Online Appendix to Effects of Rescheduling on Patient No-show Behavior in Outpatient Clinics

Due to space limit, complete online appendix can be downloaded from SSRN at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2783646`.

A Full Estimation Results

Please refer to our SSRN appendix for all the full estimation results.

B IV Specification and Validation

B.1 IV for Active Rescheduling

To formally define our IV for active rescheduling, we estimate the daily capacity, C_t, by counting the arrivals and no-shows for each clinic[1] on day t, since we do not have exact data on the physicians' shift. The number of available slots for day t observed on day s ($s < t$), denoted as $A_{t,s}$, is then computed as the difference between the capacity of day t and the number of patients already scheduled by day s for day t. Next, the *appointment availability* for day t observed on day s, denoted as $AVLB_{t,s}$, is defined as the ratio between the number of available slots and the capacity, i.e., $AVLB_{t,s} = A_{t,s}/C_t$. For each appointment, we use the same weekday 2 weeks before and after the appointment date (recall the example above), and compute the appointment availability for the 5 days (i.e., t = September 1, 8, 15, 22, and 29) 2 weeks before the appointment date (i.e., s = September 1). The *average appointment availability*, $\overline{APPAVLB}$, is then computed across these five dates. Finally, we define $APPAVLB$ as an indicator variable that equals 1 when average appointment availability is higher than 20% and 0 otherwise. Intuitively, $APPAVLB = 0$ represents the case in which the clinic is relatively congested on one's preferred days with low appointment availability.

[1]The computations hereafter are all done within the same clinic, so we omit the superscript or subscript for clinic.

B.2 Validity Tests

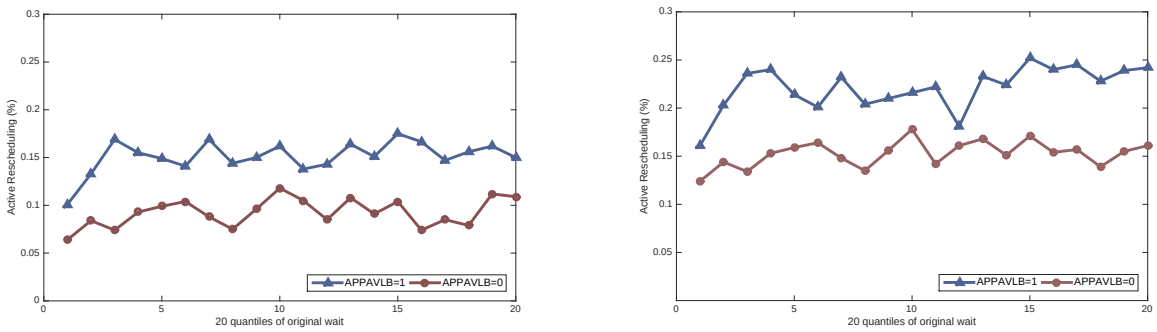
The appropriateness of $APPAVLB$ as an IV for active scheduling must be tested to satisfy two conditions: (i) it is correlated with patients' decision to reschedule, and (ii) it is independent of unobservable factors that may influence no-show behavior.

We first test the condition that $APPAVLB$ is correlated with the decision to actively reschedule. In Figure 1, we divide patients into 20 groups based on the lengths of their original waits and compare the percentage of active reschedulings by patients who observed a high level of appointment availability versus those who observed a low level. We find that based on this coarse comparison, greater appointment availability is associated with an increased fraction of rescheduled appointments. Using a series of Probit regression models, we also find, at the 0.1% significance level, that patients are more likely to reschedule their appointments when appointment availability is high.

Furthermore, the second condition requires that a valid IV is uncorrelated with *unobserved* factors that may influence no-show behavior. While this condition cannot be test statistically, we provide some evidence for the validity of IV by demonstrating that *observed* factors such as patients' age and treatment cycle (which could be considered as proxies for their severity) are equally distributed between the two levels of $APPAVLB$ (0 and 1). We also use a patient's previous rescheduling records as a proxy for the patient's attitude/personality and compute its correlation with the instrument. The resulting correlation coefficient is not significantly different from zero.

Figure 1: Appointment availability and active rescheduling

(a) New patients **(b) Follow-up patients**



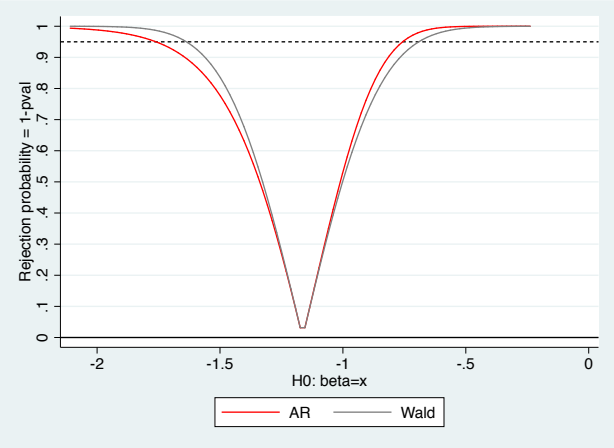
To perform formal tests for underidentification and weak identification, we note that the majority of "weak IV" tests are based on a linear IV regression model where the dependent variable in the outcome equation and the endogenous variable are continuous. Following Freeman et al. (2016), we first treat both no-show and active rescheduling as continuous and estimate the model via *ivreg2* command in Stata 14.0 (Baum et al. 2002). Note that the coefficients estimated using continuous model specification are qualitatively consistent with our main results.

For the above model, the first-stage Wald statistics is 8.30 (p-value = 0.0040). This shows strong evidence to reject the null hypothesis of under-identification at the 1% significance level and therefore the excluded instrument can be considered as "relevant". Turning next to the issue of weak identification, Stock and Yogo (2005) tabulated critical values for the first-stage F-statistic to test whether instruments are weak. For a single endogenous regressor,

assuming the model to be estimated under limited information maximum likelihood, the critical F-values are 16.38, 8.96, and 6.66 for maximum biases of 10%, 15%, and 20%, respectively. In our case, the estimated F-statistic is 46.92, indicating a maximum bias of lower than 10%.

We then create a confidence interval and p-values based on inverting the Anderson-Rubin test statistic (Mikusheva et al. 2006, Finlay et al. 2016). This confidence interval is robust to weak instruments and optimal in the case of a single instrument (Moreira 2009). Figure 2 shows the confidence intervals under Wald and Anderson-Rubin test statistics: A 95% Anderson-Rubin confidence interval excludes zero, indicating statistical significance at the 5% level. This greatly alleviates our concern of weak instrument.

Figure 2: Anderson-Rubin confidence interval



C Robustness Checks

We present the detailed analysis of robustness check in this section.

C.1 Propensity Score Matching

One could argue that patients who choose to reschedule their appointments may self-select to be different from those who do not reschedule. For instance, patients' active rescheduling may signal their strong willingness to attend the appointments, and this self-selection effect could confound our results. To reduce the bias, we use the propensity score matching (PSM) method (Rosenbaum and Rubin 1983) to estimate the effect of rescheduling. Please refer to our SSRN appendix for more details of this approach.

Table 1 shows the estimated average treatment effect on the treated (ATT) of rescheduling for both new patients and follow-ups. For new patients, the impact of active rescheduling is reduced significantly after matching. For follow-up patients, the ATT after matching is significant at the 1% level across all three matching methods, varying from -0.097 to -0.107. Passive rescheduling—after matching—increases no-show probability by 5.2 to 5.8 percentage for follow-up patients, while the effects are insignificant for new patients.

Table 1: Estimation results using the PSM approach

	New		Follow-up	
	Active	Passive	Active	Passive
Pre-matching	$-0.075^{***}(0.007)$	$-0.124^{***}(0.004)$	0.035 (0.023)	$0.069^{***}(0.014)$
Nearest-neighbor matching	-0.025^{**} (0.011)	$-0.107^{***}(0.005)$	0.027 (0.033)	$0.058^{***}(0.020)$
Radius matching	$-0.022^{***}(0.008)$	$-0.097^{***}(0.006)$	0.022 (0.029)	$0.053^{***}(0.021)$
Kernel matching	-0.021^{**} (0.009)	$-0.104^{***}(0.005)$	0.021 (0.031)	0.052^{**} (0.024)

Notes. The estimated coefficients are average treatment effect on the treated (ATT). Standard errors are shown in parentheses.
"Pre-matching" refers to the sample without matching the active group and control group.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

These results are consistent with our main findings. In addition, we add the propensity scores as a control variable to the bivariate probit model used in the paper, and find that the main results remain qualitatively the same even when this new control variable is included.

C.2 Near-Far Matching

Since Propensity score mataching is solely based on observable characteristics, one might be also concerned about the unobserved heterogeneity. In this section we draw upon the literature on design of observational studies (Rosenbaum 2010) and use recent advancements in the methodology of near-far matching (Baiocchi et al. 2010, Hu et al. 2017).

In instrumental variable settings, the goal of matching is to find a matched sample that is balanced on the observed covariates and imbalanced (or separated) on the instrument. By combining IV with this matched-pairs design, we improve the equality of matched groups, and thus reduce model dependence and potentially strengthen the instrument (see Lu et al. 2011, Lorch et al. 2012, Yang et al. 2014). Please refer to our SSRN appendix for more details of this approach.

Table 2 summarizes the estimation results after near-far matching. For follow-up patients, no-show probability decreases by 10.5 percentage points if the appointment is actively rescheduled, while for new patients, the reduction in no-show probability is only 1.7 percentage points. These results are consistent with our main analysis.

Table 2: Estimation results using the strengthened IV after matching

NS	IV	AME on $Pr(Active)$	ρ	Active	AME on $Pr(NS)$
New	$0.67^{***}(0.022)$	$0.35^{***}(0.007)$	$-0.025(0.036)$	$-0.078^{***}(0.026)$	$-0.017^{***}(0.006)$
Follow-up	$0.82^{***}(0.019)$	$0.39^{***}(0.005)$	$0.021(0.022)$	$-0.478^{***}(0.075)$	$-0.105^{***}(0.008)$

Notes. AME is the average marginal effect. Standard errors are shown in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C.3 Alternative IV Specification

Similar to Kim et al. (2014), Hu et al. (2017), Chan et al. (2017), we use the binary instrument variable to handle the endogeniety problem in our main paper. We now consider different

IV specifications for active rescheduling, $APPAVLB$.

Specifically, we consider: (1) estimation of appointment availability at 3 days and 1 week before the original appointment date and (2) the use of different appointment availability thresholds (15% and 25%). The results are qualitatively robust over these alternative cut-off values.

We also conduct the same analysis using a continuous version of the IV, i.e., $\overline{APPAVLB}$. The results are presented in Table 3 together with the original indicator version. We can observe that the estimated AMEs are similar from these two versions.

Table 3: Continuous and indicator IVs

	IV	Model	Estimate	AME	ρ	Test $\rho = 0$
New	No IV	Probit	-0.695^{***} (0.080)	-0.095		
	Continuous IV	BiProbit	-0.123^{*} (0.068)	-0.062	-0.027^{**} (0.011)	0.00
	Indicator IV	BiProbit	-0.108^{***} (0.035)	-0.056	-0.053^{***} (0.014)	0.01
Follow-up	No IV	Probit	-0.745^{***} (0.024)	-0.121		
	Continuous IV	BiProbit	-0.103^{***} (0.031)	-0.083	-0.025^{*} (0.014)	0.07
	Indicator IV	BiProbit	-0.485^{***} (0.068)	-0.109	-0.081^{**} (0.036)	0.03

Notes. AME is the average marginal effect. Standard errors are shown in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C.4 Other Tests and Alternative Explanations

We explore different definitions of no-show by treating cancellations and reschedulings that occur within 1 day of the original appointment date as no-shows, because a 1-day period is generally inadequate to rebook the freed up slot. We repeat the analytical procedures, and key results remain consistent. We also relabel cancellations and reschedulings that occur within other periods before the original appointment (2-5 days) as no-shows, and qualitative results are unchanged. In addition, we analyze different subsamples—high-volume specialties, appointments with the most common treatment cycles, different age groups, and different government-subsidy conditions—and our main results still hold.

Two alternative explanations for our results are considered: patients who actively rescheduled their appointments could be those who had more severe conditions, or those who were getting worse and thus called for a more urgent treatment. We test and rule out the two possibilities by examining the treatment cycles and days changed, respectively.

Please refer to our SSRN appendix for more results and discussion on these robustness tests and alternative explanations.

D Additional Simulations

In this section, we present additional counterfactual analysis using our simulation model. First, we investigate what occurs if the mix of new and follow-up patients changes. Recall that in our data, 75% of clinic patients are follow-ups. Tables 4 and 5 show the results when the percentage of follow-up patients decreases to 60% and increases to 90%, respectively.

In general, the key trade-off between allowing and not allowing active rescheduling is consistent with the case with 75% follow-up patients. When the number of follow-up patients

Table 4: Percentage of follow-up patients is 60%

	Rescheduling	No rescheduling	Rescheduling ahead 1 week	Rescheduling ahead 2 weeks	Increse capacity (5%)
Arrival	52.08% (51.70%, 52.46%)	55.66% (55.35%, 55.98%)	53.28% (52.97%, 53.6%)	53.51% (53.27%, 53.75%)	53.14% (52.89%, 53.40%)
No-show	20.13% (19.75%, 20.51%)	21.65% (21.32%, 21.98%)	19.47% (19.23%, 19.71%)	19.26% (19.07%, 19.45%)	19.65% (19.44%, 19.85%)
Cancellation	10.88% (10.75%, 11.01%)	12.62% (12.38%, 12.85%)	11.29% (11.09%, 11.5%)	11.17% (11.02%, 11.32%)	11.21% (11.09%, 11.33%)
Rescheduling	16.91% (16.70%, 17.12%)	10.03% (9.88%, 10.19%)	15.96% (15.73%, 16.18%)	16.06% (15.93%, 16.19%)	16.00% (15.83%, 16.17%)
- Active	7.39% (7.21%, 7.57%)	0% (0%, 0%)	6.35% (6.18%, 6.53%)	6.78% (6.64%, 6.91%)	7.08% (6.96%, 7.19%)
- Passive	9.52% (9.38%, 9.66%)	10.03% (9.88%, 10.19%)	9.60% (9.46%, 9.74%)	9.29% (9.16%, 9.41%)	8.93% (8.77%, 9.08%)
Average wait time					
All	89.48 (88.98, 89.98)	84.78 (84.38, 85.18)	85.99 (85.65, 86.33)	85.35 (84.93, 85.77)	85.58 (85.19, 85.97)
New	52.21 (51.26, 53.15)	43.42 (42.73, 44.11)	45.24 (44.43, 46.06)	42.99 (42.40, 43.59)	42.36 (41.80, 42.93)
Follow-up	117.73 (117.36, 118.09)	116.01 (115.61, 116.41)	116.42 (116.1, 116.75)	116.61 (116.06, 117.16)	118.09 (117.71, 118.46)
Utilization	87.58% (87.03%, 88.13%)	86.59% (86.08%, 87.11%)	88.84% (88.36%, 89.33%)	89.20% (88.80%, 89.60%)	84.85% (84.48%, 85.22%)

Notes. Numbers inside the parentheses indicate 95% confidence intervals.

Table 5: Percentage of follow-up patients is 90%

	Rescheduling	No rescheduling	Rescheduling ahead 1 week	Rescheduling ahead 2 weeks	Increse capacity (5%)
Arrival	51.63% (51.52%, 51.74%)	55.04% (54.92%, 55.17%)	52.13% (52.04%, 52.22%)	52.26% (52.16%, 52.36%)	52.03% (51.96%, 52.11%)
No-show	17.64% (17.56%, 17.73%)	19.21% (19.11%, 19.31%)	17.12% (17.05%, 17.19%)	16.97% (16.88%, 17.05%)	17.33% (17.26%, 17.40%)
Cancellation	13.37% (13.32%, 13.43%)	14.88% (14.8%, 14.96%)	13.35% (13.29%, 13.42%)	13.37% (13.29%, 13.45%)	13.26% (13.19%, 13.32%)
Rescheduling	17.36% (17.29%, 17.42%)	10.83% (10.77%, 10.89%)	17.39% (17.32%, 17.46%)	17.40% (17.34%, 17.47%)	17.38% (17.33%, 17.43%)
- Active	7.25% (7.20%, 7.30%)	0% (0%, 0%)	7.36% (7.30%, 7.42%)	7.44% (7.39%, 7.49%)	7.33% (7.28%, 7.38%)
- Passive	10.11% (10.04%, 10.17%)	10.83% (10.77%, 10.89%)	10.03% (9.97%, 10.09%)	9.97% (9.92%, 10.02%)	10.05% (10.00%, 10.10%)
Average wait time					
All	107.55 (107.25, 107.84)	105.35 (105.03, 105.67)	105.56 (105.28, 105.83)	105.79 (105.46, 106.11)	106.94 (106.62, 107.26)
New	27.03 (26.70, 27.36)	21.96 (21.59, 22.34)	22.90 (22.53, 23.27)	21.71 (21.38, 22.04)	22.85 (22.51, 23.19)
Follow-up	117.98 (117.66, 118.29)	116.35 (116.02, 116.67)	116.32 (116.02, 116.61)	116.54 (116.18, 116.9)	117.90 (117.56, 118.23)
Utilization	87.40% (87.13%, 87.67%)	86.44% (86.14%, 86.73%)	88.19% (87.98%, 88.39%)	88.49% (88.29%, 88.7%)	84.64% (84.39%, 84.9%)

Notes. Numbers inside the parentheses indicate 95% confidence intervals.

increases, the impact of early active rescheduling is more significant. When the percentage of follow-up patients is 60%, early rescheduling must occur two weeks in advance to achieve similar benefits to increasing capacity by 5%. In contrast, if the percentage of follow-up patients is 90%, the benefits of increasing capacity by 5% can be obtained by ensuring that all active rescheduling occurs only one week ahead.

Next, we investigate what occurs if there is more active rescheduling. From Table 6, when the number of actively rescheduled appointments increases by 50%, the no-show rate decreases from 19.3% (95%CI: 19.26%, 19.35%) to 18.17% (95%CI: 18.02%, 18.33%), while the average wait time for new patients increases from 37.38 days (95%CI: 37.07, 37.74) to 40.54 days (95%CI: 40.05, 41.04). When the number of actively rescheduled appointments doubles, the no-show rate further decreases to 16.93% (95%CI: 16.73%, 17.13%), and the average wait time for new patients increases to 44.62 days (95%CI: 43.86, 45.38). Recall our findings that new patients care more about waiting time and are insensitive to passive rescheduling: As long as their appointments are moved forward, they are more likely to show up, even if they are rescheduled by the clinic. Therefore, we simulate the case in which the clinic proactively offers freed-up slots to new patients with the longest wait times. This provides an optimistic estimate of the benefits of such an allocation strategy (see Table 6). We observe that actively allocating freed-up slots to new patients can indeed reduce their average wait time, and the benefit is more significant when the system has more actively rescheduled appointments.

Table 6: More active rescheduling

	Current System	More rescheduling			
		1.5 times		2 times	
		Without allocation of freed-up slots	With allocation of freed-up slots	Without allocation of freed-up slots	With allocation of freed-up slots
Arrival	52.03% (51.97%, 52.09%)	51.04% (50.88%, 51.2%)	51.06% (50.83%, 51.28%)	50.04% (49.82%, 50.25%)	49.98% (49.8%, 50.17%)
No-show	19.3% (19.26%, 19.35%)	18.17% (18.02%, 18.33%)	18.29% (18.13%, 18.45%)	16.93% (16.74%, 17.13%)	17.20% (17.04%, 17.37%)
Cancellation	11.96% (11.92%, 12.00%)	11.59% (11.5%, 11.68%)	11.41% (11.31%, 11.51%)	11.32% (11.2%, 11.45%)	11.21% (11.09%, 11.32%)
Rescheduling	16.71% (16.68%, 16.74%)	19.20% (19.1%, 19.3%)	19.24% (19.14%, 19.35%)	21.71% (21.58%, 21.84%)	21.61% (21.51%, 21.71%)
- Active	7.45% (7.41%, 7.49%)	10.47% (10.37%, 10.56%)	10.6% (10.5%, 10.7%)	13.41% (13.29%, 13.52%)	13.46% (13.37%, 13.55%)
- Passive	9.26% (9.22%, 9.29%)	8.73% (8.64%, 8.82%)	8.65% (8.56%, 8.73%)	8.30% (8.21%, 8.39%)	8.15% (8.06%, 8.24%)
Average wait time					
All	95.21 (95.02, 95.39)	97.35 (97.06, 97.64)	96.80 (96.46, 97.14)	100.07 (99.70, 100.45)	98.82 (98.46, 99.18)
New	37.38 (37.01, 37.74)	40.54 (40.05, 41.04)	39.62 (39.24, 40)	44.62 (43.86, 45.38)	41.77 (41.19, 42.35)
Follow-up	117.92 (117.76, 118.07)	119.23 (118.90, 119.56)	118.97 (118.58, 119.37)	120.98 (120.67, 121.30)	120.52 (120.07, 120.97)
Utilization	87.32% (87.06%, 87.58%)	88.53% (88.15%, 88.92%)	88.32% (87.97%, 88.67%)	89.65% (89.23%, 90.07%)	89.07% (88.69%, 89.44%)

Notes. Numbers inside the parentheses indicate 95% confidence intervals.

References

Baiocchi, Mike, Dylan S Small, Scott Lorch, Paul R Rosenbaum. 2010. Building a stronger instrument in an observational study of perinatal care for premature infants. *Journal of the American Statistical Association* **105**(492) 1285–1296.

Baum, CF, ME Schaffer, Steven Stillman. 2002. IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation. *Statistical Software Components* .

Chan, Carri W, Linda V Green, Lijian Lu, Suparerk Lekwijit, Gabriel J Escobar. 2017. Assessing the impact of service intensity on customers: An empirical investigation of hospital step-down units. *Working Paper* .

Finlay, Keith, Leandro Magnusson, Mark E Schaffer, et al. 2016. WEAKIV: Stata module to perform weak-instrument-robust tests and confidence intervals for instrumental-variable (IV) estimation of linear, probit and tobit models. *Statistical Software Components* .

Freeman, Michael, Nicos Savva, Stefan Scholtes. 2016. Gatekeepers at work: An empirical analysis of a maternity unit. *Management Science, forthcoming* .

Hu, Wenqi, Carri W Chan, José R Zubizarreta, Gabriel J Escobar. 2017. An examination of early transfers to the ICU based on a physiologic risk score. *Manufacturing & Service Operations Management, forthcoming* .

Kim, Song-Hee, Carri W Chan, Marcelo Olivares, Gabriel Escobar. 2014. ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science* **61**(1) 19–38.

Lorch, Scott A, Michael Baiocchi, Corinne E Ahlberg, Dylan S Small. 2012. The differential impact of delivery hospital on the outcomes of premature infants. *Pediatrics* **130**(2) 270–278.

Lu, Bo, Robert Greevy, Xinyi Xu, Cole Beck. 2011. Optimal nonbipartite matching and its statistical applications. *The American Statistician* **65**(1) 21–30.

Mikusheva, Anna, Brian P Poi, et al. 2006. Tests and confidence sets with correct size when instruments are potentially weak. *Stata Journal* **6**(3) 335–347.

Moreira, Marcelo J. 2009. Tests with correct size when instruments can be arbitrarily weak. *Journal of Econometrics* **152**(2) 131–140.

Rosenbaum, Paul R. 2010. *Design of Observational Studies*. Springer.

Rosenbaum, Paul R, Donald B Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**(1) 41–55.

Stock, James H, Motohiro Yogo. 2005. *Testing for weak instruments in linear IV regression*. Cambridge University Press.

Yang, Fan, José R Zubizarreta, Dylan S Small, Scott Lorch, Paul R Rosenbaum. 2014. Dissonant conclusions when testing the validity of an instrumental variable. *The American Statistician* **68**(4) 253–263.