomizations and select the one that performs best in terms of balance on two covariates available to the research team (tenure and skill group).

2.3   Data collection

2.3.1   Instruments

We use the following data collection instruments:

- Main worker survey

- Follow-up worker survey

The main worker survey   has a total of 8 sections:

1. Core demographics, health, and well-being

2. Understanding and awareness of reporting methods in firm

3. Job satisfaction and relationship with management

4. Observed harassment of co-workers and barriers to reporting

5. Respondents' experience of harassment [Reporting experiment], respondents' preferred reporting system, comprehension of different rates, and concerns with reporting

6. Beliefs regarding privacy of responses

7. Additional questions for separate experiment

8. Interviewer's comments

The experimental conditions apply to two sections of the main worker survey:

1. Survey elicitation method (DE, HG-high, HG-low): experimentally varies method for questions in *Section 5*

2. Explaining HG to DE: treats a fraction of individuals in the DE arm in *Section 5*, after the reporting experiment

The follow-up worker survey has one section split in two:

1. Mental well-being, locus of control, job-satisfaction

2. For workers who left factory, reasons for leaving factory and current employment

2.3.2 Data Collection

We plan to conduct the **main worker survey** from late December 2025 - February 2026. We roll out the survey by factory and production floor in order to avoid spillovers as workers could discuss the survey details with other workers before they have been treated. The **follow-up worker survey** is planned 7-10 days after a worker completes the main survey.

2.3.3 Data Processing

Data privacy and protection of human subjects: Participation in the survey may involve some risks. First, workers may feel they owe to the factory that they need to participate in the survey. To address this issue, we inform all subjects that participation in the surveys is entirely voluntary. A surveyor will read out loud an informed consent form that is provided to each worker surveyed and if they decide to participate, they will be asked them to give verbal consent before doing so.

Second, collecting survey data comes with privacy concerns. To address this, we recruit participants and conduct the survey by phone, outside of working hours. We encourage participants to find a time and location when they can speak privately

before conducting the survey. We also ensure the confidentiality of any personal information acquired. The survey data will be stored on a password protected server and will not be shared with any factories or the buyer during the survey implementation. Third, most journals require the publication of such data to ensure transparency and replicability of research. Upon this research's publication in a peer-reviewed journal, we will publish de-identified data. To ensure effective anonymity, we will aggregate variables as needed to ensure individuals are not identifiable.

3   Empirical Analysis

3.1   Variables

3.1.1   Primary Outcomes

The key outcome variables are the reporting of severe labor issues, summarized as defined in Table 2.

3.1.2   Secondary outcomes

Additionally, as secondary outcomes, we study how increased reporting affects worker welfare in the short run. We measure well-being using an index of mental health survey questions, locus of control and job satisfaction. Further, we evaluate workers' beliefs about the data being leaked, their comprehension, and preferred reporting system. We also evaluate whether the survey method impacts the reported incidence of fatal accidents (expecting no effect).[2] We also

---

[2]Has there been an accident at your factory that killed 1 or more workers?

Table 2: Primary outcome variables

Variable Name	Variable Definition
Physical harassment	Your supervisor has taken one or more of the following actions toward you against your will: Hit, slapped, or punched; Cut or stabbed; Tripped; Otherwise intentionally caused physical harm.
Sexual harassment	Your supervisor has taken one or more of the following actions toward you against your will: Made remarks about you in a sexual manner; Asked you to enter into a love or sexual relationship; Asked or forced you to perform sexual favors; Asked or forced you to meet outside of the factory or meet them alone in a way that made you feel uncomfortable; Touched you in a sexual manner or in a way that made you feel uncomfortable or scared; Shown you pictures of sexual activities.

Notes: The final statement of the sexual harassment list of actions ("Shown you pictures of sexual activities") is randomized across participants with 50% probability.

3.2 Regression Model

3.2.1 Main regression model

$$\hat{r}_{is} = \alpha\, HG_i^{.2} + \beta\, HG_i^{.01} + \mu_s + \theta X_i + \epsilon_{is} \tag{1}$$

where $\hat{r}_{is}$ is the normalized intended reporting outcome of interest for individual i in stratum s.[3] $HG_i^{.2}$ and $HG_i^{.01}$ are hard-garbling treatment arms with different garbling rates respectively (equation 1).[4] We control for stratum fixed-effects μ_s. We will also present results including individuals' characteristics X_i (selected using the post double selection (PDS) lasso (**?**)). $\hat{r}_{is}$ is identified with a mean-zero heteroskedastic error term, which we address by reporting robust standard errors.

---

[3]See Boudreau et al. (2023) for an explanation of how we identify individuals' intended reports in the HG arms.

[4]If we do not find statistically significant differences between impacts of different garbling rates, we will also show results for the pooled regression, with HG_i as the hard-garbling treatment arm, and take this as baseline.

Controls: The stratum fixed effects absorb factory, production section or team and gender fixed effects. In addition, we plan to include other characteristics as controls using a lasso-based approach.

Standard error adjustments: As discussed in Boudreau et al. (2023), with blocked garbling, which is implemented for the HG20 arm, standard errors are negatively correlated within blocks and uncorrelated across blocks. Hence, we report standard errors clustered by HG block (HG20) or respondent (other arms).

3.2.2 Additional regression models

In addition, we are interested in the following regression models:

- Heterogeneous effects by gender of the respondent.

- Heterogeneous effects by level of comprehension of fractions.

- Heterogeneous effects by level of comprehension of HG mechanism.

- We are also interested in impacts of HG on secondary outcomes. We will use regression model 1 to estimate the reduced form effects of HG on worker-wellbeing and job satisfaction. For these outcomes, we will also run a 2SLS model with this regression as the first stage and predicted reports of harassment as the explanatory variable of interest in the second stage regression.

4   Research Team

The Principal investigators are:

Laura Boudreau, Columbia Business School

Sylvain Chassang, Princeton University

Ada González-Torres, Ben-Gurion University

Rachel Heath, University of Washington

References

AMITI, M. AND D. R. DAVIS (2012): "Trade, firms, and wages: Theory and evidence," *The Review of economic studies*, 79, 1–36.

ATKIN, D. (2009): "Working for the future: Female factory work and child health in Mexico," *Unpublished Manuscript, Yale University*, 150.

BEGUM, F., R. ALI, M. HOSSAIN, AND S. B. SHAHID (2010): "Harassment of women garment workers in Bangladesh," *Journal of the Bangladesh Agricultural University*, 8.

BLATTMAN, C. AND S. DERCON (2018): "The impacts of industrial and entrepreneurial work on income and health: Experimental evidence from Ethiopia," *American Economic Journal: Applied Economics*, 10, 1–38.

BLATTMAN, C., S. DERCON, AND S. FRANKLIN (2019): "Impacts of Industrial and Entrepreneurial Jobs on Youth: Five-Year Experimental Evidence on Factory Job Offers and Cash Grants in Ethiopia," *Working Paper No. 2019–65*.

BOUDREAU, L. E., S. CHASSANG, AND A. GONZÁLEZ-TORRES (2021): "Escrow Mechanisms for Group-based Reporting: Evidence from Bangladesh's Garment Secto," *OSF Registry https://osf.io/p8qfk*.

BOUDREAU, L. E., S. CHASSANG, A. GONZÁLEZ-TORRES, AND R. HEATH (2023): "Monitoring harassment in organizations," *National Bureau of Economic Research*.

CHASSANG, S. AND G. PADRÓ I MIQUEL (2018): "Crime, Intimidation, and Whistleblowing: A Theory of Inference from Unverifiable Reports," *Review of Economic Studies*, 86, 2530–2553.

CHASSANG, S. AND C. ZEHNDER (2024): "Secure survey design in organizations: Theory and experiments," *American Economic Journal: Microeconomics*, 16, 371–405.

FRÍAS, J. A., D. S. KAPLAN, AND E. VERHOOGEN (2012): "Exports and within-plant wage distributions: Evidence from Mexico," *American Economic Review*, 102, 435–440.

GIBBS, A., R. JEWKES, S. WILLAN, M. AL MAMUN, K. PARVIN, M. YU, AND R. NAVED (2019): "Workplace violence in Bangladesh's garment industry," *Social Science & Medicine*, 235, 112383.

HEATH, R. AND A. M. MOBARAK (2015): "Manufacturing growth and the lives of Bangladeshi women," *Journal of development Economics*, 115, 1–15.

KHOSLA, N. (2009): "The ready-made garments industry in Bangladesh: A means to reducing gender-based social exclusion of women?" *Journal of International Women's Studies*, 11, 289–303.

QIAN, N. (2008): "Missing women and the price of tea in China: The effect of sex-specific earnings on sex imbalance," *The Quarterly journal of economics*, 123, 1251–1285.

STANDING, G. (1999): "Global feminization through flexible labor: A theme revisited," *World development*, 27, 583–602.

SUBRAMANIAN, N. (2020): "Workplace Attributes and Women's Labor Supply Decisions: Evidence from a Randomized Experiment," .

VERHOOGEN, E. A. (2008): "Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector," *The Quarterly Journal of Economics*, 123, 489–530.

A   Appendix

A.1   Scripts for Direct Elicitation and Hard Garbling

Direct-Elicitation (DE):

We are now going to ask you several more questions about your factory. For instance: "Has a supervisor or manager shouted at you in the last 12 months? Yes or No?" Each of the questions has a Yes or No answer.

Once the surveying is completed, no personally identifying information, like names or phone numbers, is kept. Only the anonymous version of the data is kept.

Your answers will be recorded as you go, but we can chat about them before we record them for good.

The researchers and [the buyer] are only interested in learning about overall patterns from workers' responses. They can use the information from the survey to figure this out. Please be assured that your responses to the surveys will be kept as confidential as possible.

HG Script:

We are now going to ask you several more questions about your factory. For instance: "Has a supervisor or manager shouted at you in the past year? Yes or No?" Each of the questions has a Yes or No answer.

Once the surveying is completed, no personally identifying information, like names or phone numbers, is kept. Only the anonymous version of the data is kept.

In addition, our system is setup so that if you respond YES, aside from me, no one will ever be able to know that this was your answer, not even the researchers.

Here is how our system works:

If you choose to respond YES, there is an issue, our system will record it as a YES for sure. Importantly, if someone responds NO, the system will sometimes record the response as YES. Our system is set-up such that:

For every 100 interviews with workers, our system will assign by lottery a YES to at least 20 (1) of these responses.

This means that if you respond YES, we can guarantee that you won't be the only person saying YES.

The researchers and [the buyer] are only interested in learning about overall patterns from workers' responses. They can use the information from the system to figure this out without knowing for sure who actually responded yes. Please be assured that your responses to the surveys will be kept as confidential as possible.

A.2 Power calculation assumptions

Assumptions:

- Power: $\beta = 0.9$.

- Significance level: We set $\alpha = 0.10$ and conservatively apply the Bonferroni adjustment with 3 primary outcomes of $\frac{0.10}{2} = 0.025$.

- Within-factory component of the variance (residual variance of outcome estimated from a regression with factory FE and demographic controls) = 0.008 for both physical and sexual harassment.

- Share of sample in treatment group: 2/3.

- Hypothesis test: 1-sided. This is justified based on the theoretical prediction that HG increases willingness to report compared to DE.