

Online Appendices: Technology-Mediated Services in Health Care: Evaluating Telehealth Delivery Models for Underserved Populations

Appendix A. Coding Rules for Telehealth Models

Telehealth adoption measures are drawn from the HRSA Uniform Data System for reporting years 2016 to 2019. Telehealth items were first systematically collected in 2016. Because telehealth use was recorded in 2016 without precise adoption timing, CHCs reporting telehealth use in 2016 are excluded to avoid left censoring of treatment timing. The treatment window therefore begins in 2017.

The UDS reports telehealth practices by service category. CHCs indicate: (1) the service lines for which telehealth is used (e.g., primary care, mental health, provider-to-provider telehealth, etc.); and (2) the type of communication enabled via telehealth, including communication with (i) patients at remote locations and (ii) specialists outside the organization.

Although HRSA began collecting general telehealth information in 2016, detailed communication-type information used to classify delivery models is consistently available beginning in 2018. To construct model indicators for 2017, we implement two imputation strategies. The primary approach uses backward filling: if a CHC reports a specific delivery configuration in 2018 and reports mental telehealth use in 2017, we assign the same configuration to 2017. We implement alternative treatment coding as a robustness test in Online Appendix E. Results under alternative coding assumptions are substantively unchanged from our main results.

Our analysis distinguishes between two telehealth delivery models for mental health services: the specialist-facilitated model and the remote model. A CHC-year is coded as *SpecialistFacilitatedModel*_{*i,t*} = 1 if: (1) the CHC reports telehealth use for mental health services; (2) it reports communication via telehealth with “specialists outside the organization”; and (3) it does **not** report communication with “patients at remote locations”. This configuration captures arrangements in which the CHC coordinates access to external specialists, while patients interface with telehealth within the CHC’s physical setting. Similarly, a CHC-year is coded as *RemoteModel*_{*i,t*} = 1 if: (1) the CHC reports telehealth use for mental health services;

(2) it reports communication via telehealth with “patients at remote locations”; and (3) it does **not** report communication with “specialists outside the organization”. This configuration captures telehealth encounters delivered by CHC-affiliated clinicians directly to patients at remote locations. CHC-affiliated providers include primary care physicians, clinical social workers, counselors, marriage and family therapists, substance-use treatment providers, and other behavioral health professionals operating within the CHC’s clinical workflows, if the CHC has any such personnel. CHCs reporting no telehealth use for any service line are coded as non-adopters. CHC-year observations reporting both communication with patients at remote locations and communication with specialists outside the organization (2.9% of CHC-years) are excluded from the main analysis. In these cases, attribution to a single delivery configuration is not possible. Restricting the sample to single-model adopters ensures clean identification of model-specific effects.

Appendix B. Description of Variables and Correlation Matrix

Table A1. Description of Variables						
Variables	Description	N	Mean	Std. Dev	Min	Max
<i>Ln(MentalPatients)</i>	Log number of unique mental health patients in the CHC	6941	8.04	1.16	0	11.5
<i>Ln(MentalVisits)</i>	Log number of mental health visits in the CHC	6941	8.91	1.33	0	13.4
<i>ProtocolAdherence</i>	Percentage of patients aged 12 years and older screened for depression on the date of the visit using an age-appropriate standardized depression screening tool, <i>and</i> if positive, a follow-up plan is documented on the date of the positive screen	6050	48.2	33.3	0	100
<i>SpecialistFacilitatedTelehealth</i>	1 if CHC implemented specialist-facilitated telehealth for mental health, 0 otherwise	6941	0.04	0.20	0	1
<i>RemoteTelehealth</i>	1 if CHC implemented remote telehealth for mental health, 0 otherwise	6941	0.04	0.20	0	1
<i>SpecialistRatio</i>	Mental health specialists per 100 population in the CHC's service area	6850	0.17	0.16	0	2.01
<i>TopQuality</i>	1 if CHC is recognized as a top-quality performer in the previous year	6941	0.02	0.16	0	1
<i>PctMedicaid</i>	Percentage of patients at the CHC who are enrolled in Medicaid	6934	42.2	18.5	0	93.9
<i>PctHomeless</i>	Percentage of patients at the CHC who are homeless.	6934	4.05	8.76	0	100
<i>HasPediatric</i>	1 if the CHC provides pediatric services, 0 otherwise	6941	0.98	0.10	0	1
<i>HasDental</i>	1 if the CHC provides dental services, 0 otherwise	6941	0.82	0.38	0	1
<i>HasEHR</i>	1 if the CHC uses Electronic Health Records, 0 otherwise	6941	0.96	0.18	0	1
<i>TelehealthforNonMentalCare</i>	1 if CHC provided telehealth services for specialties other than mental health, 0 otherwise	6941	0.05	0.22	0	1
<i>PctRural</i>	Percentage of the CHC's service locations situated in rural areas	6918	0.44	0.49	0	1
<i>NumSites</i>	The number of satellite sites for the CHC	6928	7.37	8.47	1	135
<i>HasBehavGrant</i>	1 if CHC received federal funding targeted to behavioral health, 0 otherwise	6941	0.01	0.09	0	1
<i>PctHasDevice</i>	Percentage of households with access to at least one computing device (e.g., desktop, laptop, smartphone, or tablet) in the CHC's service area	6941	88.5	4.28	73.7	96.9

Table A2. Summary Statistics for Cost-Environment Proxies (Matched Samples)

Variable	Description	N	Mean	Std. Dev	Min	Max
Panel A. Specialist-Facilitated Telehealth Sample						
Specialist availability	Local supply of psychiatrists relative to population in the CHC's service area. Lower provider availability indicates higher search and verification costs for matching patients with appropriate specialists.	1167	0.02	0.02	0	0.18
Panel B. Remote Telehealth Sample						
Device access	The share of households with access to a computing device. Lower device access reflects greater setup costs for remote telehealth visits.	1292	0.87	0.03	0.79	0.93
Vehicle access	The share of households with access to a vehicle. Lower vehicle access implies higher transportation costs for in-person visits.	1292	0.90	0.10	0.23	0.98
Missing income share	The share of patients with missing income information in the UDS reports. Because income is a required field for reporting, higher missing-income rates indicate weaker documentation processes and greater workflow fragmentation in patient record management.	1292	0.30	0.22	0	0.93

Notes: Cost-environment proxies are measured using pre-adoption baseline values. Each panel reports the distribution of the proxies used in the corresponding heterogeneity specifications.

Appendix C. Propensity Score Matching

By comparing the outcomes of CHCs before and after adopting different telehealth delivery models while also contrasting treatment CHCs with control CHCs, we aim to isolate the impact of mental telehealth implementation on healthcare access and process quality. Nonetheless, inherent differences between CHCs that adopt a specific delivery model and those that do not adopt telehealth at all could introduce confounding factors in our difference-in-differences estimation, potentially leading to biased results. To address this, we apply Propensity Score Matching (PSM) to enhance comparability between treatment and control groups by constructing matched samples for each delivery model based on pre-treatment characteristics that are likely to influence telehealth adoption and the dependent variables.

Matching is conducted separately for each delivery model. For the specialist-facilitated model, the treatment group consists of CHCs that adopted this model between 2017 and 2019, while the control group includes CHCs that did not adopt telehealth. A similar approach is applied to the remote telehealth model.

We use two sets of variables in the matching process. The first set captures CHC capacity and technological readiness, including the number of satellite sites, grant funding, use of other health IT (e.g., patient portals, kiosks), and the total number of patients served in the past year. The second set includes socioeconomic factors, such as the percentage of Medicaid patients, median income, rural designation, the local mental health specialist to population ratio, internet access, educational attainment, and whether the CHC is located in a state with live video telehealth reimbursement policies.

Some variables, such as the mental health professional ratio, internet penetration rate, median income, and educational attainment, are derived at the service area level, representing the locality where most CHC patients reside¹. Following HRSA guidelines, we define service areas at the county level and calculate demographic variables using weighted averages across counties. The weights are based on the proportion of CHC locations in each county relative to the total number of CHC locations. Mathematically, this is expressed as: *Service Area Value* = $\sum_i \left(\frac{n_i}{N} \cdot x_i \right)$, where i represents the index of counties in the CHC's

¹ <https://bphc.hrsa.gov/sites/default/files/bphc/compliance/pin-2007-09.pdf>

service area, n_i represents the number of the CHC's locations in county i , N represents the total number of CHC locations, and x_i represents the value of the variable for county i .

One-to-one matching without replacement generates matched pairs for 89 treated and 89 control CHCs in the specialist-facilitated model and 109 treated and 109 control CHCs in the remote telehealth model. Online Appendix Tables A3 and A4 present the mean differences in covariates between the treated and control samples. Post-matching, the mean absolute standardized bias significantly decreases – from 17.7 to 4.8 for the specialist-facilitated model and from 22.2 to 4.4 for the remote telehealth model. Similarly, the median absolute standardized bias decreases from 15.6 to 3.7 and from 14.6 to 3.7 for the two models, respectively.

Across all matched samples, differences in matching covariates between treatment and control groups are not statistically significant, with p-values consistently exceeding 0.1. We additionally assess post-matching balance in pre-treatment local mental health need using the prevalence of any mental illness and serious mental illness. The absolute standardized mean differences are 0.087 and 0.016, respectively, for the specialist-facilitated sample and 0.017 and 0.005 for the remote sample, all below the conventional 0.10 threshold. The corresponding differences in means are also statistically insignificant.

Online Appendix Figure A1 presents propensity score distributions for treatment and control groups before and after matching. Post-matching, the distributions appear more similar, indicating that the matching procedure effectively balanced both the means and the overall distribution of propensity scores. These results confirm that our matching approach successfully improved the comparability of the treatment and control groups.

Figure A1. Distribution of Propensity Scores Before and After Matching

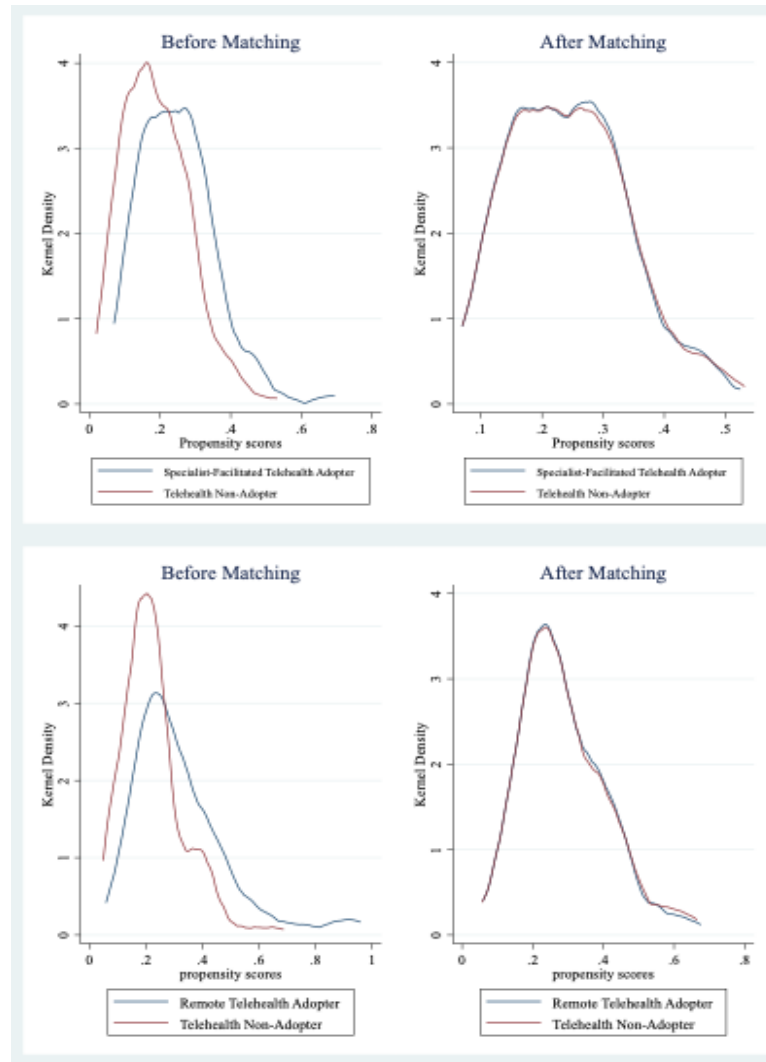


Table A3. Descriptive Statistics of Treated and Control Centers Before and After Matching for Specialist-Facilitated Telehealth Model

	Before Matching			After Matching		
	Treated (n=90)	Control (n=357)	p> t	Treated (n=89)	Control (n=89)	p> t
<i>NumSites</i>	5.944	5.389	0.391	5.676	5.583	0.909
<i>Ln(GrantAmount)</i>	14.832	14.887	0.579	14.806	14.808	0.993
<i>HasHIT</i>	0.9222	0.843	0.054	0.921	0.910	0.789
<i>Ln(TotalPatients)</i>	9.126	9.096	0.805	9.099	9.014	0.593
<i>Ln(MedianIncome)</i>	10.678	10.615	0.075	10.672	10.700	0.491
<i>PctMedicaid</i>	34.840	40.456	0.008	34.564	35.052	0.856
<i>PctRural</i>	0.556	0.412	0.014	0.561	0.606	0.546
<i>SpecialistRatio</i>	0.098	0.111	0.196	0.096	0.093	0.804
<i>PctHighSchool</i>	82.497	82.189	0.747	82.432	82.748	0.803
<i>PctWithInternet</i>	76.983	79.073	0.020	76.920	76.620	0.814
<i>Reimbursement</i>	0.978	0.952	0.287	0.978	0.966	0.652

Notes: HIT=Health Information Technology.

Table A4. Descriptive Statistics of Treated and Control Centers Before and After Matching for Remote Telehealth Model						
	Before Matching			After Matching		
	Treated (n=119)	Control (n=357)	p> t	Treated (n=109)	Control (n=109)	p> t
<i>NumSites</i>	9.3655	5.3894	0.000	7.195	7.618	0.641
<i>Ln(GrantAmount)</i>	15.287	14.887	0.000	15.137	15.170	0.767
<i>HasHIT</i>	0.9411	0.8431	0.006	0.935	0.889	0.232
<i>Ln(TotalPatients)</i>	9.623	9.096	0.000	9.466	9.359	0.464
<i>Ln(MedianIncome)</i>	10.672	10.615	0.061	10.653	10.675	0.519
<i>PctMedicaidPatients</i>	37.951	40.456	0.180	37.332	36.727	0.797
<i>PctRural</i>	0.4117	0.4117	1.000	0.422	0.403	0.784
<i>SpecialistRatio</i>	0.102	0.111	0.339	0.102	0.103	0.921
<i>PctHighSchool</i>	81.272	82.189	0.259	81.303	81.345	0.968
<i>PctWithInternet</i>	79.185	79.073	0.884	78.969	79.385	0.675
<i>Reimbursement</i>	0.9749	0.9523	0.292	0.972	0.972	1.000

Notes: HIT=Health Information Technology.

Appendix D. Additional Analyses Supporting Main Results

Table A5. Effects of Specialist-Facilitated Telehealth on Mental Health Access

	(1) <i>Ln(MentalPatients)</i>	(2) <i>Ln(MentalVisits)</i>
<i>SpecialistFacilitated Model</i>	0.116** (0.051)	0.182*** (0.066)
<i>SpecialistRatio</i>	0.034 (0.270)	-0.020 (0.373)
<i>TopQuality</i>	-0.041 (0.074)	-0.005 (0.072)
<i>PctMedicaid</i>	0.441 (0.513)	0.807 (0.651)
<i>PctHomeless</i>	1.406** (0.589)	1.468** (0.687)
<i>HasPediatric</i>	1.159 (1.060)	1.056 (1.150)
<i>HasDental</i>	0.156* (0.092)	0.192* (0.106)
<i>HasEHR</i>	-0.114 (0.098)	-0.106 (0.122)
<i>Telehealthfor NonMentalCare</i>	0.006 (0.052)	0.030 (0.066)
<i>PctRural</i>	0.151*** (0.044)	0.186*** (0.052)
<i>NumSites</i>	0.026*** (0.006)	0.031*** (0.007)
<i>HasBehavGrant</i>	0.170 (0.129)	0.157 (0.116)
<i>PctHasDevice</i>	-1.557 (2.218)	-3.569 (2.615)
CHC F.E.s	Yes	Yes
Year F.E.s	Yes	Yes
Observations	1189	1189
Adj. R2	0.880	0.855

Notes: Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

Table A6. Effects of Remote Telehealth on Mental Health Access and Protocol Adherence

	(1) <i>Ln(MentalPatients)</i>	(2) <i>Ln(MentalVisits)</i>	(3) <i>ProtocolAdherence</i>
<i>RemoteTelehealth</i>	-0.014 (0.043)	0.021 (0.055)	0.061** (0.025)
<i>SpecialistRatio</i>	-0.236 (0.163)	-0.164 (0.182)	0.126 (0.090)
<i>TopQuality</i>	0.053 (0.092)	0.019 (0.087)	-0.067 (0.047)
<i>PctMedicaid</i>	-0.062 (0.358)	-0.073 (0.407)	-0.036 (0.148)

<i>PctHomeless</i>	0.909 (0.857)	0.776 (0.839)	0.102 (0.362)
<i>HasPediatric</i>	0.246 (0.325)	0.508 (0.446)	-0.028 (0.098)
<i>HasDental</i>	0.323*** (0.122)	0.434*** (0.156)	0.072 (0.049)
<i>HasEHR</i>	0.010 (0.107)	0.070 (0.110)	0.042 (0.061)
<i>Telehealthfor NonMentalCare</i>	0.048 (0.045)	0.025 (0.052)	0.005 (0.032)
<i>PctRural</i>	0.020 (0.031)	0.060 (0.037)	-0.016 (0.029)
<i>NumSites</i>	0.029*** (0.007)	0.032*** (0.009)	-0.004 (0.003)
<i>HasBehavGrant</i>	0.285 (0.208)	0.445 (0.342)	-0.108 (0.086)
<i>PctHasDevice</i>	-0.108 (1.911)	-0.599 (2.255)	2.142* (1.249)
CHC F.E.s	Yes	Yes	Yes
Year F.E.s	Yes	Yes	Yes
Observations	1483	1483	1292
Adj. R2	0.897	0.874	0.644

Notes: Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

Impact of Specialist-Facilitated Telehealth Model in High and Low Mental Illness Prevalence Areas

To assess whether the observed increase in mental health visit volume reflects improved access rather than supply-driven utilization, we examine whether the effect of the specialist-facilitated model varies with local mental health need. If telehealth primarily activates previously unmet demand, its impact should be larger in areas with higher baseline prevalence of mental illness, where access constraints are more severe.

We measure local mental health need using data from the 2014-2016 National Survey on Drug Use and Health (NSDUH)², which reports the percentage of adults aged 18 years or older experiencing serious mental illness (SMI) within the past year at the substate level. SMI includes severe mental disorders—such as schizophrenia, bipolar disorder, and severe major depression—that cause substantial functional impairment and typically require specialized care. Because individuals with SMI often require specialist services, this measure provides a useful proxy for underlying demand for mental health treatment. As an

² <https://www.samhsa.gov/data/nsduh/2014-2016-substate-reports>

additional robustness check, we replicate the analysis using the percentage of adults experiencing a past-year major depressive episode.

We map the NSDUH substate-level prevalence rates to counties by assigning each county the prevalence estimate of its corresponding substate region. We then construct CHC service-area measures based on the counties in which each center operates. For CHCs spanning multiple counties, we compute a weighted average prevalence rate based on the number of service sites in each county. On average, 3.88% of adults in a CHC's service area experience SMI and 6.03% experience a major depressive episode. Using the median value of each measure, we classify CHCs into high- and low-need areas.

We then estimate heterogeneous treatment effects by interacting the specialist-facilitated telehealth adoption indicator with each baseline prevalence measure. Because the baseline analysis shows that only the specialist-facilitated model significantly affects access outcomes, we focus on this model.

As shown in Table A7, adoption of the specialist-facilitated model significantly increases mental health visit volume in high-need areas but not in low-need areas, under both prevalence measures. The interaction terms are positive and statistically significant. These results indicate that the utilization gains from telehealth adoption are concentrated in areas with greater underlying mental health need, consistent with the interpretation that the specialist-facilitated model improves access by activating previously unmet demand rather than inducing supply-driven utilization.

Table A7. Impact of Specialist-Facilitated Telehealth Model in High and Low Clinical Need Areas								
	Mental illness prevalence				Depressive episode			
	High	Low	High	Low	High	Low	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Ln(Mental Patients)</i>	<i>Ln(Mental Patients)</i>	<i>Ln(Mental Visits)</i>	<i>Ln(Mental Visits)</i>	<i>Ln(Mental Patients)</i>	<i>Ln(Mental Patients)</i>	<i>Ln(Mental Visits)</i>	<i>Ln(Mental Visits)</i>
<i>Specialist Facilitated Model</i>	0.236*** (0.078)	-0.033 (0.070)	0.340*** (0.099)	-0.026 (0.093)	0.234*** (0.090)	-0.002 (0.055)	0.340*** (0.117)	0.016 (0.076)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CHC F.E.s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	575	549	575	549	547	577	547	577

Notes: We leverage substate-level data from the National Survey on Drug Use and Health (NSDUH) to map the prevalence of serious mental illness and depressive episodes to each CHC's service area. Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless

patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.
*p<0.1, **p<0.05; ***p<0.01.

We conduct an analogous analysis for the remote telehealth model as a falsification test (Table A8). If the observed pattern merely reflected broader demand growth or time-varying shocks in high-need areas, we would expect similar moderation effects for the remote model. Re-estimating the specification on the remote telehealth sample yields no statistically significant interaction with either SMI prevalence or major depressive episode prevalence. This contrast reinforces the interpretation that the utilization gains are specific to the specialist-facilitated model rather than reflecting a general trend toward higher utilization in high-need areas.

Table A8. Impact of Remote Telehealth Model in High and Low Mental Illness Prevalence Areas				
	Full Sample		Matched Sample	
	High mental illness prevalence areas	Low mental illness prevalence areas	High mental illness prevalence areas	Low mental illness prevalence areas
	(1)	(2)	(3)	(4)
	<i>Ln(MentalVisits)</i>	<i>Ln(MentalVisits)</i>	<i>Ln(MentalVisits)</i>	<i>Ln(MentalVisits)</i>
<i>RemoteModel</i>	0.006 (0.079)	0.0140 (0.088)	0.078 (0.091)	-0.008 (0.102)
Controls	Yes	Yes	Yes	Yes
CHC F.E.s	Yes	Yes	Yes	Yes
Year F.E.s	Yes	Yes	Yes	Yes
Observations	1488	1510	682	695

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

Alternative DVs: Penetration and Reach

In the main text, we operationalize access using CHC-level visit volume, consistent with the realized access framework (Andersen 1995). While visit volume is widely used in safety-net research, it does not distinguish between different spatial margins of service utilization. To strengthen construct validity and assess whether our findings are robust to alternative operationalizations of access, we examine two complementary outcomes that capture distinct spatial dimensions of utilization.

Specifically, visit growth may arise from (i) deeper utilization within a CHC’s existing catchment area or (ii) geographic expansion into new markets. These have different implications. If the specialist-facilitated model primarily increases utilization among the existing safety-net population, we would observe increased penetration within established service areas. If instead the model leads the CHC to draw patients from new geographic areas, we may observe greater reach, as patients from previously unserved areas begin to utilize the CHC. Examining these two dimensions clarifies whether observed volume changes reflect deeper service use within the intended population or spatial expansion into new markets.

We construct both measures using CHC-level patient ZIP code data, which allow us to observe the geographic distribution of each CHC’s patient base in a given year. To measure **penetration**, we use patient residence ZIP codes to identify the ZCTAs in which a CHC’s patients live in a given year. At the CHC-ZCTA-year level, we compute the share of the local target population that received care from the CHC, defined as the number of CHC patients residing in the ZCTA divided by the ZCTA-level target population³. The measure is expressed as a percentage and capped at 100 percent to avoid implausible values. We then aggregate capped patient counts and corresponding target populations across all served ZCTAs to construct a CHC-year penetration measure used in the regression analysis. To measure **reach**, we count the number of distinct residential ZIP codes represented in a CHC’s patient population each year. This measure captures the geographic breadth of a CHC’s patient base and reflects whether utilization extends to additional areas rather than intensifying within existing ones.

For the penetration outcome, we estimate weighted least square models with CHC and year fixed effects, weighting observations by the CHC-year target population (total population below 200% FPL across served ZCTAs). Because penetration is a coverage rate with denominators that vary substantially across CHCs and over time, weighting by the target population addresses heteroskedasticity and gives greater weight to CHCs

³ We define the ZCTA-level target population as the residents with household income below 200% of the Federal Poverty Level (FPL), based on ACS estimates. Approximately 90% of health center patients have incomes at or below 200% FPL, and CHCs are required to provide income-based sliding fee discounts. Accordingly, the below-200% FPL population serves as a practical proxy for the CHC’s low-income service population, and the denominator reflects this group rather than the total ZCTA population.

serving larger low-income populations. For reach, we estimate linear models with CHC and year fixed effects. In all specifications, standard errors are clustered at the CHC level. We also interact *SpecialistFacilitatedModel* with local specialist supply measures (*psychologist_ratio* and *psychiatrist_ratio*) to examine whether the effect on penetration varies with the availability of mental health specialists.

As shown in Table A9, adoption of the specialist-facilitated model is associated with a significant increase in penetration (Columns 1, 3, and 5) while having no statistically significant effect on reach (Columns 2, 4, and 6). The negative interaction terms in Columns (3) and (5) indicate that the effect on penetration is larger in areas with lower specialist supply and diminishes as local supply increases. Taken together, these results indicate that the model primarily increases the intensity of service use within existing service areas rather than expanding into new geographic areas.

Table A9. Impact of Specialist-Facilitated Model on Spatial Dimensions of Access (Penetration and Reach)						
	Baseline		× Psychiatrist supply		× Psychologist supply	
	(1)	(2)	(3)	(4)	(5)	(6)
	<i>penetration</i>	<i>reach</i>	<i>penetration</i>	<i>reach</i>	<i>penetration</i>	<i>reach</i>
<i>SpecialistFacilitated Model</i>	0.395** (0.188)	-2.001 (3.041)	1.013*** (0.385)	-2.377 (3.142)	0.916*** (0.340)	-4.710 (3.895)
<i>SpecialistFacilitated Model × SpecialistSupply</i>			-0.281** (0.131)	0.229 (1.445)	-0.191** (0.090)	1.297 (0.890)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
CHC F.E.s	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.s	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1189	1189	1167	1167	1167	1167

Notes: Time-varying controls include a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects. *p<0.1, **p<0.05; ***p<0.01

Decomposition analysis

We conduct a decomposition analysis to examine whether the observed increase in total visits reflects expansion in the patient base or increased engagement with existing patients. Specifically, we examine the extent to which the effect of specialist-facilitated telehealth adoption on visit volume is mediated by changes in the number of patients served. Using a bootstrap procedure with 1,000 replications clustered at the CHC level, we estimate indirect and direct effects within the DID framework. The indirect effect through patient volume is 0.135 log points, while the direct effect is 0.047 log points, relative to a total

effect of 0.182 log points. These estimates imply that the majority of the increase in visits operates through expansion in the number of patients served.

Table A10. Direct and Indirect Effects of Specialist-Facilitated Telehealth Model on Mental Health Visits			
	Dependent Variable: $\ln(\text{MentalVisits})$		
	(1)	(2)	(3)
	Total Effect	Direct Effect (controlling for $\ln(\text{MentalPatients})$)	Indirect Effect (via $\ln(\text{MentalPatients})$)
<i>SpecialistFacilitated Model</i>	0.182** (0.063)	0.047** (0.022)	0.135** (0.056)
Controls	Yes	Yes	Yes
CHC F.E.s	Yes	Yes	Yes
Year F.E.s	Yes	Yes	Yes
Observations	1189	1189	1189

Notes: Bootstrapped with 1,000 samples. Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

The Effects of Remote Telehealth on Process Quality Measures

To further unpack the effect on protocol adherence, we examine the numerator and denominator components of the screening measure separately. The numerator captures the number of patients who (1) were screened for depression during their visit using an age-appropriate standardized tool and (2) had a follow-up visit documented if they screened positive. The denominator is the number of patients aged 12 years and older with at least one visit during the measurement period, which represents those eligible for routine depression screening.

Because both variables are non-negative counts and exhibit overdispersion, we estimate fixed-effects negative binomial models for each component. To account for within-CHC correlation and potential small-sample concerns, we compute bootstrap standard errors clustered at the CHC level with 500 replications.

As shown in Table A11, adoption of the remote telehealth model does not significantly impact the number of patients eligible for screening (i.e., denominator). In contrast, it has a positive and significant effect on the number of patients who were screened and had a follow-up visit documented after screening positive (i.e., numerator). These results indicate that the improvement in protocol adherence is driven by

enhanced screening and documentation processes rather than changes in the size of the eligible patient population. The absence of a significant effect on the denominator indicates that the number of patients with at least one visit during the measurement period does not change following remote telehealth adoption. This suggests that the pool of patients eligible for screening, and therefore the overall visit composition, remains largely stable. Instead, the effect operates through an increase in the number of patients who are screened and documented follow-up visits when screening positive. The pattern is consistent with the mechanism proposed in the main analysis: remote telehealth reduces provider-side tracking and documentation frictions, increasing compliance with standardized screening and follow-up documentation procedures.

Table A11. Decomposition of Protocol Adherence: Effects on Screening Eligibility and Screening with Follow-up Documentation		
	(1) <i>NumEligible</i>	(2) <i>NumScreenedFollowup</i>
<i>RemoteTelehealth</i>	0.023 (0.025)	0.144* (0.086)
<i>Constant</i>	-2.233 (2.780)	-4.722** (2.301)
Controls	Yes	Yes
CHC F.E.s	Yes	Yes
Year F.E.s	Yes	Yes
Observations	1294	1294

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05, ***p<0.01

Appendix E. Robustness Checks

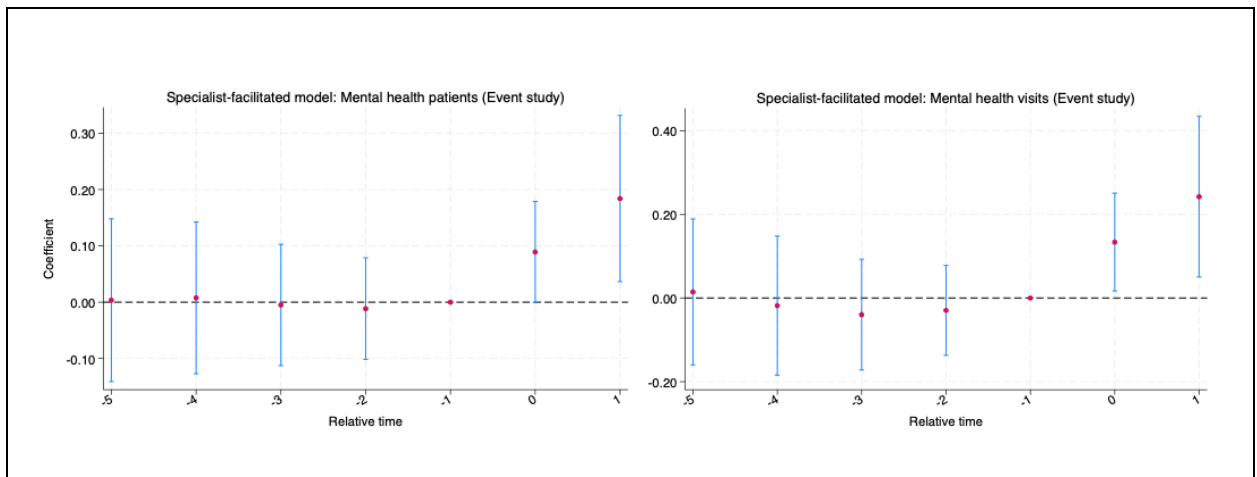
Event-Study Analysis of Pre-Trends

Our empirical strategy relies on the parallel trends assumption that absent telehealth adoption, outcomes at treated and untreated CHCs would have followed similar trajectories. To assess this assumption, we estimate an event-study specification that augments Equation (1) with a set of relative time indicators capturing years before and after telehealth adoption, for each telehealth delivery model:

$$y_{i,t} = \sum_{k \neq -1} \gamma_k \text{RelativeTime}_{i,k,t} + \text{Controls}_{i,t} + \text{CHCFE} + \text{YearFE} + \varepsilon_{i,t}$$

Here, $\text{RelativeTime}_{i,k,t}$ equals 1 if CHC i is in relative year k with respect to its adoption year and 0 otherwise. The year immediately prior to adoption ($k=-1$) is omitted as the reference period. The coefficients γ_k capture the dynamic path of outcomes relative to this pre-adoption baseline.

Under the parallel trends assumption, the pre-treatment coefficients ($k < 0$) should be jointly indistinguishable from zero. We formally test this restriction using joint Wald tests of the lead coefficients. Across all specifications, we fail to reject the null that the pre-treatment coefficients are jointly equal to zero. The corresponding F-statistics range from 0.04 to 1.31 (all p-values > 0.10). Event study plots (Figure A2) similarly show no systematic pre-trend. These results support the parallel trends assumption underlying our difference-in-differences design for both telehealth delivery models and all outcomes.



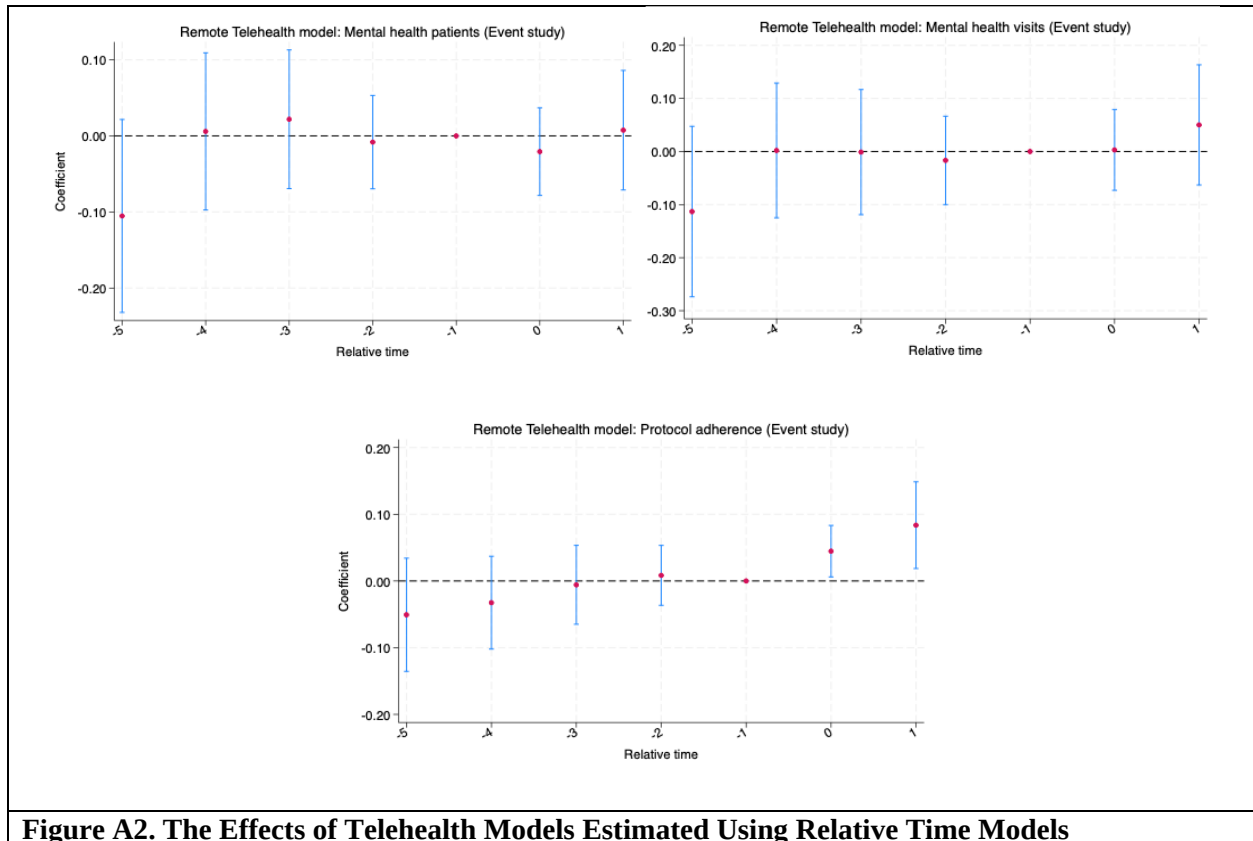


Figure A2. The Effects of Telehealth Models Estimated Using Relative Time Models

Note: Vertical bars indicate 90% confidence intervals.

Stacked Difference-in-Differences

To address potential biases arising from staggered treatment timing in two-way fixed effects (TWFE) DID models, we implement a stacked DID approach following Cengiz et al. (2019) and Miller et al. (2021). In settings with staggered adoption, conventional TWFE estimators may implicitly compare newly treated units to already-treated units, potentially introducing bias when treatment effects are heterogeneous over time. The stacked design mitigates this concern by restricting identification to comparisons between treated cohorts and never-treated cohorts.

Specifically, we construct separate cohort-specific datasets, in which each adoption cohort (defined by the year a CHC first implements mental telehealth) is analyzed relative to a set of “clean controls”, consisting of CHCs that do not adopt telehealth during the study period. For each cohort-specific dataset, we estimate a DID specification using only that cohort and the clean controls. These cohort-level datasets

are then stacked into a single pooled dataset, and we estimate the treatment effect with CHC and year fixed effects.

This structure ensures that already-treated units are never used as controls for later-treated cohorts, thereby eliminating contamination from dynamic treatment effects. Standard errors are clustered at the CHC level.

We implement this procedure separately for the specialist-facilitated and remote telehealth models, consistent with the main analysis. Table A12 and A13 report the stacked DID estimates of the average treatment effect. The results are quantitatively similar to the baseline TWFE estimates, indicating that our findings are not driven by biases associated with staggered adoption timing.

Table A12. Stacked DID Estimates of Specialist-Facilitated Telehealth		
	(1)	(2)
	<i>Ln(MentalPatients)</i>	<i>Ln(MentalVisits)</i>
<i>SpecialistFacilitatedModel</i>	0.125** (0.052)	0.199*** (0.068)
Controls	Yes	Yes
CHC F.E.s	Yes	Yes
Year F.E.s	Yes	Yes

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

Table A13. Stacked DID Estimates of Remote Telehealth			
	(1)	(2)	(3)
	<i>Ln(MentalPatients)</i>	<i>Ln(MentalVisits)</i>	<i>ProtocolAdherence</i>
<i>RemoteTelehealth</i>	-0.008 (0.046)	0.033 (0.060)	0.067*** (0.026)
Controls	Yes	Yes	Yes
CHC F.E.s	Yes	Yes	Yes
Year F.E.s	Yes	Yes	Yes

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

Instrumental Variable and Control-Function Specification

We conduct additional analyses to assess whether our findings are sensitive to selection on unobservables in telehealth adoption. Although our baseline DID with propensity-score matched samples includes CHC and year fixed effects, unobserved factors may still influence both the timing of adoption and outcome trajectories. To address this concern, we implement a two-step control function approach analogous to a Heckman-style selection correction.

In the first stage, we model each CHC's adoption decision using a probit specification in which adoption of the telehealth model is instrumented by a lagged, leave-one-out peer adoption at the Hospital Referral Region (HRR) level. For each CHC in year t , the instrument is defined as the share of other CHCs within the same HRR that had adopted telehealth by year $t-1$, excluding the focal CHC itself. This leave-one-out structure eliminates mechanical correlation between the instrument and the focal CHC's adoption status. The HRR level reflects the relevant referral market, capturing local information spillovers, learning effects, and institutional diffusion processes that plausibly influence adoption decisions.

In the first stage, we estimate a probit model of telehealth adoption as a function of the lagged HRR peer adoption measure. The instrument strongly predicts adoption in both models (Table A14). Linear first-stage diagnostics yield F-statistic of 276.63 (specialist) and 374.15 (remote), well above the conventional weak-instrument thresholds. The Kleibergen-Paap rk LM statistics (97.12 and 101.93, respectively) and the corresponding Wald F-statistics similarly indicate strong identification.

Table A14. First-Stage Probit Estimates of Telehealth Model Adoption		
	(1)	(2)
	<i>SpecialistModelAdoption</i>	<i>RemoteModelAdoption</i>
<i>HRRPeerAdoption_{t-1}</i>	4.773*** (0.504)	4.594*** (0.370)
Controls	Yes	Yes
Year F.E.s	Yes	Yes
Observations	502	631
Pseudo. R2	0.661	0.625

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

* $p < 0.1$, ** $p < 0.05$; *** $p < 0.01$

Using the probit estimates, we compute the inverse Mills Ratio (IMR) and include it in the second-stage outcome regressions. The second stage mirrors our baseline DID framework, with CHC and year fixed effects and clustered standard errors at the CHC level. The inclusion of the IMR controls for potential correlation between unobserved determinants of telehealth adoption and unobserved determinants of outcomes.

Second-stage results are reported in Tables A15 and A16. For the specialist-facilitated telehealth model, the estimated effects on mental health utilization remain positive and statistically significant after correcting selection. For the remote telehealth model, estimated effects on utilization are small and not significant. For protocol adherence, remote telehealth adoption is positively associated, while the IMR term remains statistically insignificant.

To probe the exclusion restriction, we conduct additional checks. First, we verify that lagged HRR peer adoption dates do not predict pre-adoption outcome trends. Second, we include the peer adoption measure directly in the second stage outcome regressions; results are substantively unchanged.

Taken together, these results strengthen the causal interpretation of our findings. First, they indicate that the findings are not driven by endogenous adoption timing, as the instrument strongly predicts telehealth adoption and correcting selection does not overturn the main estimates. Second, they address concerns about time-varying omitted variables at the CHC level, such as leadership changes, staffing expansions or local demand shocks that could jointly influence adoption and outcomes. The control function correction allows adoption to be correlated with unobserved outcome determinants. Importantly, the persistence of the main effects suggests that the baseline estimates are not driven by these confounders. Third, the results help distinguish organizational capability bias from treatment effects. By accounting for latent adoption propensity, the control function approach separates selection on unobserved factors from the impact of telehealth adoption itself. Overall, the instrument strongly predicts telehealth adoption in both models, and correcting selection on unobservables does not alter the qualitative pattern of results.

Table A15. Control Function Estimates: Specialist-Facilitated Telehealth Model

(1)	(2)
<i>Ln(MentalPatients)</i>	<i>Ln(MentalVisits)</i>

<i>SpecialistFacilitated</i>	0.154**	0.243***
<i>Model</i>	(0.065)	(0.084)
<i>InverseMillsRatio</i>	-0.079**	-0.120**
	(0.039)	(0.053)
Controls	Yes	Yes
CHC F.E.s	Yes	Yes
Year F.E.s	Yes	Yes
Observations	1169	1169
Adj. R2	0.890	0.865

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

Table A16. Control Function Estimates: Remote Telehealth Model			
	(1)	(2)	(3)
	<i>Ln(MentalPatients)</i>	<i>Ln(MentalVisits)</i>	<i>ProtocolAdherence</i>
<i>RemoteModel</i>	-0.073	-0.033	0.069*
	(0.062)	(0.080)	(0.039)
<i>InverseMillsRatio</i>	0.069	0.066	-0.008
	(0.042)	(0.054)	(0.030)
Controls	Yes	Yes	Yes
CHC F.E.s	Yes	Yes	Yes
Year F.E.s	Yes	Yes	Yes
Observations	1460	1460	1269
Adj. R2	0.899	0.877	0.641

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

Placebo Tests with Randomized Adoption Timing

To assess whether our estimates could be driven by endogenous adoption timing, we implement a placebo timing permutation test. We preserve the empirical structure of adoption intensity by holding fixed the annual number of telehealth adopters and randomly reassigning adoption years across treated CHCs. Specifically, for each calendar year, we maintain the observed number of adopters but randomly reallocate those adoption years among CHCs that adopt during the sample period. This procedure preserves aggregate adoption pressure by year while severing the link between actual adoption timing and CHC-specific characteristics.

For each placebo draw, we construct a placebo treatment indicator equal to 1 for CHC-year observations occurring on or after the assigned placebo adoption year and 0 otherwise. We then re-estimate the main

specification, replacing the actual treatment indicator with the placebo indicator and using the same set of controls, fixed effects, and clustered standard errors as in the baseline model. We repeat this reassignment procedure 500 times and collect the estimated placebo treatment coefficients to form the null distribution.

Figure A3 plots the distribution of placebo coefficients for each outcome, with the vertical line indicating the estimate obtained under the true adoption timing. In all cases, the estimate based on the actual adoption timing lies in the extreme upper tail of the placebo distribution.

These results indicate that the observed effects are unlikely to be generated by arbitrary reassignment of adoption years or from unobserved shocks that merely coincide with adoption timing. This finding substantially reduces concerns that timing-aligned unobservables alone drive the main findings.

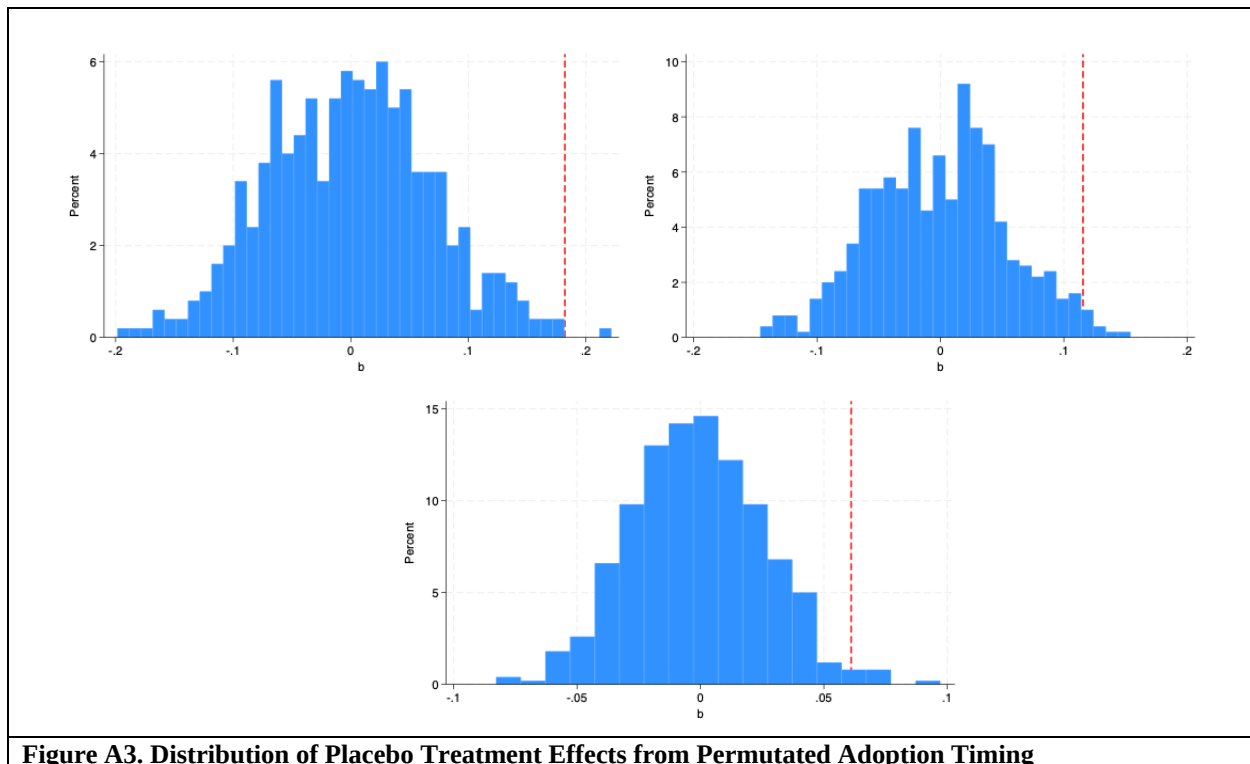


Figure A3. Distribution of Placebo Treatment Effects from Permuted Adoption Timing

Notes: Each panel shows the distribution of placebo treatment effects from 500 random reassignments of adoption timing that preserve the annual number of adopters. The top panel reports results from the specialist-facilitated model for number of mental health patients and visits, and the bottom panel reports results from the remote telehealth model for protocol adherence.

Decomposition of Treatment Effect Variation (Goodman-Bacon)

The Bacon decomposition, introduced by Goodman-Bacon (2021), breaks down the two-way fixed effects (TWFE) DID estimator into a weighted average of 2×2 group comparisons, thereby clarifying the sources

of identifying variation in treatment effects. Each pairwise comparison is assigned a weight that reflects its contribution to the overall estimate. As shown in Goodman-Bacon (2021), bias may arise in staggered adoption settings when treatment effects are heterogeneous over time, particularly if already-treated units serve as controls for later-treated units.

We implement the Bacon decomposition to assess the distribution of weights across comparison types. Because the decomposition assumes a balanced panel, we exclude unbalanced units from this analysis. Our findings indicate that potential heterogeneity in treatment effects over time does not significantly distort the DID estimates. Specifically, as reported in Tables A17 and A18, approximately 70-77% of the identifying weight comes from comparisons between treated and never-treated CHCs, 19-20% from comparisons between early- and later treated CHCs, and 4-9% from within-treated comparisons.

Importantly, comparisons in which treated units serve as controls receive relatively small weights. Although some of these within-treated comparisons yield negative estimates, their limited weight implies that they contribute minimally to the aggregate TWFE coefficient. Moreover, the consistency in sign and direction across both the treated-never-treated and timing-group comparisons suggests that the estimated effects are not driven by any single comparison type, reinforcing the stability of the TWFE results.

Table A17. Goodman-Bacon Decomposition of Specialist-Facilitated Model Adoption			
	Weight	<i>Ln(MentalPatients)</i>	<i>Ln(MentalVisits)</i>
<i>SpecialistFacilitatedModel</i>		0.130** (0.059)	0.203*** (0.077)
Timing-group	0.192	0.131	0.161
Treated vs. Never-treated	0.770	0.163	0.259
Within-treated	0.038	-0.533	-0.681

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

Table A18. Goodman-Bacon Decomposition of Remote Telehealth Model Adoption				
	Weight	<i>Ln(MentalPatients)</i>	<i>Ln(MentalVisits)</i>	<i>ProtocolAdherence</i>
<i>RemoteModel</i>		-0.013 (0.048)	0.022 (0.061)	0.068** (0.027)

Timing-group	0.206	0.008	0.059	0.072
Treated vs. Never-treated	0.705	0.061	0.093	0.065
Within-treated	0.089	-0.660	-0.638	0.087

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

Robustness to Alternative Outcome Distributions (Count Data Models)

We re-estimate our main access specifications using count data models that do not rely on log-transforming outcomes. Specifically, we estimate Negative Binomial models and Poisson pseudo maximum likelihood (PPML) models, both of which naturally accommodate zero counts and relax log-linear functional form assumptions.

Table A19 reports the results. For the specialist-facilitated model, the estimated coefficients remain positive and statistically significant across both PPML and Negative Binomial specifications for mental health patients and visits. In contrast, the effects of the remote telehealth model remain statistically insignificant, consistent with the log-linear DID estimates.

The stability of sign and statistical significance across log-linear, Negative Binomial, and PPML models indicates that our conclusions are not driven by functional-form assumptions or the treatment of zero counts.

Table A19. Robustness Check Using Count Data Models								
	Specialist-facilitated model				Remote model			
	NB		PPML		NB		PPML	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Mental Patients</i>	<i>Mental Visits</i>	<i>Mental Patients</i>	<i>Mental Visits</i>	<i>Mental Patients</i>	<i>Mental Visits</i>	<i>Mental Patients</i>	<i>Mental Visits</i>
Telehealth Adoption	0.092*** (0.037)	0.148*** (0.048)	0.071** (0.036)	0.115* (0.064)	-0.007 (0.036)	0.007 (0.042)	-0.007 (0.036)	0.036 (0.070)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CHC F.E.s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1189	1189	1189	1189	1483	1483	1483	1483

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors are bootstrapped

with 500 replications for the Negative Binomial (NB) models and clustered at the CHC level for the Poisson Pseudo-Maximum Likelihood (PPML) models. F.E.s = Fixed Effects.

*p<0.1, **p<0.05; ***p<0.01

Robustness to Organizational Scale Adjustment

To assess whether our results reflect changes in mental health utilization relative to overall CHC scale, we estimate alternative specifications that account for total patient volume. While our main analysis uses level outcomes to capture realized access, an alternative approach is to normalize utilization by CHC size. We incorporate organizational scale by including the natural log of total patients (and alternatively, the natural log of total visits) as covariates in the regression.

Table A20 reports the results. The estimated effects of the specialist-facilitated model on mental health patients and visits remain positive and statistically significant. These findings indicate that the observed effects persist conditional on CHC scale and therefore reflect a relative expansion in mental health service delivery, rather than differences in overall organizational size.

Table A20. Effect of Specialist-Facilitated Telehealth Controlling for Clinic Scale		
	(1)	(2)
	<i>Ln(MentalPatients)</i>	<i>Ln(MentalVisits)</i>
<i>SpecialistFacilitated</i>	0.072*	0.100**
<i>Model</i>	(0.037)	(0.048)
<i>Ln(NumTotalPatients)</i>	1.051***	
	(0.108)	
<i>Ln(NumTotalVisits)</i>		1.327***
		(0.104)
CHC F.E.s	Yes	Yes
Year F.E.s	Yes	Yes
Observations	1189	1189

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors in parentheses are clustered at the center level. F.E.s = Fixed Effects.

* p<0.1, ** p<0.05, *** p<0.01

Robustness to Outliers (Winsorized Outcomes)

In CHC-level panel data, a small number of unusually large or rapidly expanding centers could exert disproportionate leverage in TWFE regressions. To assess influence sensitivity, we winsorize each dependent variable at the 1st and 99th percentiles and, as a more conservative test, at the 5th and 95th percentiles. Across both specifications, the estimated treatment effects remain quantitatively similar and

statistically significant (Tables A21 and A22). This stability indicates that the main findings are not driven by a small number of extreme CHC-year observations.

Table A21. Specialist-Facilitated Telehealth Model: Robustness to Winsorization of Outcomes

	Baseline		Winsor 1-99		Winsor 5-95	
	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Ln(Mental Patients)</i>	<i>Ln(Mental Visits)</i>	<i>Ln(Mental Patients)</i>	<i>Ln(Mental Visits)</i>	<i>Ln(Mental Patients)</i>	<i>Ln(Mental Visits)</i>
<i>SpecialistFacilitated Telehealth</i>	0.116** (0.051)	0.182*** (0.066)	0.113*** (0.042)	0.180*** (0.056)	0.082** (0.036)	0.134*** (0.048)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
CHC F.E.s	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.s	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1189	1189	1189	1189	1189	1189

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior-year top-quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors are clustered at the CHC level. F.E.s = fixed effects.

*p<0.1, **p<0.05; ***p<0.01

Table A22. Remote Telehealth Model: Robustness to Winsorization of Outcomes

	Baseline	Winsor 1-99	Winsor 5-95
	(1)	(2)	(3)
	<i>ProtocolAdherence</i>	<i>ProtocolAdherence</i>	<i>ProtocolAdherence</i>
<i>RemoteTelehealth</i>	0.061** (0.025)	0.061** (0.024)	0.061** (0.025)
Controls	Yes	Yes	Yes
CHC F.E.s	Yes	Yes	Yes
Year F.E.s	Yes	Yes	Yes
Observations	1462	1462	1274

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior-year top-quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors are clustered at the CHC level. F.E.s = fixed effects.

*p<0.1, **p<0.05; ***p<0.01

Robustness to Alternative Treatment Coding Assumptions

Because delivery model information is first observed in 2018, the 2017 model-type indicators are partially unobserved. In the baseline specification, we implement a backfilling rule based on observed telehealth use. To assess whether our findings depend on this coding choice, we conduct a series of conservative sensitivity analyses.

Specifically, we reassign all ambiguous 2017 delivery model indicators to zero and, alternatively, to one. By ambiguous, we refer to CHC-year observations in which the CHC reported using mental telehealth

in 2017, but the delivery model is unknown. CHCs that reported not using mental telehealth in 2017 are not considered ambiguous. These extreme specifications provide upper and lower bounds on the estimated treatment effects under maximally conservative coding assumptions.

For the specialist-facilitated model, the baseline estimates of 0.116 (s.e. = 0.051) for mental health patients and 0.182 (s.e. = 0.066) for visits remain stable under both extreme assumptions. When all ambiguous observations are set to zero, the corresponding estimates are 0.120 and 0.193; when set to one, they are 0.109 and 0.180. Magnitudes vary by less than 0.02 log points, and statistical significance is unchanged. For the remote telehealth model, the baseline estimate of 0.061 (s.e. = 0.025) for protocol adherence is virtually identical under both extreme assumptions, yielding estimates of 0.060 and 0.062. The point estimates differ by at most 0.001.

Ambiguous 2017 observations account for only 17 CHC-years in the specialist-facilitated telehealth sample and 26 CHC-years in the remote telehealth sample. The stability of the estimates across extreme coding assumptions indicates that the main findings are not driven by first-year model classification.

Validation of the Tracking-Cost Proxy

Our tracking cost proxy, measured using the pre-adoption share of patients with missing income information, captures organizational documentation and workflow frictions. We assess whether this proxy reflects preexisting differences in organizational quality rather than documentation frictions specifically. We focus this validation on the tracking cost proxy because unlike the other three proxies which are drawn from independent administrative and demographic sources (NPI Registry, American Community Survey), it is constructed from CHC-reported documentation data and is therefore more directly susceptible to measurement concerns.

First, we compare baseline protocol adherence across high- and low-tracking cost environments prior to remote telehealth adoption. Using the pre-adoption observations from our main estimation sample, we regress protocol adherence on *HighTrackingCost*, controlling for the same covariates and year fixed effects as in Table 10, with standard errors clustered at the CHC level. As shown in Table A23, the coefficient on

HighTrackingCost is small and statistically insignificant, indicating no significant difference in baseline protocol adherence between high- and low-tracking cost CHCs prior to adoption.

Table A23. Baseline Protocol Adherence by Tracking Cost Environment

	<i>ProtocolAdherence</i>
<i>HighTrackingCost</i>	-0.031 (0.031)
Controls	Yes
Year F.E.s	Yes
Observations	1114

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior-year top-quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors are clustered at the CHC level. F.E.s = fixed effects.

*p<0.1, **p<0.05; ***p<0.01

Second, we assess whether high- and low-tracking cost CHCs exhibited differential pre-treatment trends in protocol adherence. We augment the event-study specification with interactions between relative-time indicators and *HighTrackingCost*, and jointly test the pre-treatment interaction terms. As shown in Table A24, these terms are jointly insignificant ($\chi^2(4)=5.95$, $p=0.203$), indicating no evidence of differential pre-adoption trends across tracking cost environments.

Table A24. Differential Pre-Trends in Protocol Adherence by Tracking Cost Environment

	<i>ProtocolAdherence</i>
<i>k=-5</i>	0.011 (0.090)
<i>k=-4</i>	-0.144* (0.075)
<i>k=-3</i>	-0.113 (0.075)
<i>k=-2</i>	-0.022 (0.054)
<i>k=-1</i>	(Baseline omitted)
<i>k=0</i>	0.013 (0.048)
<i>k=1</i>	0.043 (0.070)
<i>k=2</i>	0.082 (0.096)
Controls	Yes
CHC F.E.s	Yes
Year F.E.s	Yes
Observations	1292

Notes: *HighTrackingCost* is a binary indicator equal to one if the share of missing income data in the CHC is above the median, and zero otherwise. Time-varying controls include the local mental health specialist to population ratio, a prior-year top-quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites,

receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors are clustered at the CHC level and are reported in parentheses. F.E.s = Fixed Effects.
 *p<0.1, **p<0.05; ***p<0.01

These two validation analyses support the interpretation that our tracking cost proxy is unlikely to simply reflect preexisting differences in organizational documentation quality or protocol adherence.

Placebo Outcome Tests (Domain Specificity)

To examine whether our estimates reflect domain-specific treatment effects rather than broader time-varying shocks or organizational changes, we conduct placebo outcome tests. Specifically, we re-estimate our baseline DID specifications using outcomes that should not be directly affected by mental telehealth adoption but are measured using the same data source, aggregation level, and reporting structure.

For the specialist-facilitated model, we examine non-mental health access outcomes as placebos. These include patient counts and visits for conditions such as infectious and parasitic diseases (e.g., HIV, tuberculosis), respiratory diseases (e.g., asthma, chronic lower respiratory diseases), and other chronic medical conditions (e.g., diabetes, hypertension, overweight). If the estimated mental health access effects simply reflected general expansion in CHC capacity or unobserved demand shocks, we would expect similar increases across these unrelated service lines. Instead, the placebo coefficients in Table A25 are statistically insignificant and smaller in magnitude than the corresponding mental health access effects. This pattern indicates that the access gains associated with the specialist-facilitated model are concentrated in mental health services rather than reflecting system-wide utilization growth.

Table A25. Placebo Tests: Specialist-Facilitated Telehealth and Non-Mental Health Utilization

	(1) <i>Ln(NonMentalPatients)</i>	(2) <i>Ln(NonMentalVisits)</i>
<i>SpecialistFacilitated Model</i>	0.062 (0.044)	0.046 (0.037)
Controls	Yes	Yes
CHC F.E.s	Yes	Yes
Year F.E.s	Yes	Yes
Observations	1189	1189

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior-year top-quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors are clustered at

the CHC level. F.E.s = fixed effects.
 *p<0.1, **p<0.05; ***p<0.01

For the remote telehealth model, we implement placebo measures of protocol adherence that share the same structure as our primary outcome (screening and follow-up documentation) but operate in distinct clinical domains. Specifically, we examine adherence to cervical cancer screening and BMI documentation with follow-up planning. Re-estimating Equation (2) with these outcomes yields small and statistically insignificant coefficients (Table A26). The absence of effects on these measures suggests that the improvements in depression screening follow-up are not driven by generalized improvements in documentation practices or reporting intensity but are specific to the mental health workflows affected by remote telehealth adoption.

Table A26. Placebo Tests: Remote Telehealth and Non-Mental Health Protocol Adherence		
	(1)	(2)
	<i>ProtocolAdherence_CervicalCancer</i>	<i>ProtocolAdherence_BMI</i>
<i>RemoteTelehealth Model</i>	0.001 (0.014)	-0.010 (0.020)
Controls	Yes	Yes
CHC F.E.s	Yes	Yes
Year F.E.s	Yes	Yes
Observations	1480	1482

Notes: Time-varying controls include the local mental health specialist to population ratio, a prior year top-quality indicator, the percentage of Medicaid patients, the percentage of homeless patients, indicators for provision of pediatric and dental services, electronic health record adoption, telehealth use in other service lines, rural location, number of service sites, receipt of behavioral health grants, and the local percentage of households with computing device access. Standard errors are clustered at the CHC level. F.E.s = fixed effects.
 *p<0.1, **p<0.05; ***p<0.01

Together, these placebo tests support the interpretation that telehealth effects are model-specific and domain-specific. The effects of telehealth adoption are aligned with the service area and workflow targeted by each model, reducing concerns that our main results are driven by unobserved shocks or mechanical changes in reporting behavior.