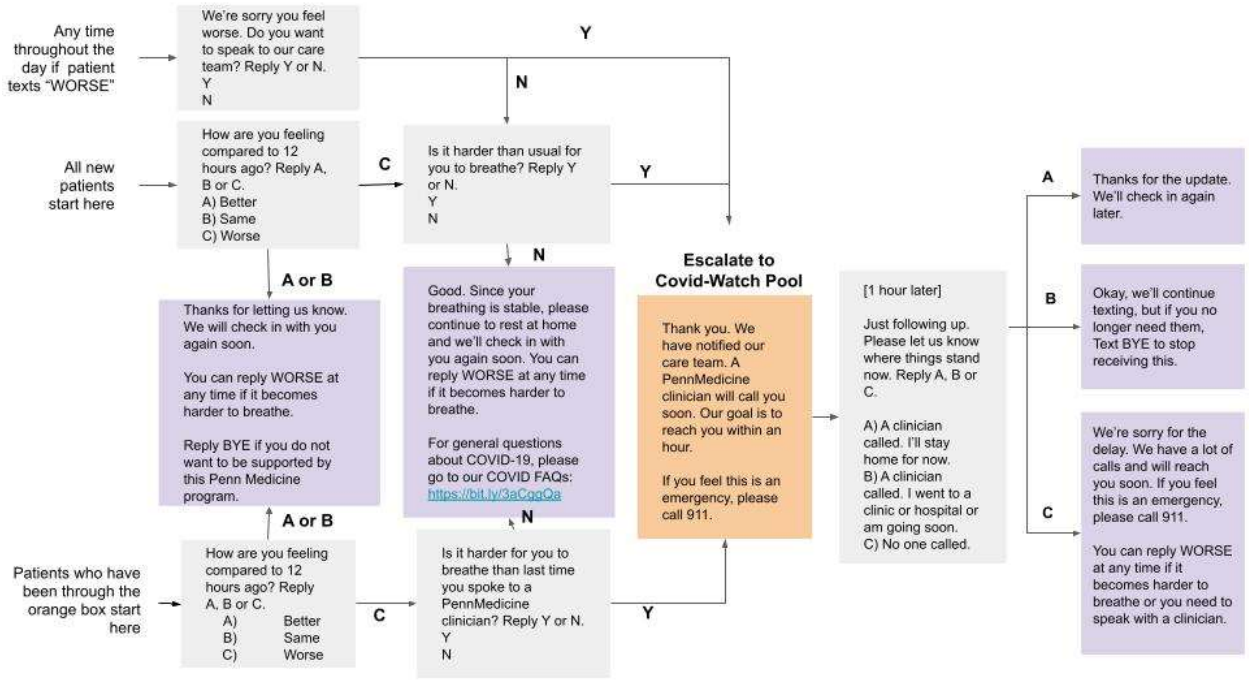
Supplementary Material*

Delgado MK, Morgan AU, Asch DA, et al. Comparative effectiveness of an automated text messaging service for monitoring COVID-19 at home. Ann Intern Med. 16 November 2021. [Epub ahead of print]. doi:10.7326/M21-2019

* This supplementary material was provided by the authors to give readers further details on their article. The material was reviewed but not copyedited.

Supplement Figure 1. COVID Watch Text Messaging and Clinical Escalation Algorithm



Supplement Figure 2. COVID Watch Patient Instructions

Additional Methodological Details

1. Selection of Outcomes and Follow up Period

Our IRB protocol submitted in July 2020 (available on the Annals of Internal Medicine website) pre-specifies the main outcomes reported in Table 3 and Table 4. After our Patient Centered Outcomes Research Institute (PCORI) grant was funded (August 18, 2020) we met with our patient and stakeholder advisory panel August 25, 2020 (members noted in Acknowledgements section). Stakeholders identified survival as the outcome most important to them and thus we selected 30-day mortality as our primary outcome. Stakeholder engagement in driving the research question is important to our team of investigators and is required of this grant.

Given the chance to revise and resubmit our work allowing additional time for ascertaining follow-up outcomes, at the request of the Editors and peer-reviewers, we added 60-day outcomes as a secondary analysis. Outcomes presented in this manuscript were pulled in July 2021, allowing seven months for out-of-hospital death notifications to make it into the electronic health record. National Death Index data were not yet available at the time of manuscript resubmission.

2. Calculation of Days Alive and Out of Hospital (DAOH)

Traditional calculation of DAOH factors inpatient hospital days and mortality over a given period of follow-up (see https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6347414/#R9 and https://jamanetwork.com/journals/jama/fullarticle/2775280). However, given the nature of our intervention to send patients to the ED when needed and avoid ED visits if not needed,

we sought to capture trips to the ED as an event in which the patient is not "out of hospital."

DAOH Algorithm:

- Start of the 30-day period (Day 0):
    - For patients tested in outpatient setting: date of COVID test collection.
    - For patients tested in the hospital (ED and inpatient discharges): the date of discharge.
- Day 1: the calendar date after the index event. Events counted between Day 1 to Day 30:
    - ED days
    - Inpatient days
    - Days after death
- Any subsequent ED or hospital contact counts for the entire day of that contact.
    - Do not double-count a day because it has an ED event and a hospital event.

3. Propensity Score Modeling and Outcomes Estimation

It is common in practice to specify a fully parametric model (such as logistic regression) to obtain predictions such as the propensity score. However, with a large number of covariates, machine learning approaches such as random forests and neural nets are often preferred because of their flexibility to better discover complex relationships including interactions, higher-order terms, and non-linear functions. The challenge, in practice, is to decide which machine learning approach to use. We therefore began our propensity score modeling using ensemble machine learning approach called SuperLearner which weights both parametric and nonparametric models to obtain optimal predictions using cross-validation.[1,2]

The advantage to ensemble machine learning approaches, such as SuperLearner, is that the analyst does not have to choose one over the others. SuperLearner combines multiple algorithms (learners) into a single algorithm and provides a prediction function. SuperLearner utilizes cross-validation to obtain an optimal prediction by minimizing the loss function (such as

the best cross-validated mean squared error). SuperLearner allows a wide range of parametric, semiparametric, and nonparametric algorithms/models/learners/wrappers including generalized linear models, BART, random forests, boosting, bagging, neural nets, ridge regression, LASSO, and support vector machines just to name a few. In our analysis, we specified logistic regression, BART, random forests, support vector machines, and caret models simultaneously in the model fitting process.

We found that logistic regression performed the best and received almost full weight compared to the other algorithms. Hence, we presented our results based on propensity score weights estimated from logistic regression (which are nearly identical to our results using SuperLearner). Propensity score distributions were evaluated for non-overlap to ensure that there were no serious positivity violations. Propensity scores and weights obtained with SuperLearner are presented in Supplement Figures 3 and 4.

In our final models, we only used inverse probability weights and did not further covariate adjust. Our reasoning for this approach was threefold. First, we used propensity score weighting to achieve balance between the treatment arms (which we achieved). Second, we did not further adjust for covariates because of the challenge of having a small number of outcomes (which we anticipated). We felt that any adjustment of covariates would lead to dangerous extrapolations of the model to covariate spaces that were not supported by the data. Third, we targeted a marginal estimand because of its relevance to policy making rather than a conditional treatment effect which would have resulted from adjusting for further covariates. As a sensitivity analysis, we did further adjust for key covariates that could impact

the outcome but found nearly identical results. For example, for difference in death rate at 30 days:

- Primary analysis with propensity score weighting: -0.19% (95% CI: -0.32, - 0.06%), p =0.005
- Primary analysis with propensity score weighting plus regression adjustment for age categories, sex, place of residence, month, income: -0.18% (95% CI: -0.31, - 0.05%), p =0.005

And for difference in death rate at 60 days:

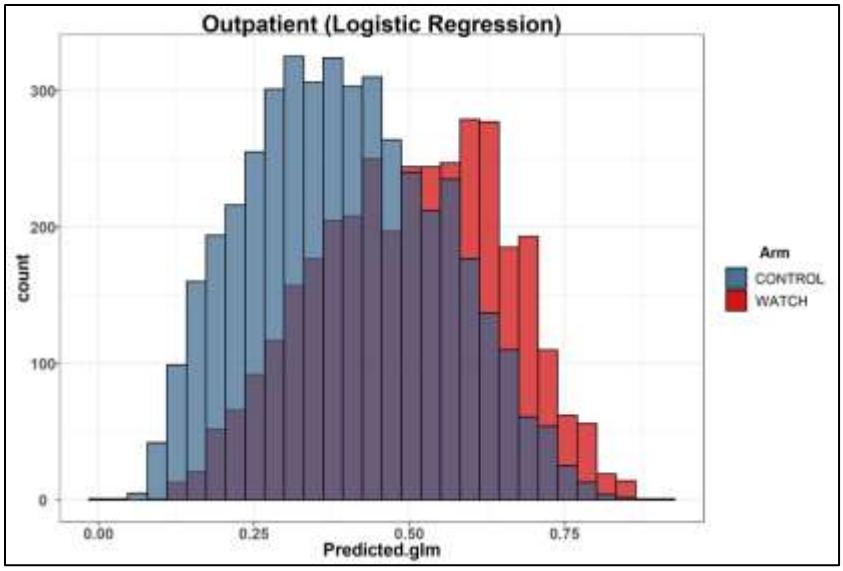
- Primary analysis with propensity score weighting: -0.25% (95% CI: -0.41, - 0.10%), p =0.002
- Primary analysis with propensity score weighting plus regression adjustment for age categories, sex, place of residence, month, income: -0.25% (95% CI: -0.40, - 0.06%), p =0.002

References

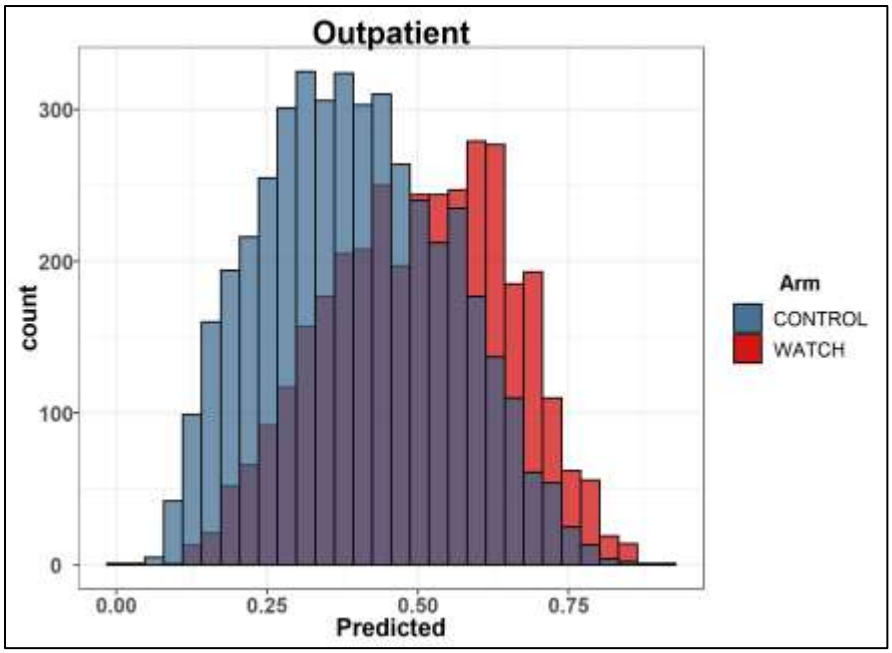
1. Van der Laan M, Polley E, Hubbard A. Super learner. Statistical applications in genetics and molecular biology. *Super learner Statistical applications in genetics and molecular biology.* 2007;6(1).
2. Van der Laan MJ, Rose S. *Targeted learning: causal inference for observational and experimental data.* Springer Science & Business Media; 2011.

Supplement Figure 3. Overlap of Propensity Scores for Enrollment in COVID Watch by Treatment Group

A) Overlap of propensity scores obtained with logistic regression (primary analysis)

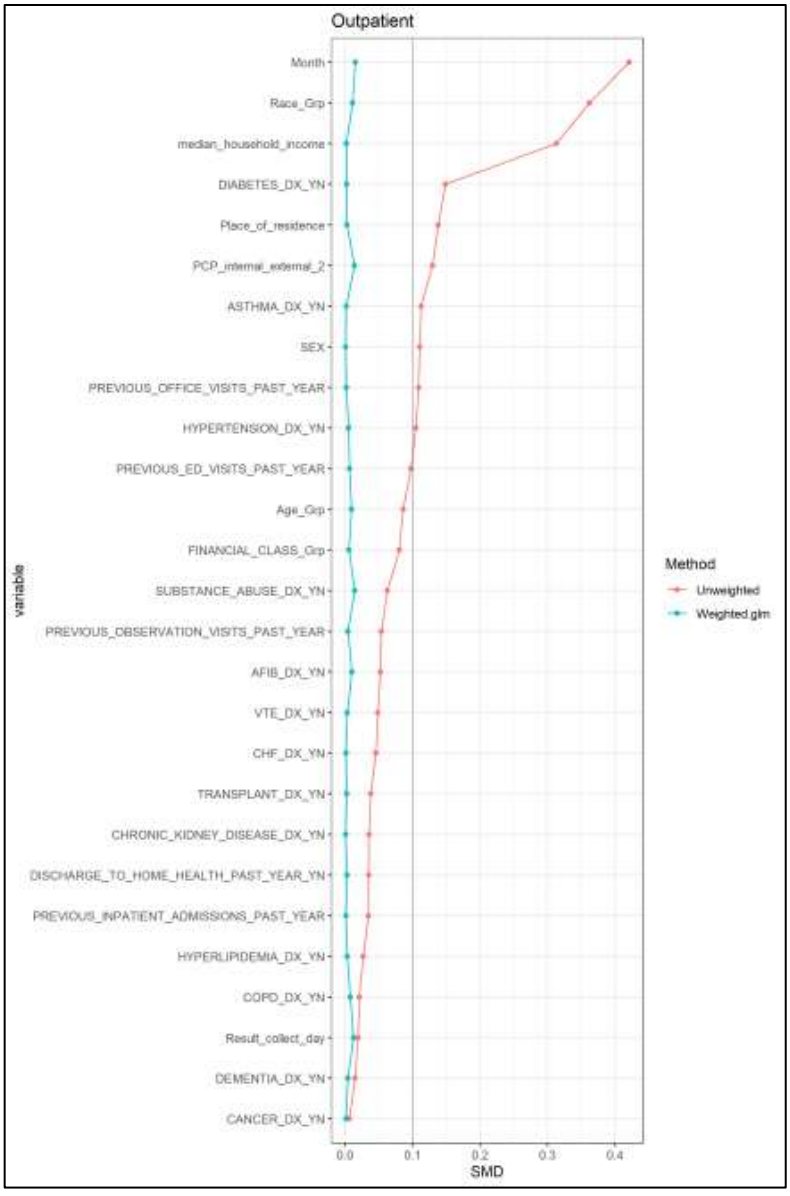


B) Overlap of propensity scores obtained with ensemble machine learning approach (SuperLearner) for comparison



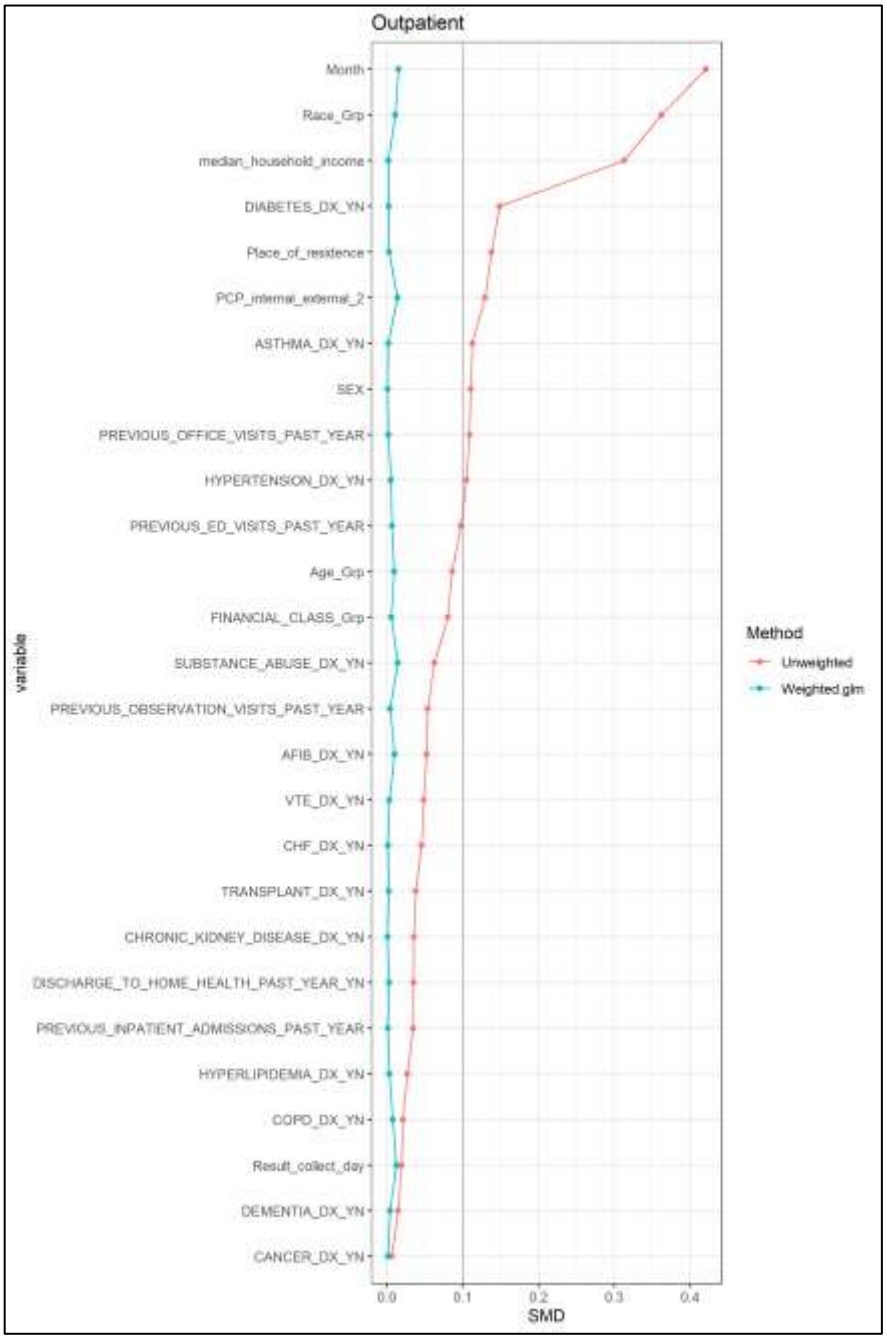
Supplement Figure 4. Change in Standard Mean Differences Before and After Propensity Score Weighting

A) Primary analysis using logistic regression



SMD: Standardized mean differences

B) Ensemble machine learning approach (SuperLearner) for comparison



SMD: Standardized mean differences

Supplement Table 1. COVID Watch Program Metrics

General Enrollment Process Metrics	Overall
Total Enrolled	3488
Non-responsive, n (%)	460 (13.2)
Responsive and engaged with program, n (%)	3028 (86.8)
Extended from 14 to 21 days of monitoring, n (%)	588 (19.4)
Total text responses, mean (SD)	23.18 (12.40)
Texts responses per day, mean (SD)	2.08 (0.90)
Days enrolled in COVID Watch, mean (SD)	11.79 (7.08)
Escalation Metrics	
Escalated to an RN, n (%)	434 (14.3)
Escalations to an RN/per patient, mean (SD)	0.20 (0.58)
Escalations (n)	595
Response time in minutes, mean (SD)	24.26 (40.48)
Clinical recommendations and response	
Continue to self-monitor	228 (38.3)
Telemedicine appointment within 24 hours	69 (11.6)
Contact PCP within next 24-48 hours	11 (1.8)
Urgent telemedicine appointment within 6 hours	157 (26.3)
Direct ED referral	90 (15.1)
No response to contact attempts	41 (6.9)

Supplement Table 2. Sensitivity Analysis for Unobserved Confounding

To assess the sensitivity of our 60-day mortality results to potential unmeasured confounding, we conducted a sensitivity analysis based on the Rosenbaum gamma approach using the R package 'rbounds' (2.1). Our table below demonstrates that there would need to be a substantial amount of unaccounted for confounding to reverse our statistically significant findings. Specifically, there would need to be an 1.8 times greater odds of differential assignment to Watch versus the control arm that was attributable to unobserved factors (upper bound >0.05). Given the large number of important covariates we have accounted for in our analysis, it is unlikely that such an impactful covariate was not included; hence, we feel that our results are robust to potential unmeasured confounding.

Gamma	Lower bound	Upper bound
1.0	0	0.0000
1.1	0	0.0000
1.2	0	0.0000
1.3	0	0.0001
1.4	0	0.0008
1.5	0	0.0032
1.6	0	0.0099
1.7	0	0.0253
1.8	0	0.0545
1.9	0	0.1019
2.0	0	0.1694
Gamma is odds of differential assignment to COVID WATCH due to unobserved factors		

Supplement Table 3. **Sensitivity Analysis for Deaths Within 30 Days Among those Excluded from Primary Analysis**

We also assessed the effect of re-including patients with DNR/DNI directives in place prior to COVID-19 testing, and patients with any prior evidence of acute rehab, long term care facility, or skilled nursing facility services even if these patients had no documentation of actively using these services at the time of Covid-19 testing given these were not exclusion criteria for COVID Watch enrollment. Re-including patients in the sample who were DNR/DNI (COVID Watch n=3, usual care =19) or who had any prior evidence of long term care in the past year (COVID Watch n=24, usual care n=64) only strengthened that primary finding that COVID Watch enrolled patients had significantly lower mortality rates compared with usual care (table below).

		Usual Care	Covid Watch	
	N	4450	3516	
All patients, but re-including DNR/DNI & patients who had previously been in long term care but were now living at home (N=7,987)	All deaths, unadjusted, No. (%)	23 (0.52)	6 (0.17)	0.018
	In-hospital deaths, No. (%)	15 (65.2)	5 (83.3)	0.633
	Out-of-hospital deaths, No. (%)	8 (34.8)	1 (16.7)	
	All deaths, adjusted difference relative to usual care (95% CI)	-3.8 deaths per 1,000 (-5.6, -2.0)		< 0.001
Patients with missing data (N=21)	All deaths, No. (%)	0	0	

DNR: Do Not Resuscitate; DNI: Do Not Intubate

Supplement Table 4. "Per Protocol" Sensitivity Analyses of Outcomes by COVID Watch Engagement

Unweighted Outcomes of Engaged COVID Watch patients vs. Usual Care				
	Overall	**Usual Care**	**Covid Watch**	**p**
n	7405	4377	3028	
Deaths at 30 days	14 (0.19)	12 (0.27)	2 (0.07)	0.055
Days Alive and Out of Hospital at 30 days, mean (SD)	29.65 (2.00)	29.71 (1.98)	29.56 (2.03)	0.001
Deaths at 60 days	19 (0.26)	16 (0.37)	3 (0.10)	0.033
Days Alive and Out of Hospital at 60 days, mean (SD)	59.50 (3.39)	59.54 (3.66)	59.45 (2.96)	0.288

Weighted Outcomes of Engaged COVID Watch patients vs. Usual Care				
	Difference between COVID Watch and usual care (95% CI)	**p**	**Odds Ratio (95% CI)**	**p**
Deaths at 30 days	-2.0 per 1,000 (3.3, -0.7)	0.002	0.25 (0.08, 0.62)	0.006
Days Alive and Out of Hospital at 30 days, mean (SD)	-0.16 (-0.25, -0.07)	<0.001		
Deaths at 60 days	-2.8 per 1,000 (-4.3, -1.3)	<0.001	0.25 (0.10, 0.55)	0.001
Days Alive and Out of Hospital at 60 days, mean (SD)	-0.09 (-0.25, 0.06)	0.249		

Unadjusted Outcomes of COVID Watch Patients by Engagement			
	Responsive	**Non-responsive**	**p**
n	3028	460	
Deaths at 30 days	2 (0.07)	1 (0.22)	0.346
Days Alive and Out of Hospital at 30 days, mean (SD)	29.6 (2.03)	29.2 (3.33)	0.002
ED visits within 30 days	368 (12.15)	52 (11.30)	0.657
Deaths at 60 days	3 (0.10)	2 (0.43)	0.132
Days Alive and Out of Hospital at 60 days, mean (SD)	59.5 (2.96)	58.9 (5.35)	0.001

Supplement Table 5. Baseline Characteristics of Patients Who Died Within 60 Days

	Overall	Usual Care	Covid Watch	p
n	21	16	5	
Age, mean years (standard deviation, SD)	67.90 (15.07)	68.19 (15.81)	67.00 (14.04)	0.882
Age group, years				
<40	1 (4.8)	1 (6.2)	0 (0.0)	1
40-49	0 (0)	0 (0)	0 (0)	NA
50-59	3 (14.3)	2 (12.5)	1 (20.0)	1
60-69	8 (38.1)	5 (31.2)	3 (60.0)	0.53
70 or older	9 (42.9)	8 (50.0)	1 (20.0)	0.506
Male, No. (%)	14 (66.7)	12 (75.0)	2 (40.0)	0.365
Race and ethnicity, No. (%)				0.296
Hispanic	4 (19.0)	3 (18.8)	1 (20.0)	
Non-Hispanic Black	6 (28.6)	3 (18.8)	3 (60.0)	
Non-Hispanic White	9 (42.9)	8 (50.0)	1 (20.0)	
Other/Unknown	2 (9.5)	2 (12.5)	0 (0.0)	
Insurance Coverage, No. (%)				0.779
Blue Cross, Blue Shield, Commercial, or Managed Care	9 (42.9)	7 (43.8)	2 (40.0)	
Medicaid or Managed Medicaid	2 (9.5)	1 (6.2)	1 (20.0)	
Medicare or Managed Medicare	9 (42.9)	7 (43.8)	2 (40.0)	
Other	1 (4.8)	1 (6.2)	0 (0.0)	
Source of Primary Care, No. (%)				0.457
Non-Penn Medicine primary care	6 (28.6)	4 (25.0)	2 (40.0)	
Penn Medicine primary care	13 (61.9)	11 (68.8)	2 (40.0)	
No PCP	2 (9.5)	1 (6.2)	1 (20.0)	
Body mass index, mean (SD)	32.3 (7.55)	32.8 (7.56)	29.7 (8.52)	0.538
Pre-existing conditions, No. (%)				
Hypertension	13 (61.9)	10 (62.5)	3 (60.0)	1
Diabetes	7 (33.3)	6 (37.5)	1 (20.0)	0.856
Hyperlipidemia	9 (42.9)	7 (43.8)	2 (40.0)	1
Chronic heart failure	2 (9.5)	2 (12.5)	0 (0.0)	1
Atrial fibrillation	5 (23.8)	5 (31.2)	0 (0.0)	0.406
Chronic obstructive pulmonary disease	8 (38.1)	7 (43.8)	1 (20.0)	0.669
Asthma	1 (4.8)	1 (6.2)	0 (0.0)	1
Chronic kidney disease	0 (0)	0 (0)	0 (0)	NA
Cancer	2 (9.5)	1 (6.2)	1 (20.0)	0.967
Venous thromboembolism	1 (4.8)	0 (0.0)	1 (20.0)	0.529

Received a transplant	7 (33.3)	6 (37.5)	1 (20.0)	0.856
Dementia	1 (4.8)	1 (6.2)	0 (0.0)	1
Substance use diagnosis	1 (4.8)	1 (6.2)	0 (0.0)	1
State of residence (%)				0.359
New Jersey	4 (9.0)	4 (25.0)	0 (0.0)	
Other	1 (4.8)	1 (6.2)	0 (0.0)	
Pennsylvania	16 (76.2)	11 (68.8)	5 (100.0)	
Median household income of home ZIP code, mean U.S. dollars $ (SD)	71575.9 (35463.1)	76918.9 (35186.2)	54478.2 (34143.7)	0.226
Hospital and office visits 1 year prior to COVID testing, mean (SD)				
ED visit	0.10 (0.44)	0.12 (0.50)	0.00 (0.00)	0.589
Admission - observation	0.05 (0.22)	0.06 (0.25)	0.00 (0.00)	0.589
Admission - inpatient	0.57 (1.54)	0.31 (0.48)	1.40 (3.13)	0.173
Office visit	5.14 (5.99)	5.31 (5.45)	4.60 (8.23)	0.823
Prior discharge to home health, no longer on home health, No. (%)	3 (14.3)	2 (12.5)	1 (20.0)	1
Days between COVID test and result, mean (SD)	1.14 (1.46)	1.19 (1.64)	1.00 (0.71)	0.809
Month of index positive test (%)				0.076
March	0 (0)	0 (0)	0 (0)	
April	5 (23.8)	2 (12.5)	3 (60.0)	
May	5 (23.8)	4 (25.0)	1 (20.0)	
June	0 (0)	0 (0)	0 (0)	
July	0 (0)	0 (0)	0 (0)	
August	1 (4.8)	1 (6.2)	0 (0.0)	
September	1 (4.8)	1 (6.2)	0 (0.0)	
October	1 (4.8)	0 (0.0)	1 (20.0)	
November	8 (38.1)	8 (50.0)	0 (0.0)	

Supplement Table 6. Coded Diagnoses of In-Hospital Deaths

Within Health System Hospital Deaths, (n=6), primary diagnoses	n
Acute kidney failure with tubular necrosis	1
COVID-19	2
Other specified sepsis	3
Outside Hospital Deaths (n=9), top 10 most common diagnoses*	n
Other viral pneumonia	6
COVID-19	5
Hypoxemia	4
Acidosis	2
Acute kidney failure, unspecified	2
Acute respiratory failure with hypoxia	2
Cough	2
Essential (primary) hypertension	2
Shortness of Breath	2
Weakness	2

*Health information exchange data provide all diagnosis codes associated with encounter, but in no particular order