Web Appendix for "Empowering Inclusive Work"

by Yanyou Chen, Mitchell Hoffman, Huilan Xu, and Zhe Yuan

The Web Appendix is organized as follows. Appendix A provides additional figures and tables (with a guide at the start). Appendix B presents additional analysis: effects on good ratings (Section B.1), DHH worker entry (Section B.2), and a mediation analysis (Section B.3). Appendix C describes hearing loss levels in China. Appendix D discusses potential nondisclosure of DHH status. Appendix E provides details on the AI tool. Appendix F presents qualitative data on DHH workers.

A Appendix Figures and Tables

Organization. Appendix A figures and tables are grouped into six themes. The themes roughly follow the order of the main text.

Gaps between DHH and non-disabled workers in the US	
Table A.1	DHH employment and wage gaps, U.S. labor market
Summary statistics	
Tables A.2–A.4	By DHH status; by work model (gig vs. team); by DHH severity
Figure A.1	Raw means over time, DHH vs. non-disabled workers
Figure A.2	Missingness rates for hours data over time
Pre-AI differences in outcomes (Section 2)	
Figure A.3	Pre-AI differences by work model
Figure A.4	Pre-AI differences by severity of hearing loss
Table A.5	Pre-AI differences in tabular form
AI treatment effects and event studies (Section 3.2)	
Figure A.5	Main DiD estimates by work model
Table A.6	Extended DiD results: subsamples and specifications
Figure A.6	Event studies by work model
Figure A.7	Event study extended through month 15 post-adoption (months 0–15)
Heterogeneous treatment effects of the AI tool (Section 3.4)	
Figure A.8	By pre-AI tenure
Figure A.9	By pre-AI productivity
Figure A.10	Robustness on HTE by disability severity: dosage-response model
Figure A.12	By regional unemployment
Figures A.11, A.13	Robustness on HTE by disability severity and regional unemployment: controlling for interactions with age and gender
Robustness checks on impact of AI tool (Section 3.5)	
Figure A.14	Hours sample, matched sample, and additional control variables
Figure A.15	Differential attrition
Figure A.16	Longer-run effects and external validity
Tables A.7, A.8, A.9	Additional results and robustness: bad ratings; late deliveries; attrition
Figure A.17	Tenure-productivity curves for bad ratings, counts instead of ratings per 10k orders

Table A.1: DHH Employment and Wage Gaps in the U.S. General Labor Market
(CPS ASEC 2008–2023)

	(1) No controls	(2) Demog. + Year FE	(3) Demog. + Year + State FE	(4) Demog. + Year + State + Occ. + Ind. FE
Panel A: Employment (Linear Probability Model)				
DHH	-0.2665^{***}	-0.2445^{***}	-0.2454^{***}	—
	(0.0041)	(0.0039)	(0.0039)	
Observations			1,549,931	
Panel B: Log Hourly Wage (OLS, conditional on employment)				
DHH	0.0000	-0.0619^{***}	-0.0567^{***}	-0.0369^{***}
	(0.0086)	(0.0077)	(0.0077)	(0.0069)
Observations			985,211	

Notes: CPS ASEC microdata from 2008–2023, accessed via IPUMS (Flood et al., 2024); sample restricted to working-age adults aged 18–64. Each observation is a worker (one per person per survey year). DHH equals 1 if the respondent reports hearing difficulty (`DIFFHEAR`), and 0 for the non-disabled comparison group (those reporting no difficulty on any of the six CPS disability questions, `DIFFANY` = No). Panel A estimates a linear probability model for employment. In Panel B, hourly wages are constructed as annual wage and salary income divided by weeks worked times usual weekly hours; the top and bottom 1% of the weighted hourly wage distribution are trimmed before taking logs. Column (1) has no controls; column (2) adds age, age squared, gender, education categories, race/ethnicity indicators, and year fixed effects; column (3) further adds state fixed effects; column (4) further adds occupation and industry fixed effects. Column (4) is omitted for Panel A because industry is only observed for employed workers, making industry fixed effects inapplicable in an employment regression. All regressions are weighted by CPS ASEC person weights (`ASECWT`) to produce population-representative estimates. The near-zero unadjusted wage gap in Panel B, Column (1) reflects DHH workers being substantially older on average (mean age 49.7 vs. 40.2 for non-disabled workers); once age is controlled for, a sizable gap emerges. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Summary Statistics for Key Outcomes at Worker-week Level: DHH vs. Non-disabled Workers

	DHH Workers				Non-disabled Workers			
	Mean (1)	Median (2)	SD (3)	Obs. (4)	Mean (5)	Median (6)	SD (7)	Obs. (8)
Panel A: Efficiency/productivity: Quantity and Quality								
Order Acceptance Time, in minutes	3.96	2.98	10.34	69,444	3.27	2.38	8.97	395,273
Pickup Time, in minutes	11.47	10.94	4.13	69,462	12.05	11.55	7.23	395,388
Delivery Time, in minutes	14.42	13.85	4.93	69,498	14.14	13.36	5.13	396,048
Late Deliveries per 10k orders	426.81	147.06	811.01	70,301	414.67	131.93	817.34	409,769
Bad Ratings per 10k orders	11.53	0.00	83.23	70,301	9.30	0.00	95.88	409,769
Week with a Bad Rating, 0 or 1	0.09	0.00	0.29	70,301	0.07	0.00	0.26	409,769
Panel B: Labor Supply on the Platform								
Hours Online, in hours per week	45.02	46.07	29.05	54,984	39.54	36.16	29.13	288,749
Work Days per week	5.57	7.00	2.18	70,301	5.26	6.00	2.12	409,769
Driver attrites from Platform, 0 or 1	0.05	0.00	0.22	66,622	0.08	0.00	0.27	389,570
Panel C: Total Output								
Total Orders	161.10	141.00	126.66	70,301	150.90	124.00	143.86	409,769
Peak Hour Orders	96.19	85.00	77.27	70,301	90.02	74.00	90.16	409,769
Panel D: Income, Wages, and Profits								
Driver Income, in CNY per week	Confid.	Confid.	Confid.	70,301	Confid.	Confid.	Confid.	409,769
Hourly Wages	Confid.	Confid.	Confid.	54,675	Confid.	Confid.	Confid.	287,386
Platform Profits from a driver	Confid.	Confid.	Confid.	70,301	Confid.	Confid.	Confid.	409,769
Workers	4,284				39,321			

Notes: This table is similar to Table 2 in the main text, but provides summary statistics separately for DHH workers and for non-disabled workers. Summary statistics use the full sample period, including both pre- and post-AI observations.

Table A.3: Summary Statistics Comparing Gig vs. Team Workers

	Gig Workers				Team Workers			
	Mean (1)	Median (2)	SD (3)	Obs. (4)	Mean (5)	Median (6)	SD (7)	Obs. (8)
Panel A: Efficiency/productivity: Quantity and Quality								
Order Acceptance Time, in minutes	4.17	3.29	3.65	206,131	2.74	2.01	11.85	258,586
Pickup Time, in minutes	11.06	10.65	3.55	206,201	12.69	12.07	8.57	258,649
Delivery Time, in minutes	16.69	16.24	5.45	206,237	12.19	11.84	3.76	259,309
Late Deliveries per 10k orders	517.33	140.85	981.70	207,084	339.91	132.16	653.86	272,986
Bad Ratings per 10k orders	15.60	0.00	135.09	207,084	5.09	0.00	41.15	272,986
Week with a Bad Rating, 0 or 1	0.07	0.00	0.26	207,084	0.07	0.00	0.26	272,986
Panel B: Labor Supply on the Platform								
Hours Online, in hours per week	22.83	15.55	21.54	172,695	58.17	61.83	24.83	171,038
Work Days per week	4.25	4.00	2.23	207,084	6.11	7.00	1.65	272,986
Driver attrites from Platform, 0 or 1	0.12	0.00	0.32	195,834	0.05	0.00	0.21	260,358
Panel C: Total Output								
Total Orders	73.35	42.00	85.28	207,084	212.35	205.00	146.18	272,986
Peak Hour Orders	42.87	22.00	54.47	207,084	127.38	125.00	91.75	272,986
Panel D: Income, Wages, and Profits								
Driver Income, in CNY per week	Confid.	Confid.	Confid.	207,084	Confid.	Confid.	Confid.	272,986
Hourly Wages	Confid.	Confid.	Confid.	171,184	Confid.	Confid.	Confid.	170,877
Platform Profits from a driver	Confid.	Confid.	Confid.	207,084	Confid.	Confid.	Confid.	272,986
Panel E: Worker Characteristics								
Age		30.8				30.1		
Male		90.2%				92.3%		
Driver is in a big city		34.8%				26.4%		
Workers		27,777				15,828		
DHH Workers		2,769				1,515		
Non-disabled Workers		25,008				14,313		

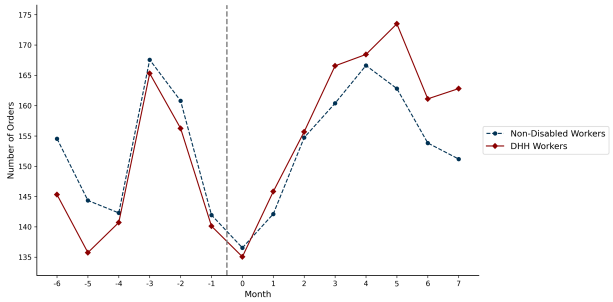
Notes: Panels A–D are similar to Table 2 in the main text, reporting worker-week-level summary statistics separately for gig and team workers. The "Confid." marking indicates that certain statistics cannot be shared due to confidentiality requirements. Panel E reports worker-level characteristics, similar to Table 1 in the main text, averaged across workers within each group. Summary statistics use the full sample period, including both pre- and post-AI observations.

Table A.4: Summary Statistics Comparing Profound DHH Workers vs. Non-Profound DHH Workers

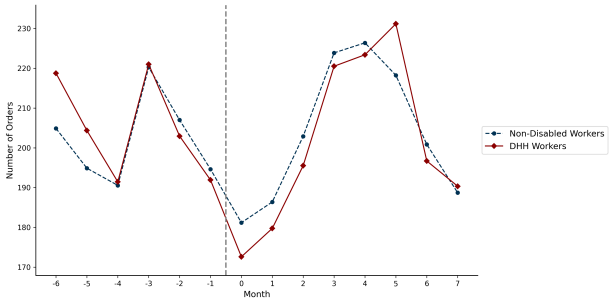
	Profound DHH Workers				Non-profound DHH Workers			
	Mean (1)	Median (2)	SD (3)	Obs. (4)	Mean (5)	Median (6)	SD (7)	Obs. (8)
Panel A: Efficiency/productivity: Quantity and Quality								
Order Acceptance Time, in minutes	3.72	3.11	2.97	34,218	4.19	2.85	14.22	35,226
Pickup Time, in minutes	11.69	11.15	3.74	34,224	11.25	10.75	4.46	35,238
Delivery Time, in minutes	14.98	14.30	4.83	34,244	13.87	13.33	4.97	35,254
Late Deliveries per 10k orders	448.67	161.29	827.42	34,444	405.80	132.45	794.38	35,857
Bad Ratings per 10k orders	12.29	0.00	91.38	34,444	10.80	0.00	74.55	35,857
Week with a Bad Rating, 0 or 1	0.10	0.00	0.30	34,444	0.08	0.00	0.28	35,857
Panel B: Labor Supply on the Platform								
Hours Online, in hours per week	46.47	50.20	28.88	26,430	43.68	41.98	29.15	28,554
Work Days per week	5.52	7.00	2.08	34,444	5.63	7.00	2.26	35,857
Driver attrites from Platform, 0 or 1	0.06	0.00	0.23	32,759	0.04	0.00	0.21	33,863
Panel C: Total Output								
Total Orders	170.65	160.00	126.29	34,444	151.92	122.00	126.35	35,857
Peak Hour Orders	104.88	100.00	78.88	34,444	87.85	71.00	74.75	35,857
Panel D: Income, Wages, and Profits								
Driver Income, in CNY per week	Confid.	Confid.	Confid.	34,444	Confid.	Confid.	Confid.	35,857
Hourly Wages	Confid.	Confid.	Confid.	25,068	Confid.	Confid.	Confid.	27,207
Platform Profits from a driver	Confid.	Confid.	Confid.	34,444	Confid.	Confid.	Confid.	35,857
Panel E: Worker Characteristics								
Age		31.1				34.9		
Male		90.4%				95.0%		
Driver is in a big city		42.2%				31.4%		
Workers		2,295				1,989		

Notes: Panels A–D are similar to Table 2 in the main text, reporting worker-week-level summary statistics separately for DHH workers with profound hearing loss (Level 1; >90 dB) and non-profound hearing loss (Levels 2–4), who together constitute the full DHH sample. The "Confid." marking indicates that certain statistics cannot be shared due to confidentiality requirements. Panel E reports worker-level characteristics, similar to Table 1 in the main text, averaged across workers within each group. Summary statistics use the full sample period, including both pre- and post-AI observations.

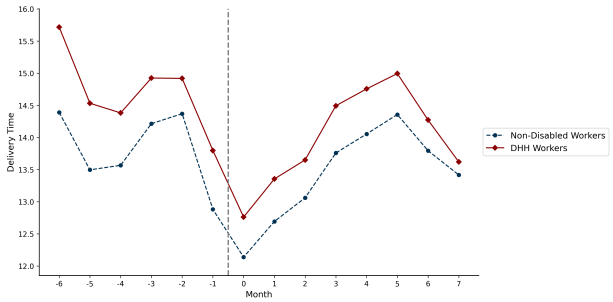
Figure A.1: Comparison of Raw Means Between DHH and Non-disabled Workers



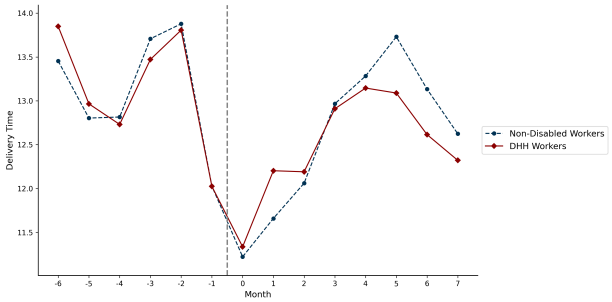
(a) Orders Completed (All Workers)



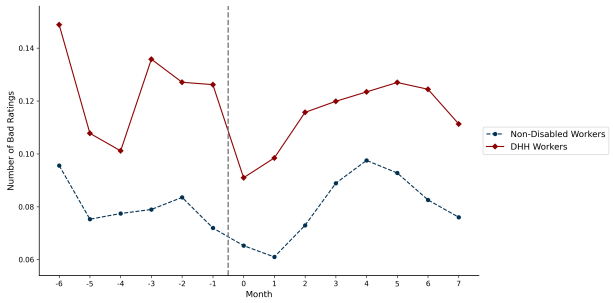
(b) Orders Completed (Balanced Panel)



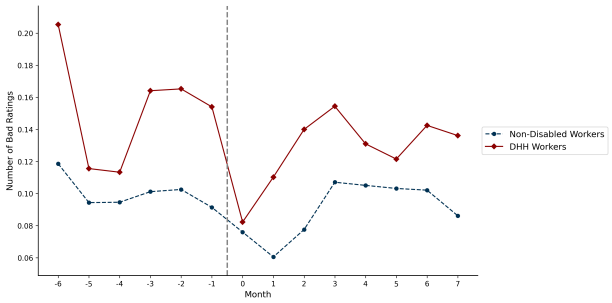
(c) Delivery Time (All Workers)



(d) Delivery Time (Balanced Panel)



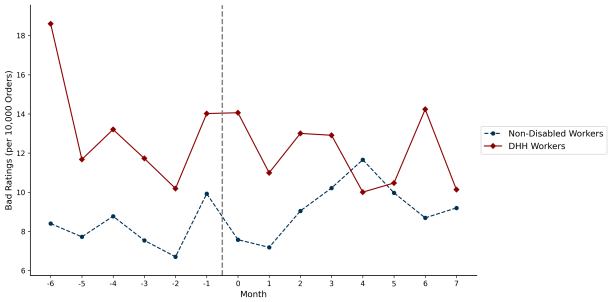
(e) Bad Ratings (Count) (All Workers)



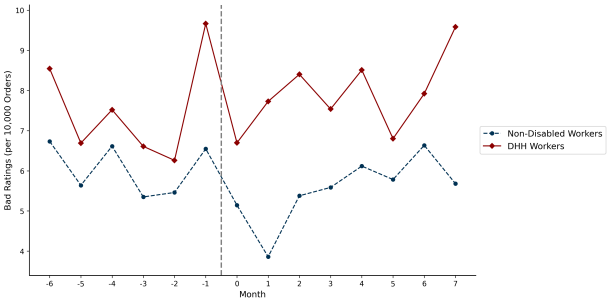
(f) Bad Ratings (Count) (Balanced Panel)

Figure continues on the next page.

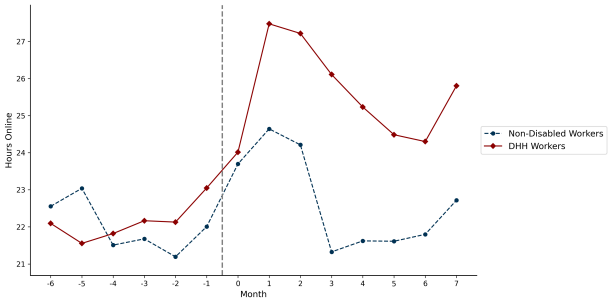
Figure A.1: Comparison of Raw Means Between DHH and Non-disabled Workers (*continued*)



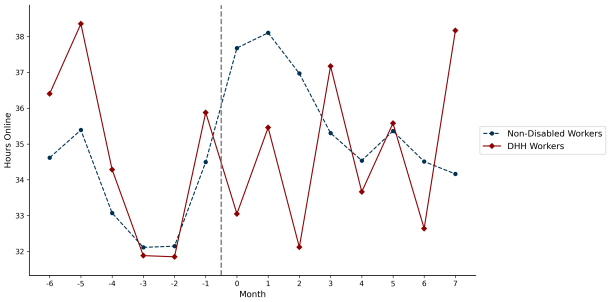
(g) Bad Ratings (per 10,000 Orders) (All Workers)



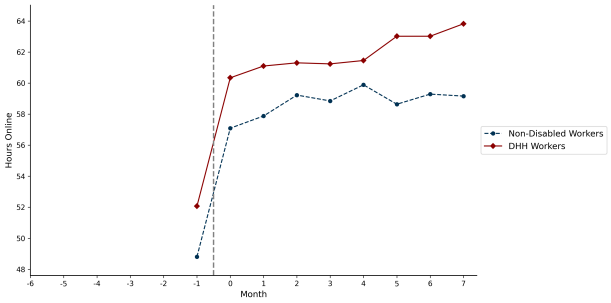
(h) Bad Ratings (per 10,000 Orders) (Balanced Panel)



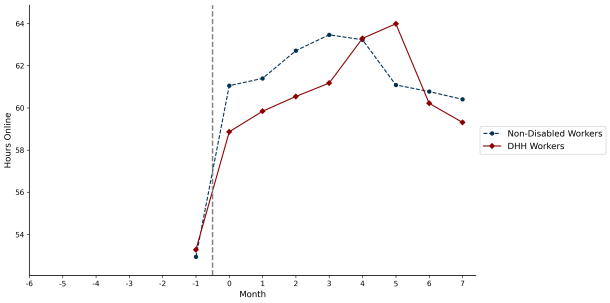
(i) Hours Online – Gig Workers (All Workers)



(j) Hours Online – Gig Workers (Balanced Panel)



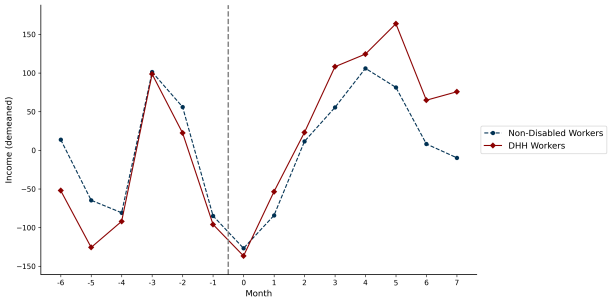
(k) Hours Online – Team Workers (All Workers)



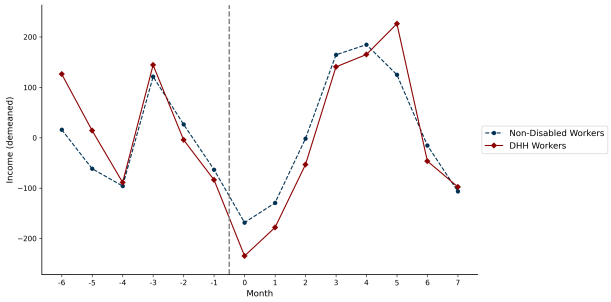
(l) Hours Online – Team Workers (Balanced Panel)

Figure continues on the next page.

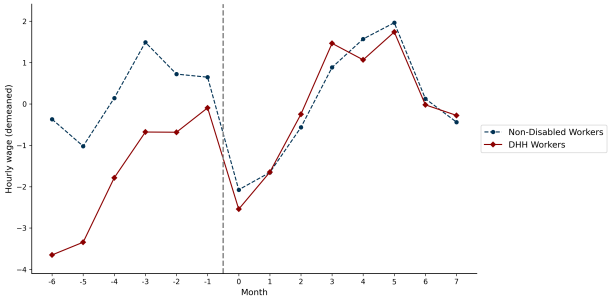
Figure A.1: Comparison of Raw Means Between DHH and Non-disabled Workers (*continued*)



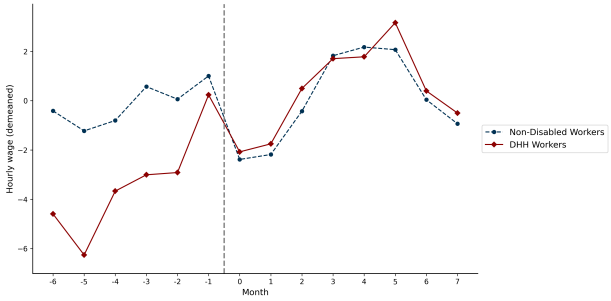
(m) Income
(All Workers)



(n) Income
(Balanced Panel)



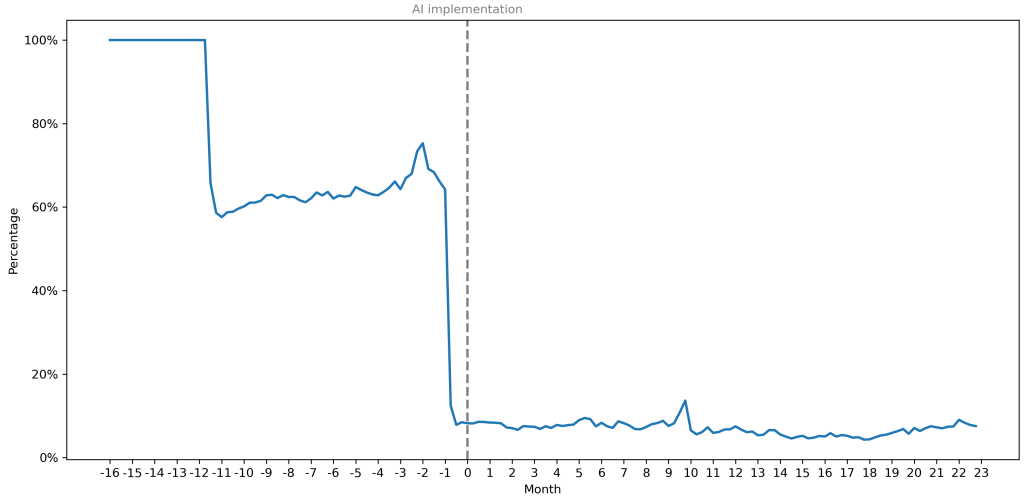
(o) Hourly Wage
(All Workers)



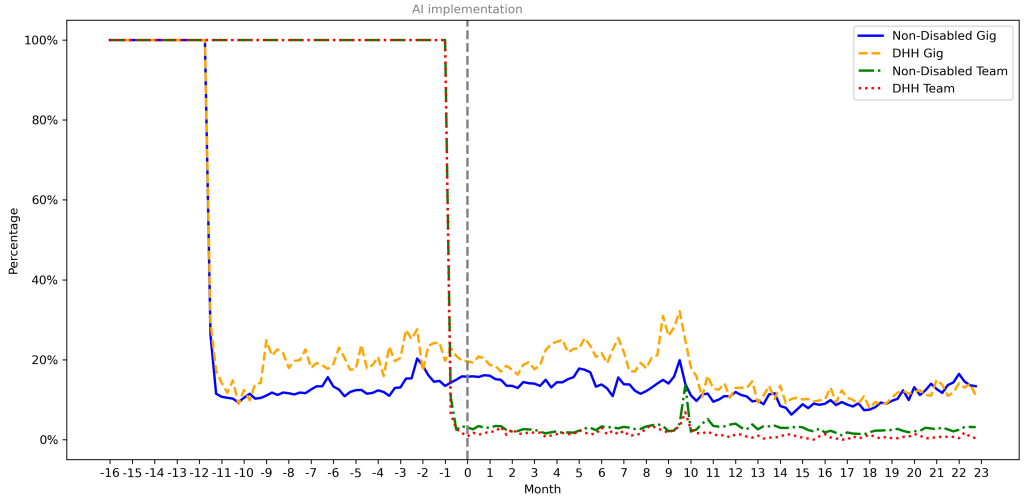
(p) Hourly Wage
(Balanced Panel)

Notes: This figure compares raw means of outcomes for DHH workers (red) and non-disabled workers (dark blue, dashed). The left panels show results for all workers. There are 43,605 workers in total, including 4,284 DHH workers and 39,321 non-disabled workers. The right panels show results for workers who remained continuously active on the platform (i.e., did not exit) from month -6 before the AI implementation to month 7 after the AI implementation. This sample includes 779 workers in total, with 107 DHH workers and 672 non-disabled workers. Bad ratings are shown in two ways (following footnote 20 in the main text) as a weekly count (panels (e) and (f)), consistent with the main analysis, and as bad ratings per 10,000 orders (panels (g) and (h)), which normalizes by order volume. Since these are raw means, we show Hours Online separately for gig and team workers (panels (i)–(l)). Hours data become available for gig workers at week $t_0 - 46$ and for team workers only at week $t_0 - 4$, where t_0 is the week of AI implementation (see Figure A.2). Showing the two groups in a single pooled series would therefore conflate differences in outcome levels with differences in coverage periods. This is not a concern for our DiD analyses, where work model fixed effects ensure the comparison is always made within work model. The apparent increase in hours around t_0 for team workers (panels (k) and (l)) reflects that hours data are only available for the final few weeks of month -1, so the pre-AI data point is an average over an incomplete month rather than a true behavioral change. For confidentiality, the income variable (panels (m) and (n)) and the hourly wage variable (panels (o) and (p)) are demeaned using the sample average, with the same value subtracted for both DHH and non-disabled workers. Income is in CNY per week and hourly wage is in CNY per hour.

Figure A.2: The Share of Worker-Weeks with Missing Hours Data Over Time



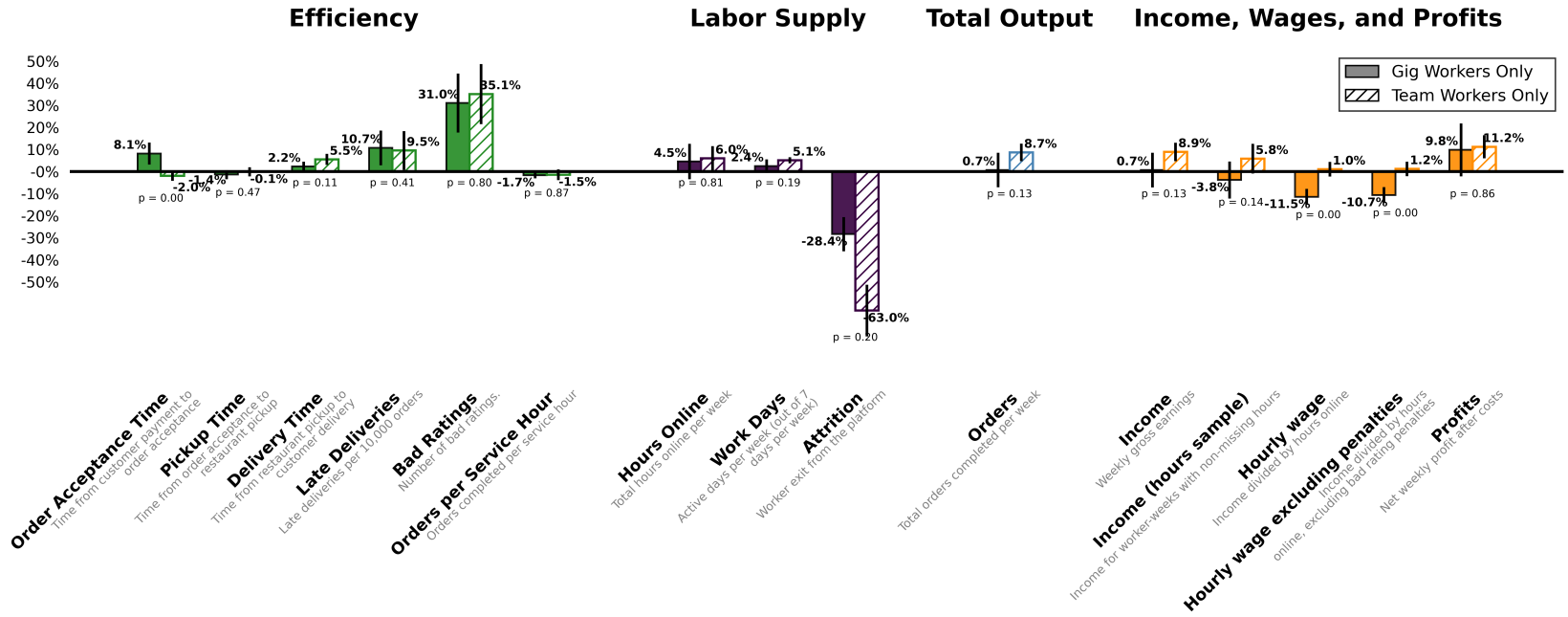
(a) Share of worker-weeks with missing hours over time



(b) Share of worker-weeks with missing hours, split by work model and DHH status

Notes: This figure shows the share of worker-weeks with missing hours data over time. The dotted vertical line shows when the AI tool is implemented. Panel (a) shows the overall share of missing hours data over time. As seen in panel (b), which shows missing rate broken down by work model (gig or team) and DHH status, hours data become available for gig workers at week $t_0 - 46$, and then for team workers at week $t_0 - 4$, where t_0 is the week of AI implementation.

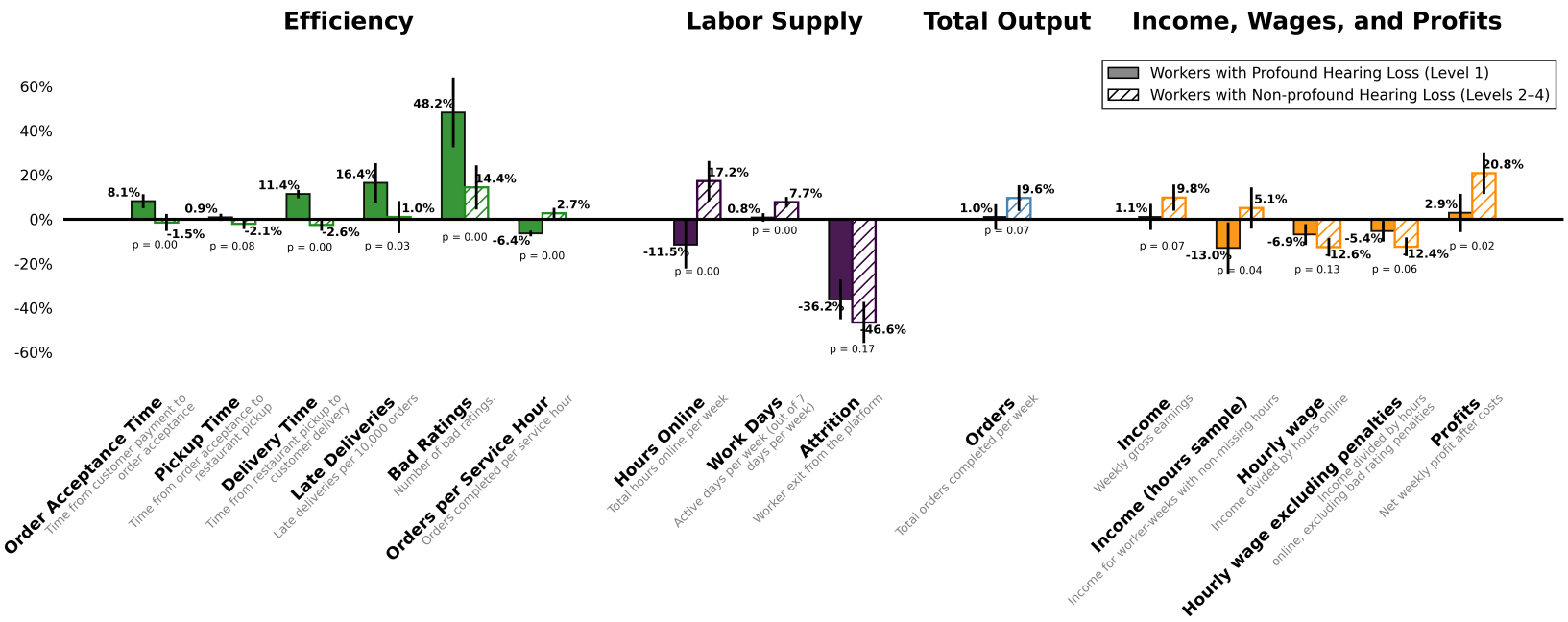
Figure A.3: Pre-AI Differences Between DHH and Non-disabled Workers: Gig and Team Workers Separately



Notes: This figure reproduces the analysis in Figure 1, but splits the sample by gig workers and team workers. As discussed in the main text, gig workers operate individually, exercise full discretion over their working hours, and decide which orders to accept. We plot pre-AI differences for DHH gig workers vs. non-disabled gig workers, as well as pre-AI differences for DHH team workers vs. non-disabled team workers. Each bar is estimated from a separate regression in the corresponding subsample (gig or team workers). The p-values shown are from tests of equality between gig and team workers. For example, looking at orders, among gig workers, DHH workers do 0.7% more orders than non-disabled workers, whereas among team workers, DHH workers do 8.7% more orders than non-disabled workers, and the difference is not statistically significant ($p = 0.13$).

The p-values are based on pooled regressions with interaction terms that test equality of the underlying OLS coefficients across groups. For outcomes estimated using IHS transformation (orders, delivery time, hours online, income), the tests are conducted on IHS coefficients; for outcomes estimated in levels (late deliveries, bad ratings, and attrition), the tests are conducted on level coefficients. Because the plotted coefficients are scaled by group-specific pre-AI means, the p-values need not correspond exactly to visual differences in the bars.

Figure A.4: Pre-AI Differences Between DHH and Non-disabled Workers: Splitting by Severity of Disability



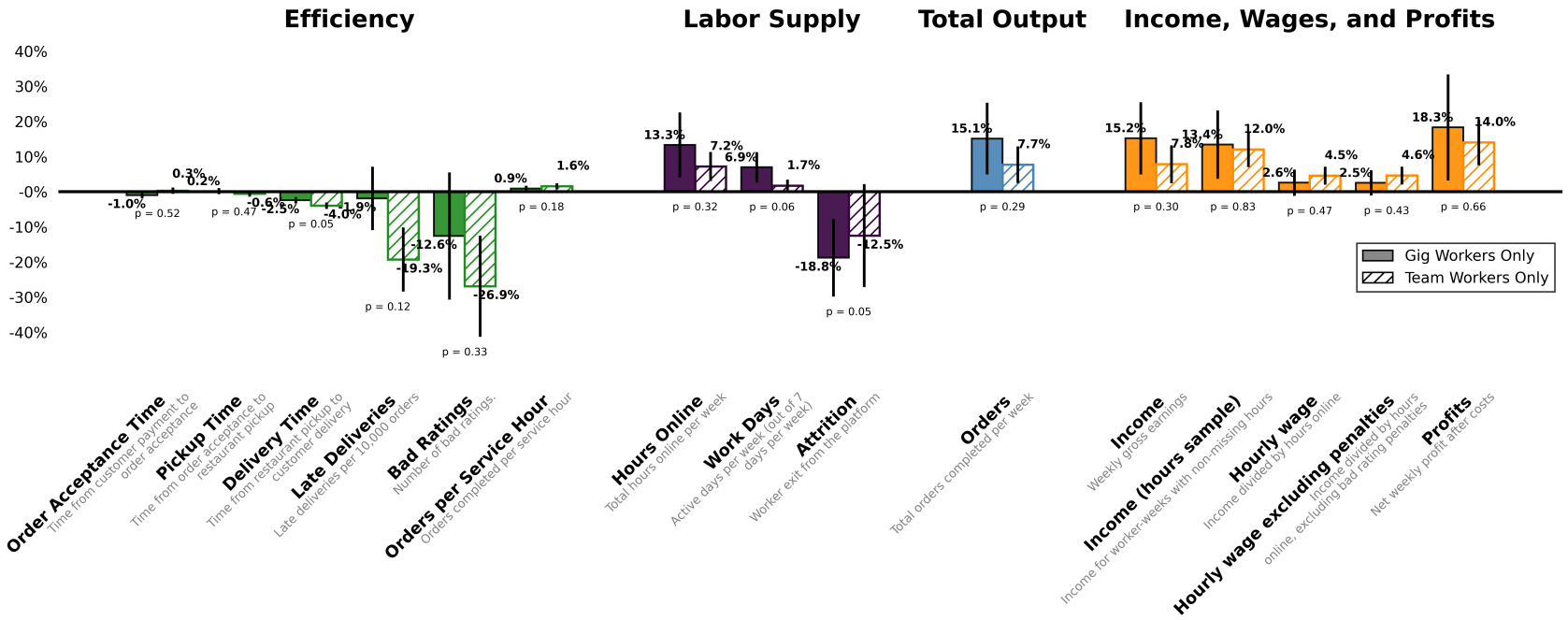
Notes: This figure reproduces the analysis in Figure 1, but splits DHH differences by severity of disability. As discussed in Section 1 of the main text, profound hearing loss (Level 1 in the Chinese system) is loss of more than 90 decibels. Non-profound hearing loss is loss of 41-90 decibels. The p-values shown are from tests of equality between profound and non-profound HL workers. We divide the raw coefficients by the pre-AI DHH mean within each severity level to obtain the percentages shown.

Table A.5: Pre-AI Differences Between DHH and Non-disabled Workers

	(1)	(2)	(3)	(4)
	Baseline	Only Gig	Only Team	Control for Avg. Distance from Restaurant to Customer
Panel A: Efficiency				
Order Acceptance Time	0.034**	0.081***	-0.020	0.032**
	(0.016)	(0.030)	(0.014)	(0.015)
Pickup Time	-0.006	-0.014	-0.001	-0.007
	(0.009)	(0.013)	(0.013)	(0.009)
Delivery Time	0.045***	0.022	0.055***	0.041***
	(0.011)	(0.014)	(0.015)	(0.010)
Late Deliveries	0.087**	0.107**	0.095*	0.087**
	(0.035)	(0.048)	(0.053)	(0.035)
Bad Ratings	0.313***	0.310***	0.351***	0.313***
	(0.058)	(0.081)	(0.083)	(0.058)
Orders per Service Hour	-0.019**	-0.017**	-0.015	-0.017**
	(0.009)	(0.008)	(0.015)	(0.009)
Panel B: Labor Supply				
Hours Online	0.049	0.045	0.060*	0.049
	(0.044)	(0.049)	(0.034)	(0.044)
Work Days	0.043***	0.024	0.051***	0.043***
	(0.010)	(0.019)	(0.008)	(0.010)
Attrition	-0.414***	-0.284***	-0.630***	-0.414***
	(0.041)	(0.047)	(0.071)	(0.041)
Panel C: Total Output				
Orders	0.053**	0.007	0.087***	0.053**
	(0.026)	(0.048)	(0.024)	(0.026)
Panel D: Income, Wages, and Profits				
Income	0.054**	0.007	0.089***	0.054**
	(0.026)	(0.048)	(0.026)	(0.026)
Wage rate no penalty	-0.094***	-0.107***	0.012	-0.080***
	(0.020)	(0.022)	(0.020)	(0.019)
Wage rate with penalty	-0.102***	-0.115***	0.010	-0.088***
	(0.020)	(0.022)	(0.020)	(0.019)
Profits	0.118***	0.098	0.112***	0.118***
	(0.040)	(0.073)	(0.032)	(0.040)

Notes: Standard errors clustered by worker in parentheses. Each coefficient here is from a separate regression. Each regression follows the procedure in Section 2.1. For all outcomes besides late deliveries, bad ratings, and attrition, we use an inverse-hyperbolic sine transform on y. For late deliveries, bad ratings, and attrition, the coefficients shown here are the OLS regression coefficients divided by the pre-AI DHH mean, so that the coefficients here can be interpreted as percentage differences. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Column 1 is the baseline specification using the full panel. Column 2 restricts the sample to gig workers only. Column 3 restricts the sample to team workers only. Column 4 uses the same specification as Column 1 but additionally controls for the average distance from restaurant to customer, where average distance is measured at the worker-week level.

Figure A.5: Impact of the AI Tool on DHH Worker Outcomes: Gig and Team Workers Separately



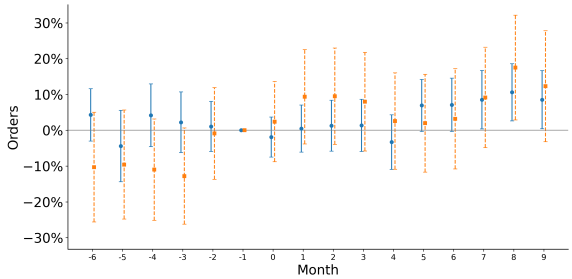
Notes: This figure reproduces the analysis in Figure 3, focusing exclusively on gig or team workers. As discussed in the main text, gig workers operate individually, exercise full discretion over their working hours, and decide which orders to accept. The p-values shown are from tests of equality between gig and team workers in the impact of the AI tool. For example, the AI tool raises the orders of gig workers by 15.1%, raises the orders of team workers by 7.7%, and the difference is statistically insignificant ($p = 0.29$).

Table A.6: The Impact of the AI Tool on the Outcomes of DHH Workers: Extended Results

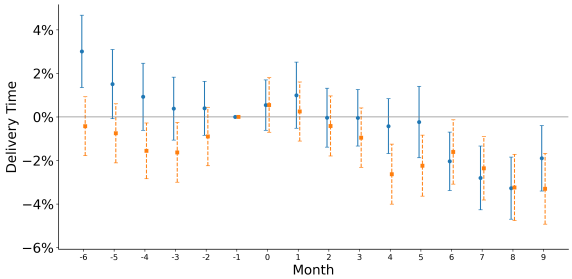
	Baseline (1)	Gig Workers (2)	Team Workers (3)	Exclude First 4 Months Post-AI (4)	Workers Present Both Pre-AI & Post-AI (5)	Restrict to Observations with Hours Data (6)	Control for Avg. Distance from Restaurant to Customer (7)
Panel A: Efficiency							
Order Acceptance Time	-0.008**	-0.010*	0.003	-0.010*	-0.006	-0.001	-0.009**
	(0.004)	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)
Pickup Time	-0.004	0.002	-0.006	-0.005	-0.006	0.001	-0.005
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Delivery Time	-0.035***	-0.025***	-0.040***	-0.051***	-0.035***	-0.026***	-0.037***
	(0.004)	(0.006)	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)
Late Deliveries	-0.112***	-0.019	-0.193***	-0.148***	-0.135***	-0.011	-0.112***
	(0.038)	(0.055)	(0.055)	(0.044)	(0.045)	(0.036)	(0.038)
Bad Ratings	-0.218***	-0.126	-0.269***	-0.252***	-0.197***	-0.086	-0.218***
	(0.068)	(0.110)	(0.088)	(0.083)	(0.063)	(0.073)	(0.068)
Orders per Service Hour	0.016***	0.009*	0.016***	0.021***	0.015***	0.009**	0.017***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Panel B: Labor Supply							
Hours Online	0.106***	0.133**	0.072***	0.136***	0.100***	0.106***	0.106***
	(0.034)	(0.056)	(0.025)	(0.046)	(0.035)	(0.034)	(0.034)
Work Days	0.035***	0.069***	0.017	0.040***	0.033***	0.051***	0.035***
	(0.012)	(0.026)	(0.011)	(0.015)	(0.012)	(0.015)	(0.012)
Attrition	-0.210***	-0.188***	-0.125	-0.193***	Not app-	-0.119**	-0.210***
	(0.052)	(0.067)	(0.089)	(0.054)	licable	(0.053)	(0.052)
Panel C: Total Output							
Orders	0.105***	0.151**	0.077**	0.129***	0.101***	0.128***	0.105***
	(0.031)	(0.062)	(0.032)	(0.038)	(0.031)	(0.036)	(0.031)
Panel D: Income, Wages, and Profits							
Income	0.106***	0.152**	0.078**	0.132***	0.102***	0.130***	0.106***
	(0.031)	(0.062)	(0.033)	(0.039)	(0.032)	(0.036)	(0.031)
Wage rate no penalty	0.036***	0.025	0.046***	0.033*	0.035**	0.036***	0.036***
	(0.014)	(0.022)	(0.015)	(0.019)	(0.014)	(0.014)	(0.014)
Wage rate with penalty	0.036**	0.026	0.046***	0.027	0.034**	0.036**	0.036**
	(0.014)	(0.023)	(0.016)	(0.019)	(0.015)	(0.014)	(0.014)
Profits	0.168***	0.183**	0.140***	0.244***	0.184***	0.200***	0.168***
	(0.043)	(0.092)	(0.039)	(0.051)	(0.040)	(0.058)	(0.043)
Maximum worker-weeks in column	480,070	207,084	272,986	427,612	184,450	343,733	480,070
Workers	43,605	27,777	15,828	40,821	3,397	34,965	43,605

Notes: This table provides extended results on the impact of the AI tool on DHH workers. Each coefficient reported is from a separate DiD regression. Standard errors clustered by worker in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. N varies by outcome; see Table 2 for outcome-specific observations. Column 1 is the baseline specification that appears in Figure 3. Column 2 restricts the sample to gig workers only. Column 3 restricts the sample to team workers only. Column 4 excludes the first four months of data after the introduction of the AI tool. Column 5 restricts the sample to workers who have at least one week of data both pre- and post-AI; we do not analyze attrition in column 5 because we are analyzing a sample that is present both before and after the AI. Column 6 restricts to observations with hours data. Column 7 adds a control for average distance between restaurant and customer, where average distance is measured at the worker-week level.

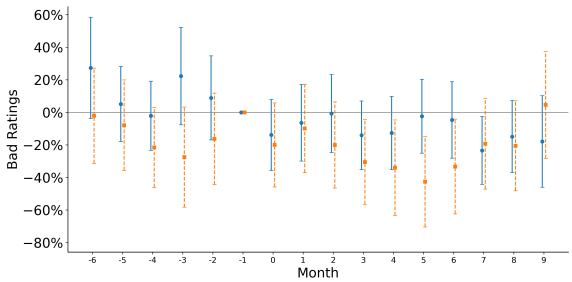
Figure A.6: Event Study Analysis: Splitting by Gig vs. Team Workers



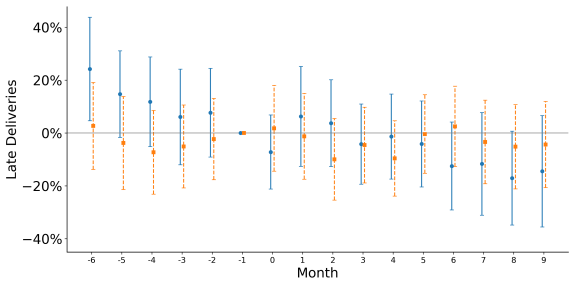
(a) Number of Orders Completed



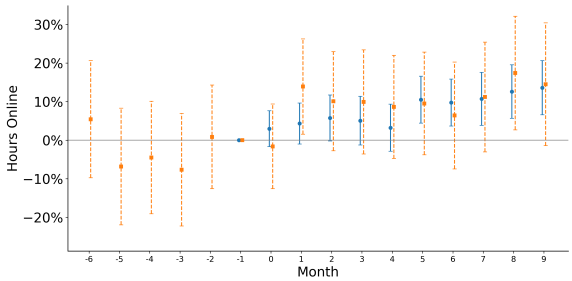
(b) Delivery Time



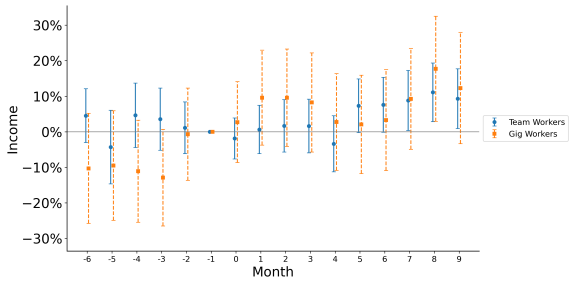
(c) Bad Ratings



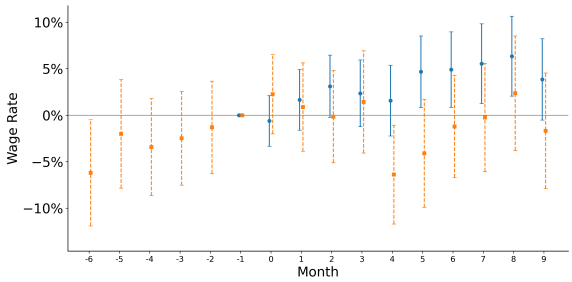
(d) Late Deliveries



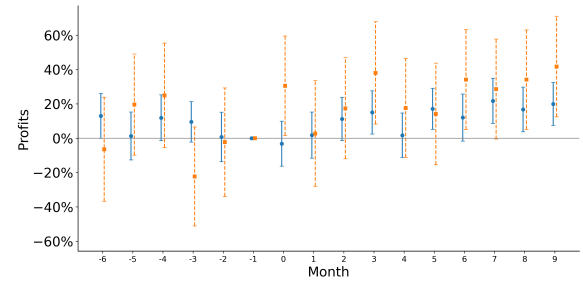
(e) Hours Worked



(f) Income



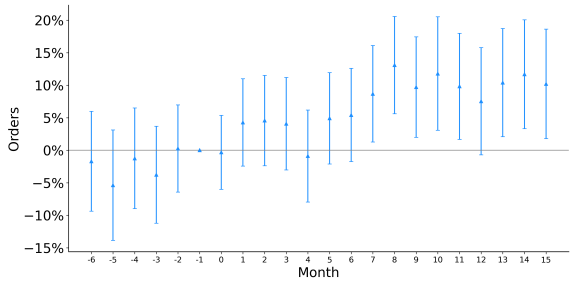
(g) Hourly Wage



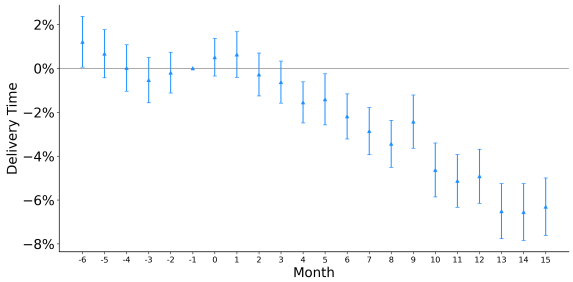
(h) Profit

Notes: This figure replicates the event studies in Figure 4 and Figure 5, but conducts the analysis separately for team workers and gig workers. Dots show the event-study coefficients, and whiskers show 95% confidence intervals clustered at the worker level. The blue dots are for team workers, and the orange squares are for gig workers. Hours worked data are missing for team workers prior to 1 month before the AI tool introduction.

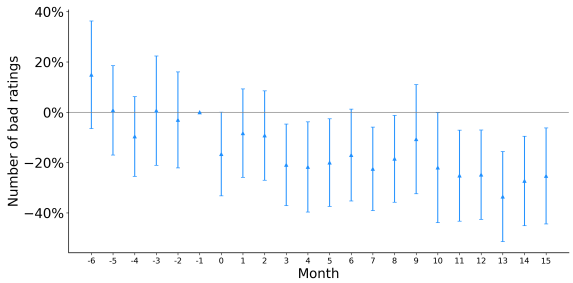
Figure A.7: Event Study Analysis: Robustness to Extended Post-Period



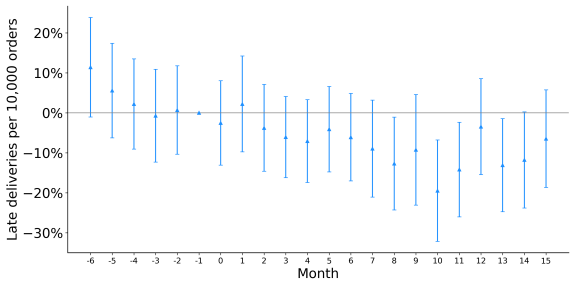
(a) Number of Orders Completed



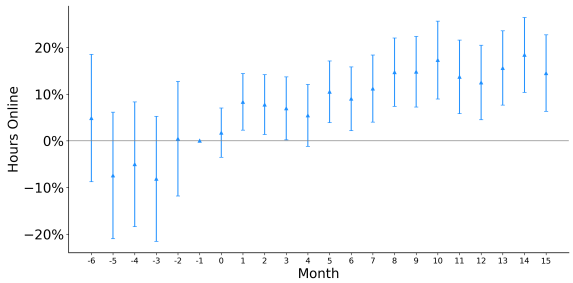
(b) Delivery Time



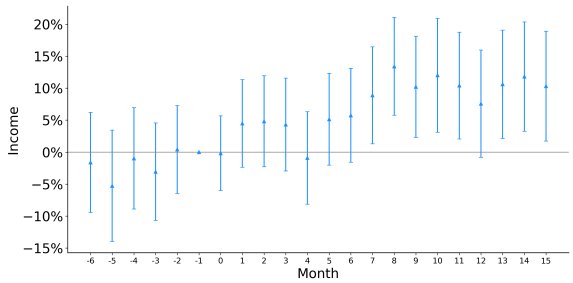
(c) Bad Ratings



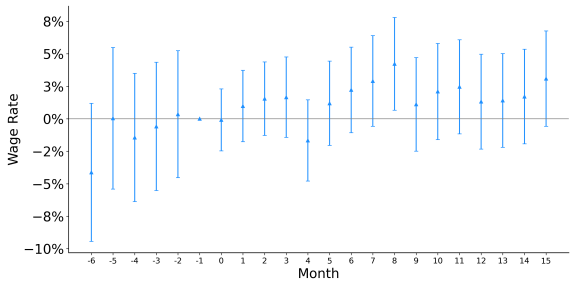
(d) Late Deliveries



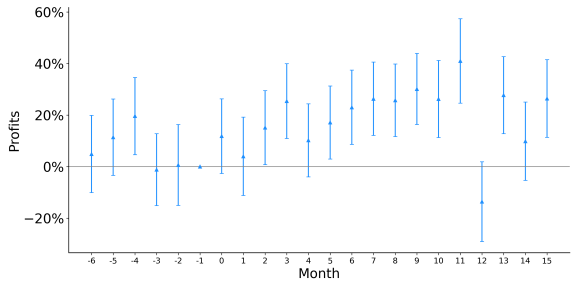
(e) Hours Worked



(f) Income



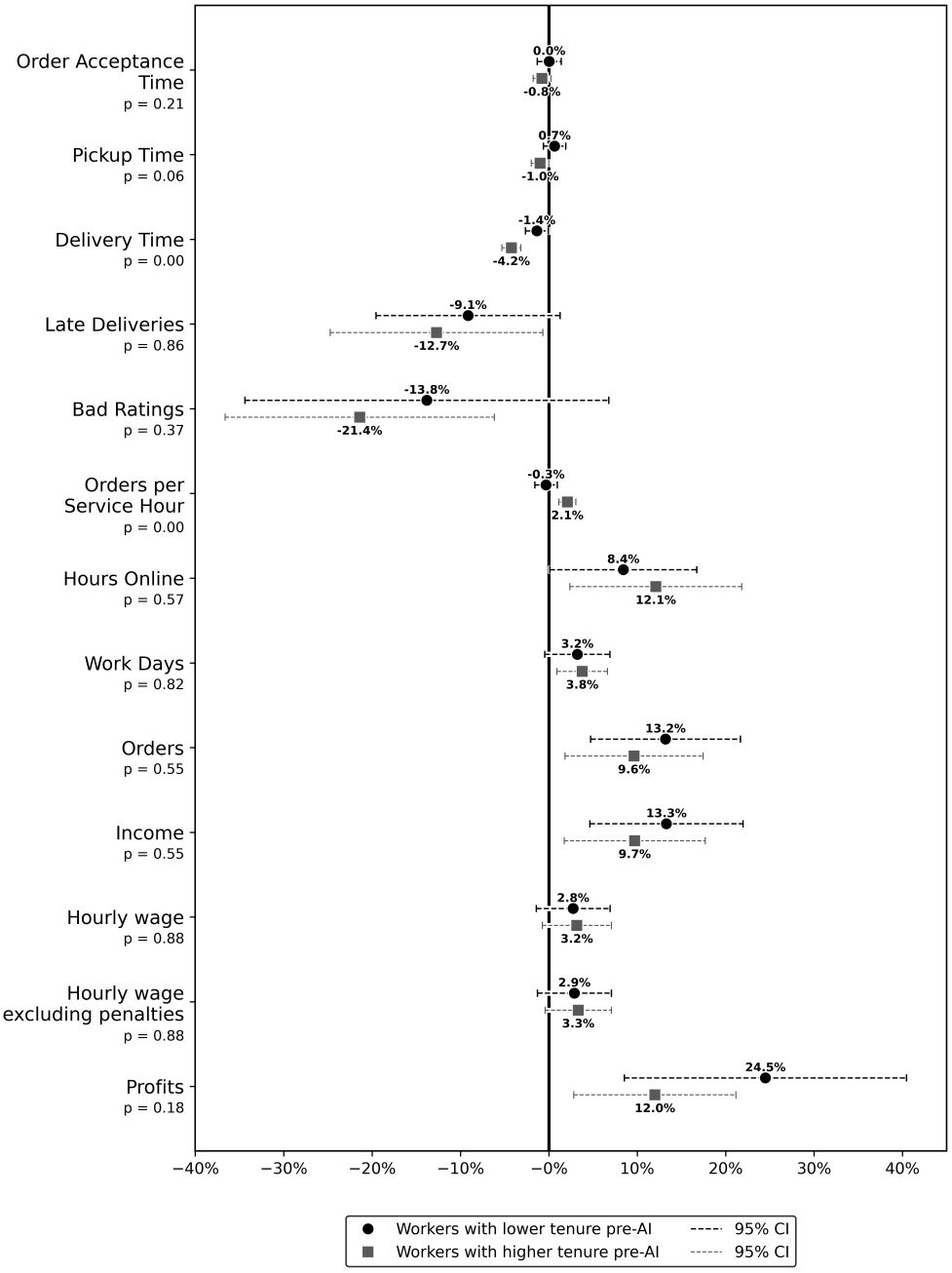
(g) Hourly Wage



(h) Profit

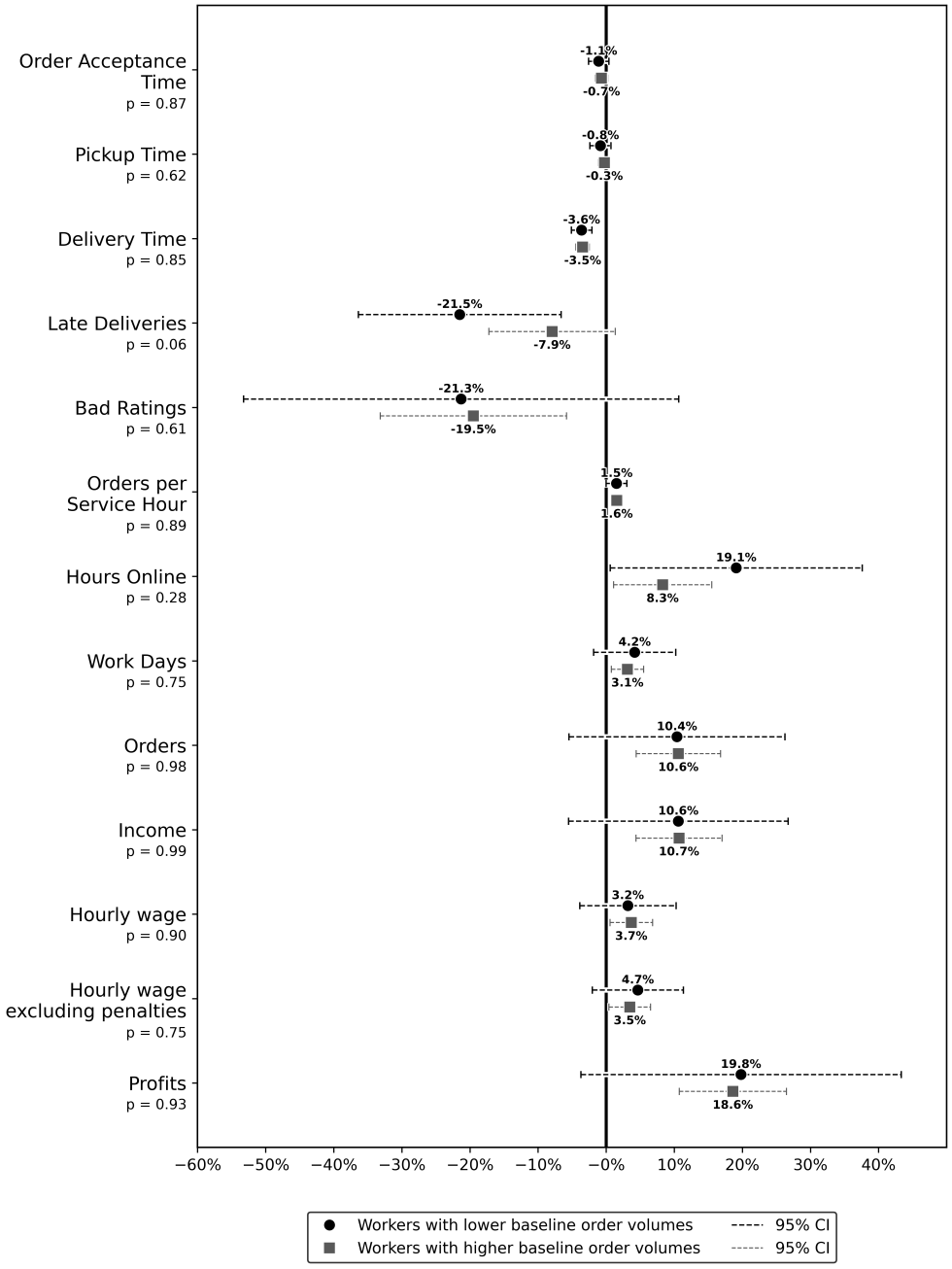
Notes: This figure replicates the event studies in Figure 4 and Figure 5 by extending the analysis through month 15 post-adoption (months 0–15). Dots show the event study coefficients, and whiskers show the 95% confidence intervals accounting for clustering by driver.

Figure A.8: Heterogeneity in Effects of AI Tool by Pre-AI Tenure



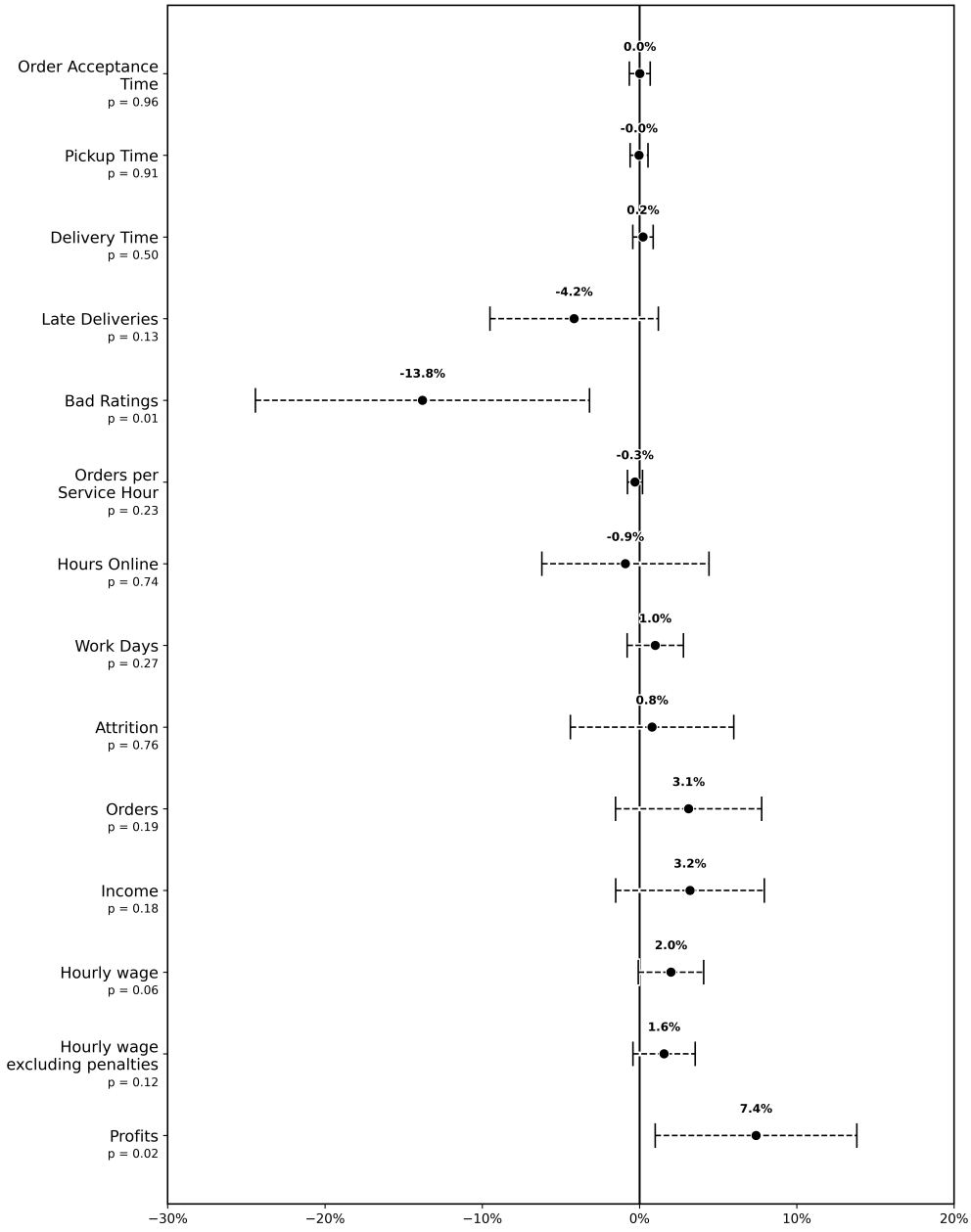
Notes: This figure presents heterogeneous treatment effects of the AI tool by pre-AI worker tenure. For each outcome, we estimate our main DiD model separately for workers within that group. The p-values shown are tests for equality of treatment effects between groups and are estimated using a standard triple-interaction model (including all three double interactions). The p-value is a test of whether the coefficient on $DHH_i \times Post_t \times z_i$ equals zero, where z_i is the heterogeneity characteristic of interest, in this case, pre-AI tenure. Pre-AI tenure is categorized as Low or High using a median split across workers, with tenure measured as of the date of AI implementation, t_0. The sample consists of workers who are present at t_0. There are observations from 482 DHH workers and 2,995 non-disabled workers.

Figure A.9: Heterogeneity in Effects of AI Tool by Pre-AI Productivity



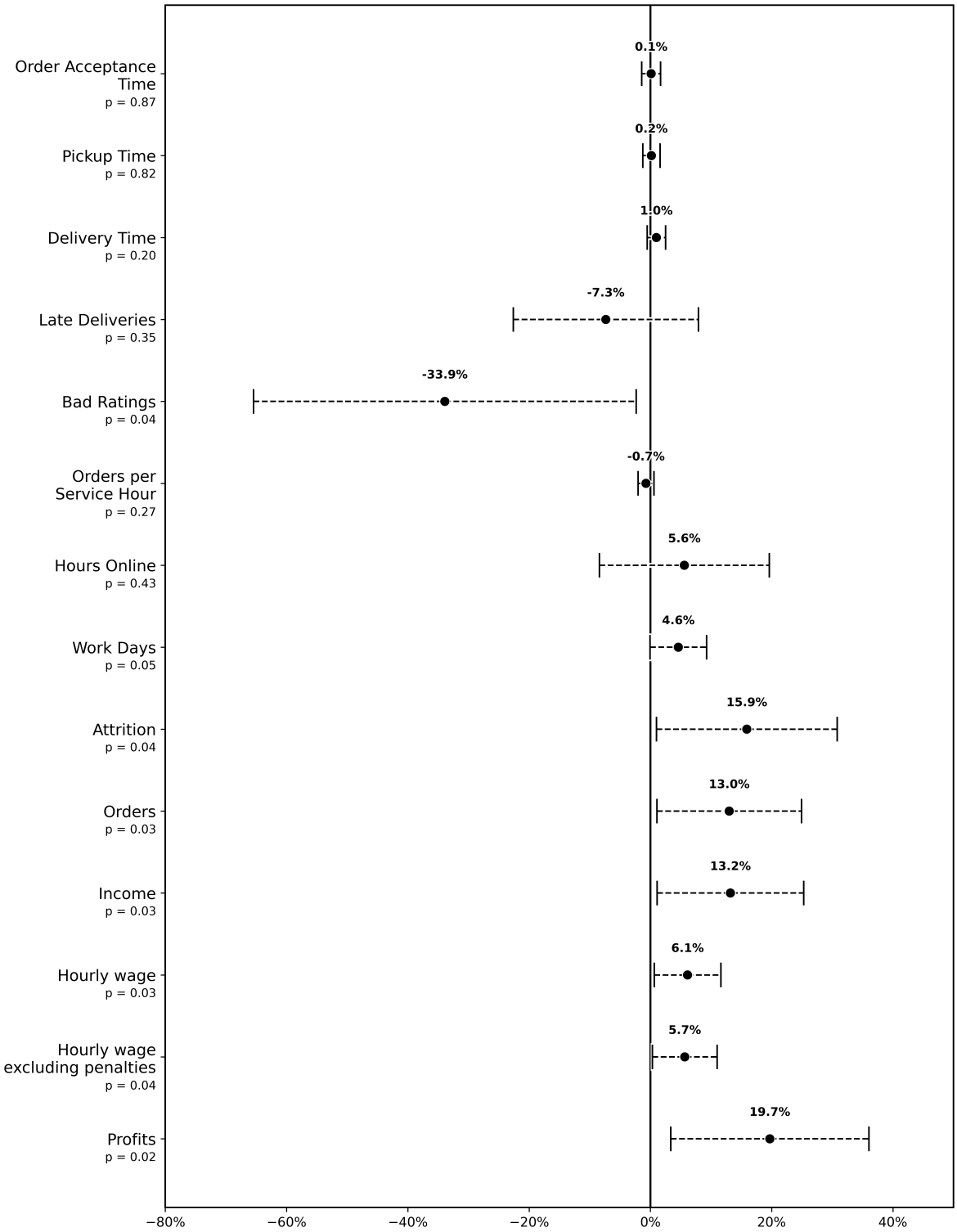
Notes: This figure is similar to Figure A.8, but splits by pre-AI productivity. To construct the productivity measure, we first estimate worker fixed effects from a regression of the inverse hyperbolic sine of weekly orders on worker and year-week fixed effects, using only pre-AI observations. We then regress these estimated worker fixed effects on an indicator for team (versus gig) configuration, so that the resulting residuals capture productivity variation net of systematic differences between team and gig workers. Workers are classified as high- (low-) productivity if their residual falls above (below) the sample median. The sample consists of worker-months from 1,515 DHH and 15,397 non-disabled workers.

Figure A.10: Robustness on Heterogeneity in the Effects of AI Tool by Disability Severity: Dosage Response Model



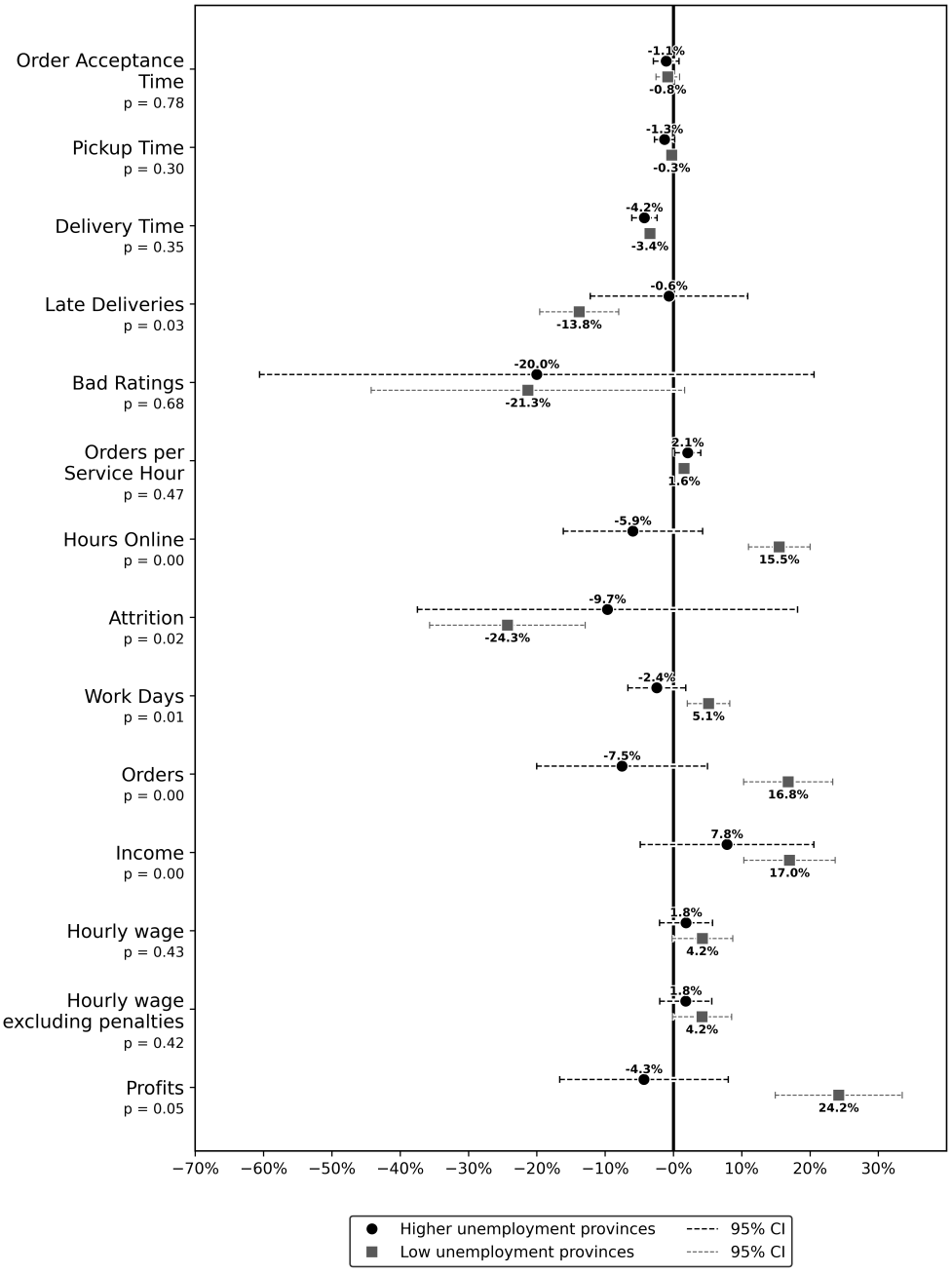
Notes: This figure is similar to Figure 9 in the main text. The difference is that instead of splitting up DHH workers by profound and non-profound hearing loss, we estimate a dosage response model. Each dot shows the dosage-response coefficient; the p-value tests whether it equals zero. The original Level 1 (most severe, $>$ 90 dB hearing loss) to Level 4 (least severe, 41–60 dB) is transformed using Severity $= 5 -$ Level. The figure plots the coefficient on the triple interaction of $DHH_i \times Post_t \times \text{Severity}_i$. Whiskers indicate 95% confidence intervals clustered by driver.

Figure A.11: Robustness on Heterogeneity By Disability Severity: Control for Other Interactions



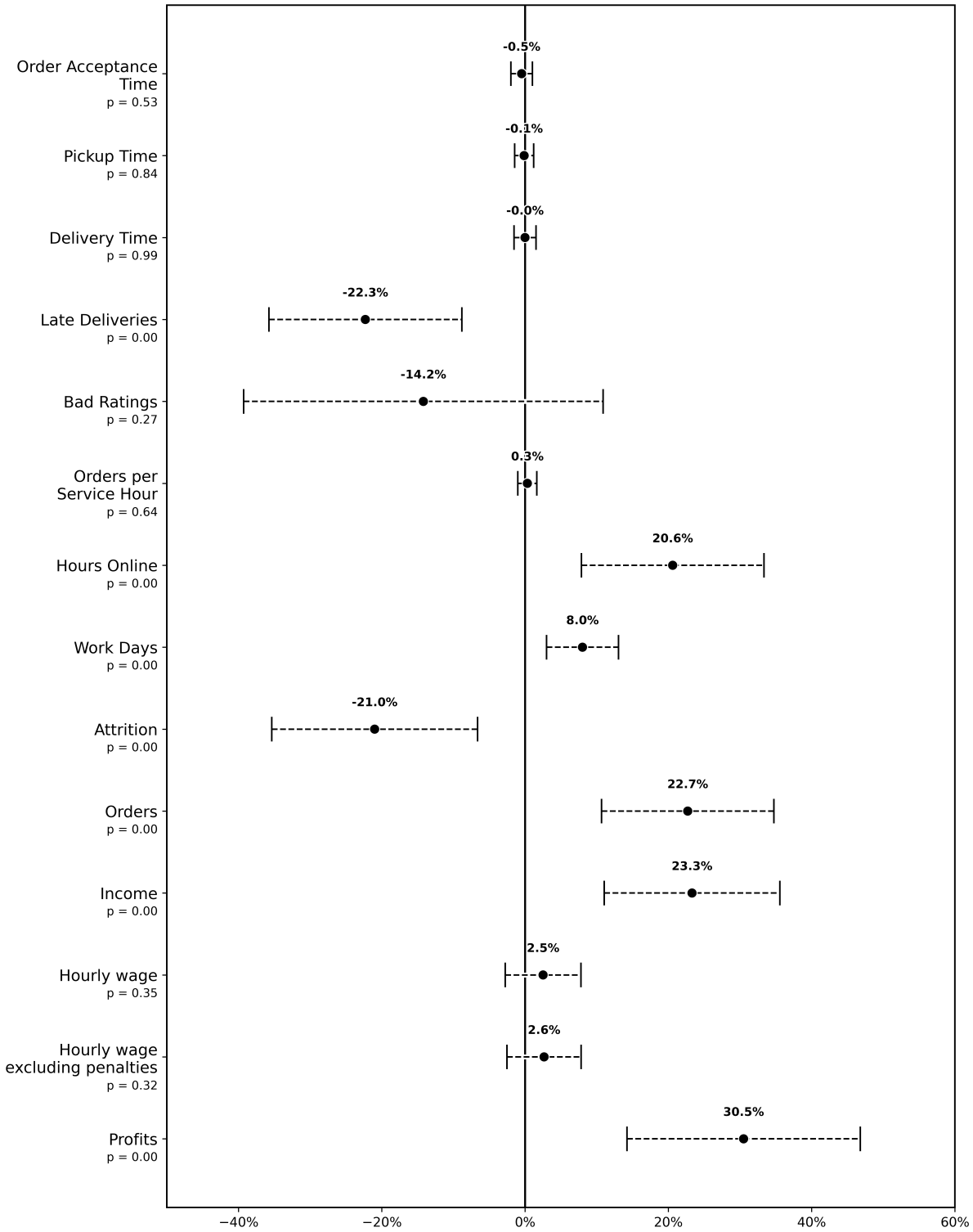
Notes: This figure presents a robustness check on heterogeneity by disability severity. It presents analyses similar to those in Figure 9, but also controls for triple interactions, $DHH \times Post \times Female$ and $DHH \times Post \times Age$, as well as relevant double interactions. This helps address whether our heterogeneity by severity is standing in for heterogeneity by demographics. The coefficients plotted here are the triple interactions $DHH_i \times Post_t \times \text{Profound}_i$. A coefficient different from 0 indicates a differential effect by disability severity.

Figure A.12: Heterogeneity in the Effects of the AI Tool by Regional Unemployment Level



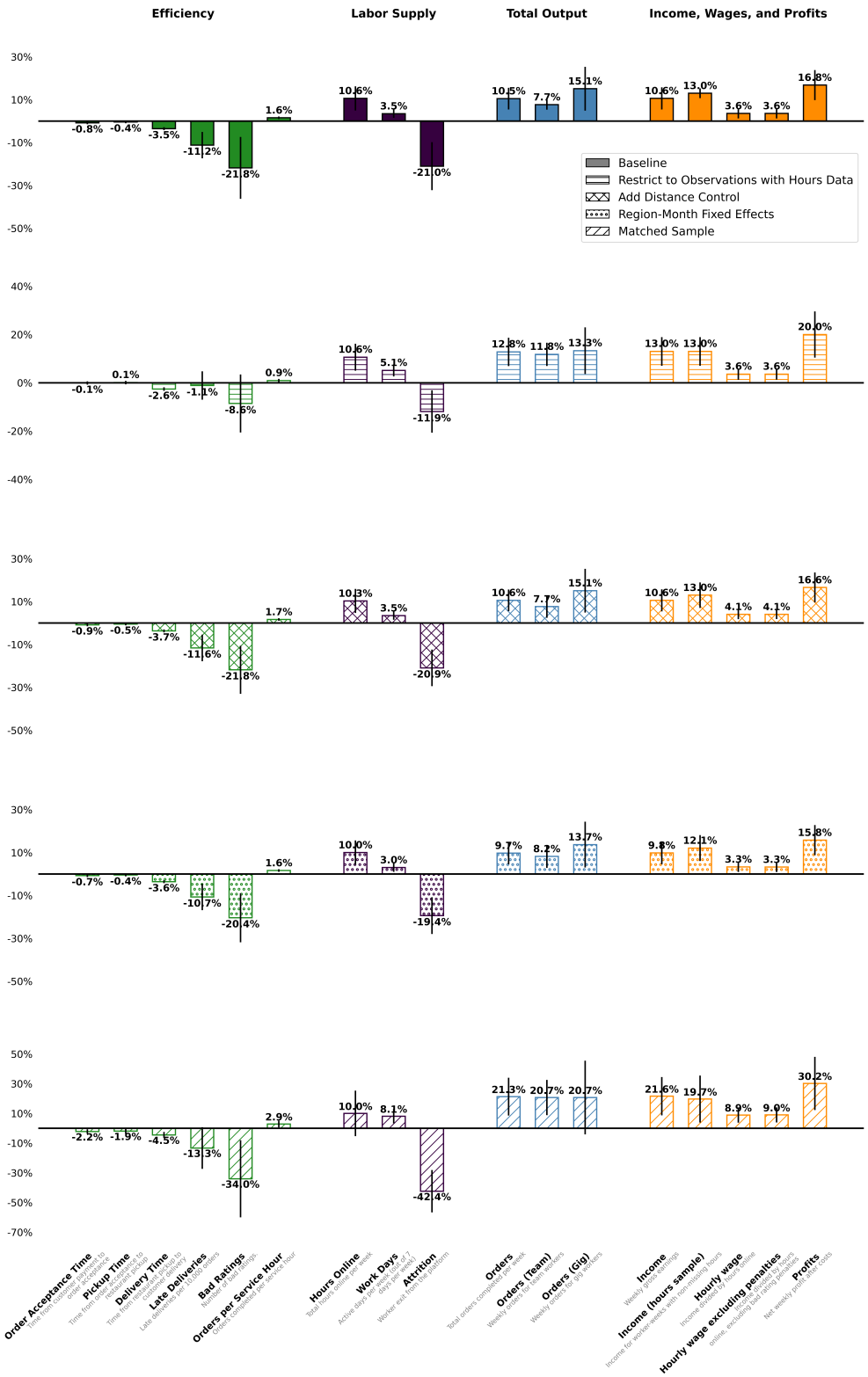
Notes: This figure presents heterogeneous treatment effects by regional unemployment level. It is similar to Figure A.8, but instead splits the sample by regional unemployment rate. Regions are divided into high- and low-unemployment regions using the median unemployment rate as the cutoff. The sample includes 31 provincial-level administrative regions ("regions") in mainland China. Unlike for other heterogeneity analyses, standard errors are clustered by region, reflecting that we study a characteristic that varies by region. The sample consists of worker-weeks from 3,982 DHH and 36,719 non-disabled workers. The sample size is smaller than in our main analyses because region is missing for some workers. Information on unemployment is from the 2021 China Labour Statistical Yearbook. Low-unemployment regions are Qinghai, Xinjiang, Shanxi, Hunan, Jiangsu, Anhui, Guangdong, Guangxi, Zhejiang, Tibet, Shanghai, Jiangxi, Shandong, Chongqing, and Hubei. High-unemployment regions are Hebei, Hainan, Beijing, Heilongjiang, Jilin, Fujian, Henan, Gansu, Shaanxi, Sichuan, Tianjin, Inner Mongolia, Yunnan, Ningxia, Liaoning, and Guizhou.

Figure A.13: Robustness on Heterogeneity in the Effects of the AI Tool by Regional Unemployment Level: Control for Other Interactions



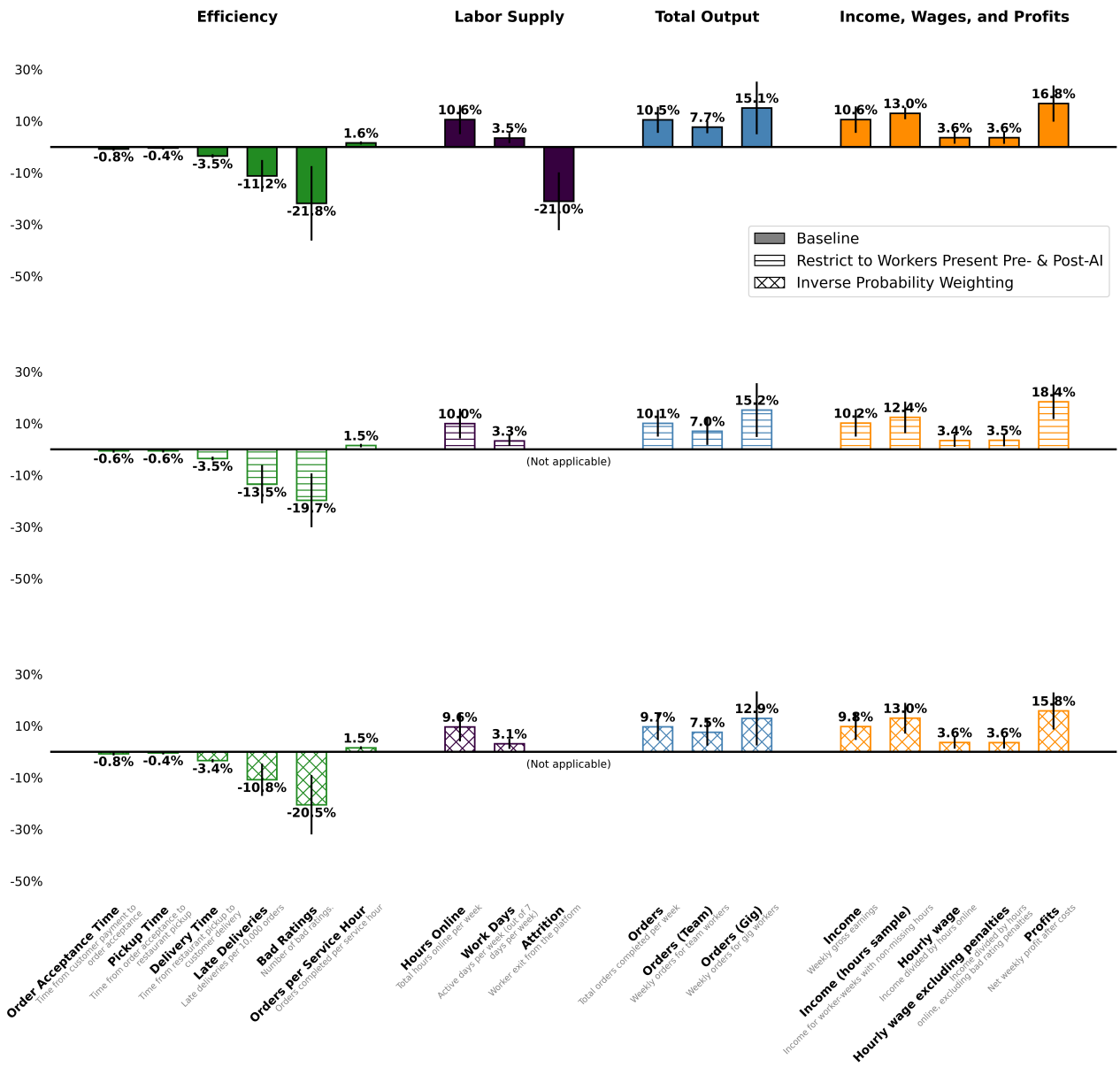
Notes: This figure presents a robustness check on heterogeneity by regional unemployment. It presents analyses similar to those in Figure A.12, but also controls for triple interactions, $DHH \times Post \times Female$ and $DHH \times Post \times Age$, as well as relevant double interactions. This helps address whether our heterogeneity by regional unemployment is standing in for heterogeneity by demographics. The coefficients plotted here are the triple interactions $DHH_i \times Post_t \times$ Lower Unemployment Region$_i$, where Low-Unemployment Region$_i$ is a dummy for being in a lower unemployment region. A coefficient different from 0 indicates a differential effect by regional unemployment level.

Figure A.14: Figure 3 Robustness: Checks on Hours Sample, Matched Sample, and Controls



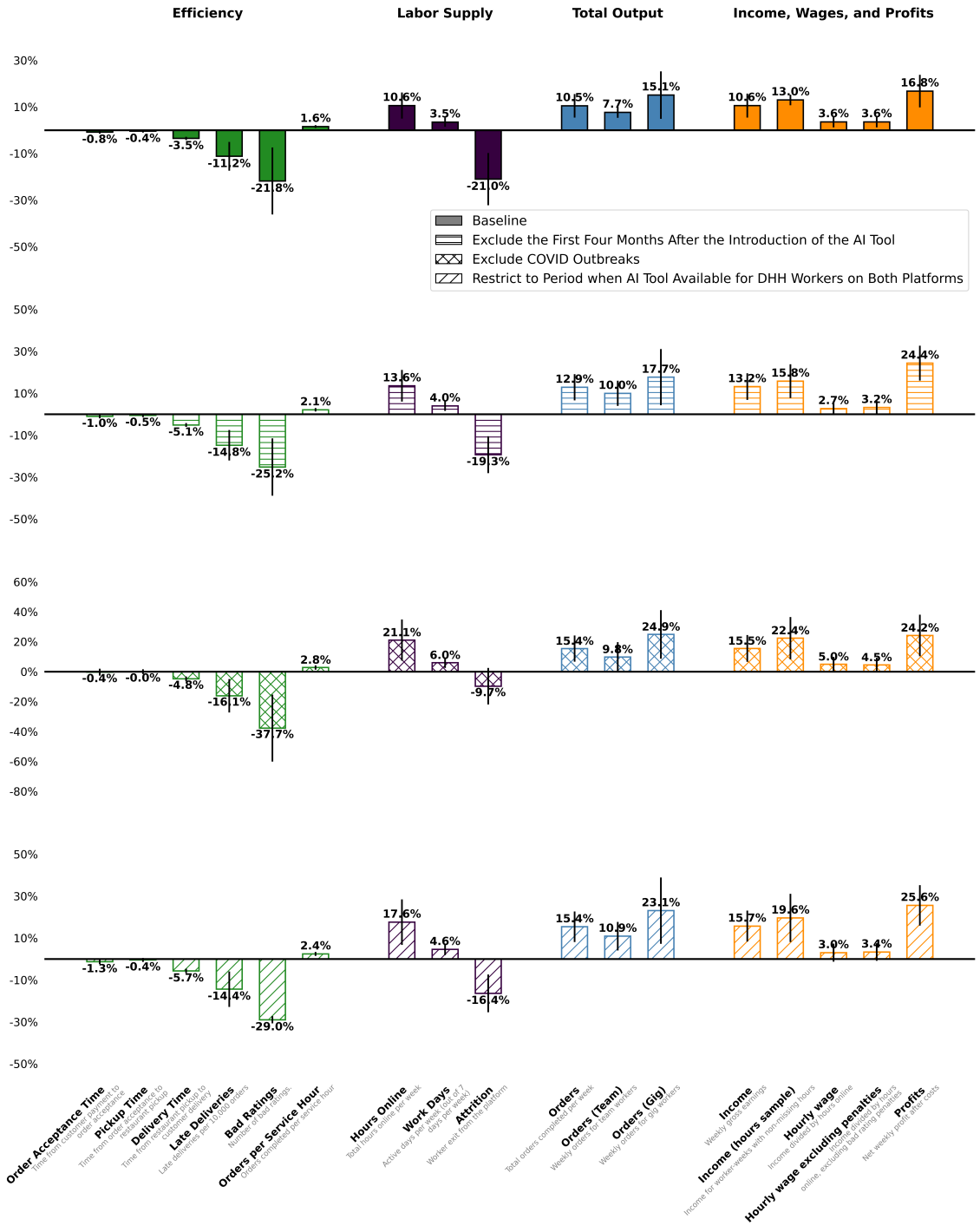
Notes: The top panel replicates the baseline analyses in Figure 3. The 2nd panel restricts the sample to observations with hours data, since hours are missing for many worker-weeks (see Figure A.2); the resulting sample contains 343,733 worker-week observations from 34,965 workers. The 3rd panel adds a worker-week-level control for the average restaurant-to-customer distance per order. This variable proxies for order difficulty, since longer-distance orders may require more travel time, involve greater delivery-time uncertainty, and take workers farther from dense restaurant areas or into less familiar delivery locations. This specification uses the same sample as the baseline. The 4th panel adds region-month dummies. The 5th panel performs the DiD analysis using exact matching, matching each DHH worker to one non-disabled worker based on entry year, work model (gig or team), city, 5-year age bin, and gender, leaving 2,354 DHH and 2,354 non-disabled workers.

Figure A.15: Robustness of Figure 3: Robustness Checks on Differential Attrition



Notes: The top panel replicates the baseline analyses in Figure 3. The middle panel restricts the sample to drivers observed for at least one week both before and after the introduction of the AI tool; this restricted sample includes 184,450 worker-week observations from 3,397 workers. The bottom panel estimates the main DiD specification using inverse probability weights to account for differential attrition. Specifically, we first estimate each worker's probability of remaining in the sample using a logit model as a function of time-invariant worker characteristics, namely city-of-work fixed effects, age fixed effects, gender, and team status (gig vs. team), as well as year-week fixed effects. We then weight observations by the inverse of the predicted probability. The inverse probability weighting specification uses the baseline sample of 480,070 worker-week observations from 43,605 workers. Attrition is not analyzed in the middle or bottom panels, as we use strategies there to address concerns about differential attrition, and analyzing attrition there would be circular.

Figure A.16: Robustness of Figure 3: Robustness Checks Related to Longer-run Effects and External Validity



Notes: The top panel replicates the baseline analyses in Figure 3. The second panel excludes the first four months after the introduction of the AI tool, yielding 427,612 worker-week observations from 40,821 workers. The third panel excludes worker-week observations from regions-weeks affected by COVID outbreaks, leaving 234,004 worker-week observations from 28,303 workers. Under China's "Dynamic Zero-COVID" policy, restrictions were localized rather than synchronized nationwide, so their timing and stringency varied across regions and within regions. Oftentimes, only part of a region would have a COVID outbreak, so we are being conservative in excluding region-weeks with any outbreak. We identify affected region-weeks using official government announcements and verify the timing with contemporaneous news reports. The bottom panel excludes the period when the AI tool was available on our platform but not yet on the other major food delivery platform. Since China's food delivery market is largely dominated by these two platforms, this restriction addresses the possibility that temporary cross-platform spillovers, such as worker or customer reallocation, affect our estimates. This restricted sample includes 391,839 worker-week observations from 38,164 workers.

Table A.7: Bad Ratings: Pre-AI Differences, DiD Estimates, and DiD Robustness Checks

	(1)	(2)	(3)
Panel A: Pre-AI Means			
	Non-disabled Workers	DHH Workers	
Mean	0.10	0.15	
Panel B: Pre-AI Difference, Raw and Scaled			
	OLS, raw coefficient	OLS coefficient divided by non-disabled mean	OLS coefficient divided by DHH mean
DHH	0.047***	0.470***	0.313***
	(0.009)	(0.087)	(0.058)
Panel C: DiD Estimate, Raw and Scaled			
	OLS, raw coefficient	OLS coefficient divided by DHH mean	
$DHH \times Post$	−0.032***	−0.218***	
	(0.010)	(0.068)	
Panel D: DiD Robustness, Dependent Variable = Bad Ratings per Week			
	OLS; coefficient divided by pre-AI DHH mean	Poisson	OLS; coefficient divided by pre-AI DHH mean, controlling for completed orders
$DHH \times Post$	−0.218***	−0.185***	−0.239***
	(0.068)	(0.071)	(0.068)
Panel E: DiD Robustness, Dependent Variable = Bad Ratings per 10,000 Orders			
	OLS	OLS, order-weighted	
$DHH \times Post$	−2.350**	−2.197***	
	(1.024)	(0.568)	

Notes: Panel A reports pre-AI means for non-disabled and DHH workers. Panel B reports the regression-adjusted pre-AI difference (DHH vs. non-disabled) from estimating equation (2): the raw OLS coefficient in levels, the coefficient divided by the non-disabled pre-AI mean, and the coefficient divided by the DHH pre-AI mean. Panel C reports the baseline DiD estimate from equation (3): the raw OLS coefficient and the coefficient divided by the DHH pre-AI mean. Panels D and E examine robustness of the DiD results to alternative estimation methods, functional forms, and control variables. Each DiD coefficient is from a difference-in-differences specification using the same fixed effects as in the main text. Standard errors clustered by worker in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A.8: Late Deliveries: Pre-AI Differences, DiD Estimates, and DiD Robustness Checks

	(1)	(2)	(3)
Panel A: Pre-AI Means			
	Non-disabled Workers	DHH Workers	
Mean	4.79	5.11	
Panel B: Pre-AI Difference, Raw and Scaled			
	OLS, raw coefficient	OLS coefficient divided by non-disabled mean	OLS coefficient divided by DHH mean
DHH	0.497*	0.104**	0.097**
	(0.275)	(0.057)	(0.054)
Panel C: DiD Estimate, Raw and Scaled			
	OLS, raw coefficient	OLS coefficient divided by DHH mean	
$DHH \times Post$	−0.853*	−0.167*	
	(0.497)	(0.097)	
Panel D: DiD Robustness, Dependent Variable = Late Deliveries per Week			
	OLS; coefficient divided by pre-AI DHH mean	Poisson	OLS; coefficient divided by pre-AI DHH mean, controlling for completed orders
$DHH \times Post$	−0.167*	−0.164*	−0.255**
	(0.097)	(0.094)	(0.103)
Panel E: DiD Robustness, Dependent Variable = Late Deliveries per 10,000 Orders			
	OLS	OLS, order-weighted	
$DHH \times Post$	−51.15***	−40.60**	
	(17.09)	(17.28)	

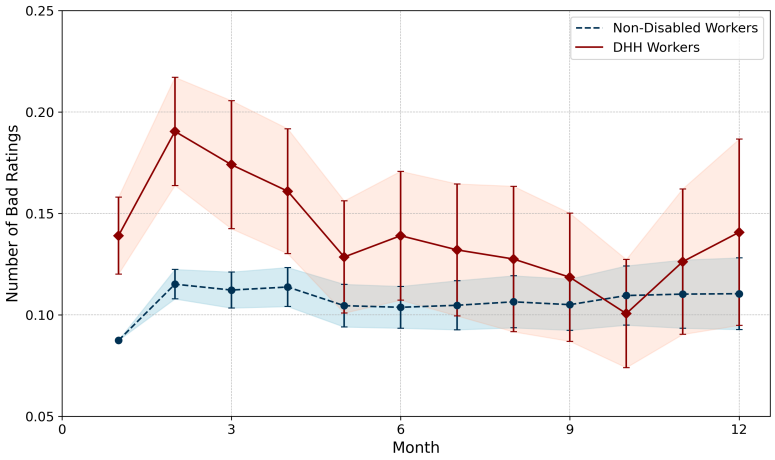
Notes: Panel A reports pre-AI means for non-disabled and DHH workers. Panel B reports the regression-adjusted pre-AI difference (DHH vs. non-disabled) from estimating equation (2): the raw OLS coefficient in levels, the coefficient divided by the non-disabled pre-AI mean, and the coefficient divided by the DHH pre-AI mean. Panel C reports the baseline DiD estimate from equation (3): the raw OLS coefficient and the coefficient divided by the DHH pre-AI mean. Panels D and E examine robustness of the DiD results to alternative estimation methods, functional forms, and control variables. Each DiD coefficient is from a difference-in-differences specification using the same fixed effects as in the main text. Standard errors clustered by worker in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A.9: Attrition: Pre-AI Differences and DiD Estimates

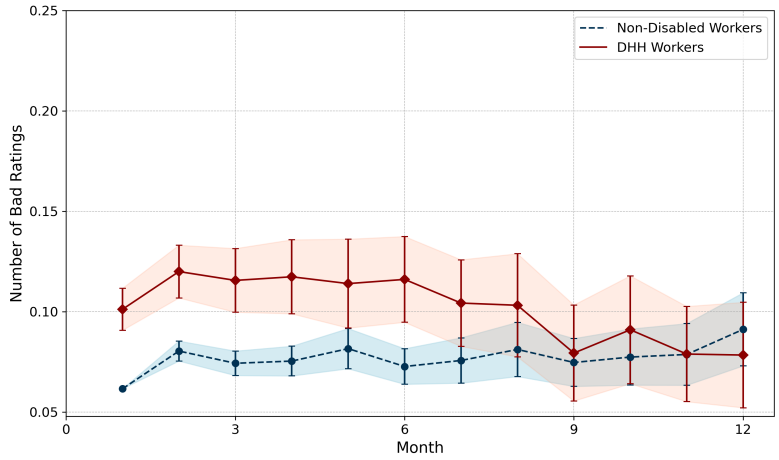
	(1)	(2)	(3)
Panel A: Pre-AI Means			
	Non-disabled Workers	DHH Workers	
Mean	0.076	0.056	
Panel B: Pre-AI Difference, Raw and Scaled			
	OLS, raw coefficient	OLS coefficient divided by non-disabled mean	OLS coefficient divided by DHH mean
DHH	−0.023***	−0.301***	−0.414***
	(0.002)	(0.030)	(0.041)
Panel C: DiD Estimate, Raw and Scaled			
	OLS, raw coefficient	OLS coefficient divided by DHH mean	
$DHH \times Post$	−0.012***	−0.210***	
	(0.003)	(0.052)	

Notes: Panel A reports pre-AI means for non-disabled and DHH workers. Panel B reports the regression-adjusted pre-AI difference (DHH vs. non-disabled) from estimating equation (2): the raw OLS coefficient in levels, the coefficient divided by the non-disabled pre-AI mean, and the coefficient divided by the DHH pre-AI mean. Panel C reports the baseline DiD estimate from equation (3): the raw OLS coefficient and the coefficient divided by the DHH pre-AI mean. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Figure A.17: Robustness of Figure 7: Tenure-Productivity Curves in Bad Ratings, Before and After the AI Tool (Count of Bad Ratings instead of Ratings per 10,000 Orders)



(a) Pre-AI tool



(b) Post-AI tool

Notes: This figure is similar to Figure 7, but the outcome is number of bad ratings. Error bars show 95 percent confidence intervals accounting for clustering by worker. The samples in each panel are the same as in Figure 7.

B Additional Discussion and Analysis

B.1 Impact of AI Tool on Good Ratings of DHH Workers

This section examines how the AI tool affects the number of good ratings received by DHH workers. As discussed in the main text, our main text analyses focus on bad ratings because bad ratings are much more important to the Platform and have monetary consequences for workers. By contrast, based on our interviews with the Platform, good ratings do not generate direct practical benefits for workers. We apply the same empirical framework used in the main text: column (1) reports the pre-AI difference specification from equation (2), and column (2) reports the overall treatment effect using the difference-in-differences specification from equation (3). Table B.1 presents the results.

Table B.1: Pre-AI Differences and Impact of AI Tool on Good Ratings of DHH Workers

	(1)	(2)
	Pre-AI Difference	DiD Result
DHH_i	0.165***	
	(0.038)	
$DHH_i \times Post_t$		0.006
		(0.051)
Observations	180,375	480,070

Notes: Column (1) reports pre-AI differences in good ratings between DHH and non-disabled workers. Column (2) reports DiD estimates of the impact of the AI tool on good ratings for DHH workers relative to non-disabled workers. As for bad ratings, we estimate non-transformed OLS models and divide by the pre-AI DHH mean to calculate percentage differences. Standard errors clustered by worker in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Column (1) shows that, prior to the introduction of the AI tool, DHH workers receive 16.5% more good ratings than non-disabled workers in the pre-AI period. In comparison, our main results on bad ratings show that DHH workers also received 31.3% more bad ratings before the AI tool was introduced. Together, these findings suggest that DHH workers receive more customer feedback of both types, which likely partly reflects that they complete more orders. Column (2) shows that the AI tool did not have a statistically significant effect on good ratings for DHH workers relative to non-DHH workers.

Taken together, these results suggest that while DHH workers were already receiving more good ratings prior to the AI tool, the tool itself did not further improve this dimension of customer feedback. This contrasts with our findings for bad ratings in the main text, where the AI tool meaningfully reduced negative feedback. One interpretation is that the AI tool mainly operates by reducing downside communication failures, rather than shifting the upper tail of customer satisfaction. Importantly, the null effect of the AI tool on good

ratings (column 2) suggests that customer sympathy for DHH workers is not driving our main results: if sympathy were the mechanism, we would expect good ratings to improve alongside the reduction in bad ratings, which we do not find.

B.2 Impact of AI Tool on Entry of DHH Workers

This section examines how the introduction of the AI tool affects the entry of DHH workers onto the Platform. Our analysis in the main text is based on workers who join the Platform and complete at least one order. However, if the AI tool makes work on the Platform more attractive for DHH workers, it could also spur more DHH workers to join.

We cannot observe the set of DHH workers potentially interested in working on the Platform. Instead, we collapse the data by disability type (DHH or non-disabled) and year-week. We have 3 years of data, so 3*52 = 156 year-weeks, and there are two disability types, so we have 312 observations. We then estimate a model very similar to equation (3):

$$Y_{gt} = \alpha + \gamma \cdot DHH_g \times Post_t + \delta \cdot DHH_g + \xi_t + \varepsilon_{gt}. \tag{B.1}$$

where g denotes group (i.e., DHH or non-disabled) and t denotes time in year-weeks; Y_{gt} is the inverse hyperbolic sine of the number of new entrants on the Platform in group g during year-week t; and DHH_g is a dummy for being DHH. We are interested in estimates of γ.

Table B.2: Impact of AI Tool on Entry of DHH Workers: DiD Results

	(1)
$DHH_g \times Post_t$	0.275**
	(0.110)
Observations	312

Notes: An observation is a year-week × disability type (i.e., DHH or non-disabled). Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.2 presents the regression results. The AI tool increases the number of DHH registrations on the Platform by 27.5%. This is a sizable recruitment effect.

B.3 Mediation Analysis

We run a mediation analysis to examine whether the treatment's effect on order volume operates through algorithmic reallocation. As discussed in the main text, the Platform's algorithm looks favorably upon drivers with fewer bad ratings and fewer late deliveries. If this is the primary channel, the AI tool would raise orders mainly by improving late deliveries and bad ratings, which the algorithm then rewards—with observed gains in order volume reflecting task reallocation rather than a direct increase in worker capacity. We examine two

potential mediating channels: late deliveries and bad ratings.[1] For each channel, we estimate (i) the total effect of the AI tool on orders, (ii) the effect of the AI tool on the mediator, and (iii) the effect of the AI tool on orders when controlling for the mediator. Results are reported in Table B.3.

Table B.3: Mediation Analysis: Is the Effect on Orders Driven by Algorithmic Reallocation?

Dep. Var.:	Orders	Late Deliveries	Orders	Bad Ratings	Orders
	(1)	(2)	(3)	(4)	(5)
$DHH_i \times Post_t$	0.105^{***}	-51.15^{***}	0.103^{***}	-2.350^{**}	0.105^{***}
	(0.0307)	(17.09)	(0.0307)	(1.024)	(0.0307)
Late Deliveries (per 10k orders)			$-3.24 \times 10^{-5***}$		
			(4.30×10^{-6})		
Bad Ratings (per 10k orders)					$-8.68 \times 10^{-5***}$
					(3.34×10^{-5})

Notes: Standard errors clustered by worker in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Column 1 reports the baseline DiD treatment effect on orders following equation (3). Columns 2 and 4 use late deliveries and bad ratings, respectively, as the dependent variable with the same controls. Columns 3 and 5 re-estimate the baseline orders regression adding late deliveries and bad ratings, respectively, as additional controls. Late deliveries and bad ratings are measured per 10,000 orders.

Column 1 of Table B.3 gives the baseline total effect: the treatment increases the IHS-transformed number of orders by 0.105. Column 2 shows that the treatment reduces late deliveries by 51.15 per 10,000 orders, and column 4 shows that it reduces bad ratings by 2.350 per 10,000 orders. Columns 3 and 5 re-estimate the baseline orders regression controlling for each mediator in turn. The treatment coefficient attenuates only slightly to 0.103 when controlling for late deliveries (column 3), implying that the late delivery channel accounts for about 1.6% of the total effect. The coefficient is essentially unchanged at 0.105 when controlling for bad ratings (column 5), implying that the bad ratings channel accounts for less than 0.2% of the total effect.

Consistent with the algorithm penalizing late deliveries and bad ratings, both mediators are negatively associated with orders (columns 3 and 5). However, controlling for either of these does little to diminish the treatment effect of the AI tool on orders. The dominant mechanism linking the AI tool to higher order volume therefore lies outside of bad ratings and late deliveries. This supports that algorithmic reallocation is not an important mechanism for our treatment effect on orders.

[1] We analyze late deliveries and bad ratings each in terms of a ratio per 10,000 orders, doing so to avoid a "reverse causality" issue where someone has lots of bad ratings due to working more orders in a week.

C Further Information on Hearing Loss Levels

Below is an English translation of Section 5.3 of GB/T 26341-2010, done by ChatGPT on 1/25/26. Note that drivers with HL of 40 dB or less are not issued a certificate of disability.

- *5.3.2 Hearing Disability Level I. The structure and function of the auditory system are extremely severely impaired. The average hearing loss of the better ear is greater than 90 dB HL. Speech communication cannot be carried out relying on hearing. There are extremely severe limitations in activities such as understanding and communication, and extremely serious barriers exist in participation in social life.*

- *5.3.3 Hearing Disability Level II. The structure and function of the auditory system are severely impaired. The average hearing loss of the better ear is between 81 and 90 dB HL. There are severe limitations in activities such as understanding and communication, and serious barriers exist in participation in social life.*

- *5.3.4 Hearing Disability Level III. The structure and function of the auditory system are moderately to severely impaired. The average hearing loss of the better ear is between 61 and 80 dB HL. There are moderate limitations in activities such as understanding and communication, and moderate barriers exist in participation in social life.*

- *5.3.5 Hearing Disability Level IV. The structure and function of the auditory system are moderately impaired. The average hearing loss of the better ear is between 41 and 60 dB HL. There are mild limitations in activities such as understanding and communication, and mild barriers exist in participation in social life.*

D Potential Nondisclosure of DHH Status

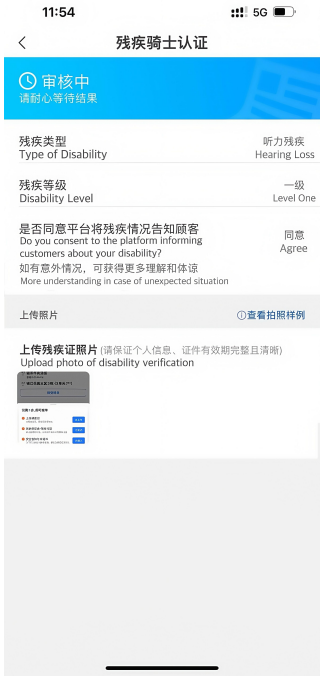
This Appendix discusses the issue of potential nondisclosure of DHH status by workers to the Platform, which determines whether a worker is classified as DHH in our data.

Why workers disclose. As noted in Section 1 of the main text, based on our qualitative work and understanding of the Platform, essentially all certified DHH workers disclose their disability to the Platform, for two main reasons.

First, it is difficult to hide, especially for team workers. In the team work model, drivers have managers, and a worker's manager would need to know if they are DHH. Drivers in both work models (gig and team) also generally receive calls from customer service or customers, and the Platform must decide whether those calls should go through.

Second, DHH workers have a strong personal incentive to disclose their disability so that the Platform can accommodate them. People with disabilities in China are accustomed to disclosing their disability to employers and the government in order to receive potential accommodations. For example, if DHH workers do not disclose their status to the Platform, customers cannot be informed of their disability and may be less understanding in unexpected situations. Figure D.1 shows the screen from the registration process where workers can disclose their DHH status.

Figure D.1: Screen During Registration Process where Workers Indicate Disability Status



Notes: This figure shows the original text, as well as text translated into English.

Comparison of severity between our data and the broader population. Table D.1 compares the distribution of hearing loss severity in our Platform sample to that among registered disability certificate holders in China. Workers with profound hearing loss (Level 1) account for 54% of DHH workers on the Platform, compared to 26% among certificate holders overall in 2012.[2]

Table D.1: Distribution of Hearing Loss Severity

Source	Level 1	Level 2	Level 3	Level 4
Registered certificate holders (2012)	26%	24%	23%	26%
Our Platform sample	54%	16%	14%	16%

Notes: This table compares disability severity prevalence between the overall Chinese population and our data, which covers new DHH workers on the Platform between 2021-2023.

In principle, an analyst might try to use quantitative tests to detect nondisclosure of DHH status on the Platform, but this is extremely difficult. One might compare the share of DHH workers in our data to population shares, overall or by severity. However, such

[2]The 2012 data are from the *China Statistical Yearbook on the Work for Persons with Disabilities*. In more recent publications, severity levels are reported for all disability types combined rather than separately for hearing loss.

comparisons also reflect selection of DHH workers onto the Platform. While the share of drivers with profound HL is higher in our sample than in the population, we believe this primarily reflects selection based on outside options. As in many countries, DHH workers face substantial discrimination and barriers to employment, particularly those with profound HL, leading many to gravitate toward gig work.

Implications of potential nondisclosure. How would any nondisclosure of DHH status affect interpretation of our results? Any nondisclosure that is fixed over time (presumably the most natural kind) would not affect the main DiD results on the impact of the AI tool, since these specifications include worker fixed effects. Moreover, undisclosed DHH workers would not receive the AI tool—which was rolled out only to workers who disclosed their DHH status—and would thus remain in the control group throughout, creating no treatment misclassification. It would only affect the pre-AI comparisons, potentially making DHH workers appear less productive if higher-performing DHH workers were less likely to disclose their status.[3]

A different issue would arise if the AI tool itself affected disclosure of DHH status. While this is theoretically possible, none of our qualitative data or conversations with the Platform suggest that the AI tool made workers more likely to disclose their disability.[4]

E Details on the AI Tool: Intelligent Outbound Call

The IOC feature is an in-app tool designed to empower DHH workers by facilitating effective communication with customers through automated voice calls. Previously, these workers relied solely on text messaging within the app, which customers often overlooked, leading to misunderstandings and delays. This technology bridges a critical communication gap by converting selected text prompts into automated voice calls.

The IOC feature is integrated into the delivery app. Workers can select from predefined default prompts or input customized prompts relevant to specific situations. These text messages are converted into audio through an advanced text-to-speech system and sent to the customer via a cellphone call that appears to originate from the worker's phone number.

Panel (a) of Figure E.1 displays the default prompt menu provided by the app. Workers can quickly select an appropriate option, preview the automatically generated message, and initiate the automated call instantly.

For situations beyond standard prompts, workers have the flexibility to create and store personalized messages. Panel (b) of Figure E.1 illustrates the menu with customized additions, showing prompts a worker has added alongside the defaults to address real-world delivery challenges.

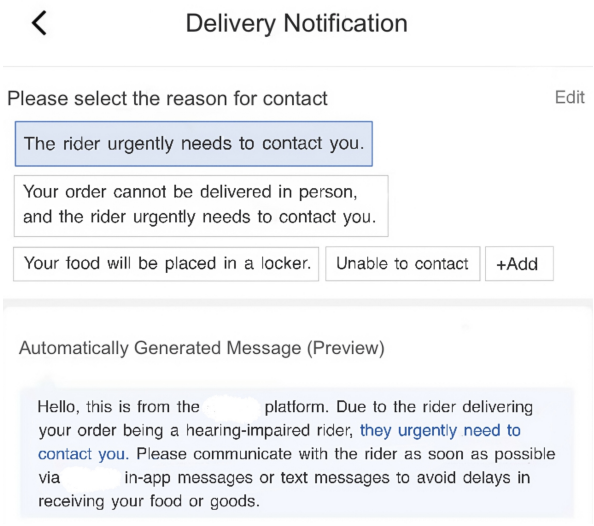
The IOC feature significantly enhances the delivery experience in three key ways:

- Improved Customer Responsiveness — Customers are more likely to notice and respond to automated voice calls compared to in-app text notifications, leading to faster

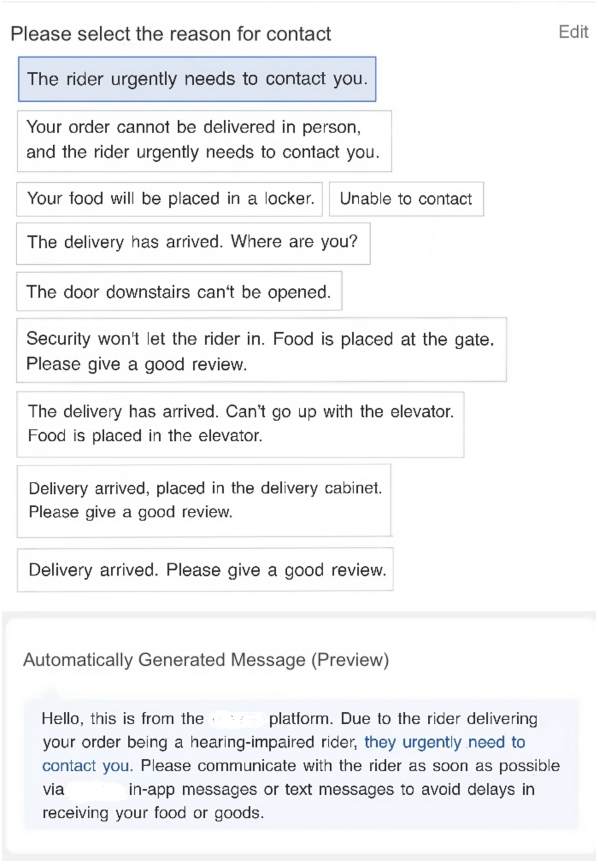
[3] That is, it would work against finding that DHH do well for the Platform in terms of output and profits.

[4] In our data, DHH status is measured as of a person's last record in the dataset. If someone disclosed their DHH status as a result of the AI tool, they would be classified as DHH for the full period of our data. Our effects are identified from within-person changes in outcomes for DHH workers.

Figure E.1: Screenshot of the Intelligent Outbound Call Feature



(a) Default Prompt Menu



(b) Menu with Customized Additions

Notes: Panel (a) shows the default prompt menu for the Intelligent Outbound Call feature. Panel (b) shows the menu with customized additions from one of the DHH workers we worked with, illustrating prompts they added alongside the defaults. The request for "Please give a good review" is likely a polite way of asking to avoid a bad rating.

resolutions and fewer misunderstandings.

- Enhanced Efficiency — By addressing communication bottlenecks through automated messages, workers with disabilities can complete more deliveries with fewer delays, significantly improving productivity.
- Flexible and Personalized Communication — Workers can easily adapt their messages to specific circumstances, clearly communicate unique instructions, or politely request customer assistance or feedback, thereby enhancing the overall delivery service quality.

In essence, the IOC technology represents a significant advancement in inclusive communication practices within delivery services. It empowers workers with disabilities to operate efficiently and confidently while ensuring high-quality service for customers.

F Qualitative Data

F.1 Interviews

We sought to better understand how DHH workers interact with the AI tool, how frequently they use it, and, more broadly, their daily routines, work practices, and the challenges they encounter. Thus, we conducted a group interview in summer 2024 with seven DHH workers, all team workers from a station in Hangzhou. Workers were recruited through the station manager, who asked for volunteers to discuss the experiences of DHH workers in food delivery. To facilitate communication with DHH workers, we hired a translator proficient in sign language. Workers were seated around a long table. Two coauthors took notes, along with representatives of the Platform. Workers received compensation from the Platform, around 300-400 CNY for their time. The group interview lasted around 1.5 hours.

Table F.1 summarizes the demographics of the workers interviewed. As shown in Table 1, most workers on the Platform, both DHH and non-disabled, are men. Our interview sample also consists entirely of men, but it is somewhat younger than the overall DHH worker population on the Platform. We therefore view it as broadly similar in gender composition, while noting that it skews younger in age. The DHH workers interviewed varied in severity — some had profound hearing loss and some had non-profound hearing loss — though we did not record the exact severity level for each interviewee. These workers had disclosed their disability to the Platform.

None of the 7 drivers wore hearing aids to the interview. This is consistent with our general observation that drivers on the Platform do not typically wear hearing aids while working. This is for multiple reasons, including: (1) hearing aids being costly, (2) hearing aids amplifying background noise (e.g., wind and traffic) and being uncomfortable while working, and (3) hearing aids being less helpful for some drivers with profound hearing loss.

AI tool usage and user experience. During the interviews, a representative of the Platform asked if workers were aware of the AI-based IOC tool and did they use it. All workers answered Yes to both questions. DHH workers were also asked what they thought about the tool, and they reported that the tool is useful for them.

Table F.1: Summary of Interviewee Demographics

Age	Hometown	Tenure on Platform	Marital Status	Gender
22	Fujian Province	1+ years	Single	Male
26	Chongqing Municipality	3 years	Single	Male
27	Chongqing Municipality	4 years	Single	Male
24	Henan Province	3 months	Single	Male
25	Henan Province	3+ months	Single	Male
24	Henan Province	1+ years	Single	Male
23	Henan Province	1+ years	Single	Male

DHH workers reported that the AI tool was especially useful for questions of where to leave the food, how to get into the building, and whether to knock on the customer's door (e.g., some customers could penalize drivers for unwanted knocking with a bad rating). This reduced misunderstandings and incidents, and it also reduced time waiting around for a customer's response. Workers are often quite anxious about bad ratings and late deliveries, and helping with these outcomes makes workers feel more comfortable and reduce stress related to these outcomes.

While some drivers interviewed had joined the Platform after the introduction of the AI tool, two drivers had started pre-AI. These drivers reported that there were significant bottlenecks pre-AI, such as not being able to reach out to customers upon delivery.

Outside job opportunities . The interviews suggested to us that DHH workers often lack strong outside options. When we asked the DHH workers interviewed about what other jobs they could do, they mentioned jobs such as assembly line work and food service work. Some explicitly said that there were not many other jobs available to them. A few also mentioned that they hoped to save money from food-delivery work so that they could eventually open a small business, such as a grocery store, in their hometown.

In addition to asking about AI usage and outside options, we list five other main questions we asked, as well as summarize the answers of the 7 DHH worker respondents. These are based on the notes taken by coauthors.

Q: What specific challenges have you encountered during deliveries?
A: Workers highlighted several challenges:

- One worker described an incident where a customer interaction required mediation by the station manager due to communication barriers. Despite the manager's assistance, the worker received a negative review, significantly affecting his morale and leading him to consider leaving the job.
- Workers reported examples of situations where customers left bad ratings after communication issues, citing reasons such as "not calling" or "not speaking clearly."

- Another worker expressed frustration with having multiple deliveries classified as significantly late deliveries (i.e., beyond the expected time windows). Such labels negatively impacted his performance record.

- A worker mentioned taking proactive measures during deliveries to reduce confusion, such as sending a pre-written explanatory message to customers with the station's contact information to address issues caused by delays or miscommunications.

Q: During pickup, do you encounter communication barriers?

A: In general, DHH workers felt much better about pickup relative to delivery. However, workers reported experiencing occasional communication barriers during pickups. Four respondents described needing to repeatedly remind restaurant staff—often two to three times—due to communication difficulties, while non-disabled riders typically resolve issues swiftly by shouting a single instruction that kitchen staff immediately understand. To remind kitchen staff about orders they needed to get, some DHH workers would bang on a counter and show their phones to the workers. Some workers perceived that non-disabled riders tend to receive meals faster because their vocal presence exerts greater urgency on kitchen staff.[5]

Q: Have you experienced forced stops during deliveries?

A: Four workers reported that switching between Platform apps, such as moving from navigation to the order interface, frequently results in forced stops. One rider specifically noted that managing two phones simultaneously—one dedicated to navigation and another to orders—necessitated frequent stops, causing significant disruptions.

Q: Do you receive advance notifications about potentially overdue orders?

A: All workers acknowledged receiving in-app reminders, but due to the nature and pace of their work, they rarely had sufficient time to read or react to these notifications.

Q: Do you have access to rest facilities or water during work?

A: Workers reported using specialized restroom-locating maps to find nearby public facilities for quick rest and water. They seldom return to the station for breaks.

F.2 Fieldwork

To deepen our understanding of challenges faced by DHH delivery workers, we undertook immersive fieldwork alongside them. Since the field notes cover some topics not relevant to the present research (e.g., how workers swap electric motorcycle batteries, discounts restaurants offer to delivery workers), for brevity, we omit some of these from the Appendix. We conducted two practical explorations:

1. One coauthor spent several days directly observing 3 DHH workers by accompanying them full-time, seated on the back seat of an electric motorcycle during their delivery shifts. This hands-on approach enabled us to closely observe and document DHH workers' AI use, daily routines, and difficulties they encounter. The first was a pilot,

[5]While workers perceived communication barriers during pickup, this did not translate into a statistically significant pickup-time difference in our quantitative analysis (see Figure 1).

during which we worked for around an hour as a starting point, whereas the other two sessions lasted about 11-12 hours each.

2. To gain first-hand insight into the communication challenges faced by DHH workers, a different coauthor conducted his delivery work wearing earplugs to mimic hearing impairment (see Panel (b) of Figure F.1), at the invitation and encouragement of the station manager.

While waiting for orders during off-peak hours, workers would often gather in certain areas, such as a specific place in the plaza. Among the DHH workers we met there, both gig workers and team workers had disclosed their disability status to the Platform.

In the text that follows, the first person "I" refers to the perspective of a coauthor instead of a subject. All photographs were taken by the author group. Images were converted to black-and-white to anonymize the color of the Platform's uniforms, and all Platform-identifying information was removed to preserve anonymity. To protect their anonymity, worker faces are obscured, whereas coauthor faces are retained.

Fieldwork Records, July 1, 2024, Hangzhou

Subject: Male, 26 years old, 3 years of employment, non-profound hearing loss
Duration: Around 1 hour (3:00 PM to 4:00 PM); 3 orders delivered
Observations:

1. During the pilot, we observed that upon arrival at, or near, the delivery destination, the worker usually used the AI tool to place a call to the customer. Sometimes the screen displayed a notification that the customer did not answer the call; in those situations, the worker repeatedly used the function "initiate IOC" to send the message again.

2. In cases where there were multiple orders to be delivered to one place, the worker could use the AI tool to simultaneously send out IOCs for multiple orders.

3. There was one case in which we noticed that the worker paid particular attention to traffic safety. During a delivery, we were about to enter a small tunnel. The worker stopped at the entrance, honked the horn first, even though he himself could not hear it, and then entered the tunnel cautiously.

Fieldwork Records, July 2, 2024, Hangzhou

Subject: Male, 27 years old, 3 years of employment, non-profound hearing loss
Duration: 10 hours 50 minutes (10:00 AM – 8:50 PM), 33 orders delivered
Observations:

1. In total, the worker delivered around 33 orders during the hours observed, and I observed that he used the AI tool for all deliveries. He was quite familiar with the tool and used it smoothly. I also briefly discussed his experience with the AI tool with him, communicating through a chat app similar to WhatsApp. The worker explained

Figure F.1: Immersive Fieldwork with Delivery Workers



(a) Coauthor accompanying a DHH worker during deliveries



(b) Coauthor delivering orders with earplugs to simulate hearing impairment

that communicating with customers at the final delivery stage had become much more effective. In most cases, after using the intelligent outbound calling feature, he only needed to drop off the food at the designated location, without further interaction or communication.

2. Delivery workers tend to gather at a common meeting point after completing their orders. For instance, we were near Hangzhou 303 Life Plaza, where many workers frequently returned. Workers tend to avoid orders exceeding 4 kilometers, and after delivering such "distant orders," they usually navigate back to common gathering points.

3. A common tactic I noticed was workers pressing multiple elevator buttons when making deliveries. For instance, if a worker had to deliver to the 20th floor, he would also press buttons for the 21st and 22nd floors. This strategy allows them to reduce waiting time on the way down.

4. Over the course of a day, I noticed that many of the deliveries were from the same restaurants or to the same apartment complexes. The worker often made repeated trips to the same locations throughout the day.

5. During the lunchtime peak hours, the worker was handling six orders simultaneously, which led to a delay in one of the deliveries. After this delay, the number of orders assigned to them decreased significantly. One reason for the delay was that the second order was held up due to the restaurant's slow preparation and packaging process. The authors confirmed with the platform that no additional time is granted for subsequent orders the worker is simultaneously carrying in such situations. Another cause was the worker's decision to wait for the delayed order instead of collecting food for the other orders from nearby restaurants in the same mall. The worker could have optimized their pickup sequence to avoid unnecessary waiting.

6. The top-performing worker at this station managed to complete approximately 2,823 orders in a month, earning around 20,000 CNY ($2,800 USD). However, this required working over 16 hours a day. At times, the worker would fall asleep on their electric motorcycle between orders, waking up only when the next one was assigned. Occasionally, they would even fall asleep while waiting at traffic lights.

Fieldwork Records, July 7, 2024, Hangzhou

Subject: Male, 24 years old, 1 year of employment, profound hearing loss
Duration: 12 hours 20 minutes (7:45AM – 8:05PM), 75 orders delivered
Observations:

1. Among all the orders delivered, I observed that the worker used the AI tool for most orders. For some orders, customers had left specific remarks when placing the order, such as "no need to contact me, just leave the order at the front door." (around 2 - 5% based on my observation). In these cases, the worker did not use the IOC feature.

Instead, he simply took a photo of the food upon delivery and uploaded it to the app as proof of delivery.

2. At 7:45 AM, we started the day. The worker provided me with a full set of sun protection gear, including a sun-protection suit and a neck gaiter. The temperature that day ranged from 38-40C. Figure F.3 shows the equipment required for working outdoors.

3. Around 9 AM, we received an order for a large watermelon. However, the mall where the order was placed didn't open until 10 AM, so we had to pick it up from a locker in the underground parking lot. The security guard wouldn't allow us to ride the motorcycle down to the parking lot, so we had to run from the garage entrance to the basement, grab the watermelon, and run back up to avoid a late delivery.

4. Later in the morning, we received an order for McDonald's. The Platform indicated that the food should be picked up from a collection locker. When we entered the store, a staff member in the kitchen shouted, "The food is on the counter," but the worker didn't hear and went straight to the locker. After unsuccessfully entering the pick-up code several times, the worker finally went to the counter to communicate with the staff and retrieve the meal.

5. The intelligent outbound call is frequently used when workers request customers to open doors, and it is generally very effective. However, there was one exception during our deliveries. In this case, when we arrived at the destination, the worker followed the customer's instructions to contact the apartment complex management via the intercom. After the system connected, no one spoke on the other side. Although I heard the door unlocking sound, the door did not automatically pop open, and there was no visible indication that it had unlocked. The worker continued waiting by the intercom until I eventually walked over and pulled the door open.

6. One order was delivered to an urban village [chengzhongcun in Mandarin], but the address was inaccurate. Upon arriving, the worker used the intelligent outbound call to contact the customer, who messaged back, "I'm upstairs." However, it was unclear what the customer meant by "upstairs." I then heard the customer shouting from inside a factory behind us, but by that time, the worker had already started riding in the opposite direction. I stopped the worker, informing him that the customer was inside the factory, and we managed to find the customer. In such situations, a worker with a hearing disability could waste significant time locating the exact address. Later that afternoon, we encountered a similar issue with an order delivered to a construction site.

7. There was another incident where a restaurant incorrectly packed an order. The customer had ordered noodles and fresh tomatoes, but the restaurant forgot to include the tomatoes. The restaurant called the worker during delivery; as the worker could not hear the caller, the phone was passed to the accompanying coauthor to handle the conversation. We were already halfway to the destination, and returning would have delayed the delivery, so I explained to the restaurant that we were near the destination

and couldn't return, and suggested they contact the customer directly to explain the situation.

8. In general, when a situation requires in-depth communication with restaurants or customers, a worker with a hearing disability still faces difficulties and takes significantly more time to resolve the issue compared to a non-disabled worker. While these types of problems are difficult to address, they occur rarely. In most cases, the intelligent outbound calls are sufficient to assist workers.

9. The worker I worked with mentioned, "I typically start running orders at 7 AM and take a break around 3 PM. Then I go back online from 4 PM until 8 or 9 PM." My observations and experiences working alongside different full-time workers were similar. According to them, to earn a decent income, they need to work 10-13 hours a day.

10. Below is a summary of orders that I found particularly difficult for workers to deliver:

    - Some residential complexes have very slow elevators (there are many residents, but very few elevators).
    - Some older complexes do not have elevators.
    - Security guards at certain complexes do not allow workers to enter with electric motorcycles.
    - Apartments in large residential complexes can be difficult to navigate.
    - Delivering to urban villages is particularly challenging, as the locations are difficult to find and often require climbing stairs without elevators.

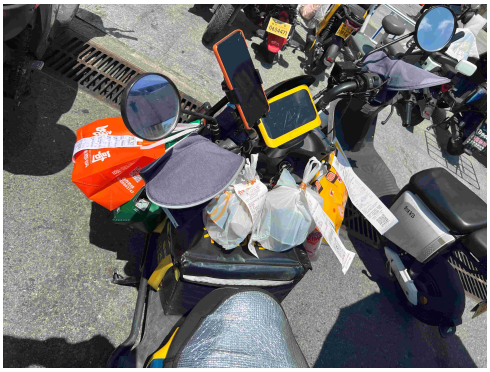
Fieldwork Records, July 2024, Hangzhou

Subject: None. Coauthor delivering by himself wearing earplugs
Duration: Part-time, multiple days; 3–5 orders per session
Observations: For the coauthor who delivered orders while wearing earplugs, he worked part-time over multiple days and delivered 3–5 orders each time entirely on his own. He used the AI tool for all orders and found it highly effective.

Because customers expect to receive phone calls upon delivery, the inability to communicate by phone can create significant difficulties for DHH workers. Thus, to better understand this challenge, on some early orders the coauthor also tried both in-app messaging and direct phone calls without speaking, intentionally simulating the communication constraints faced by DHH workers before the AI implementation. In-app messages were often ignored by customers or answered only after a delay, usually several minutes, which was particularly frustrating because the worker needed to continue delivering other orders. Direct phone calls without speaking were also uncomfortable for the coauthor. Customers would often say "hello?" multiple times and, based on their tone, appeared to become increasingly impatient or irritated. After AI implementation, the direct in-app phone call feature was replaced by the AI tool for DHH workers; thus, the coauthor used direct phone calls only to understand the communication environment faced by DHH workers before the AI tool became available.

In this context, the IOC feature was very helpful and easy to use at the delivery stage. As a new worker, the coauthor was unfamiliar with many delivery destinations and often arrived close to the expected arrival time. As a result, saving even one minute, or sometimes just a few seconds, by using the AI tool instead of waiting for in-app message responses or making direct phone calls significantly reduced stress and helped prevent late deliveries.

Figure F.2: Concurrent Orders During Peak Hours



Notes: This figure illustrates multiple orders being simultaneously handled by a delivery worker during peak hours. Experienced workers typically manage 6-10 concurrent orders during peak hours. The maximum number of orders a rider can handle simultaneously is set by the Platform's algorithm, considering factors like tenure and past performance. Further details on the assignment algorithm are discussed in Section 1.

Figure F.3: Essential Equipment for Outdoor Work



Notes: Figure F.3 illustrates the necessary protective gear for outdoor delivery work, including a sun-protection suit, helmet, and neck gaiter. Typical summer temperatures in the focal city range from 38C-40C. Complete sun-protection equipment is essential, and workers must remain well-hydrated to avoid heat-related illnesses.

Appendix References

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Westberry, "IPUMS CPS: Version 12.0 [dataset]," 2024. `https://doi.org/10.18128/D030.V12.0`.