

Appendix

Table A1: JHIC Compliance Categories with Warnings and Sanctions as per 2016 Regulation

Facility score (% of maximum score)	Compliance category	Warning/sanction/reward
0-10% *(or absence of licenses)	Non-Compliant	Close facility. Recommend prosecution where requisite registration and licensing is absent (as specified by the respective legislation) Closed facilities that have a license, may ask for a re-inspection after corrections for purposes of reopening and re-categorization, subject to the criteria established by the competent authority.
11-40%	Minimally compliant	3-month notice for correction + re-inspection Facility will be closed if it does not score over 40% of the maximum score in the 3 rd inspection. Facility may ask for re-inspection for purposes of re-categorization after corrections
Facility score (% of maximum score)	Compliance category	Warning/sanction/reward
41-60%	Partially Compliant	6-month notice for correction + re-inspection Facility will be closed if it does not score over 60% of the maximum score in the 3 rd inspection. Facility may ask for re-inspection for purposes of re-categorization after corrections
61-75%	Substantially Compliant	12-month notice for correction + re-inspection Facility may ask for re-inspection for purposes of re-categorization after corrections
>75%	Fully Compliant	No re-inspection for two years

* Other incentives under consideration include client signaling (quality mark and/or SMS feedback mechanism) and insurance recommendation

Source. Ministry of Health (2015).

Table A2: Balance Checks

	Unweighted		Weighted		Obs.
	(C) Control Mean	(T-C) Adj. Diff.	(C) Control Mean	(T-C) Adj. Diff.	
	(1)	(2)	(3)	(4)	(5)
Panel A: Balance using baseline sample					
JHIC Score (% of max)	35.539 (10.412)	0.631 (0.947) [0.506]	42.781 (11.906)	0.436 (1.775) [0.806]	1027
OOP (USD PPP)	4.525 (7.136)	1.023* (0.545) [0.062]	3.493 (7.171)	-0.072 (0.578) [0.901]	958
Daily Outpatients	24.817 (30.961)	-0.397 (2.041) [0.846]	24.817 (30.961)	-0.397 (2.041) [0.846]	1025
Compliance with IPC Practices (Patient-HCW indication level)	0.318 (0.466)	-0.001 (0.010) [0.900]	0.198 (0.399)	-0.009 (0.010) [0.367]	105869 (928 HFs)
IPC Knowledge (HCW level)	0.735 (0.098)	0.017*** (0.007) [0.010]	0.732 (0.082)	0.023** (0.009) [0.012]	1624 (972 HFs)
IPC Supplies (Site level)	0.639 (0.188)	-0.003 (0.012) [0.834]	0.621 (0.204)	0.005 (0.014) [0.741]	1885 (1005 HFs)
Public	0.350 (0.478)	-0.024 (0.028) [0.405]	0.350 (0.478)	-0.024 (0.028) [0.405]	1104
Level 2	0.824 (0.382)	0.020 (0.023) [0.389]	0.824 (0.382)	0.020 (0.023) [0.389]	1104
Level 3	0.133 (0.340)	-0.023 (0.022) [0.290]	0.133 (0.340)	-0.023 (0.022) [0.290]	1104
F-test from regression of treatment on all outcome variables listed above			1.020		
P-value			0.425		
Panel B: Balance using randomization sample (select variables)					
Public	0.308 (0.462)	-0.013 (0.024) [0.607]	0.308 (0.462)	-0.013 (0.024) [0.607]	1348
Level 2	0.855 (0.352)	0.007 (0.019) [0.729]	0.855 (0.352)	0.007 (0.019) [0.729]	1348
Level 3	0.108 (0.311)	-0.005 (0.019) [0.811]	0.108 (0.311)	-0.005 (0.019) [0.811]	1348
Unlicensed (Private only)	0.571 (0.496)	-0.055* (0.033) [0.100]	0.571 (0.496)	-0.055* (0.033) [0.100]	944

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Columns 2 and 4 present adjusted differences between the means for the treatment markets and the control group. These differences include controls for the strata included in the randomization (by county and market size). HF = health facility; JHIC = Joint Health Inspection Checklist; OOP = out-of-pocket payments; PPP = purchasing power parity, IPC = infection prevention and control; HCW = health care worker.

Table A3: Baseline Quality Association with OOP, by JHIC Functional Category
(Private Facilities)

	OOP (USD PPP)						
	Infrastructure	Equipment	Supplies (Low Cost)	Supplies (Medium Cost)	Management	Medical Records	SOPs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Item Adherence	0.639*** (0.169)	0.427* (0.228)	0.212 (0.229)	1.995** (0.844)	0.479** (0.226)	1.212* (0.665)	0.442 (0.536)
Unlicensed	-0.467 (0.705)	-0.541 (0.767)	-1.048 (1.084)	-1.356 (1.306)	-0.997 (0.721)	-1.690 (1.116)	-1.028 (1.041)
Mean Adherence (SD)	0.45 (0.50)	0.30 (0.46)	0.40 (0.49)	0.25 (0.43)	0.25 (0.43)	0.47 (0.50)	0.08 (0.27)
Observations	22876	7350	23209	1308	25452	2567	13088
R^2	0.16	0.17	0.16	0.15	0.15	0.12	0.15
Total Controls Selected by PDS (out of 29)	6	6	5	3	8	0	4

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Controls are selected by PDSLASSO out of a list of 29 variables. The indicator for unlicensed at baseline is partialled out (imposed as controls in the regression) and presented in the table. Facility levels and strata FE at baseline are partialled out for all regressions so not included in the list of 29 variables.

Table A4: Government Closures During Implementation and Inactivity at Endline
by Baseline JHIC Score Quintile and License Status at Randomization
(Private Facilities)

	Licensed			Unlicensed				All Private			
	Control		Treatment	Control		Treatment		Control		Treatment	
	Inactive	Closed	Inactive	Closed	Inactive	Closed	Inactive	Closed	Inactive	Closed	Inactive
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
JHIC Quintile											
Lowest	0.27	0.21	0.14	0.04	0.32	0.61	0.45	0.03	0.31	0.49	0.36
2nd	0.08	0.11	0.19	0.00	0.35	0.47	0.33	0.00	0.26	0.33	0.28
3rd	0.04	0.10	0.04	0.00	0.14	0.33	0.29	0.00	0.08	0.21	0.16
4th	0.00	0.03	0.09	0.00	0.11	0.44	0.33	0.00	0.05	0.15	0.16
Top	0.05	0.01	0.04	0.00	0.06	0.11	0.11	0.00	0.05	0.03	0.05
All	0.07	0.07	0.09	0.01	0.23	0.45	0.34	0.01	0.15	0.24	0.20

Notes. JHIC quintile estimated using baseline JHIC scores by treatment group within private facilities. License status as per randomization. The estimates only include facilities for which baseline JHIC score is available. Closure indicates government enforcement of facility closure during the implementation. For Columns 1 to 7, the denominator is the number of private facilities per quintile, treatment group, and license status. There was one facility in the control group that was closed due to contamination as per Table S7 in the [Supplemental Material](#).

Table A5: Treatment Effects on JHIC Score and OOP:
Overall and Interacted with Indicators for Private and Unlicensed Health Facilities at
Randomization

	Unweighted		Weighted	
	JHIC Score (pp of max) (1)	OOP (USD PPP) (2)	JHIC Score (pp of max) (3)	OOP (USD PPP) (4)
Panel A: Overall Impact				
Treatment	5.702*** (0.773) [0.000]	0.946** (0.456) [0.039]	4.177*** (1.257) [0.001]	0.132 (0.579) [0.819]
Observations	1121	1121	1121	1121
R ²	0.315	0.137	0.524	0.186
Control Mean	36.325	3.984	42.765	3.157
Impact: {%; SD}	{16%; 0.54}	{24%; 0.19}	{10%; 0.36}	{4%; 0.03}
Panel B: Interaction with Private				
Treatment	3.111*** (1.080) [0.004]	-0.066 (0.273) [0.809]	3.242** (1.600) [0.044]	0.388 (0.270) [0.152]
Private HF at Randomization	-5.672*** (1.088) [0.000]	4.480*** (0.449) [0.000]	0.137 (2.348) [0.954]	5.681*** (1.078) [0.000]
Private HF at Randomization x T	4.104*** (1.232) [0.001]	1.519** (0.669) [0.024]	3.181 (2.506) [0.205]	-0.036 (1.191) [0.976]
Observations	1121	1121	1121	1121
R ²	0.333	0.232	0.532	0.362
Control Mean Public	39.946	0.658	42.300	0.819
Control Mean Private	34.243	5.897	43.631	7.513
Impact Public: {%; SD}	{8%; 0.35}	{-10%; -0.07}	{8%; 0.35}	{47%; 0.34}
Impact Private: {%; SD}	{21%; 0.67}	{25%; 0.28}	{15%; 0.42}	{5%; 0.06}
Test T + Private x T = 0 (p-value)	0.000	0.020	0.001	0.775
Panel C: Interaction with Unlicensed (Private and active at randomization only)				
Treatment	8.784*** (1.216) [0.000]	1.501* (0.813) [0.066]	8.452*** (2.103) [0.000]	0.036 (1.735) [0.983]
Unlicensed HF at Randomization	-2.185* (1.231) [0.078]	-0.214 (0.644) [0.740]	-1.498 (2.539) [0.556]	-1.955* (1.159) [0.093]
Unlicensed HF at Randomization x T	-3.837*** (1.406) [0.007]	-0.313 (0.739) [0.672]	-5.463* (3.065) [0.076]	0.648 (1.752) [0.712]
Observations	720	720	720	720
R ²	0.381	0.097	0.629	0.082
Control Mean Licensed	37.471	6.652	48.285	8.408
Control Mean Unlicensed	31.356	5.222	36.398	6.123
Impact Licensed: {%; SD}	{23%; 0.80}	{23%; 0.25}	{18%; 0.52}	{0%; 0.01}
Impact Unlicensed: {%; SD}	{16%; 0.50}	{23%; 0.27}	{8%; 0.28}	{11%; 0.16}
Test T + Unlicensed x T = 0 (p-value)	0.000	0.065	0.076	0.436

Notes. Robust standard errors reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Regressions include randomization strata controls (by county and market size) and health facility level controls. HF = health facility; JHIC = Joint Health Inspection Checklist; OOP = out-of-pocket; PPP = purchasing power parity.

Table A6: Average and Unconditional Quantile Treatment Effects on Outpatient Wealth Index

	Wealth Index (1)	Observations (2)
Treatment	−0.036 (0.171) [0.834]	10957
QTE		
20th	0.018 (0.040) [0.650]	10957
40th	0.028 (0.055) [0.605]	10957
60th	−0.080 (0.072) [0.265]	10957
80th	−0.009 (0.083) [0.915]	10957

Notes. Robust standard errors are reported in parentheses and p-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Wealth Index is constructed with a subset of variables taken from DHS. Adjusted with patient sampling weights. Regression controls for facility levels and strata.

Table A7: Treatment Effects on Quality Indicators Not Included in the JHIC,
Infection Prevention and Control:
Overall and Interacted with Indicators for Private and Unlicensed Health Facilities at Endline

	Practice, Knowledge, and Supplies in IPC		
	Practice (Patient-HCW indication level)	Knowledge (HCW level)	Supplies (Site level)
	(1)	(2)	(3)
Panel A: Overall Impact			
Treatment	-0.006 (0.010) {1.000}	-0.001 (0.001) {1.000}	0.001 (0.002) {1.000}
Observations	104565	2098	2644
R ²	0.011	0.383	0.182
Control Mean	0.336	0.764	0.737
Impact: {%; SD}	{-2%; -0.01}	{-0%; -0.04}	{0%; 0.03}
Panel B: Interaction with Private			
Treatment	-0.012 (0.013) {0.514}	-0.002 (0.001) {0.514}	-0.003 (0.002) {0.514}
Private HF	0.014 (0.017) [0.423]	-0.007*** (0.001) [0.000]	-0.021*** (0.003) [0.000]
Private HF x T	0.015 (0.018) {0.514}	0.002 (0.002) {0.514}	0.005 (0.003) {0.514}
Observations	104565	2098	2644
R ²	0.012	0.418	0.260
Control Mean Public	0.336	0.769	0.753
Control Mean Private	0.335	0.761	0.728
Impact Public: {%; SD}	{-3%; -0.02}	{-0%; -0.10}	{-0%; -0.13}
Impact Private: {%; SD}	{1%; 0.01}	{0%; 0.00}	{0%; 0.09}
Test T + Private x T = 0 (p-value)	0.810	0.989	0.332
Panel C: Interaction with Unlicensed (Private and active at endline only)			
Treatment	-0.000 (0.015) {1.000}	0.000 (0.002) {1.000}	0.003 (0.004) {1.000}
Unlicensed HF at Endline	-0.022 (0.017) [0.189]	-0.001 (0.001) [0.535]	-0.003 (0.004) [0.506]
Unlicensed HF at Endline x T	0.015 (0.020) {1.000}	-0.001 (0.002) {1.000}	-0.002 (0.005) {1.000}
Observations	42505	1302	1711
R ²	0.018	0.254	0.178
Control Mean Licensed	0.351	0.762	0.732
Control Mean Unlicensed	0.310	0.759	0.723
Impact Licensed: {%; SD}	{-0%; -0.00}	{0%; 0.03}	{0%; 0.11}
Impact Unlicensed: {%; SD}	{5%; 0.03}	{-0%; -0.04}	{0%; 0.06}
Test T + Unlicensed x T = 0 (p-value)	0.395	0.711	0.608

Notes. Robust standard errors are reported in parentheses and clustered at the market level. *** (**) (*) denotes significance at 1% (5%) (10%) level. Stars reported next to the estimated coefficients denote significance related to the “naive” p-value. Sharpened q-values are reported in braces, following [Benjamini et al. \(2006\)](#), with stars next to the braces. Compliance means are estimated at the indication level over 104,565 indications that required an action by the healthcare workers (HCWs) in terms of IPC practices. Regressions include randomization strata controls (by county and market size) and health facility level controls.

Table A8: Treatment Effects on Quality Indicators Not Included in the JHIC (Reported by Patients):
Overall and Interacted with Indicators for Private and Unlicensed Health Facilities at Endline

	Time with HCW and Waiting Time		Patient Satisfaction	Provider Consultation Practices		
	Minutes spent with HCW in examination	Minutes waiting before examination, laboratory, and pharmacy	Patient is satisfied or very satisfied (1-5 scale)	Physical examination (PCA index)	Prescribed or gave medicines	Referred to another HF
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Overall Impact						
Treatment	0.423* (0.246) {0.349}	1.933 (1.945) {0.670}	-0.013** (0.006) {0.349}	0.021 (0.061) {1.000}	-0.004 (0.015) {1.000}	-0.005 (0.005) {0.670}
Observations	9634	11098	11098	9649	9737	9736
R ²	0.013	0.045	0.011	0.017	0.007	0.007
Control Mean	7.760	31.992	0.926	-0.028	0.810	0.054
Impact: {%; SD}	{5%; 0.06}	{6%; 0.04}	{-1%; -0.05}	{-75%; 0.01}	{-1%; -0.01}	{-9%; -0.02}
Panel B: Interaction with Private						
Treatment	0.562** (0.263) {0.444}	3.090 (2.715) {0.745}	-0.019* (0.010) {0.444}	0.025 (0.064) {1.000}	-0.000 (0.020) {1.000}	-0.009 (0.007) {0.745}
Private HF	2.119*** (0.388) [0.000]	-20.770*** (2.811) [0.000]	0.045*** (0.010) [0.000]	0.780*** (0.105) [0.000]	-0.051** (0.022) [0.024]	-0.013 (0.009) [0.140]
Private HF x T	-0.185 (0.475) {1.000}	-4.330 (3.522) {0.745}	0.019 (0.012) {0.669}	0.062 (0.117) {1.000}	-0.017 (0.027) {1.000}	0.011 (0.011) {0.818}
Observations	9634	11098	11098	9649	9737	9736
R ²	0.024	0.092	0.020	0.069	0.011	0.007
Control Mean Public	6.897	40.711	0.906	-0.329	0.829	0.056
Control Mean Private	9.133	17.841	0.958	0.449	0.781	0.051
Impact Public: {%; SD}	{8%; 0.09}	{8%; 0.06}	{-2%; -0.06}	{-8%; 0.02}	{-0%; -0.00}	{-16%; -0.04}
Impact Private: {%; SD}	{4%; 0.05}	{-7%; -0.05}	{0%; 0.00}	{19%; 0.05}	{-2%; -0.04}	{5%; 0.01}
Test T + Private x T = 0 (p-value)	0.357	0.569	0.942	0.392	0.424	0.780
Panel C: Interaction with Unlicensed (Private and active at endline only)						
Treatment	0.653 (0.425) {0.610}	-1.320 (2.371) {1.000}	0.010 (0.010) {0.975}	0.225** (0.106) {0.610}	-0.002 (0.026) {1.000}	0.009 (0.012) {1.000}
Unlicensed HF at Endline	0.791 (0.760) [0.299]	-2.521 (3.082) [0.414]	0.021* (0.012) [0.075]	0.140 (0.187) [0.454]	0.035 (0.039) [0.364]	0.016 (0.016) [0.310]
Unlicensed HF at Endline x T	-0.608 (0.900) {1.000}	-1.574 (3.468) {1.000}	-0.026* (0.015) {0.610}	-0.194 (0.216) {0.975}	-0.070 (0.045) {0.610}	-0.008 (0.021) {1.000}
Observations	3127	3681	3681	3132	3166	3165
R ²	0.013	0.085	0.011	0.024	0.009	0.020
Control Mean Licensed	8.944	21.443	0.947	0.390	0.772	0.046
Control Mean Unlicensed	9.765	14.722	0.973	0.512	0.808	0.064
Impact Licensed: {%; SD}	{7%; 0.10}	{-6%; -0.04}	{1%; 0.04}	{58%; 0.13}	{-0%; -0.00}	{19%; 0.04}
Impact Unlicensed: {%; SD}	{0%; 0.01}	{-20%; -0.12}	{-2%; -0.10}	{6%; 0.02}	{-9%; -0.18}	{2%; 0.00}
Test T + Unlicensed x T = 0 (p-value)	0.957	0.202	0.130	0.877	0.048	0.951

Notes. Robust standard errors are reported in parentheses and clustered at the market level. *** (**) (*) denotes significance at 1% (5%) (10%) level. Stars reported next to the estimated coefficients denote significance related to the “naive” p-value. Sharpened q-values are reported in braces, following Benjamini et al. (2006), with stars next to the braces. Regressions include randomization strata controls (by county and market size) and health facility level controls.

Table A9: Treatment Effects on Quality Indicators Not Included in the JHIC (Healthcare Staff): Overall and Interacted with Indicators for Private and Unlicensed Health Facilities at Endline

	Ratio of healthcare workers to total staff	Healthcare workers per outpa- tient*	Monthly total staff cost (USD PPP)	Monthly staff cost per staff (USD PPP)	Staff cost per out- patient* (USD PPP)
	(1)	(2)	(3)	(4)	(5)
Panel A: Overall Impact					
Treatment	0.006 (0.011) {0.313}	0.009 (0.008) {0.250}	1416.664** (707.763) {0.250}	23.945 (15.347) {0.250}	6.085 (3.834) {0.250}
Observations	1284	1273	1284	1284	1273
R ²	0.090	0.011	0.363	0.087	0.014
Control Mean	0.660	0.031	4022.330	379.387	13.595
Impact: {%; SD}	{1%; 0.03}	{29%; 0.09}	{35%; 0.15}	{6%; 0.10}	{45%; 0.21}
Panel B: Interaction with Private					
Treatment	0.035* (0.020) {0.341}	-0.001 (0.002) {0.607}	3393.735* (1891.131) {0.341}	-1.334 (17.738) {0.607}	-0.595 (1.288) {0.607}
Private HF	0.075*** (0.022) [0.001]	0.022** (0.009) [0.015]	-3191.873*** (1159.038) [0.006]	-34.757 (21.102) [0.101]	6.692** (2.989) [0.026]
Private HF x T	-0.043 (0.026) {0.341}	0.014 (0.013) {0.341}	-2923.509 (2411.402) {0.341}	37.416 (24.448) {0.341}	9.928* (5.420) {0.341}
Observations	1284	1273	1284	1284	1273
R ²	0.100	0.018	0.375	0.088	0.019
Control Mean Public	0.600	0.012	7340.894	385.531	6.483
Control Mean Private	0.689	0.040	2443.708	376.465	17.038
Impact Public: {%; SD}	{6%; 0.18}	{-6%; -0.05}	{46%; 0.25}	{-0%; -0.01}	{-9%; -0.08}
Impact Private: {%; SD}	{-1%; -0.04}	{34%; 0.11}	{19%; 0.07}	{10%; 0.14}	{55%; 0.27}
Test T + Private x T = 0 (p-value)	0.575	0.281	0.629	0.066	0.087
Panel C: Interaction with Unlicensed (Private and active at endline only)					
Treatment	0.004 (0.022) {0.287}	0.039** (0.020) {0.213}	1627.777 (992.866) {0.213}	36.110 (22.164) {0.213}	19.612* (10.312) {0.213}
Unlicensed HF at Endline	0.034 (0.027) [0.220]	0.036* (0.018) [0.053]	430.217 (889.601) [0.629]	17.981 (38.166) [0.638]	12.032** (5.674) [0.035]
Unlicensed HF at Endline x T	-0.042 (0.031) {0.213}	-0.059** (0.028) {0.213}	-1562.430 (1113.349) {0.213}	8.990 (53.461) {0.287}	-22.993* (12.634) {0.213}
Observations	718	714	718	718	714
R ²	0.085	0.015	0.426	0.067	0.015
Control Mean Licensed	0.644	0.021	4177.144	384.911	12.117
Control Mean Unlicensed	0.713	0.058	1190.724	391.635	21.393
Impact Licensed: {%; SD}	{1%; 0.02}	{182%; 1.48}	{39%; 0.17}	{9%; 0.20}	{162%; 1.40}
Impact Unlicensed: {%; SD}	{-5%; -0.16}	{-35%; -0.11}	{5%; 0.05}	{12%; 0.13}	{-16%; -0.07}
Test T + Unlicensed x T = 0 (p-value)	0.089	0.307	0.810	0.326	0.619

Notes. *Monthly outpatients. Robust standard errors are reported in parentheses and clustered at the market level. Robust standard errors are reported in parentheses and clustered at the market level. *** (**) (*) denotes significance at 1% (5%) (10%) level. Stars reported next to the estimated coefficients denote significance related to the “naive” p-value. Sharp-ened q-values are reported in braces, following [Benjamini et al. \(2006\)](#), with stars next to the braces. Regressions include randomization strata controls (by county and market size) and health facility level controls.

Table A10: Treatment Effects on JHIC Item Compliance by Cost Categories:
Overall and Interacted with Indicators for Private and Unlicensed Health Facilities at Endline

	Cost Groups				Marginal and Fixed Costs	
	Lowest Cost (1)	Low Cost (2)	Medium Cost (3)	High Cost (4)	Marginal (5)	Fixed (6)
Panel A: Overall Impact						
Treatment	0.034*** (0.008) {0.001}***	0.061*** (0.010) {0.001}***	0.074*** (0.010) {0.001}***	0.063*** (0.012) {0.001}***	0.064*** (0.009) {0.001}***	0.047*** (0.008) {0.001}***
Observations	82979	62872	19618	51062	73329	143202
R ²	0.036	0.018	0.052	0.045	0.023	0.032
Control Mean	0.199	0.407	0.290	0.409	0.359	0.295
Impact: {%; SD}	{17%; 0.08}	{15%; 0.12}	{25%; 0.16}	{15%; 0.13}	{18%; 0.13}	{16%; 0.10}
Panel B: Interaction with Private						
Treatment	0.027** (0.012) {0.021}**	0.031** (0.013) {0.019}**	0.052*** (0.015) {0.010}***	0.035** (0.015) {0.021}**	0.040*** (0.013) {0.011}**	0.028** (0.011) {0.019}**
Private HF	-0.093*** (0.012) [0.000]	-0.030** (0.014) [0.037]	-0.037** (0.015) [0.012]	-0.093*** (0.014) [0.000]	-0.027** (0.013) [0.047]	-0.089*** (0.012) [0.000]
Private HF x T	0.009 (0.014) {0.052}*	0.046*** (0.016) {0.013}**	0.034** (0.017) {0.026}**	0.043*** (0.017) {0.019}**	0.038** (0.015) {0.019}**	0.028** (0.013) {0.026}**
Observations	82979	62872	19618	51062	73329	143202
R ²	0.045	0.019	0.052	0.048	0.023	0.036
Control Mean Public	0.271	0.425	0.322	0.482	0.377	0.363
Control Mean Private	0.157	0.397	0.272	0.370	0.348	0.258
Impact Public: {%; SD}	{10%; 0.06}	{7%; 0.06}	{16%; 0.11}	{7%; 0.07}	{11%; 0.08}	{8%; 0.06}
Impact Private: {%; SD}	{23%; 0.10}	{20%; 0.16}	{32%; 0.19}	{21%; 0.16}	{22%; 0.16}	{22%; 0.13}
Test T + Private x T = 0 (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Panel C: Interaction with Unlicensed (Private and active at endline only)						
Treatment	0.047*** (0.013) {0.001}***	0.072*** (0.015) {0.001}***	0.103*** (0.015) {0.001}***	0.079*** (0.015) {0.001}***	0.078*** (0.014) {0.001}***	0.063*** (0.013) {0.001}***
Unlicensed HF at Endline	-0.023*** (0.008) [0.007]	-0.061*** (0.018) [0.001]	-0.043*** (0.013) [0.001]	-0.068*** (0.015) [0.000]	-0.053*** (0.015) [0.001]	-0.039*** (0.011) [0.000]
Unlicensed HF at Endline x T	-0.031** (0.015) {0.024}**	0.003 (0.024) {0.298}	-0.052*** (0.020) {0.009}***	-0.016 (0.019) {0.170}	-0.011 (0.021) {0.202}	-0.024 (0.015) {0.058}*
Observations	52378	40101	12452	33213	46381	91763
R ²	0.039	0.027	0.061	0.051	0.028	0.036
Control Mean Licensed	0.185	0.430	0.315	0.424	0.381	0.294
Control Mean Unlicensed	0.117	0.348	0.215	0.304	0.300	0.209
Impact Licensed: {%; SD}	{25%; 0.12}	{17%; 0.15}	{33%; 0.22}	{19%; 0.16}	{20%; 0.16}	{22%; 0.14}
Impact Unlicensed: {%; SD}	{14%; 0.05}	{21%; 0.16}	{24%; 0.13}	{21%; 0.14}	{22%; 0.15}	{19%; 0.10}
Test T + Unlicensed x T = 0 (p-value)	0.023	0.000	0.000	0.000	0.000	0.000

Notes. Robust standard errors are reported in parentheses and clustered at the market level. *** (**) (*) denotes significance at 1% (5%) (10%) level. Stars reported next to the estimated coefficients denote significance related to the “naive” p-value. Sharpened q-values are reported in braces, following [Benjamini et al. \(2006\)](#), with stars next to the braces. Regressions include randomization strata controls (by county and market size) and health facility level controls.

Table A11: Treatment Effects on JHIC Score for Private Facilities:
Interacted with the Number of Public Facilities in Market

	JHIC Score (Private Facilities) (1)
Treatment	2.425* (1.250) [0.054]
No. Public Facilities in Market	-0.890*** (0.333) [0.008]
No. Public Facilities in Market x T	1.780*** (0.427) [0.000]
Observations	872
R ²	0.354
Control Mean	33.463
Mean No. of Public in Markets	2.234
Impact Evaluated at Mean No. of Public {%; SD}	{19%; 0.59}
T + No. of Public x T = 0 (p-value)	0.000

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Regressions include randomization strata controls (by county and market size) and health facility level controls.

Table A12: Treatment Effects on Select Intermediate Outcomes:
Intervention Awareness, Knowledge, and Perceptions

	In-Charge Level		Patient Level			
	Familiar with the New Legislation JHIC (Aware- ness)	Ever Noticed a Scorecard (Aware- ness)	Know Scorecards' Letter Ranking (Vignette: A vs C vs D)	Ever Noticed a Scorecard (Aware- ness)	Perceive Improve- ment in 2017 in HF's Quality (If opened before 2018)	Perceive Recent Govern- ment Inspection
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Unweighted						
Inspection Only (T1)	0.280*** (0.039) [0.000]	0.090*** (0.032) [0.006]	0.016 (0.014) [0.263]	0.017* (0.010) [0.098]	0.013 (0.018) [0.460]	0.005 (0.016) [0.775]
Inspections plus Information (T2)	0.321*** (0.033) [0.000]	0.576*** (0.031) [0.000]	0.006 (0.013) [0.661]	0.091*** (0.011) [0.000]	-0.008 (0.017) [0.638]	0.025* (0.014) [0.070]
Observations						
Facilities/ (Patients)	1285	1285	1213 (11098)	1213 (11098)	1145 (10165)	1213 (11098)
R ²	0.107	0.287	0.012	0.016	0.009	0.012
Control Mean	0.306	0.233	0.740	0.132	0.659	0.392
T1 (SD Control)	0.608	0.213	0.036	0.050	0.028	0.009
T1 (% Control Mean)	92%	39%	2%	13%	2%	1%
T2 (SD Control)	0.695	1.361	0.013	0.270	-0.017	0.051
T2 (% Control Mean)	105%	247%	1%	69%	-1%	6%
Test T1 = T2 (p-value)	0.309	0.000	0.465	0.000	0.157	0.185
Panel B: Weighted						
Inspection Only (T1)	0.239*** (0.051) [0.000]	0.176*** (0.062) [0.005]	-0.011 (0.020) [0.582]	0.020 (0.014) [0.143]	0.028 (0.021) [0.182]	0.002 (0.021) [0.943]
Inspections plus Information (T2)	0.337*** (0.048) [0.000]	0.651*** (0.046) [0.000]	0.006 (0.017) [0.705]	0.091*** (0.014) [0.000]	0.009 (0.021) [0.677]	0.026 (0.023) [0.259]
Observations						
Facilities/ (Patients)	1285	1285	1210 (11095)	1210 (11095)	1142 (10162)	1210 (11095)
R ²	0.165	0.347	0.017	0.015	0.011	0.016
Control Mean	0.338	0.219	0.749	0.130	0.642	0.404
T1 (SD Control)	0.504	0.426	-0.025	0.061	0.059	0.003
T1 (% Control Mean)	71%	80%	-1%	16%	4%	0%
T2 (SD Control)	0.711	1.571	0.015	0.269	0.019	0.052
T2 (% Control Mean)	100%	297%	1%	70%	1%	6%
Test T1 = T2 (p-value)	0.057	0.000	0.302	0.000	0.317	0.253

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Panel B includes weights constructed using average facility outpatients. Regressions include randomization strata controls (by county and market size) and health facility level controls.

Figure A1: Scorecards

A. Scorecards for information arm



B. Scorecards for closures in all treatment arms

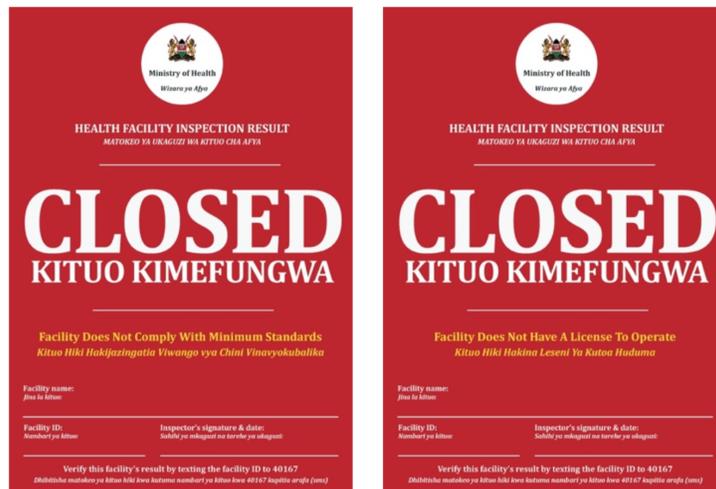


Figure A2: Scorecard Dissemination Materials

A. Description Sheet

COMPLIANCE CATEGORY	FACILITY SCORE (% OF MAX SCORE)
	<ul style="list-style-type: none"> Fully Compliant >75%
	<ul style="list-style-type: none"> Substantially Compliant 61 - 75%
	<ul style="list-style-type: none"> Partially Compliant 41 - 60%
	<ul style="list-style-type: none"> Minimally Compliant 11 - 40%
	<ul style="list-style-type: none"> Non-Compliant 0 - 10% (or absence of licenses)

B. Dissemination Flyer



MINISTRY OF HEALTH
Wizara ya Afya

Health Facility Kituo cha afya

? Why am I seeing a scorecard at this health facility?

Je kwa nini ninaona alama ya matokeo kwenye mlango wa kituo hiki cha afya?

A Ministry of Health inspector **conducted an inspection** at this health facility.
Mkaguzi kutoka wizara ya afya **amefanya ukaguzi** katika kituo hiki cha afya.
The scorecard tells you how well this facility complies with **minimum patient safety standards**.

Alama ya matokeo inakuonyesha jinsi kituo hiki cha afya kinaafikiana na kiwango cha chini cha usalama wa mgonjwa.

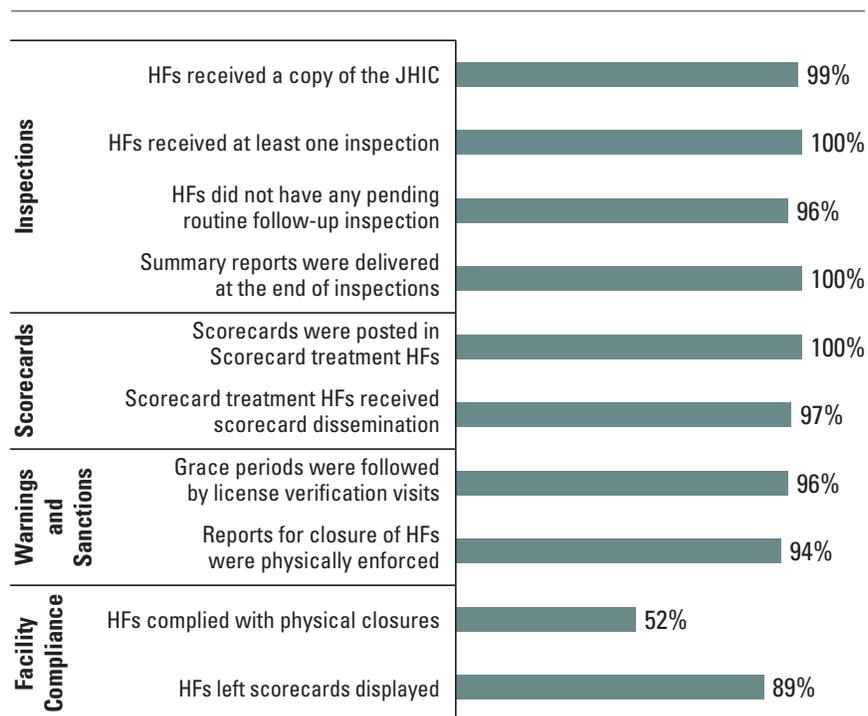
? What does this mean? Je, maana yake ni nini?

<p>If you see... <i>Ukiona kadi zifuatazo...</i></p> 	<p>The facility scored <i>Kituo cha afya kimepata alama</i></p> <ul style="list-style-type: none"> > 75% 61% - 75% 41% - 60% 11% - 40% 0% - 10% (or no license) <i>0% - 10% (au hakuna leseni)</i> 	<p>This means it is... <i>Hii ina maana kwamba kituo hiki kina...</i></p> <ul style="list-style-type: none"> Fully Compliant <i>Uzingatiaji kamilifu</i> Substantially Compliant <i>Uzingatiaji wa kiwango cha juu</i> Partially Compliant <i>Uzingatiaji wa kiwango cha wastani</i> Minimally Compliant <i>Uzingatiaji wa kiwango cha chini</i> Non-Compliant <i>Hakuna uzingatiaji wowote</i>
---	--	---

**? How can I find out more?
Jinsi gani naweza kupata maelezo zaidi?**

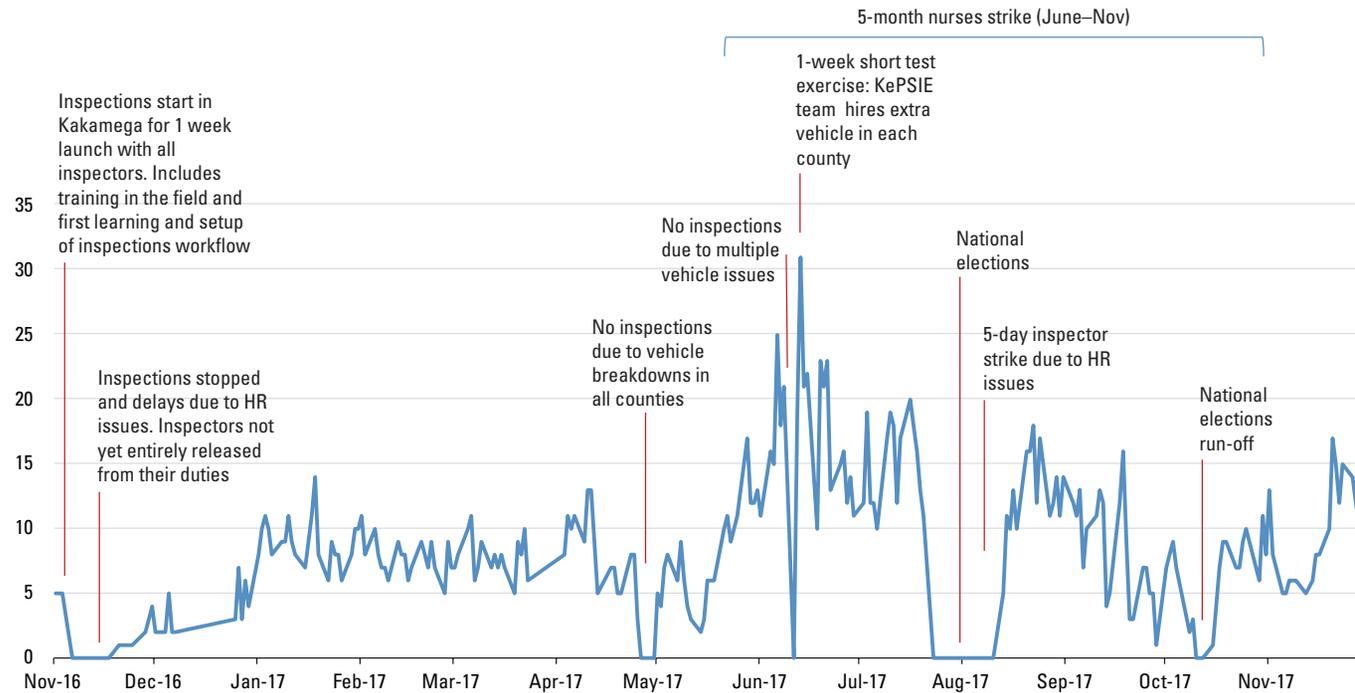
For verification / Dhibitisho kamili
Call / Simu:
SMS Facility ID (Free) / Tuma ujumbe mfupi wa nambari ya kituo cha afya bila malipo kwa:

Figure A3: Compliance with Intervention Components



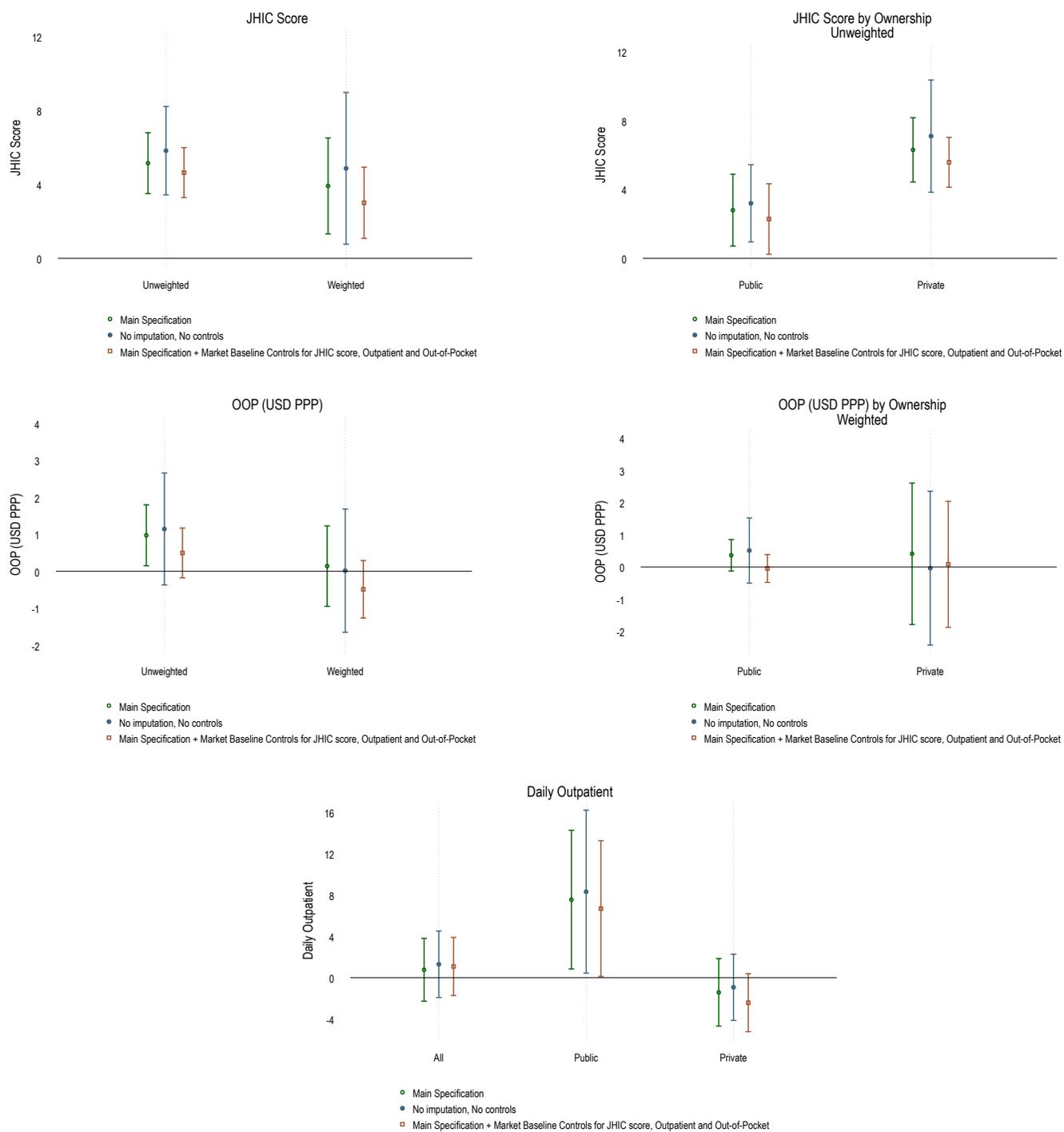
Notes. Source: [Bedoya et al. \(2020\)](#). HF = health facility. Facility compliance with physical closures is based on quality checks on average 3 months after the inspection. Facility compliance with scorecards displayed is based on quality checks on average 2 months after closure.

Figure A4: Inspection Visits by Day and Select Events



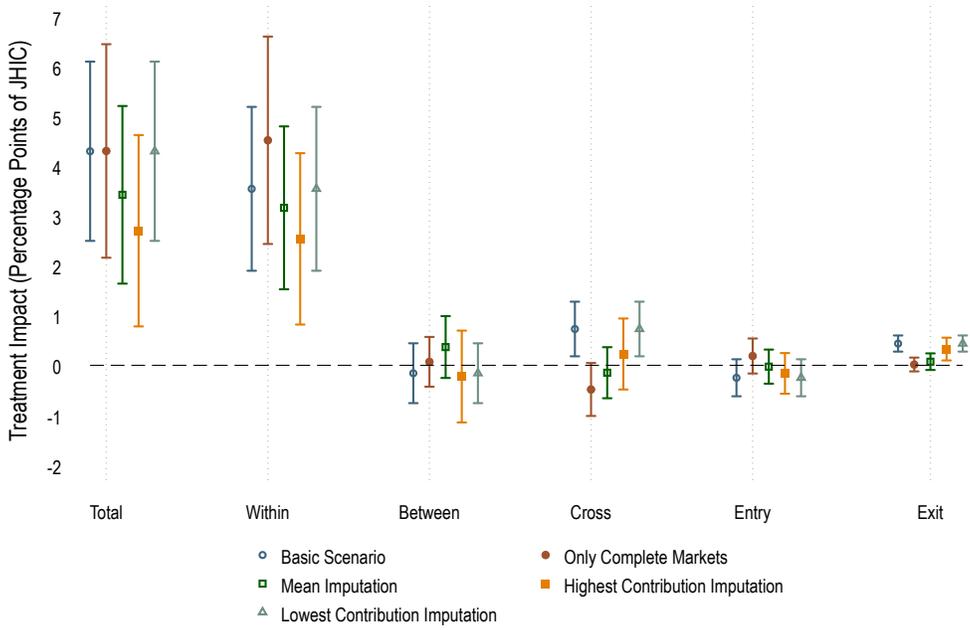
Notes. Source: [Bedoya et al. \(2020\)](#). Vehicle issues include breakdowns/maintenance, no fuel due to payment delays, and vehicles being used by county government.

Figure A5: Main Outcomes: Robustness Checks



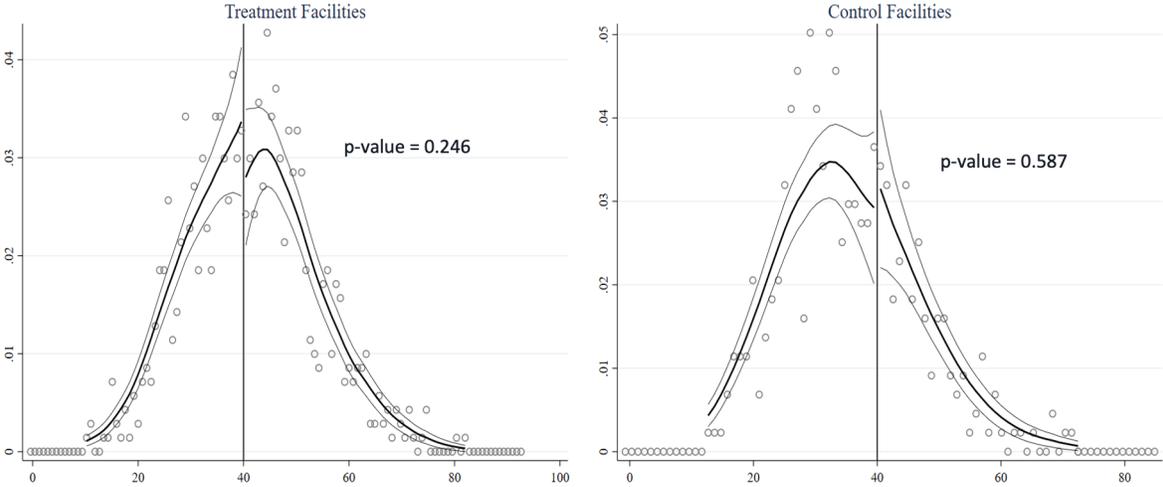
Notes. Vertical lines correspond to 95% confidence intervals.

Figure A6: Treatment Effects on Market Weighted JHIC Score and Decomposition Components
Robustness to Different Scenarios



Notes. This figure corresponds to robustness scenarios for estimates in [Table 6](#). Vertical lines correspond to 95% confidence intervals following the decomposition formula given in [Equation 3](#). “Total” reports the impact on the weighted JHIC score between baseline and endline at the market level, followed by the impact on each of the five terms of the decomposition. Regressions include randomization strata controls (by county and market size) and the percentage of facilities of each level in the market. The scenarios are as follows: “Basic” refers to estimates as in [Table 6](#); “Only Complete Markets” excludes all markets in which at least one facility has missing data; “Mean Imputation” includes imputation for missing values by level, treatment status and ownership; and “Highest (Lowest) Contribution Imputation” includes imputations for the missing values with the highest (lowest) contribution to the total quality change by treatment and facility type (entering, exiting, or facilities active both at baseline and endline). For instance, we replace the missing values for entering facilities in the control group with the quality score and outpatient values of the facility with the highest (lowest) entry contribution in the control group.

Figure A7: McCrary Test of Density Discontinuity of JHIC Score at Endline, by Treatment Group



Notes. Vertical lines indicate the border between the categories “Minimally Compliant” (JHIC score <40) and “Partially Compliant” (JHIC score >40). The McCrary-type density tests the null hypothesis of continuity of the density of the JHIC score for treatment facilities around 40% of the maximum score. The p-value presented is the probability that the discontinuity is significant.

Figure A8: JHIC Score Distribution by Treatment Status

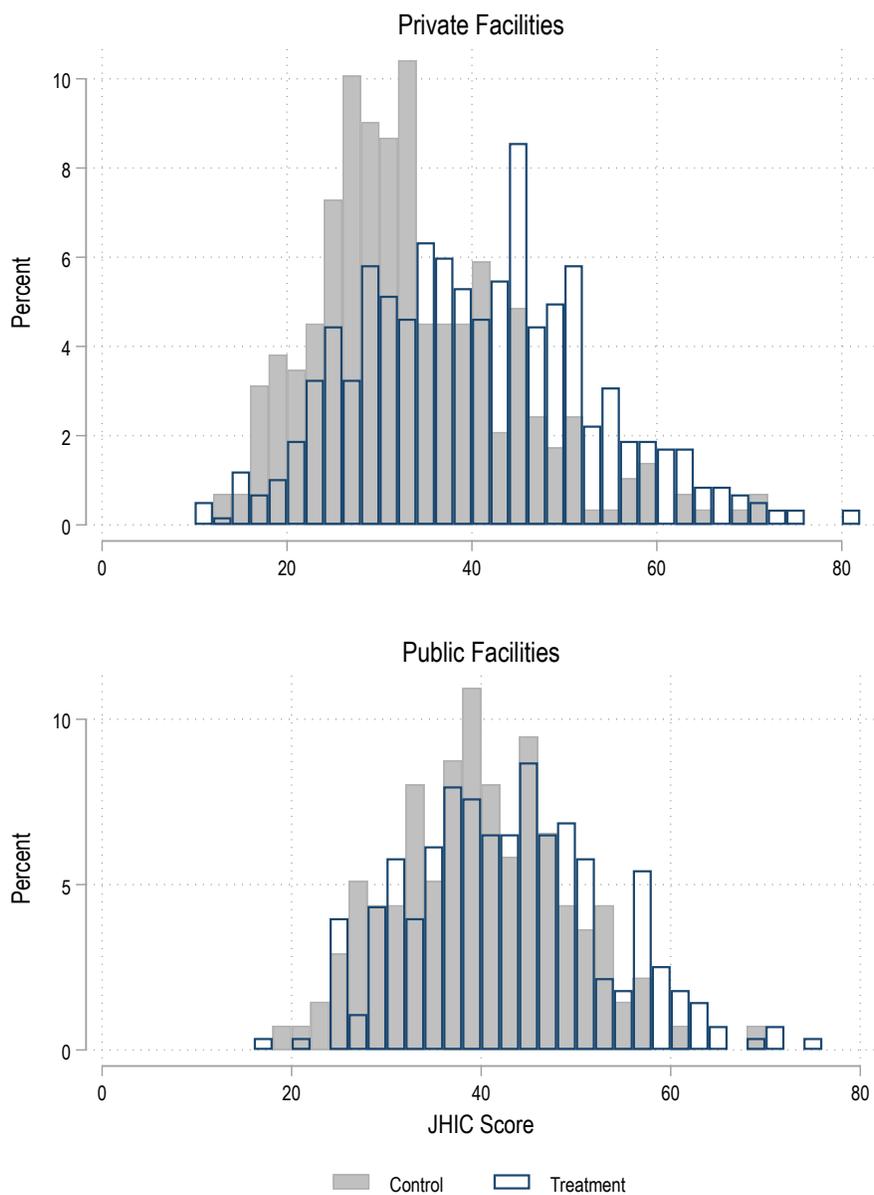
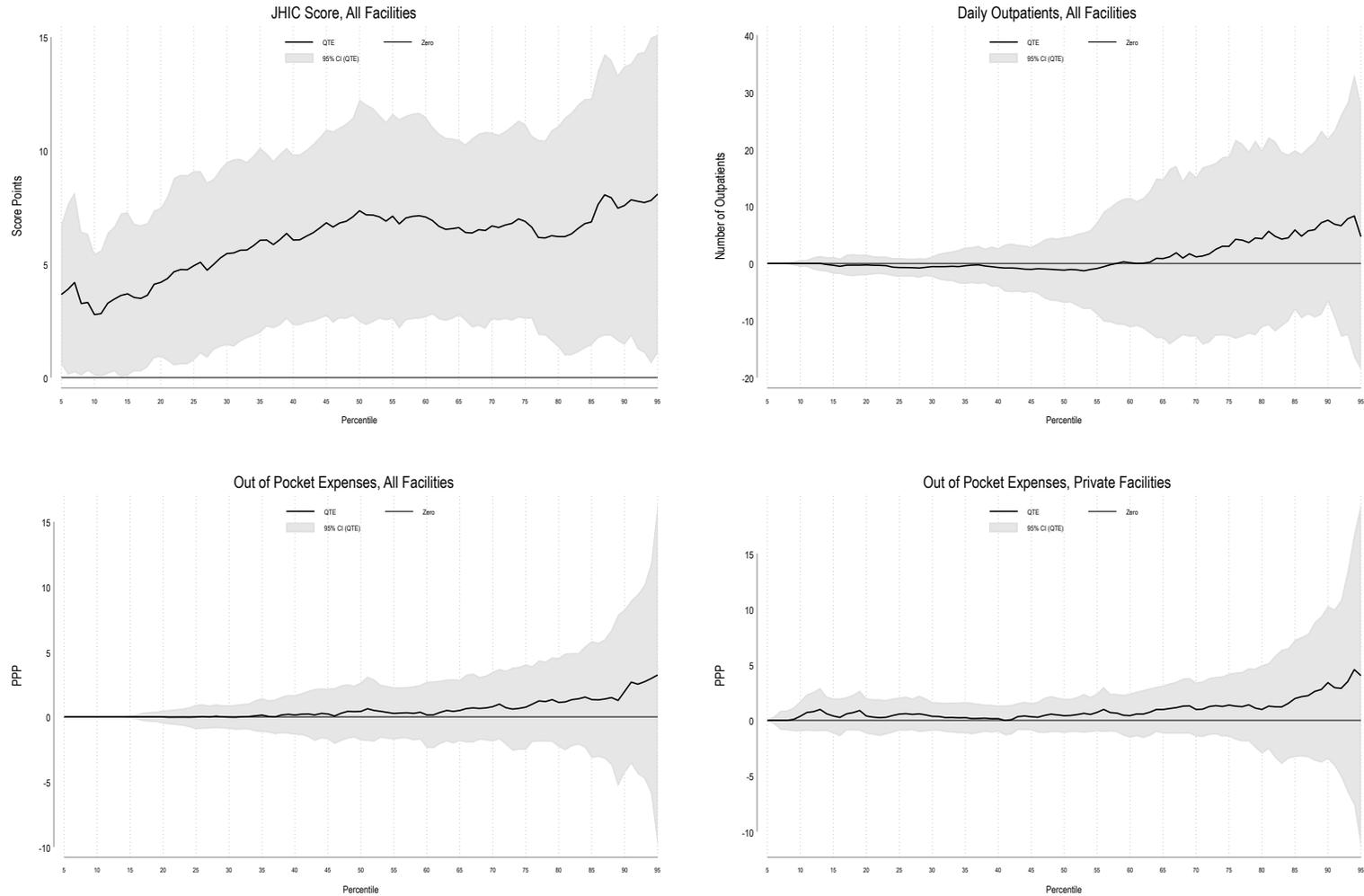


Figure A9: Unconditional Quantile Treatment Effects on JHIC Score, Outpatients and OOP by Percentiles



Notes. Estimation is made for every percentile between the 5th and 95th. Bootstrapped 95% confidence intervals (2,000 replications). The confidence intervals control for the 1-family-wise error rates (probability of at least one false rejection across tests), following Romano & Wolf (2010), using codes from Bedoya et al. (2017).

SUPPLEMENTAL MATERIAL

Randomized Regulation:

The Impact of Minimum Quality Standards on Health Markets

Version: December 4, 2022

Table of Contents: Supplemental Material

1	Health Markets Construction and Characteristics	4
2	Regulatory Reform Background	8
3	Joint Health Inspection Checklist (JHIC) Instrument	11
4	Ethical Considerations	14
5	Census, Survey Response Rates, Attrition and Accretion	15
6	Variable Definition and Construction	19
6.1	Main Outcomes	19
6.2	Secondary Outcomes	20
6.3	Other Key Variables	25
7	Additional Analyses	27
7.1	Cross-Market Externalities	27
7.2	Program Duration	28
7.3	Quantile Treatment Effects	29
8	Costing	30
9	Other Tables and Figures	32

Tables

S1	Facility Market Characteristics in KePSIE Counties and at the National Level Before the Intervention	5
S2	JHIC Sections	12
S3	Survey Response Rates at Baseline and Endline	16
S4	Census of Health Facilities at Randomization and Endline	16
S5	Randomized Markets and Corresponding Health Facilities by County	17
S6	Health Facility Response Rates by Waves and Treatment Status	17
S7	Proportion of Facilities by Market Treatment Assigned and Treatment Received	18
S8	Correlations of Select Indicators with Market Size at Randomization	32
S9	Share of Total Patients and Facilities: Facilities that Exited and Entered by Endline	32
S10	Treatment Effects on JHIC Score, OOP, and Outpatients with Baseline Market Level Controls: Overall and Interacted with Indicators for Private and Unlicensed Health Facilities at Endline	33
S11	Treatment Effects on JHIC Score, OOP, and Outpatients without Imputation and Including Strata Controls Only: Overall and Interacted with Indicators for Private and Unlicensed Health Facilities at Endline	34
S12	Spillover Effects on JHIC Score, OOP, Outpatients, Facility Inactivity, and Entry for Control Facilities	35
S13	Spillovers: Baseline Balance on Main Outcomes	35
S14	Impact on JHIC Score by Month of First Inspection and Endline Month Interaction	36
S15	Baseline Balance on JHIC Score by Month of First Inspection and Endline Month Interaction	37
S16	Unconditional QTE on JHIC Score, by Ownership and Market Size	38
S17	Treatment Effects on JHIC Score by Closure Probability Group	39
S18a	Treatment Effects on JHIC Score, OOP, Outpatients, and Entry: Interacted with and Indicator for JHIC Score at Baseline >40	40
S18b	Treatment Effects on JHIC Score, OOP, Outpatients, and Entry: Interacted with and Indicator for JHIC Score at Baseline >50	40

Figures

S1	Illustration of Facilities by Level	6
S2	Illustration of Markets in Study Areas	7
S3	JHIC Score Density in Control Facilities by Data Collection Round (% of max score x100)	41
S4	Impact on JHIC Score by Time of Market's First Inspection With Endline Time Interactions	42

1 Health Markets Construction and Characteristics

Markets are defined by geography. In the 2015 baseline, among 8,577 patients with whom we completed exit surveys, 73% lived less than 4km from the health facility they visited.¹ Using this distance, we apply a z-center clustering algorithm to the facilities. First, a randomly-chosen facility is defined as a “cluster center.” Next, every facility is matched to the market defined by the nearest cluster center. If any facility is more than 4km from its corresponding cluster center, the facility which is farthest from its cluster center becomes a new cluster center. Then the process is repeated. Our algorithm stops creating new clusters when all facilities are within the predefined distance (4km) from the cluster center which defines their market.

In our three study counties, 273 markets were identified through the clustering algorithm and were randomized into treatment and control groups. The three counties have diverse market sizes ranging from those with only one facility or singletons to markets with 2, 3, 4, 5 facilities or even large markets with 30 or more facilities. [Figure S2](#) shows examples of typical markets in these areas: a singleton, a small market, and a densely populated market. At randomization, 1,258 health facilities in those markets were assigned to these markets.² By endline, 90 facilities that were missed or classified as temporarily or permanently closed during the census were found to be operational and added to the nearest market for a total of 1,348 facilities operational at randomization. We also applied the market clustering algorithm to administrative data from the government’s Master Facility List (MFL) of August 2014 to assess how these data compare to ours in the three study counties and to compare our data with the national figures. In the administrative data we identified 2,795 markets for the 8,776 facilities listed, of which 863 facilities were located in our three counties.³

[Table S1](#) shows market characteristics in Kenya and our three study counties using administrative data (Columns 1 and 2) and for our three counties using our census and survey data (Columns 3 to 5). The administrative data over-represent public facilities and singletons and are missing small, private facilities (which include facilities operated by NGO’s and faith-based organizations). Compared to the national data, Columns 1 and 2 show that there are more private facilities in the three study counties (48% vs. 39%), more markets of size 4 or larger (65% vs. 57%) and fewer singleton markets (11% vs. 17% singletons). When using KePSIE data, these differences increase considerably. This is partly explained by the fact that KePSIE data include unlicensed informal private facilities.

¹We define a health facility as any premise offering outpatient care or primary care services. Level 2 includes Dispensaries and Clinics; Level 3 consists of Health Centers and Maternity and Nursing homes; Level 4 includes Primary Hospitals; Level 5 includes Secondary Hospitals, and Level 6 consists of Tertiary Hospitals.

²This excludes 14 facilities originally listed in the 273 randomized markets that were later identified as duplicates.

³We also found more facilities in our 2015 census than those registered in government records in February of the same year (1,115 vs. 940, or 19% more). We identified 379 new facilities (40%) that were not included in the MFL and 202 (21%) from the MFL that had gone out of business.

Table S1: Facility Market Characteristics in KePSIE Counties and at the National Level Before the Intervention

	Government Data		KePSIE Data		
	National	3 Counties	3 Counties		
	% of facilities (1)	% of facilities (2)	% of facilities (3)	% of patients (4)	% of patients (5)
Panel A. Facility level					
Facility ownership					
Public	0.61	0.52	0.30	0.71	0.67
Private	0.39	0.48	0.70	0.29	0.33
Facility level					
Level 2	0.80	0.83	0.85	0.62	0.66
Level 3	0.15	0.12	0.11	0.21	0.19
Levels 4 and 5	0.05	0.05	0.04	0.17	0.15
Proportion by market size					
1 facility	0.17	0.11	0.07	0.11	0.11
2-3 facilities	0.25	0.24	0.15	0.19	0.18
4+ facilities	0.57	0.65	0.79	0.70	0.71
Observations	8776	863	1348	1025	1348
Panel B. Market level					
	% of markets		% of markets	% of patients	% of patients
Singleton markets	0.51	0.39	0.30	0.11	0.11
Only public	0.47	0.32	0.23	0.10	0.09
Only private	0.04	0.07	0.08	0.01	0.01
Markets with at least 1 public facility	0.95	0.92	0.88	0.98	0.97
Observations	2795	239	273	273	273

Notes. Columns 1 and 2 display government data provided by Kenya’s Ministry of Health in a master health facility list available as of August 2014 (MOH, 2014) and Columns 3 to 5 report study data. In Column 5, missing values for OOP are imputed using means defined by level, ownership, treatment, license status at randomization, and daily outpatients. Markets are generated using a z-center clustering algorithm with a 4km radius for all health facilities and applied to the government data for comparable estimates at the study and national levels. Outpatient caseload uses daily outpatients from facility records at baseline.

Figure S1: Illustration of Facilities by Level

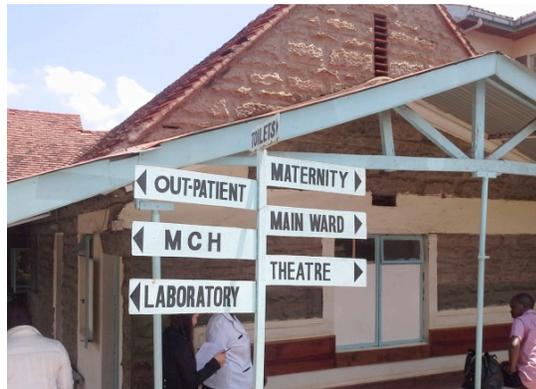
Level 2: Dispensaries and clinics



Level 3: Health centers and maternity and nursing homes



Level 4 or 5: Primary and secondary hospitals



Notes. We define a health facility as any premise offering outpatient care or primary care services. Level 2 includes Dispensaries and Clinics; Level 3 consists of Health Centers and Maternity and Nursing homes; Level 4 includes Primary Hospitals; Level 5 includes Secondary Hospitals, and Level 6 consists of Tertiary Hospitals. Facilities in our baseline sample offer the following services: consultation (100%); labor ward (31%); medical and pediatric wards (11%), theater (4%); pharmacy (59%); laboratory (58%); radiology (3%); and nutrition (5%).

Figure S2: Illustration of Markets in Study Areas



Notes. Markets are defined by geographic delimitations with a 4km radius using a z-center clustering algorithm. Panel A is an example from Kakamega County of a singleton market (30% of markets are singletons). Panel B is an example from Meru County of a small to medium-size market (around 28% of markets have between 2 and 3 facilities). Panel C is an example from Kilifi County of a large market (around 42% of markets have more than 4 facilities, and where the largest randomized market has 45 facilities).

2 Regulatory Reform Background

As part of a World Bank team, we started working with the Government of Kenya in 2012, after which the government initiated a series of reforms to improve the health inspections system. In 2012, a Joint Health Inspections Checklist (JHIC) was published in the official gazette in an effort to improve fairness, coordination, and transparency. This JHIC replaced multiple individual inspections from Kenya’s regulatory “Boards and Councils” (B&Cs), improved standardization, and reduced costs to health facilities.⁴

There are three main types of inspections conducted in Kenya by the national government: (1) inspections of new facilities before issuing licenses; (2) inspections of operating facilities due to complaints received, for instance, due to malpractice cases (reactive inspections), and (3) inspections to verify the quality of health providers. Before the reform that we evaluate, inspections in category (3), which are closest to the subject of this study, were carried out sporadically within small geographic regions (inspectors would visit these regions and inspect all health facilities in the surrounding areas based on the quota for the inspection period). The boards and councils and the Ministry of Health that are responsible for the inspections supervised them from their headquarters in Nairobi, since there was no regional network of inspectors established.

As the first step to the reform, in October 2013, all stakeholders led by the Ministry of Health (MOH) held a two-day workshop to assess progress and signed the “Windsor Agreement,” embarking on an ambitious health inspections reform to improve patient safety and health care in Kenya. This agreement became the basis for this study.⁵ The assessment identified the following challenges with the JHIC and the inspection framework:

1. Unclear and discretionary rules of the game (e.g., rating in the inspections was not consistent due to a lack of definitions on the specific items to be inspected). The 2012 version of the JHIC did not adequately define the concept of patient safety, and failed to specify consequences of inspections. The definition of patient safety was ‘fuzzy’, both in terms of concepts being measured, and how they were measured. For example, the 2012 JHIC asked inspectors to rank from 1 to 5 whether there was an emergency tray in the facility, without listing specific components of the tray that guarantee minimum patient safety, or what to do in cases where some components were present. This gave inspectors a degree of discretion that weakened the regulatory agencies’ ability to monitor facility performance. Lack of clarity on regulation was also been linked to problems of bribery and corruption in the Kenyan health sector. A third

⁴When we started this study, there were nine regulatory boards and councils: Kenya Medical Practitioners and Dentists Council, Clinical Officers Council, Nursing Council of Kenya, Radiation Protection Board, Pharmacy and Poisons Board, Council of the Institute of Nutritionists and Dieticians, Kenya Medical Laboratory Technicians and Technologists Board, and Public Health Officers and Technicians Council.

⁵The stakeholders and signatories of the Windsor Agreement included the MOH, the regulatory boards and councils, and representatives from the private health sector.

weakness was the poor dissemination of the 2012 JHIC. While the checklist was gazetted and made publicly available online, many health facilities reported that they were unaware of the current regulation, and could not, as a result, conduct self-assessments and improve voluntarily.

2. Lack of incentives to improve patient safety at different levels of compliance with the standards, with very limited sanctions and weak enforcement, except for cases of extreme malpractice. There were no clear follow-up actions to take based on inspection results. The lack of clear consequences, combined with the lack of information on part of health facilities on the consequences of inspections, weakened the capacity of the historical inspections regime to encourage improvement or enforce sanctions at the facility level.
3. Inadequate capacity to conduct inspections: Around 4% of health facilities in Kenya were inspected each year by the B&Cs and virtually none were re-inspected (as of July 2015). Reasons for the low inspection coverage include limited numbers of trained inspectors and inadequate resources to support field visits (there was only a small pool of joint inspectors, all of whom were based in Nairobi). Furthermore, the B&Cs had limited capacity to follow up on warnings and sanctions imposed by inspectors. There was also variation in the background of inspectors representing the different boards and councils. This is likely to result in some variability in the outcome of inspections.
4. Lack of an integrated information system: The 2012 JHIC was not linked to an integrated information system, therefore limiting the ability of actors to use inspections results to strategically manage inspection process and prioritize follow-up efforts. As of July 2015, the results of the inspections were physically archived in some B&C, with few analyzed reports. Not having an electronic system limited the regulators' ability to identify priority locations, reference results from previous inspections of the same facility or compare results across facilities to identify shared constraints.

Stakeholders judged this a sufficient basis for action and constituted a technical working group led by the MOH to further the reform of the regulatory framework. They also committed to conduct an evaluation to provide evidence on the impact of the inspection reform on patient safety and quality of care. The task force was appointed with the authority and mandate to oversee all aspects of the design and implementation of the pilot inspections regimes, provide support on key decisions, deal with any issues that arose during the implementation, and to play a critical role after the intervention including dissemination of results and decision-making for national policy. The task force consisted of members from the MOH, the regulatory boards councils, representatives from the private sector and county health teams.

In March 2016, an enhanced regulatory framework was published in the official gazette based on the inspection reforms developed by the task force ([Kenya Gazette, 2016](#)). The main elements of the new regulatory framework defined by Legal Notice No. 46 of the Public Health Act and the implementation guidelines prepared by the technical working group include a refined JHIC focused on fundamentals of patient safety, a scoring system that allowed facilities to be categorized according to the level of risk presented to patients, and warnings and sanctions that are triggered by the scores according to a facility's level of risk. The interventions evaluated as a part of the study, which are based on the new regulatory framework, as well as the instruments, and evaluation design, were designed through a participatory approach over a three-year process. The three counties where the study takes place, for instance, were selected by the health management representatives of the 47 Kenyan counties to represent different conditions and markets in the country. The inspection checklist was designed by the regulators and tested before application in the three study counties. Throughout, we provided technical support to the working group through pilots and development of technical systems.

3 Joint Health Inspection Checklist (JHIC) Instrument

The standardized regulatory checklist or JHIC, whose use we evaluate, largely focuses on structural measures of quality of care and patient safety. Five sections are general and represent minimum standards for all facilities in terms of infrastructure, management, and infection prevention and control. The remaining eight sections are connected to particular services that the facilities offer, including the labor ward, medical and pediatric wards, the operating theatre, pharmacy, laboratory, radiology, nutrition and kitchen services, and mortuary. In our sample at baseline, health facilities offered the following services: consultation (100%); labor ward (31%); medical and pediatric wards (11%), theater (4%); pharmacy (59%); laboratory (58%); radiology (3%); and nutrition (5%). [Table S2](#) presents the list of sections and subsections and the complete checklist can be found in the [2016 Kenya Gazette](#) Supplement No. 31 as part of Legal Notice No. 46 Public Health Act, Cap. 242.

The JHIC scores for each item are defined in the regulation using a scoring system that weights equally each section (and similarly for some subsections) with total values per section ranging from 0 to 100. The score represents the score as a percentage of maximum score. The new regulatory framework defines all items to be inspected with a specific score for each item. Therefore, any facility can predict with precision what score and in which category they would fall using the gazette checklist which includes the table with scores and corresponding warnings, sanctions, and rewards.

During an inspection, inspectors review specific elements within the structural category looking at standard procedures and policies, infrastructure, equipment and supplies, and medical records, and confirm items such as basic medicines available in an emergency tray, or that the facility has adequate storage for safe water supply. In addition to these measures, there are a few process measures of quality of care and patient safety. This may include, for example, reviewing a sample of patient charts for appropriate care, such as whether a completed partograph accurately monitors TPR/BP and codes drugs. Overall, most sections combine these elements within structural measures of quality of care.

Examples of the minimum standards for all facilities include items across administration (e.g., having licenses for the facility, departments, and staff), health facility infrastructure (e.g., having a stable source of power, and safe, clean water available from a tap or container with sufficient storage, and the presence of an emergency signage), general management and recording of information (e.g., having patient records with names and unique patient numbers, collecting and regularly reporting performance indicators on immunization and notifiable diseases, and having service contracts for maintenance and calibration of medical equipment), infection prevention and control (e.g., evidence of disinfectant use, availability of soap at hand-washing areas, and protocols for sterilization of equipment), and medical consultation (e.g., access to essential medicines including

glucose, adrenaline, sodium bicarbonate, diazepam, and phenobarbitone in an emergency tray, and a system for coding patients in triage at the first point of contact with patients).

Table S2: JHIC Sections

Section Number	Section title	Section score
Section 1	Administrative information	-
Section 2	Health Facility Infrastructure	100
Section 3	General Management and Recording of Information	100
Section 4	Infection Prevention and Control	100
Section 5	Medical and Dental Consultation Services	100
Section 6	Labour Ward	100
Section 7	Medical and Paediatric Wards	100
Section 8	Theatre	100
Section 9	Pharmacy	100
Section 10	Laboratory	100
Section 11	Radiology	100
Section 12	Nutrition and Dietetics Services Unit	100
Section 13	Mortuary	100
Section 14	Findings and Recommendations	-

Notes. Section scores represents the score as a percentage of maximum score. The maximum score per section is 100 except if the facility has a radiology department and if the facility does not produce radioactive waste, and for which the maximum score for this section is 80. Source: [Ministry of Health \(2015\)](#) and [Kenya Gazette \(2016\)](#).

Further sections are activated for facilities that provide additional services including labor ward (e.g., access to a sterile delivery set and resuscitation equipment with oxygen, suction machine, and ambu bags, evidence of monitoring labor with partographs where contraction, cervical dilation, color coding, TPR/BP, urine input/output, and drugs are properly charted and recorded, and availability of procedures for obstetrics emergencies, such as obstructed labor fetal distress, eclampsia, and APH/PPH/HELLP), medical and pediatric wards (e.g., having regular ward rounds, standardized procedures for patient admission as well as handover and discharge reports, and access to basic monitoring equipment and supplies such as pulse oximeter, BP machine, thermometer, and suction machine), theatre (e.g., presence of an operating area with adequate space, lighting, and equipment such as anesthetic machine, monitors, laryngoscope, endotracheal tubes, and a back-up oxygen source, and a scrubbing area with clean clothing, masks, caps, theatre shoes, and hand-washing area), pharmacy (e.g., having adequate storage conditions for drugs, having a system for feeding back medication errors, documentation showing where medicines are procured, and staff that are familiar with reference materials), laboratory (e.g., having procedures for handling, labeling, and storage of specimens, having quality controls for internal and external review and improvement of

lab tests, and having documentation for equipment registration, validation, and calibration), radiology (e.g., having radiation badges for monitoring, a radiation safety service provider, and safety measures with waste management), nutrition and dietetics (e.g., availability of nutrition equipment, formulations, and supplements, such as a weighing scale, glucometer, reference charts, MUAC tape, vitamin A, iron and folate and zinc, and F75 and F100 diet milks), and mortuary (e.g., having standard operating procedures for receiving, identifying, storage, release, and disposal of bodies, adequate infrastructure such as a working drainage system, and adequate security measures).

4 Ethical Considerations

The study was approved by the Ethics and Scientific Review Board at the African Medical and Research Foundation (Approval no. AMREF-ESRC P94/2013), the Kenyan Ministry of Health and authorities at participating facilities. We took care to ensure that our research satisfied the 3 basic principles advanced in the Belmont Declaration (1979) of Respect for Persons, Beneficence and Justice. We sought consent from every facility and patient who participated in the study and by working with every facility in the 3 counties (which were themselves chosen to represent the diversity of counties in the country) we ensured that the principle of justice was fulfilled. The question of whether facilities who do not have a license to practice and/or do not meet minimum patient safety requirements should be closed down, as required by the regulators, satisfied the principle of equipoise, with uncertain potential for harm or benefit. Although there is a strong belief in the global health community and among policymakers that unlicensed practitioners (who are therefore practicing illegally) should not be allowed to function, there is no prior evidence on the costs and benefits of actually doing so. If such clinics are the only option for local populations, shutting them down could decrease health care provision. Prior to the intervention, we used the full census data to plot each market and found that in Kenya, clinics that are stand-alone (that is, do not have any other clinics within 4km) were predominantly public. This meant that were unlicensed clinics, which are all private, to be shutdown, it would still allow for multiple other options in the vicinity. In our data, 97% of unlicensed facilities have at least another facility in the market, 93% have 2 or more, and 98% of patients are in a market with a public facility. In cases where clinics were closed by federal regulators, the closures were conducted with the full cooperation of the county authorities and the local community.

5 Census, Survey Response Rates, Attrition and Accretion

We conducted two complete censuses that coincide with baseline (2015) and endline (2018) data collection. The sample is very close to the census and therefore all our statistics can be interpreted as population statistics. We will refer to 1,319 as the 2018 census of facilities in this study at endline, and similarly to 1,104 as the 2015 census of facilities at baseline. The response rate was 93% at baseline and 97% at endline. [Table S3](#) shows the response rates by survey instruments.

The randomization was conducted in December 2016, more than one year after the baseline census due to delays in the start of the implementation. With high turnover in facilities, a partial update of the census was conducted between October and November 2016 in markets of size 1, 2, and 3. At randomization, 1,258 facilities were listed after the census update. Throughout the project we conducted listing updates on a rolling basis. By endline, 90 facilities that were missed or classified as temporarily or permanently closed during the census were also found operational and added to the nearest market for a total of 1,348 facilities operational at randomization. In total, between randomization and endline, 1,537 facilities are listed in the study as operational at some point, of which 1319 were active at endline ([Table S4](#)).

As [Table S5](#) indicates, we allocated all new facilities that were found by the endline census to one of the 273 fixed markets that were randomized into the treatment and control groups. These new facilities were matched using an algorithm to the nearest randomized market, according to their proximity (if they are within the established proximity). Five markets disappear because all facilities are permanently closed or out of business, for a total of 268 markets at endline.⁶ We show the facilities that were active at randomization based on the new facilities that were added to the census in the continuous updates, and given we know when they started operations. [Table S7](#) shows treatment status as per fixed randomization market versus treatment received. Note that most new facilities and the facilities missed at randomization that were allocated into the fixed treatment markets did not received the treatment, as the majority were added during endline. That is the reason why the proportion of treatment effective for the census at endline is lower than for the randomization census. [Table S6](#) shows the response rate by data collection wave.

⁶Three new facilities (out of 1322) are also excluded from the analysis as they are beyond a defined distance (4km) from any randomized market.

Table S3: Survey Response Rates at Baseline and Endline

	Baseline	Endline
	(1)	(2)
Panel A. Surveys completed		
In-charge/staff surveys	1027	1285
Patient exit surveys	8577	11098
HCW-patient observations	18698	19178
Healthcare worker surveys	1625	2098
Site observations	1886	2534
Panel B. Response rates		
Facility-level interviews	0.93	0.97
Patient exit surveys	0.88	0.85
HCW-patient observations (patient consent to observation)	0.99	1.00
Healthcare worker surveys	0.97	0.98
Site observations	1.00	1.00

Notes. Non-response includes direct refusals, which explicitly request right to not participate in the surveys, and indirect refusals, where surveyors are unable to complete surveys after multiple visits and attempts following a visits protocol. Surveyors conducted the patient-provider observations for a minimum of 3 hours in the baseline data collection, and for a minimum of 2 hours at endline.

Table S4: Census of Health Facilities at Randomization and Endline

	All	Public	Private
	(1)	(2)	(3)
a. Randomization census (Dec 2016)	1258	383	875
b. Facilities active at randomization but missed in census [1]	90	21	69
c. Facilities active at randomization (a+b) [2]	1348	404	944
d. New facilities after randomization [3]	189	12	177
e. Inactive facilities at endline	218	2	216
f. Total in study from randomization to endline (c+d)	1537	416	1121
g. Of which, active at endline (f-e)	1319	414	905

Notes. Facilities active at randomization is based on updated census data using facility self-reported year of establishment. This is our estimate of active facilities at randomization and may be inflated as new facilities were listed in a partial census update in markets of sizes 1, 2, and 3, but out of business (inactive) facilities were not able to be confirmed. New facilities includes facilities active at endline that were not active at randomization.

Table S5: Randomized Markets and Corresponding Health Facilities by County

	Markets				Facilities			
	Kakamega	Kilifi	Meru	Total	Kakamega	Kilifi	Meru	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. At Randomization [1]								
Treatment 1	30	31	29	90	125	82	202	409
Treatment 2 (Scorecards)	30	33	33	96	139	146	219	504
Control	28	31	28	87	113	91	231	435
Total	88	95	90	273	377	319	652	1348
Panel B. At Endline [2]								
Treatment 1	30	30	29	89	119	76	198	393
Treatment 2 (Scorecards)	30	33	30	93	131	140	219	490
Control	28	30	28	86	106	85	245	436
Total	88	93	87	268	356	301	662	1319

Notes. Panel A includes baseline census plus a partial census update that was conducted during the implementation. 273 markets were identified through the clustering algorithm and were randomized into the treatment and control groups. Health facilities in those markets were assigned the same treatment. Panel B includes the randomized markets that were still active at endline, or those with at least one health facility found in the market. 268 markets still remain. Five markets disappear because they permanently closed or out of business. The endline census includes 213 new facilities that were not in the randomization stage and excludes 189 facilities that were out of business. New facilities identified were matched using an algorithm to the nearest randomized market. Three of these new facilities are excluded from the analysis as they are beyond a defined distance (4km) from any randomized market.

Table S6: Health Facility Response Rates by Waves and Treatment Status

	(C) Control Mean	(T-C) Adj. Diff.	Obs.
	(1)	(2)	(3)
Response Rate at Baseline	0.902 (0.298)	0.040*** (0.015) [0.008]	1104
Surveyed Facilities at Baseline (Randomization sample)	0.717 (0.451)	0.060** (0.027) [0.025]	1348
Response Rate at Endline	0.975 (0.157)	-0.000 (0.010) [0.974]	1319

Notes. Standard errors are clustered at the market level and reported in parentheses. *** (***) (*) denotes significance at 1% (5%) (10%) level. Column (2) presents adjusted differences between the means for the treatment markets and the control group. These differences include controls for the 16 randomization strata (by county and market size). This table shows the response rate of the targeted facilities listed in the census at baseline and endline. Non-response includes direct refusals, which explicitly request right to not participate in the surveys, and indirect refusals, where surveyors are unable to complete surveys after multiple visits and attempts following a visits protocol.

Table S7: Proportion of Facilities by Market Treatment Assigned and Treatment Received

Treatment Assigned	Treatment Received			Out of Business at 1st Inspection	Total Facilities
	Treatment 1	Treatment 2 (Scorecards)	No Treatment		
	(1)	(2)	(3)	(4)	(5)
Panel A. Active at Randomization					
Treatment 1	0.90 (0.95)	0.02 (0.02)	0.03 (0.03)	0.05	409 (387)
Treatment 2 (Scorecards)	0.00 (0.00)	0.85 (0.95)	0.05 (0.05)	0.10	504 (454)
Control	0.03 (0.03)	0.00 (0.00)	0.97	NA	434
All					1347 (1275)
Panel B. Active at Endline					
Treatment 1	0.87	0.02	0.11	NA	393
Treatment 2 (Scorecards)	0.00	0.82	0.17	NA	490
Control	0.03	0.00	0.97	NA	436
All					1319

Notes. Estimates in parentheses report the proportion of facilities that received the treatment assigned not including those facilities that were out of business by the first inspection. In Panel A, treatment received differs from treatment by taking into account (i) re-allocation of some markets due to contamination (19 facilities) and (ii) the fact that some facilities included in the randomization became inactive (went out of business, were found duplicate, etc.) during implementation before they were ever inspected (66 facilities). In Panel B, the endline census includes 213 new facilities that were not in the randomization. New facilities are matched to the nearest randomized market using a clustering algorithm. Three of the new facilities are excluded from the analysis as they are beyond a defined distance (4km) from any randomized market. Column 3 includes facilities that were reported by inspectors as out of business but found to be operating during the endline census and therefore did not receive treatment (5 facilities).

6 Variable Definition and Construction

There are various units of analysis for our outcomes including health facility, health care worker (HCW), patient and HCW-patient interaction. The Joint Health Inspection Checklist (JHIC) score and the outpatient caseload are collected at the health facility level. Questions relating to out-of-pocket expenses are asked to a sample of outpatients in exit interviews. Measures of Infection Prevention and Control (IPC) practices (adherence and supplies) are collected through direct observation of healthcare worker-patient interactions. IPC knowledge surveys were applied to the healthcare workers observed, and observation of the availability of supplies required for the practices at the sites were taken while observation took place. When stated, we use weights to adjust for sampling variation and estimate population statistics at the patient, healthcare worker and site level, based on the proportion of surveyed outpatients, healthcare workers and observation sites, with respect to the total outpatient caseload, medical staff and observation sites of the census of facilities.

6.1 Main Outcomes

JHIC Score

This indicator measures patient safety as per adherence to Kenya’s Joint Health Inspection Checklist developed as part of the intervention and gazetted in March 2016 under the Public Health Act. The JHIC includes indicators of patient safety related to protocols, infrastructure, and equipment at the facility level. We included 11 JHIC sections in the survey: (1) health facility infrastructure; (2) general management and recording of information; (3) infection prevention and control; (4) medical consultation; (5) labor ward; (6) medical and pediatric wards; (7) theatre; (8) pharmacy; (9) laboratory; (10) radiology, and (11) nutrition and dietetics. The first 4 sections are common to all facilities. The JHIC scores and its aggregation is defined in the regulation’s scoring system with values ranging from 0 to 100, representing the score as a percentage of the maximum. Further details are included in [Section 3](#).

Outpatient Caseload

The daily outpatient caseload per facility. We use three main sources of data starting with monthly records and, when no records are kept, self-reported weekly and daily data. All data were collected as part of the health facility surveys. The correlation between daily outpatients from monthly records and weekly self-reported data, when data are available for both, is 0.7.

The monthly records checked at the facility level include (1) Ministry of Health (MOH) official reporting books for aggregate totals (MOH Form 717); (2) daily records by patient age (MOH under and over 5); and (3) any other available facility-based monthly patient registers. Sources (2) and (3) were only asked if (1) was not available. We also collect the number of days missing from

the registers, if any. We collect monthly data for 4 different months: November 2017 to February 2018 from any facility records in endline, and November 2014 to February 2015 in baseline. For facilities that started operating in 2018, we also collect data for March and April of that year during endline. Self-reported weekly outpatients for the previous week is collected for all facilities. Daily outpatients for the previous day and two days before the survey are collected for facilities that do not have monthly records.

For the main analysis we use the data from the month of January (2015 for baseline and 2018 for follow-up) as this month has the largest proportion of monthly data. We conduct robustness checks with other months, as well as the average across all months available and verify that the results are not affected by the selection of month.

Out-of-pocket Expenses

Total amount paid for the facility visit as self-reported by patients for all services the day of the visit, including consultation, medicines, laboratory tests, and other expenses. Monetary amounts are reported in purchasing power parity (PPP)-adjusted USD terms set at 2015 (baseline) and 2018 (endline) prices using the Kenyan CPI and PPP conversion factor from the IMF, unless otherwise stated. Figures in current USD are converted at the exchange rate for the year the data were collected with the IMF exchange rates for the corresponding year: 2015 (for baseline data). The exchange rates used are: 1 USD = 98.18 KES (2015).

6.2 Secondary Outcomes

Inactivity/Exits

Equals 1 if a facility is no longer operating as a health facility. This includes those facilities that have permanently closed (out of business) or have changed their operations and only offer specialized services (e.g., standalone laboratories or pharmacies) and no longer provide general outpatient services.

New/Entry

Equals 1 if a facility opened after randomization.

Unlicensed

Equals 1 if a private facility has no license or expired license as per the administrative licensing records provided by the Boards and Councils: Kenya Medical Practitioners and Dentists Council; Clinical Officers Council; and Nursing Council of Kenya. Administrative data is matched to study data using an algorithm to match on facility name and county, with additional manual matching conducted by field teams. When there are multiple matches from both the algorithm and manual

matching, the algorithm is given priority. Expired licenses indicate that a facility does not have a valid license in the year of interest but has a registration number in a previous year for which we have registration data available.

Infection Prevention and Control

Outcomes from this set of outcomes follow definitions as per [Bedoya et al. \(2017\)](#).

(a) Compliance with IPC Practices. Direct observation of IPC practices is based on patient-HCW interactions observed in the consultation room, laboratory, and injection room, where most outpatient care occurs. Compliance indicators are constructed based on indications and actions. Indications are the principal rationale for performing a safety action and occur when there is a risk of transmitting pathogens from one surface to another. For example, “before patient contact” is an indication as there is a risk of microbial transmission to the patient. Actions occur in response to the indications and therefore define whether the safety practice took place, for instance, “washing hands with water and soap prior to patient contact.” Compliance is estimated as the proportion of indications accompanied by the appropriate IPC action. Compliance is estimated individually for each of 20 indications that we measure across the five domains and is defined as:

$$Compliance_{iw} = Action_{iw}/Indication_{iw} \tag{1}$$

where i denotes indication (of 20 indications), and w denotes healthcare workers. Compliance equals 1 if the action corresponds to the right action for each indication, and zero otherwise.

We analyze compliance with safety practices at the indication level: as long as an indication is observed, the action taken in response by the healthcare worker is recorded, regardless of whether the indication occurred at the same time as another indication. We present the mean of the individual indicators as well as the means aggregated by domain and for all safety measures. This implies that the weight of each indicator is its frequency. Weighting across domains/indicators may rely on several possible schemes. For instance, if we wanted to compute the costs of compliance, we could weight these by the cost of the supplies for each; alternatively, if we focused on the benefits, we would weight by the mapping from compliance to the likelihood of Healthcare Acquired Infections. Unfortunately, the literature is far from establishing these likelihoods even in high-income countries, and particularly for primary care.

(b) HWC IPC Knowledge. Equal to 1 if the HCW answers correctly each question related to every indication and domain measured. To develop the measures of knowledge of healthcare workers, we selected questions that map into the practices observed in each of the five domains. We use a combination of open-ended questions such as “Can you name the most important indica-

tions where hand hygiene is recommended?” as well as questions such as “Do you agree or disagree with the following statement: Washing hands before putting on gloves to examine a patient is not necessary,” to elicit knowledge of appropriate use of gloves. We further measure knowledge on characteristics of the practice that may or may not be observed such as “Do you know what is the minimum time recommended to perform hand washing to kill most germs on the hands?” which is not observed if the health care worker never washes his hands. We triangulate some of the questions (e.g., asking a similar concept with different wording and changing the answer between “agree” and “disagree”) and find overall agreement for the questions tested.

(c) Compliance with IPC Supplies. Equal to 1 if the supplies are available based on supplies required to perform the practice for every indication and domain measured. We measure the availability of supplies, policies, and practices necessary for compliance with IPC standard procedures. Supplies were assessed through structured observation during HCW-patient interactions, and only for healthcare workers who had the relevant indications e.g., the supply of puncture-resistant sharps containers was only checked for the health care workers that had to conduct injections. Data are collected at the level of the observation site since their presence may be critical at more than one site in the facility, and as some of them are particular to the procedures performed at that site. Only a selection of practices could be mapped into this set due to difficulties in observing the availability of some supplies without asking for them if they are not visible and/or due to the HCW not performing the safety practice at the moment of observation.

Facility In-charge awareness and knowledge of the regulation and the scorecard

(a) Familiar with the new legislation JHIC (awareness)

Equals 1 if facility in-charge answered “Yes” to the following question, 0 otherwise: “Are you familiar with the new legislation, gazetted in 2016, that uses a new Joint Health Inspection Checklist to inspect public and private health facilities?”

(b) Ever noticed a scorecard (awareness)

Equals 1 if in-charge answered “Yes” to the following question, 0 otherwise: “Have you noticed a scorecard such as this one posted on any health facility?”

Patient awareness and knowledge of scorecards, perceptions about quality and government inspections

(a) Ever noticed a scorecard (awareness)

Equals 1 if patient answered “Yes” to the following question, 0 otherwise: “The government has

introduced scorecards in some health facilities. Have you ever noticed a scorecard similar to this one posted on a facility? (Show the image to the respondent/patient until he/she give you an answer.)”

(b) Know scorecards’ letter ranking (A vs. C)

Equals 1 if patient answered “A” to the following question, 0 otherwise: “Imagine there are two health facilities with two different scorecards (Note: one facility has an “A” scorecard, the other facility has a “C” scorecard), as shown below. Which one do you think is better? (A vs. C)”

(c) Know scorecards’ letter ranking (vignette: A vs. C vs. D)

Equals 1 if patient answered “I would go to Clinic Belief” to the following question, 0 otherwise: “Imagine that there are three clinics in your village/city/town: Faith, Hope, and Belief. You have always gone to Clinic Faith. Yesterday you found out that the government gave a scorecard D to your facility, a C to Clinic Hope, and an A to Clinic Belief. Based on this information, would you change the health facility you go to? (Show the image to the patient/respondent.)” (Note: An image is given to help understanding)

(d) Perceive improvement in 2017 in facility’s quality (If opened before 2018)

Equals 1 if patient answered “improved” to the following question, 0 otherwise: “Since 2017, have you perceived any change in the quality of services provided by this facility?”

(e) Perceive recent government inspection

Equals 1 if patient answered “Yes” to the following question, 0 otherwise: “Do you think this facility has been inspected recently by the government to verify the quality of the services it provides?”

Patient Satisfaction

Overall satisfaction is a rank-ordered variable where patients rate their service satisfaction according to 5 categories: very dissatisfied, dissatisfied, neither satisfied nor dissatisfied, satisfied, and very satisfied. Overall satisfaction is equal to 1 if ranked satisfied or above, and 0 otherwise.

Provider Consultation Practices

(a) Physical examination index

Physical examination index is estimated through principal components analysis (PCA) for a set of variables, including whether the provider checked pulse, blood pressure, temperature, listened (directly or with a stethoscope) to the front and back of the body, and conducted other basic checks (e.g., weight, height).

(b) Prescribed or gave medicines

Equals 1 if the patient reported that the health provider gave any prescription(s) or medicine(s) during the consultation on the day of the visit.

(c) Referred to another facility

Equals 1 if the patient reported that the health provider gave a referral to another health facility for other services during the consultation on the day of the visit.

Time with HCW and Waiting Times

(a) Minutes spent with HCW in examination

Minutes the patient reported spending with the health provider during examination on the day of the visit (if patient received an examination).

(b) Minutes waiting before examination, laboratory, and pharmacy

Minutes spent waiting for select services on the day of the visit as follows: (i) the approximate minutes spent waiting before being examined (if the patient received an examination); (ii) the approximate minutes spent waiting in line for laboratory tests (if the patient received any laboratory services); and (iii) the approximate minutes the patient spent waiting in line to buy medicines in the pharmacy at the facility (if the patient received any medicines from the pharmacy).

Health Facility Staff

(a) Staff costs

Monthly staff cost is estimated using average take-home pay or government job group reported by health facility staff, and where data is collected for all staff in small facilities (<15 staff) or a random sample of 15 staff stratified by cadre if the facility has more than 15 staff. For staff in public facilities, administrative records are used to estimate salary ranges for government job groups, and the midpoint is used for salary ranges by job group. Monthly take-home pay is imputed for 14% of staff using means by cadre, facility ownership (private/public), and facility level. Facility-level averages of take-home pay are estimated separately for clinical and non-clinical staff. Monthly staff cost is the sum of both groups multiplied by the corresponding number of staff in each group.

(b) Staff ratios

The ratio of healthcare workers to total staff is based on reported staff type, where the type includes clinical, administrative, and support staff. Healthcare workers include staff members that

report providing clinical care, including those that may have both clinical and non-clinical roles. Healthcare workers and staff costs per outpatient use monthly outpatient caseload.

JHIC Functional Categories

Indicator variables equal to 1 if the JHIC item belongs to one of the following categories: Infrastructure, equipment, supplies (low-cost and medium-cost separately), management, medical records, and standard operating procedures (SOPs). Infrastructure items include items such as adequate ventilation, lighting, water, and physical structure requirements for emergency rooms and medicine storage. Equipment includes medical devices and equipment like neonatal incubators and delivery beds. Medium-cost supplies include specialized obstetrics and medical ward supplies (e.g., drip stands), as well as radiology supplies. Low-cost supplies include hygiene supplies (disinfectant or waste bins) and personal protective equipment as well as equipment like thermometers, stethoscopes, and sphygmomanometers used to measure blood pressure. Management includes items related to staff management, quality management, and information systems such as patient register systems, equipment service contracts, and quality assurance programs. Medical records include systems to record patients' medical history and records. Standard operating procedures include facility protocols across departments, such as waste management and cleaning charts for infection prevention and control (IPC), and for the handling, labeling and storage of samples in the laboratory.

JHIC Cost Categories

JHIC compliance to items that are classified into 4 cost groups: lowest, low, medium, and high costs. For marginal and fixed costs, we apply the following general rules: (i) for items such as sterile instruments, BP machines, and healthcare worker safety boots, which require a higher number for higher patient loads but are not changed per patient, we assume the item is a fixed cost; (ii) for items where some elements of fixed and marginal costs are included, for example, suction machines where the equipment is fixed but select parts are required to be changed per patient, we classify based on the main item; in this case, the suction machine is classified as a fixed cost; and (iii) for items that may vary based on the size of the facility, for example, having either an ambulance or contact details for an ambulance available, we assume the cost to the majority of facilities in the sample, which are smaller, level 2 facilities which only require contact details for ambulance services.

6.3 Other Key Variables

Predicted Probability of Government Closure for Private Facilities

We estimate a predicted probability of being closed by the government both for control and treat-

ment facilities, based on their baseline or fixed characteristics, including: not having a license at randomization, having an expired license at randomization, the facility level, a categorical variable corresponding to when the facility was open, the type of private and the county. The sample is restricted to private facilities operating at randomization. We calibrate a logit model for treatment facilities, for which we know whether they were closed by the government. To limit the risks of over-fitting the model and biasing the prediction by using the same treatment facilities both for the logit training and predicting sets (which we call the “naive” approach), we follow [Abadie et al. \(2018\)](#) and use a leave-one-out estimator. For each treatment facility h , we use treatment facilities excluding facility h to calibrate the logit model and predict the closure probability of facility h only. For control facilities, we follow the process outlined above and use all treatment facilities to calibrate the logit model and predict the closure probability for each control facility. Facilities are then classified in three categories: “Low” probability of closure for those between 0 and less than 0.4, “Mid” probability of closure for facilities between 0.4 and 0.6, and “High” probability of closure for facilities with predicted probability higher than 0.6. Figure 4 presents results using both the leave-one-out and the “naive” estimators.

Wealth Index

Constructed asset index based on ownership of assets based on socioeconomic questions from the Kenya Demographic and Health Survey (DHS) to permit calibration against the national distribution. The variables include whether the household has electricity from the grid; running water in the house; a clock; a radio; a television; a mobile telephone; a landline; a refrigerator; a solar panel; a bicycle; a boat; a car or truck; a motorbike; an animal drawn cart; any farm animals. Also, how many of the following livestock the household owns: cows or bulls; horses; goats; sheep; chicken. Finally, the index includes the number of people per sleeping rooms; and whether the toilets are shared with other households. The select variables for our Wealth Index are highly statistically significant when regressed against the DHS wealth index, and explain 82% of its variation for the national sample.

7 Additional Analyses

7.1 Cross-Market Externalities

A potential spillover effect of interest in this setting results from cross-market externalities which may lead to impacts on control health facilities in markets located near treatment facilities. If these externalities are present, impact estimates comparing health facilities in treated and untreated markets will be biased, with the direction of bias depending on the nature of the spillover. To identify cross-market externalities, we use exogenous variation in the local density of facilities induced by the stratified market-level randomization, following a method that is similar to the method in Miguel & Kremer (2004), but focusing exclusively on the externalities for control facilities. In particular, although all treatment markets, defined as per our z-center clustering algorithm ($z=4\text{km}$) are treated, a control facility x could still be located within 4km of a treatment facility w , but would not be included in w 's market because x is more than 4km away from the center of the treated market w . Therefore, we leverage the random variation in treatment facilities close to control facilities for this estimation.⁷ Equation 2 below presents the specification we use to estimate cross-market externalities, or the impact of the program on facilities in the control group:

$$Y_{hm}^C = \alpha + \delta_{04} \cdot N_{04h}^T + \lambda_{04} \cdot N_{04h} + \sum_{j=1}^{n-1} \theta_j V_{hj} + \gamma X_h + \theta Z_{m,t-1} + \epsilon_{hjm} \quad (2)$$

where Y_{hm}^C indicates any of the outcomes of interest for health facility h in market m for control markets C ; N_{04h}^T is the number of facilities randomly assigned to treatment markets T within a 4km from facility h and N_{04h} is the number of all facilities (treatment and control) within distance d from facility h ; X_h is the level of facility h ; $Z_{m,t-1}$ is a set of market baseline controls, and V_{hj} are the randomization strata; and δ_{04} is our coefficient of interest measuring the regulation treatment externalities across markets. In this specification $\delta_{04} \cdot \bar{N}_{04h}^T$ is the average externality of inspections on control facilities, where \bar{N}_{04h}^T is the average number of treatment facilities located within 4km from control facility h .

We always control for the total density of health facilities since it may affect the outcomes of interest or the impact of the intervention, for instance, through higher competition. The identifying assumption is that conditional on the total number of facilities within a certain distance from the facility, the number of these facilities assigned to treatment is exogenous. Table S13 shows that for control facilities the main outcomes of interest at baseline are not significantly associated with the number of treatment facilities within 4km, conditional on the total local density of health facilities.

We find no statistically significant spillover impacts on the JHIC score in control facilities. Table S12 shows that the JHIC score is 0.21 percentage points (p-value = 0.50) higher in control

⁷Taking 4km as our benchmark, 31% of control facilities have no treatment facility within 4km but 44% have 1-3 treatment facilities within this distance and 25% have 4 or more treatment facilities within the same distance.

health facilities within four kilometers of treatment facilities. To see this, note that the average spillover impact on control facilities is the average number of treatment facilities located within four kilometers \bar{N}_{04h}^T times the average estimated effect of an additional facility in the treatment group located within four kilometers γ_{04} . Based on estimates in [Table S12](#) this implies that the estimated cross-facility externality in patient safety is $[\hat{\delta}_{04} \cdot \bar{N}_{04h}^T] = [0.063 \cdot 3.3] = 0.21$ percentage points. Even if we lack power to detect this impact, it represents less than 1% of the mean JHIC score in the control group of 35%, and just over 4% of the impact on treatment facilities of 5.2 percentage points ([Table 3](#)).

7.2 Program Duration

We exploit variation in the timing of inspections and endline data collection to assess the heterogeneity of impacts by length of exposure to the program. Exposure length captures both (potential) fade-out and the fact that facilities who received an inspection earlier in the cycle will have received more inspections. We use the following specification:

$$\begin{aligned} JHIC\ Score_{hm} = & \alpha + \delta_1 T_{m123} + \delta_2 T_{m456} + \beta_1 E_{h12} + \delta_3 T_{m123} * E_{h12} + \\ & \delta_4 T_{m456} * E_{h12} + \theta MainTownDistance_h + \gamma X_h + \sum_{k=1}^{n-1} \phi_k V_{hk} + \varepsilon_{hm} \end{aligned} \quad (3)$$

where $JHIC\ Score_{hm}$ represents the JHIC score for health facility h in market m at endline; T_{mj} is an indicator that equals one if the first inspection in treatment market m is in one for two groups $j=123$ (1-to-3-month group), 456 (4-to-6-month group), and zero otherwise;⁸ E_{h12} is an indicator that equals one if facility h has endline data collection in the 1-to-2-month group, and zero otherwise;⁹ δ_l are the parameters of interest. As before, X_h represents the facility level, V_h indicates the randomization strata, and ε_{hm} represents a set of random unobserved characteristics. We control for the distance to the main county town, $MainTownDistance_h$, to account for potential imbalances due to the inspections and data collection starting in facilities closer to the main county town, which tend to have higher average JHIC than facilities in more rural areas. Our main identifying assumption is that conditional on the controls, the variation in the market date of first inspection (T_{mj}) and endline date (E_{h12}) is not correlated with the JHIC score at endline. [Table S15](#) reports balance of the baseline JHIC score across the groups of interest for these estimates.

[Table S14](#) and [Figure S4](#) show that the impact of inspections on the JHIC score was 7 percentage

⁸The start of the intervention was interrupted over the first two months. For the majority of the markets (85%), the first inspection happened within 6 months between mid-January and mid-July 2017. Therefore, we use this as the main benchmark for this analysis. However, the intervention started slowly between mid-November 2016 until mid-January 2017 (7% of the markets). We include these markets in the 1-to-3-month group. For the remaining markets (8%, 2% facilities), the first inspection was spread between mid-July and November 2017. To avoid outliers, we include these markets in the 4-to-6-month group.

⁹Most of the endline data collection happened in 4 months between March and June 2018 (98% of facilities). The endline for the remaining facilities (2%) happened between July and August, so they are included in the 3-to-4-month group at endline.

points (0.65 SD, p-value < 0.005) for facilities with the longest exposure to the program in treatment markets. For this group, the average time between first (last) inspection and endline is 15 (10) months. Effects are smaller for facilities that were in the program for a shorter period, with effect sizes ranging between 3.7 (0.35 SD, p-value = 0.004) and 5.3 percentages points (0.5 SD, p-value < 0.005). The markets with the longest exposure to the program received 2.4 inspection vs. between 1.6 and 2.0 inspections for the other groups.

7.3 Quantile Treatment Effects

To estimate unconditional quantile treatment effects in [Table S16](#), we use the following specification:

$$\Delta_{QTT}(\tau) = q_1(\tau|R = 1) - q_0(\tau|R = 0) \quad (4)$$

where $q_D(\tau|R = 1)$ is the τ -th quantile of potential outcomes Y_D under treatment. This specification assumes full compliance with the random assignment to estimate the treatment-on-the-treated effects. Given the nearly universal program compliance measured for the policy, this approximation is justifiable to simplify analysis without risk of bias.

To estimate conditional quantile treatment effects in [Figure 3](#), we use conditional quantile regressions with the following specification:

$$Y_{hm}^q = \alpha + \beta_1 T_m + Level_h + FE_{counties} + \epsilon_{hm} \quad (5)$$

where $Y_{h,m}^q$ is the outcome of interest (patient safety as measured by the JHIC score, as per regulation, excluding licensing) at quantile q for health facility h in market m at follow-up; T_m is a treatment indicator at the market level that equals one for facilities belonging to the treatment group and zero for facilities in the control group; $Level_h$ and $FE_{counties}$ are two vectors of dummy variables for the health facility level and county; and $\epsilon_{h,m}$ represent unobserved characteristics. Standard errors are robust and clustered at the market level. In [Figure 3](#), we run the previous regression first on the whole sample and by ownership, ([Panel A](#)) and second by market density (1-2 health facilities, 3-10 health facilities and 11 and more health facilities in the market) in [Panels B and C](#) of the table. Note that for consistency across panels, we do not control for market size in the quantile regression.

8 Costing

There are two important components of the cost of the intervention: the investment to set up the system and the cost of conducting routine operations. We focus on the cost of routine operation activities in this section, which helps us assess the average operational cost per visit to make the system work as intended by the regulation. This section is based on [Bedoya et al. \(2020\)](#).

The pilot followed a particular model of inspections where all inspectors were located in the county headquarters and used vehicles provided by central and regional governments to visit health facilities. Inspectors were seconded from different government institutions, and most transferred from other regions. Facility closure visits required staff from the central government to travel to the regions. Additional external support was required for implementation and monitoring of the operation. This is a poor model for costs in a fully scaled-up version, where the number and location of inspectors can be flexibly determined and external support is minimized. Nevertheless, the routine pilot costs help provide a benchmark that can be improved upon using standard tools from operational research.

On average, a visit to a health facility during the pilot cost USD 165 in operational costs. Of this, USD 54 (33%) were inspector costs, including salaries, allowances and compensations for being outside of their duty station; USD 17 (10%) were transportation costs to visit each facility; and USD 13 (8%) included other costs related to office, supplies and technology. The remaining USD 81 (49%) of the total, included government management (USD 22 per visit) and external World Bank support for implementation, MIS management and inspection quality assurance (USD 59 per visit).

Several factors complicate the interpretation of this cost. First, for 28% of visits the inspector could not start the inspection and the facility required multiple visits.¹⁰ Second, there were days when vehicles were used for other government activities or were not functioning. Third, there were days when vehicles were available, but inspectors were absent. Therefore, we view this cost per visit as an upper-bound, since at least three of these problems—unsuccessful visits, non-functioning vehicles and inspector absence (during which we paid for the vehicles)—can be sharply reduced in subsequent years with more experience.¹¹ For instance, in a scaled-up version, a team of two inspectors could complete 6 inspection visits in a day (versus 3.5 during the pilot) with variation across regions based on market structures. Additionally, the World Bank support would be reduced considerably by hiring local experts. These two actions would imply a per-visit cost of USD 95. Alternate models where inspectors are either located in multiple cities in the county or have multiple

¹⁰Visits did not result in an inspection because the in-charges were absent or had run away when the inspector arrived (likely due to lack of a license). Due to vehicle constraints, inspectors in these cases were required to wait for a shared vehicle to come back after taking other inspectors to separate (sometimes distant) facilities, before proceeding to the next facility. Waiting times could be up to several hours.

¹¹For example, the government established a new protocol that would lead to closure reports for multiple unsuccessful visits. These visits are expected to decrease further as the system matures.

bases within which they travel will further decrease transport costs, and a larger scale of inspections will also decrease costs per visit related to office, supplies, technology and management.¹²

¹²We exclude from these costs the fixed costs of building the inspection system, which included the development of the enhanced regulatory framework, implementation protocols, training materials, and the electronic inspection system.

9 Other Tables and Figures

Table S8: Correlations of Select Indicators with Market Size at Randomization

	JHIC Score at Baseline at the Market Level (All Markets)	Market Size at Randomization (Treated Markets Only)	
	(1)	(2)	(3)
Market Size at Randomization	0.311 (0.212)		
Month of First Inspection Visit in Market		-0.079 (0.079)	
Average Inspections in the Market			-0.083 (0.352)
Observations	259	178	178
R ²	0.284	0.849	0.848

Notes. Robust standard errors are reported in parentheses. *** (**) (*) denotes significance at 1% (5%) (10%) level. Regressions include controls for the 16 strata included in the randomization (by county and market size) and health facility level controls (levels are 2, 3, and 4 and 5 merged together, and are positively related to the amount, and specialization of services provided) as percentage in each market. Columns 2 and 3 include controls for mean baseline JHIC score at the market level.

Table S9: Share of Total Patients and Facilities:
Facilities that Exited and Entered by Endline

	Share of patients when operational	Share of facilities when operational
	(1)	(2)
Exited by endline (active at baseline)	0.03	0.12
Entered by endline	0.12	0.24
Entered by endline (control only)	0.12	0.26

Notes. Exit equals the share of patients that the facilities that exited the market by endline were responsible for in the baseline. Enter equals the share of patients that the facilities that entered the market after baseline were responsible for in the endline. Share of facilities when operational includes those with patient data available. This table does not include facilities that exited by endline and were active after the baseline.

Table S10: Treatment Effects on JHIC Score, OOP, and Outpatients
with Baseline Market Level Controls:
Overall and Interacted with Indicators for Private and Unlicensed Health Facilities at Endline

	Unweighted		Weighted		Daily Outpatients (5)
	JHIC Score (pp of max) (1)	OOP (USD PPP) (2)	JHIC Score (pp of max) (3)	OOP (USD PPP) (4)	
Panel A: Overall Impact					
Treatment	4.640*** (0.685) {0.001}***	0.495 (0.343) {0.177}	3.675*** (1.007) {0.001}***	0.209 (0.349) {0.333}	1.789 (1.573) {0.238}
Observations	1285	1285	1285	1285	1285
R ²	0.358	0.150	0.567	0.252	0.262
Control Mean	35.493	4.069	42.526	3.136	20.793
Impact: {%; SD}	{13%; 0.44}	{12%; 0.10}	{9%; 0.31}	{7%; 0.04}	{9%; 0.06}
Panel B: Interaction with Private					
Treatment	2.303** (1.040) [0.028]	-0.367 (0.263) [0.164]	2.475 (1.534) [0.108]	0.160 (0.267) [0.549]	6.797** (3.268) [0.039]
Private HF	-6.058*** (1.014) [0.000]	4.210*** (0.306) [0.000]	-1.541 (2.130) [0.470]	4.611*** (0.523) [0.000]	-28.898*** (2.916) [0.000]
Private HF x T	3.252*** (1.120) [0.004]	1.553*** (0.507) [0.002]	3.637* (2.118) [0.087]	0.526 (0.764) [0.492]	-9.260** (4.066) [0.024]
Observations	1285	1285	1285	1285	1285
R ²	0.379	0.234	0.572	0.378	0.420
Control Mean Public	39.760	0.643	42.236	0.808	41.060
Control Mean Private	33.463	5.698	43.033	7.211	11.151
Impact Public: {%; SD}	{6%; 0.26}	{-57%; -0.41}	{6%; 0.27}	{20%; 0.14}	{17%; 0.22}
Impact Private: {%; SD}	{17%; 0.52}	{21%; 0.23}	{14%; 0.39}	{10%; 0.12}	{-22%; -0.10}
Test T + Private x T = 0 (p-value)	0.000	0.010	0.000	0.386	0.130
Panel C: Interaction with Unlicensed (Private and active at endline only)					
Treatment	6.030*** (0.933) [0.000]	0.902 (0.718) [0.211]	8.110*** (0.895) [0.000]	0.574 (0.918) [0.532]	-2.697 (2.369) [0.256]
Unlicensed at Endline	-3.280*** (0.960) [0.001]	-0.951 (0.630) [0.133]	-0.137 (1.329) [0.918]	-2.155*** (0.764) [0.005]	-3.276 (2.267) [0.150]
Unlicensed at Endline x T	-1.976 (1.316) [0.135]	0.226 (0.847) [0.790]	-5.447** (2.186) [0.014]	0.496 (1.135) [0.662]	1.263 (2.904) [0.664]
Observations	872	872	872	872	872
R ²	0.418	0.111	0.661	0.138	0.317
Control Mean Licensed	36.703	6.393	45.718	8.083	15.821
Control Mean Unlicensed	30.086	4.974	35.991	4.924	6.283
Impact Licensed: {%; SD}	{16%; 0.54}	{14%; 0.16}	{18%; 0.51}	{7%; 0.09}	{-17%; -0.08}
Impact Unlicensed: {%; SD}	{13%; 0.43}	{23%; 0.25}	{7%; 0.23}	{22%; 0.27}	{-23%; -0.13}
Test T + Unlicensed x T = 0 (p-value)	0.000	0.032	0.146	0.213	0.178

Notes. This table corresponds to Figure A5. Robust standard errors reported in parentheses and clustered at the market level. *** (**) (*) denotes significance at 1% (5%) (10%) level. “Naive” p-values are reported in brackets and sharpened q-values in braces, following ?. Stars reported next to the estimated coefficients denote significance related to the “naive” p-value. Missing values for OOP in 5.8% of observations are imputed using means defined by level, ownership, treatment, license status at randomization, and daily outpatients. Regressions include randomization strata controls (by county and market size), health facility level controls, and baseline JHIC score, OOP, and outpatient controls at the market level with missing values replaced by zero and with a dummy indicator for whether values are missing. HF = health facility; JHIC = Joint Health Inspection Checklist; OOP = out-of-pocket; PPP = purchasing power parity.

Table S11: Treatment Effects on JHIC Score, OOP, and Outpatients
without Imputation and Including Strata Controls Only:
Overall and Interacted with Indicators for Private and Unlicensed Health Facilities at Endline

	Unweighted		Weighted		Daily Outpatients (5)
	JHIC Score (pp of max) (1)	OOP (USD PPP) (2)	JHIC Score (pp of max) (3)	OOP (USD PPP) (4)	
Panel A: Overall Impact					
Treatment	5.090*** (0.838) {0.000}***	0.975** (0.423) {0.048}**	4.217** (1.808) {0.044}**	0.186 (0.594) {0.705}	1.252 (1.601) {0.402}
Observations	1285	1285	1285	1285	1285
R ²	0.193	0.125	0.235	0.162	0.034
Control Mean	35.493	4.069	42.526	3.136	20.793
Impact: {%; SD}	{14%; 0.48}	{24%; 0.20}	{10%; 0.35}	{6%; 0.04}	{6%; 0.04}
Panel B: Interaction with Private					
Treatment	2.728** (1.088) [0.013]	-0.040 (0.243) [0.868]	3.464** (1.683) [0.041]	0.483 (0.301) [0.110]	8.282** (3.478) [0.018]
Private HF	-7.782*** (1.069) [0.000]	4.209*** (0.389) [0.000]	-1.525 (3.475) [0.661]	5.317*** (1.243) [0.000]	-33.463*** (3.049) [0.000]
Private HF x T	3.515*** (1.262) [0.006]	1.484** (0.584) [0.012]	2.227 (3.867) [0.565]	-0.154 (1.317) [0.907]	-10.252** (4.418) [0.021]
Observations	1285	1285	1285	1285	1285
R ²	0.235	0.216	0.236	0.318	0.263
Control Mean Public	39.760	0.643	42.236	0.808	41.060
Control Mean Private	33.463	5.698	43.033	7.211	11.151
Impact Public: {%; SD}	{7%; 0.31}	{-6%; -0.05}	{8%; 0.38}	{60%; 0.42}	{20%; 0.27}
Impact Private: {%; SD}	{19%; 0.58}	{25%; 0.28}	{13%; 0.37}	{5%; 0.06}	{-18%; -0.08}
Test T + Private x T = 0 (p-value)	0.000	0.009	0.126	0.804	0.320
Panel C: Interaction with Unlicensed (Private and active at endline only)					
Treatment	6.292*** (1.266) [0.000]	1.062 (0.767) [0.168]	5.586 (3.975) [0.162]	-0.164 (1.506) [0.913]	-3.633 (2.806) [0.197]
Unlicensed HF at Endline	-5.527*** (1.153) [0.000]	-1.348** (0.543) [0.014]	-10.237** (3.983) [0.011]	-3.380** (1.327) [0.012]	-9.189*** (3.056) [0.003]
Unlicensed HF at Endline x T	-1.358 (1.484) [0.361]	0.526 (0.824) [0.524]	-2.449 (4.704) [0.603]	1.477 (1.573) [0.349]	3.279 (3.234) [0.312]
Observations	872	872	872	872	872
R ²	0.310	0.089	0.284	0.075	0.065
Control Mean Licensed	36.703	6.394	45.718	8.083	15.821
Control Mean Unlicensed	30.086	4.973	35.991	4.924	6.283
Impact Licensed: {%; SD}	{17%; 0.57}	{17%; 0.19}	{12%; 0.35}	{-2%; -0.03}	{-23%; -0.11}
Impact Unlicensed: {%; SD}	{16%; 0.53}	{32%; 0.35}	{9%; 0.27}	{27%; 0.34}	{-6%; -0.03}
Test T + Unlicensed x T = 0 (p-value)	0.000	0.009	0.136	0.056	0.748

Notes. This table corresponds to Figure A5. Robust standard errors reported in parentheses and clustered at the market level. *** (**) (*) denotes significance at 1% (5%) (10%) level. “Naive” p-values are reported in brackets and sharpened q-values in braces, following ?. Stars reported next to the estimated coefficients denote significance related to the “naive” p-value. Missing values for OOP in 5.8% of observations are imputed using means defined by level, ownership, treatment, license status at randomization, and daily outpatients. Regressions include randomization strata controls (by county and market size).

Table S12: Spillover Effects on JHIC Score, OOP, Outpatients, Facility Inactivity, and Entry for Control Facilities

	JHIC Score (pp of max) (1)	OOP (USD PPP) (2)	Daily Outpatients (3)	Inactive (4)	New (5)
Number of T HF's within 4km	0.063 (0.115) [0.583]	0.044 (0.055) [0.427]	0.097 (0.322) [0.764]	0.006 (0.005) [0.182]	-0.006 (0.004) [0.129]
Number of HF's within 4km	-0.095 (0.076) [0.213]	0.002 (0.037) [0.950]	-0.086 (0.166) [0.607]	-0.002 (0.002) [0.282]	0.004 (0.002) [0.110]
Observations	421	390	481	491	431
R ²	0.184	0.209	0.072	0.094	0.099
Control Mean	35.379	3.955	18.212	0.122	0.135
Mean No. T HF's in 4km	3.3	3.3	3.3	3.3	3.3

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Regressions include randomization strata controls (by county and market size), proportion of health facility level at market-level controls, and baseline JHIC score, OOP, and outpatient controls at the market level applied to all columns, with missing values replaced by zero and with a dummy indicator for whether values are missing values at the market level. 5 outliers are dropped for facilities with the number of T facilities within 4km ≥ 40 .

Table S13: Spillovers: Baseline Balance on Main Outcomes

	JHIC Score (pp of max) (1)	OOP (USD PPP) (2)	Daily Outpatients (3)
Number of T HF's within 4 kms	0.004 (0.139) [0.979]	-0.020 (0.122) [0.869]	0.764 (0.613) [0.216]
Number of HF's within 4 kms	-0.001 (0.085) [0.988]	0.053 (0.069) [0.448]	-0.132 (0.249) [0.598]
Observations	310	287	310
R ²	0.103	0.163	0.079
Control Mean	35.479	4.516	24.881
Mean No. T HF's in 4kms	3.3	3.3	3.3

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Regressions include randomization strata controls (by county and market size). 5 outliers are dropped for facilities with number of T facilities within 4km ≥ 40 .

Table S14: Impact on JHIC Score by Month of First Inspection and Endline Month Interaction

	JHIC Score (pp of max) (1)
First Market Inspection in Months 1-3 (Ta)	5.341*** (1.473) [0.000]
First Market Inspection in Months 4-6 (Tb)	3.961*** (0.999) [0.000]
Endline in Months 1-2	2.730*** (0.991) [0.006]
Endline in Months 1-2 x Ta	1.616 (1.673) [0.335]
Endline in Months 1-2 x Tb	-0.267 (1.598) [0.867]
Observations	1285
R ²	0.354
Control Mean	39.399
Estimated Coefficients for Endline in Months 3-4	
Inspection in Months 1-3	6.957*** [0.000]
p-value	
Inspection in Months 4-6	3.694*** [0.004]
p-value	
Impact {%; SD}	
Inspection in Months 1-3 and Endline in Months 1-2	{15%; 0.50}
Inspection in Months 1-3 and Endline in Months 3-4	{20%; 0.65}
Inspection in Months 4-6 and Endline in Months 1-2	{11%; 0.37}
Inspection in Months 4-6 and Endline in Months 3-4	{10%; 0.35}

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Regressions include randomization strata controls (by county and market size), health facility level controls, and distance to the main town in each county. The start of the intervention was interrupted over the first two months. For the majority of the markets (85%), the first inspection happened within 6 months between mid-January and mid-July 2017. Therefore, we use this as the main benchmark for this analysis. However, the intervention started slowly between mid-November 2016 until mid-January 2017 (7% of the markets). We include these markets in the 1-to-3-month group. For the remaining markets (8%, 2% facilities), the first inspection was spread between mid-July and November 2017. To avoid outliers, we include these markets in the 4-to-6-month group. Most of the endline data collection happened in 4 months between March and June 2018 (98% of facilities). The endline for the remaining facilities (2%) happened between July and August, so they are included in the 3-to-4-month group at endline.

Table S15: Baseline Balance on JHIC Score by Month of First Inspection and Endline Month Interaction

	JHIC Score at Baseline (pp of max) (1)
First Market Inspection in Months 1-3 (Ta)	0.153 (1.284) [0.905]
First Market Inspection in Months 4-6 (Tb)	1.447 (0.966) [0.135]
Endline in Months 1-2	3.314*** (1.101) [0.003]
Endline in Months 1-2 x Ta	1.216 (1.702) [0.476]
Endline in Months 1-2 x Tb	-1.686 (1.708) [0.325]
Observations	889
R ²	0.382
Control Mean	37.461
Estimated Coefficients for Endline in Months 3-4	
Inspection in Months 1-3	1.369
p-value	[0.200]
Inspection in Months 4-6	-0.238
p-value	[0.872]

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Regressions include randomization strata controls (by county and market size), health facility level controls, and distance to the main town in each county. The start of the intervention was interrupted over the first two months. For the majority of the markets (85%), the first inspection happened within 6 months between mid-January and mid-July 2017. Therefore, we use this as the main benchmark for this analysis. However, the intervention started slowly between mid-November 2016 until mid-January 2017 (7% of the markets). We include these markets in the 1-to-3-month group. For the remaining markets (8%, 2% facilities), the first inspection was spread between mid-July and November 2017. To avoid outliers, we include these markets in the 4-to-6-month group. Most of the endline data collection happened in 4 months between March and June 2018 (98% of facilities). The endline for the remaining facilities (2%) happened between July and August, so they are included in the 3-to-4-month group at endline.

Table S16: Unconditional QTE on JHIC Score, by Ownership and Market Size

	By Ownership			By Market Size Group			By Market Size Group (Private)		
	All	Private	Public	1-2 HF's	3-10 HF's	11+ HF's	1-2 HF's	3-10 HF's	11+ HF's
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Percentile									
10th	2.780** (1.162)	4.450*** (1.569)	2.560 (2.022)	2.610 (3.029)	4.430*** (1.460)	2.970 (2.140)	4.670 (7.550)	4.780*** (1.539)	2.380 (2.093)
25th	4.970*** (1.284)	4.860*** (1.542)	2.330 (1.515)	3.160 (2.803)	4.580*** (1.647)	7.340*** (2.361)	5.610 (5.332)	3.950** (1.676)	6.760*** (2.406)
50th	7.400*** (1.376)	8.620*** (1.790)	2.910* (1.523)	1.300 (2.377)	6.200*** (1.552)	10.860*** (2.831)	10.320** (4.678)	7.180*** (1.998)	10.390*** (3.284)
75th	6.860*** (1.320)	8.760*** (2.194)	3.840*** (1.428)	1.650 (2.098)	6.520*** (1.380)	11.540*** (3.227)	10.110** (4.505)	8.090*** (1.960)	9.740*** (3.136)
90th	7.550*** (2.006)	9.640*** (3.019)	5.710*** (1.861)	3.320 (2.155)	4.820** (2.068)	11.150*** (4.060)	4.040 (5.600)	7.440*** (2.654)	11.150** (4.633)
Observations	1285	872	413	191	638	456	66	413	393
Test 10th = 90th	0.001	0.018	0.169	0.833	0.869	0.010	0.927	0.319	0.012

Notes. Robust standard errors are clustered at the market level and reported in parentheses. *** (**) (*) denotes significance at 1% (5%) (10%) level. Regressions include controls for county, health facility level, and JHIC score quartile dummies at baseline, where missing values are replaced by zero and an indicator for missing. Conditional quantile treatment effect are estimated using the qreg2 stata command, which allows to estimate robust standard errors clustered at the market level but does not allow to test the equality of the 10th and 90th coefficient. For the test, we run simultaneously the five quantile regressions using the stata command sqreg, which allows us to test the equality of percentile coefficients but does not estimate robust standard errors accounting for intra cluster correlation. In this case, standard errors are estimated after performing 1,000 bootstrap replications.

Table S17: Treatment Effects on JHIC Score by Closure Probability Group

	JHIC Score
	(1)
Treatment	7.556*** (0.955) [0.000]
Mid prob. (0.4,0.6)	-1.934 (1.669) [0.248]
Mid prob. (0.4,0.6) x T	-2.676 (2.002) [0.183]
High prob. ≥ 0.6	-3.832*** (1.390) [0.006]
High prob. ≥ 0.6 x T	-2.092 (1.651) [0.207]
Observations	719
R ²	0.422
Mean Control Low ≤ 0.4	36.409
Mean Control Mid (0.4,0.6)	32.193
Mean Control High ≥ 0.6	28.868
Impact Low ≤ 0.4 {% change; SD}	{21%; 0.72}
Impact Mid (0.4,0.6) {% change; SD}	{15%; 0.42}
Impact High ≥ 0.6 {% change; SD}	{19%; 0.63}
Test (T + Mid x T = 0) p-value	0.013
Test (T + High x T = 0) p-value	0.000

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (**) (*) denotes significance at 1% (5%) (10%) level. Sample: Private facilities active at randomization. Regressions include 16 randomization strata controls (by county and market size), baseline market controls for JHIC, OOP, and outpatients, and health facility level controls.

Table S18a: Treatment Effects on JHIC Score, OOP, Outpatients, and Entry:
Interacted with and Indicator for JHIC Score at Baseline >40

	Unweighted		Weighted		Daily Outpatients (5)	New (6)
	JHIC Score (pp of max) (1)	OOP (USD PPP) (2)	JHIC Score (pp of max) (3)	OOP (USD PPP) (4)		
Treatment	5.472*** (0.941) [0.000]	0.779 (0.528) [0.141]	3.702*** (1.152) [0.001]	0.363 (0.478) [0.447]	0.123 (2.117) [0.954]	0.000 (0.000) [.]
JHIC Score at Baseline >40	7.760*** (1.167) [0.000]	-0.316 (0.565) [0.576]	5.643*** (1.438) [0.000]	0.647 (0.737) [0.381]	11.605** (5.036) [0.022]	0.000 (0.000) [.]
JHIC Score at Baseline >40 x T	0.548 (1.343) [0.684]	0.840 (0.839) [0.318]	1.339 (1.857) [0.472]	-0.152 (0.819) [0.853]	1.164 (6.007) [0.847]	0.000 (0.000) [.]
Observations	889	889	889	889	889	901
R ²	0.408	0.152	0.574	0.203	0.262	.
Control Mean JHIC <=40	33.625	4.005	36.685	2.562	17.494	0.000
Control Mean JHIC >40	44.091	2.950	48.406	3.511	44.610	0.000
Impact JHIC <=40: {%; SD}	{16%; 0.64}	{19%; 0.16}	{10%; 0.48}	{14%; 0.09}	{1%; 0.01}	{.%; .}
Impact JHIC >40: {%; SD}	{14%; 0.60}	{55%; 0.38}	{10%; 0.43}	{6%; 0.04}	{3%; 0.03}	{.%; .}
Test T + JHIC >40 x T = 0 (p-value)	0.000	0.061	0.003	0.814	0.810	.

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (***) (*) denotes significance at 1% (5%) (10%) level. Sample: facilities active at baseline and endline. Missing values for OOP are imputed using means defined by level, ownership, treatment, license status at randomization, and daily outpatients. Regressions include randomization strata controls (by county and market size).

Table S18b: Treatment Effects on JHIC Score, OOP, Outpatients, and Entry:
Interacted with and Indicator for JHIC Score at Baseline >50

	Unweighted		Weighted		Daily Outpatients (5)	New (6)
	JHIC Score (pp of max) (1)	OOP (USD PPP) (2)	JHIC Score (pp of max) (3)	OOP (USD PPP) (4)		
Treatment	5.289*** (0.799) [0.000]	0.874* (0.511) [0.088]	3.739*** (0.889) [0.000]	0.369 (0.387) [0.342]	0.761 (1.963) [0.699]	0.000 (0.000) [.]
JHIC Score at Baseline >50	7.175*** (2.242) [0.002]	0.099 (0.869) [0.909]	4.426 (3.136) [0.159]	2.082 (1.484) [0.162]	23.862** (10.107) [0.019]	0.000 (0.000) [.]
JHIC Score at Baseline >50 x T	1.632 (2.489) [0.513]	1.420 (1.025) [0.167]	1.629 (3.762) [0.665]	-0.504 (1.734) [0.771]	-6.100 (12.638) [0.630]	0.000 (0.000) [.]
Observations	889	889	889	889	889	901
R ²	0.364	0.154	0.552	0.213	0.266	.
Control Mean JHIC <=50	33.625	4.005	36.685	2.562	17.494	0.000
Control Mean JHIC >50	44.091	2.950	48.406	3.511	44.610	0.000
Impact JHIC <=50: {%; SD}	{16%; 0.64}	{19%; 0.16}	{10%; 0.48}	{14%; 0.09}	{1%; 0.01}	{.%; .}
Impact JHIC >50: {%; SD}	{14%; 0.60}	{55%; 0.38}	{10%; 0.43}	{6%; 0.04}	{3%; 0.03}	{.%; .}
Test T + JHIC >50 x T = 0 (p-value)	0.003	0.039	0.140	0.938	0.667	.

Notes. Robust standard errors are reported in parentheses and clustered at the market level. P-values are reported in brackets. *** (***) (*) denotes significance at 1% (5%) (10%) level. Sample: facilities active at baseline and endline. Missing values for OOP are imputed using means defined by level, ownership, treatment, license status at randomization, and daily outpatients. Regressions include randomization strata controls (by county and market size).

Figure S3: JHIC Score Density in Control Facilities
by Data Collection Round (% of max score x100)

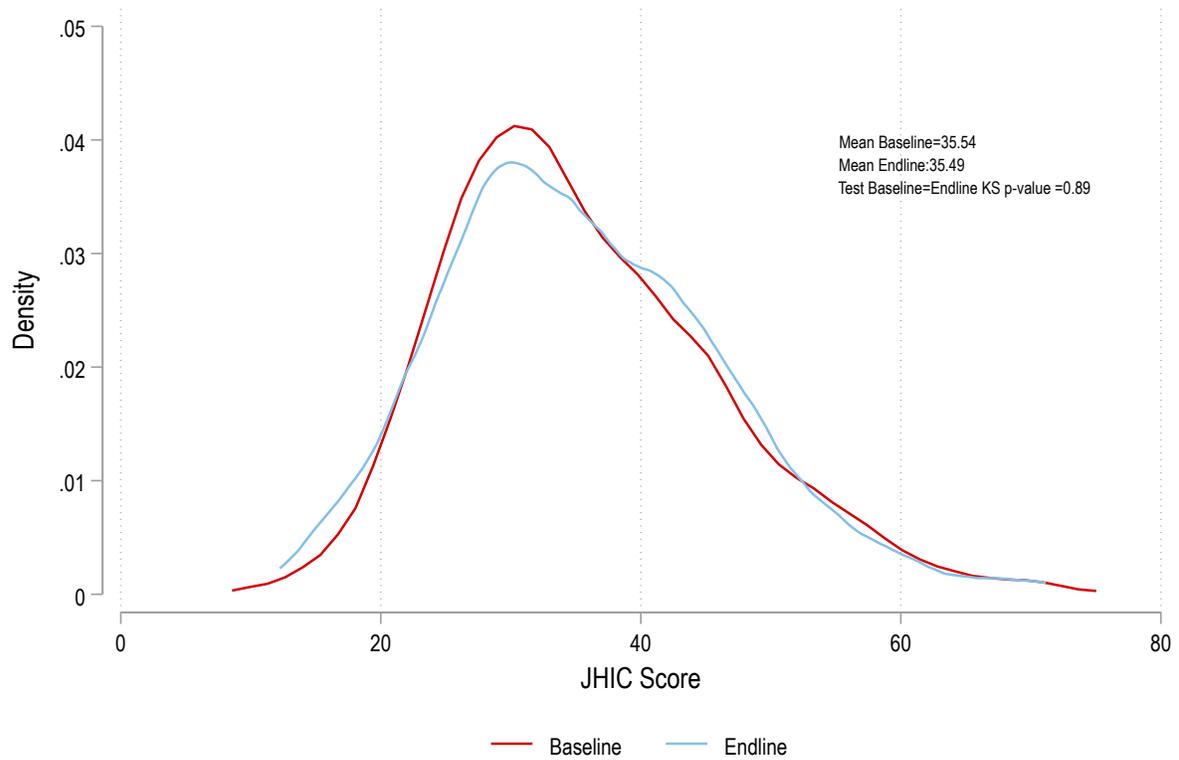
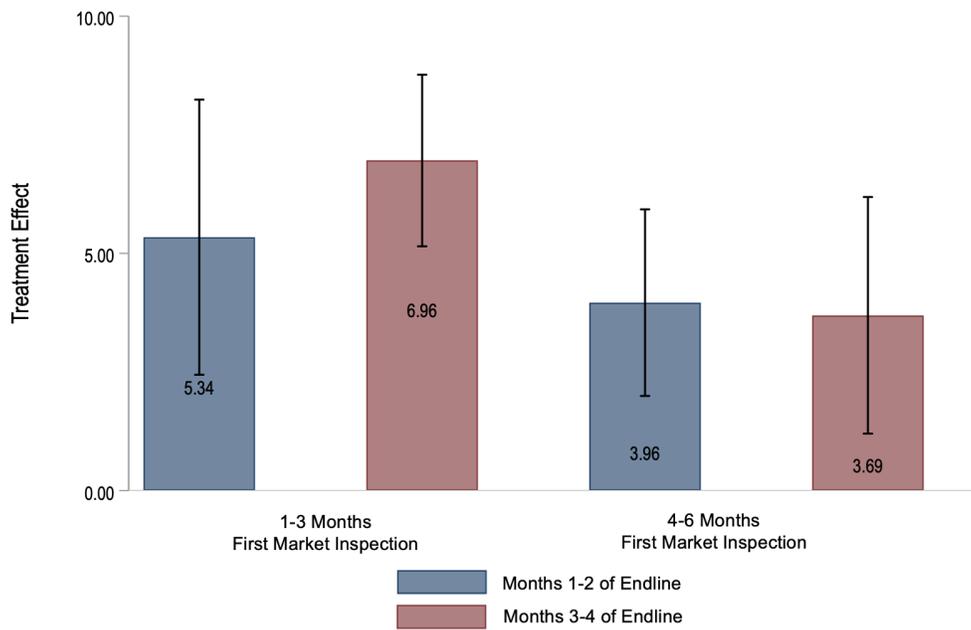


Figure S4: Impact on JHIC Score by Time of Market’s First Inspection
With Endline Time Interactions



Notes. Robust standard errors are clustered at the market level. Regressions include randomization strata controls (by county and market size) and health facility level controls. The start of the intervention was interrupted over the first two months. These facilities are included in Months 1-3. 18 facilities had their first market inspection after month 6 (including delays in the first two months of the intervention). These facilities are included in Months 4-6.

References

- Abadie, A., Chingos, M. M., & West, M. R. (2018). Endogenous stratification in randomized experiments. *Review of Economics and Statistics*, 100(4), 567–580.
- Bedoya, Das, J., Dolinger, A., De Guttry, R., Hur, Y. S., & Lee, J. Y. (2020). Regulation for Safety and Quality of Care : A Process Evaluation of the Health Inspection Pilots of the Kenya Patient Safety Impact Evaluation. *World Bank*.
- Bedoya, G., Dolinger, A., Rogo, K., Mwaura, N., Wafula, F., Coarasa, J., ... Das, J. (2017, July). Observations of infection prevention and control practices in primary health care, Kenya. *Bulletin of the World Health Organization*, 95(7), 503–516.
- Kenya Gazette. (2016). *Checklist for inspections of public and private providers by health regulatory bodies under the ministry of health* (Vols. Supplement No. 31, Legislative Supplement No. 25, Legal Notice No. 46 in the Public Health Act, Cap. 242).
- Miguel, E., & Kremer, M. (2004, January). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1), 159–217.
- Ministry of Health. (2015). *Ministry of Health Implementation Guidelines for the Joint Health Inspection Checklist*.
- MOH. (2014). *Kenya master health facility list*.