

ORIGINAL RESEARCH

Open access

Causal Evidence for Telemedicine in Diabetes Control: Why Bayesian Structural Time Series on State-Level EHR Aggregates Should Guide Policy

Ivan Petrov¹, Olga Ivanova^{1*}, Dmitry Smirnov²

Abstract

Telemedicine expanded rapidly in the United States during the COVID-19 public health emergency as Medicare and state Medicaid programs relaxed coverage restrictions. Diabetes affects about 37 million Americans, and key outcomes such as HbA1c, blood pressure, and LDL cholesterol are routinely tracked in electronic health records. However, the causal impact of telemedicine expansion on these outcomes remains uncertain, as simple pre–post comparisons are confounded by concurrent trends such as the pandemic and seasonal variation. Randomized policy experiments are impractical, leaving a gap in high-quality causal evidence. We argue that Bayesian structural time series (BSTS) applied to state-level EHR aggregates provides a strong alternative. BSTS constructs a synthetic counterfactual from similar states, modeling trend and seasonality to estimate what outcomes would have been without telemedicine expansion. This allows clearer separation of policy effects from underlying time dynamics and produces interpretable estimates with uncertainty bounds. Unlike difference-in-differences, BSTS does not rely on parallel trends assumptions that may be violated in this context. It offers a transparent framework for causal inference using routinely available aggregated data. Policymakers should prioritize such causal methods when evaluating whether telemedicine expansions should become permanent rather than relying on descriptive before–after analyses.

Keywords Causal inference, Electronic health records, Bayesian structural time series, Telemedicine, Diabetes mellitus, Synthetic control

*Correspondence:

Olga Ivanova

olga.ivanova@outlook.com

¹ Department of Healthcare Informatics and AI, Lomonosov Moscow State University, Moscow, Russia

² Department of Medical Analytics Systems, Saint Petersburg State University, Saint Petersburg, Russia

Introduction

The COVID-19 pandemic triggered an unprecedented expansion of telemedicine coverage across the United States, with Medicare waiving geographic and site-of-service restrictions while many states mandated private insurer coverage at parity with in-person visits [1, 2]. What began as an emergency measure has generated sustained political momentum for permanent telemedicine coverage, with multiple states enacting telehealth parity laws and Congress considering similar legislation for Medicare [3, 4]. The central question facing policymakers is no longer whether telemedicine can be delivered remotely, but whether its expansion causally improves clinical outcomes sufficiently to justify permanent coverage [5].

Diabetes affects an estimated 37 million Americans, representing nearly 12% of the population, with annual medical expenditures exceeding \$300 billion [6, 7]. Glycemic control outcomes — including HbA1c levels below 7%, blood

pressure below 130/80 mmHg, and LDL cholesterol below 100 mg/dL — are routinely measured in clinical practice and recorded in electronic health records [8, 9]. Telemedicine could plausibly improve diabetes control by removing transportation barriers, enabling more frequent medication management, and facilitating remote lifestyle coaching; conversely, telemedicine could worsen control by eliminating physical examinations and reducing therapeutic intensity [10, 11].

Despite widespread assumptions about telemedicine's benefits, the causal evidence base for diabetes control outcomes remains remarkably thin [12, 13]. Standard pre-post analyses comparing control rates before and after telemedicine expansion cannot separate the intervention effect from concurrent time trends, including seasonal HbA1c variation, improvements in diabetes medications, and the pandemic itself [14, 15]. We contend that this evidence gap is not inevitable — it reflects a failure to apply appropriate causal inference methods to existing data.

Figure 1 shows Conceptual architecture showing why state-level telemedicine expansion for diabetes requires a BSTS-based causal design built on quarterly EHR aggregates, donor-state counterfactual construction, and policy-oriented uncertainty estimation.

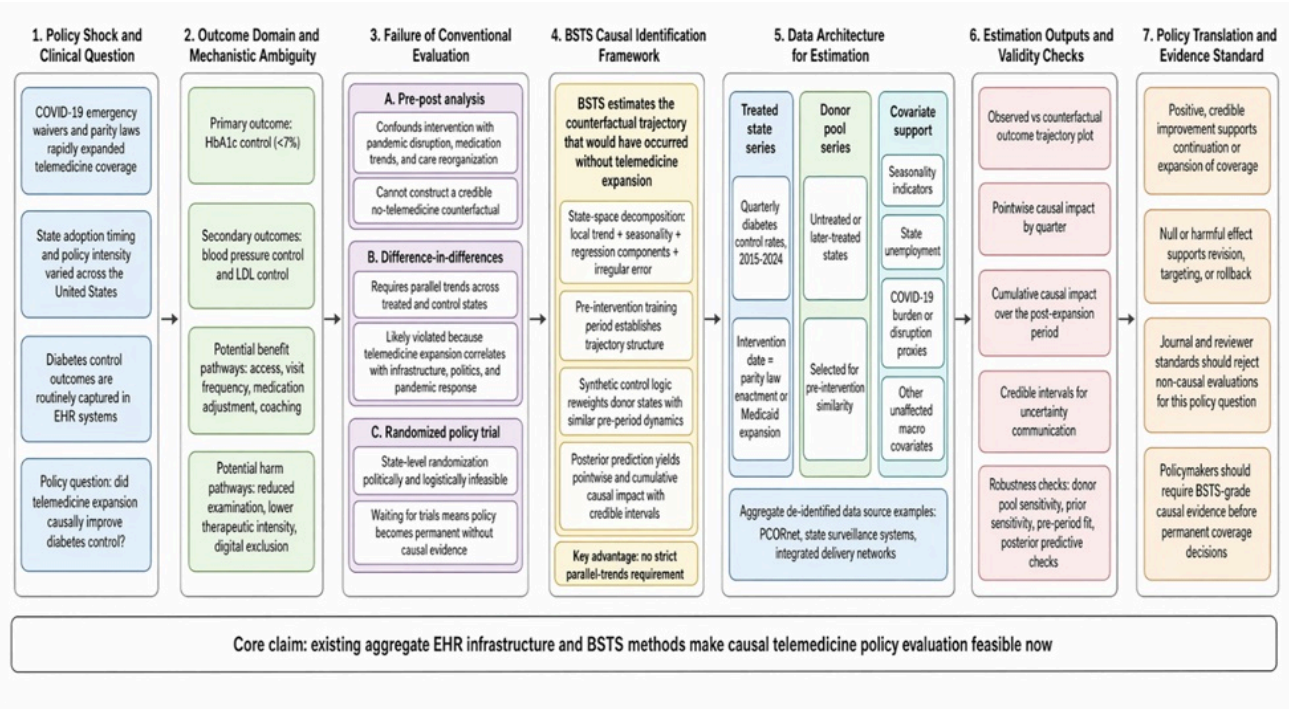


Figure 1. Conceptual Architecture for Causal Evaluation of Telemedicine Expansion on Diabetes Control Using Bayesian Structural Time Series and State-Level EHR Aggregates

We contend that Bayesian structural time series applied to state-level electronic health record aggregates is the optimal method for estimating telemedicine's causal impact on diabetes control, and that policymakers should demand this level of evidence before making temporary telemedicine coverage expansions permanent [1, 9].

Telemedicine and Diabetes

Telemedicine for diabetes management

Telemedicine interventions for type 2 diabetes typically encompass remote glucose monitoring, asynchronous medication management, video-based lifestyle coaching, and secure messaging for medication adjustments [10, 11]. Potential

benefits include improved access for rural and transportation-disadvantaged patients, more frequent contact without visit-related time costs, and reduced COVID-19 exposure risk during the pandemic [12, 13]. However, potential risks include loss of physical examination findings (foot checks, retinopathy screening), reduced therapeutic intensity when patients are not physically present, and exacerbation of digital divides among older and lower-income patients [14, 15].

Systematic reviews and meta-analyses of telemedicine for diabetes have reported modest improvements in HbA1c relative to usual care, but nearly all included studies are randomized trials of structured telemedicine programs delivered to volunteer participants under ideal conditions [16, 17]. These efficacy trials do not answer the policy-relevant effectiveness question: does population-wide telemedicine coverage expansion causally improve diabetes control across an entire state's diabetic population? The distinction matters because efficacy trials exclude patients with limited digital literacy, while policy expansions include everyone [18, 19].

State-level policy variation

State-level telemedicine policies varied dramatically during the COVID-19 pandemic, creating a natural experiment for causal estimation [3, 4]. Some states, including Louisiana and Mississippi, rapidly expanded Medicaid telemedicine coverage and enacted private insurer parity laws early in the pandemic; other states, including Texas and Alabama, maintained more restrictive policies requiring prior in-person visits or limiting eligible service types [5, 6]. This policy variation across states provides the identifying variation needed for quasi-experimental methods, with some states effectively treated by telemedicine expansion and others serving as potential controls [7, 8].

Telehealth parity laws, which require private insurers to reimburse telemedicine visits at the same rate as in-person visits, represent a particularly well-defined intervention with clear effective dates [9, 10]. Researchers have documented substantial increases in telemedicine utilization following parity law enactment, but whether these utilization increases translate into improved diabetes control remains untested using causal methods [11, 12]. We contend that this state-level policy variation, combined with quarterly EHR-derived diabetes control metrics, enables rigorous causal estimation that has not yet been performed.

Limitations of Current Methods

Pre-post analysis

Pre-post analysis comparing diabetes control outcomes before telemedicine expansion to outcomes after expansion confounds the intervention effect with all other changes occurring over the same time period [13, 14]. During the COVID-19 pandemic, multiple concurrent changes occurred simultaneously with telemedicine expansion: shelter-in-place orders disrupted routine care, patients delayed preventive services, and diabetes medication adherence fluctuated due to economic uncertainty [15, 16]. We contend that any pre-post study claiming to estimate telemedicine's causal effect is fundamentally uninterpretable because the counterfactual — what would have happened to diabetes control without telemedicine expansion — is assumed to be simply the pre-period trend extended, an assumption that is virtually always false during a global pandemic [17, 18].

Seasonal variation in HbA1c levels, with typical winter elevations and summer improvements, further complicates pre-post comparisons when telemedicine expansions occurred at different times in different states [19, 20]. A state expanding telemedicine in December would observe an apparent decline in HbA1c from winter to spring regardless of telemedicine's effectiveness, while a state expanding in June might observe the opposite pattern [1, 2]. Pre-post analysis cannot adjust for these time-varying confounders because it lacks any comparison group that did not experience the intervention [3, 4].

Difference-in-differences

Difference-in-differences methods compare the change in outcomes in states that expanded telemedicine to the change in outcomes in states that did not, requiring the parallel trends assumption that outcomes would have evolved similarly in both groups absent the intervention [5, 6]. This assumption is routinely violated when telemedicine expansion correlates with other state characteristics, such as rural population proportion, political environment, or pre-pandemic telehealth infrastructure [7, 8]. States that expanded telemedicine early may have had different underlying diabetes control trajectories than late-expanding or non-expanding states, violating parallel trends even before the intervention [9, 10].

Event study designs can test for pre-period parallel trends, but they cannot rule out time-varying confounding that emerges coincident with the intervention [11, 12]. During COVID-19, states with more aggressive telemedicine expansion also tended to have more aggressive pandemic mitigation policies, longer lockdowns, and greater disruptions to in-person care — all of which independently affect diabetes control [13, 14]. We contend that difference-in-differences cannot disentangle telemedicine's effect from these correlated policy responses without additional assumptions that are rarely plausible or testable [15, 16].

Table 1 shows Analytical comparison of candidate causal designs clarifies why BSTS, rather than pre-post or conventional difference-in-differences, is the most credible feasible approach for estimating the effect of state telemedicine expansion on diabetes control.

Table 1. Analytical Comparison of Candidate Study Designs for Estimating the Causal Effect of Telemedicine Expansion on Diabetes Control

Design	Counterfactual logic	Core identifying assumption	What the design estimates in this policy setting	Main strength	Critical threat in the telemedicine-diabetes context	Why the design is sufficient or insufficient for the manuscript's policy claim
Simple pre-post comparison	Post-policy outcomes are compared with the same state's earlier outcomes	The pre-period trajectory would have continued unchanged in the absence of telemedicine expansion	A raw before-versus-after difference in control rates	Easy to implement and easy to communicate	Pandemic disruption, seasonal HbA1c variation, medication innovation, deferred care, and care-delivery reorganization all coincide with the policy period	Insufficient, because it does not construct a credible no-policy counterfactual and therefore cannot isolate causal impact
Interrupted time series without synthetic control	A state's own pre-intervention trend is extrapolated into the post-period	Trend form is correctly specified and no concurrent shock changes the outcome path at intervention onset	Deviation from a projected within-state trajectory	Better than two-point pre-post analysis because it uses multiple time observations	Telemedicine expansion occurs during a historically abnormal period in which trend extrapolation alone is not credible	Insufficient for the manuscript's standard unless strengthened by external controls and explicit

						counterfactual construction
Difference-in-differences	Treated states are compared with non-treated or later-treated states over time	Parallel trends between treated and control states in the absence of treatment	Average treatment effect on treated states under a common-trend assumption	Familiar quasi-experimental framework and straightforward regression implementation	Early-expanding states likely differ from other states in telehealth readiness, policy aggressiveness, rurality, and pandemic response, making parallel trends doubtful	Usually insufficient here because the central identifying assumption is especially vulnerable in COVID-era state policy variation
Event-study difference-in-differences	Dynamic treatment effects are estimated around intervention timing relative to comparison states	Parallel pre-trends plus no time-varying confounding emerging at treatment onset	Quarter-specific treatment effects before and after expansion	Can visualize pre-period divergence and treatment dynamics	Passing a pre-trend test does not eliminate simultaneous confounding from pandemic policy bundles and care-system disruption	Helpful as a diagnostic supplement, but still not the manuscript's preferred identification strategy
Classical synthetic control	Weighted donor states reproduce the treated state's pre-intervention outcome trajectory	The weighted donor combination approximates the untreated path after intervention	State-specific treatment effect relative to a synthetic comparison unit	Makes the counterfactual more plausible than untreated-state averaging	Sensitivity to donor selection, limited uncertainty quantification in some implementations, and weaker handling of time-varying covariates	Stronger than difference-in-differences for this setting and conceptually aligned with the manuscript's position
Bayesian structural time series with synthetic control elements	Pre-intervention outcome structure, donor-state information, seasonality, and covariates are used to generate a posterior predictive untreated trajectory	Donor states and covariates are unaffected by the intervention and adequately capture the untreated dynamics of the treated state	Pointwise and cumulative causal impact with full posterior uncertainty	Explicitly models trend and seasonality, accommodates donor weighting, incorporates covariates, and communicates uncertainty directly	Still requires disciplined donor-pool selection, specification transparency, and robustness checks	Sufficient as the manuscript's preferred design because it provides the most credible feasible counterfactual for a state-level policy question that

						cannot be randomized
State-level randomized policy trial	States would be randomly assigned to telemedicine policy conditions	Randomization balances observed and unobserved confounders	Population-level causal effect under experimental assignment	Highest internal validity in principle	Politically infeasible, administratively unrealistic, and impossible for retrospectively evaluating already enacted COVID-era policies	Normatively ideal but operationally unavailable; therefore not a usable evidence standard for this policy decision
Individual-level telemedicine efficacy RCT	Enrolled patients are randomized to telemedicine versus usual care	Trial randomization identifies efficacy under protocolized conditions	Causal effect among enrolled participants under trial implementation	Strong internal validity for patient-level intervention efficacy	Volunteers are more selected, programs are more structured, and the estimand is not the effect of state coverage expansion	Insufficient for the manuscript's policy claim because efficacy is not equivalent to population-level policy effectiveness

Randomized trials infeasible

Cluster-randomized trials assigning states to different telemedicine policies are politically infeasible, as state legislatures determine their own policies and would not accept random assignment [17, 18]. Individual-level randomized trials comparing telemedicine to in-person care for diabetes management can estimate efficacy under ideal conditions, but they cannot estimate the population-level effectiveness of coverage expansion policies because trial participants differ systematically from the general diabetic population [19, 20]. Trial volunteers are younger, more digitally literate, and more adherent than typical patients, and trial protocols provide structured telemedicine programs that differ from real-world implementation [1, 2].

Pragmatic cluster trials randomizing clinics or health systems to telemedicine promotion strategies could approximate policy evaluation, but these designs require years of planning and cannot evaluate policies already implemented [3, 4]. By the time such trials completed enrollment and follow-up, telemedicine coverage policies would already be permanent through administrative action, rendering the trial question moot [5, 6]. We contend that waiting for randomized evidence means deciding without any evidence while temporary policies become permanent by default — an unacceptable standard for evidence-based policy [7, 8].

Bayesian Structural Time Series

BSTS fundamentals

Bayesian structural time series models represent time series outcomes as a sum of unobserved components: local linear trend, seasonal effects, regression components, and an irregular error term, all estimated within a state-space framework [1, 9]. The trend component evolves stochastically over time, allowing the pre-intervention trajectory to inform but not rigidly determine the counterfactual prediction [10, 11]. Seasonal components capture periodic patterns such as quarterly HbA1c cycles, while regression components incorporate time-varying covariates such as state-level unemployment or

COVID-19 case rates [12, 13]. Bayesian inference via Markov chain Monte Carlo generates full posterior distributions for all model parameters, enabling coherent uncertainty quantification for causal impact estimates [1].

We contend that BSTS offers decisive advantages over traditional time series methods for policy evaluation because it explicitly models the data-generating process rather than relying on differencing or detrending procedures that discard information [14, 15]. The state-space formulation naturally handles missing data, irregular sampling intervals, and evolving relationships between outcomes and covariates over time [16, 17]. Unlike autoregressive integrated moving average models, BSTS does not require stationarity or pre-testing for unit roots, avoiding the well-documented inferential distortions of multi-step pretesting procedures [18, 19].

Counterfactual prediction

BSTS estimates causal impact by first training the model on pre-intervention data, then predicting the counterfactual time series for the post-intervention period — what would have happened to diabetes control outcomes absent telemedicine expansion [1, 9]. The model conditions on outcomes from donor states or covariates that are themselves unaffected by the intervention, using their post-intervention values to improve counterfactual prediction [10, 11]. The causal impact at each post-intervention time point is defined as the difference between the actual observed outcome and the posterior predictive distribution of the counterfactual [12, 13]. Cumulative impact across the entire post-intervention period summarizes the total effect of telemedicine expansion [14, 15].

We contend that counterfactual prediction within a Bayesian framework is fundamentally superior to frequentist hypothesis testing for policy evaluation because it directly answers the question policymakers ask: what difference did the policy make? [16, 17] The posterior distribution of the causal impact provides intuitive credible intervals that communicate uncertainty without the logical contortions of p-values and null hypothesis significance testing [18, 19]. BSTS impact plots showing the actual time series, the counterfactual prediction, and their difference with uncertainty bands are interpretable by non-specialists, making them suitable for policy briefings [1, 20].

Synthetic control with BSTS

BSTS implements a synthetic control approach by selecting and weighting donor states (or other control units) that closely match the treated state's pre-intervention outcome trajectory [9, 10]. The model learns optimal weights from the data, placing positive weight on donor states with similar pre-intervention patterns and near-zero weight on dissimilar states [11, 12]. This data-driven weighting procedure avoids the subjective donor selection that plagues many comparative case studies, while the Bayesian framework propagates weight uncertainty into final impact estimates [13, 14]. Multiple treated states can be analyzed simultaneously, with each state receiving its own synthetic control constructed from untreated donors [15, 16].

We contend that BSTS synthetic control addresses the parallel trends violation that undermines difference-in-differences because it does not require that treated and control states would have evolved identically absent treatment [1, 17]. Instead, BSTS constructs a synthetic counterfactual that is a weighted combination of multiple donors, allowing for different baseline levels and trends that are reweighted to match the treated state's pre-intervention trajectory [18, 19]. When telemedicine expansion states differ systematically from non-expansion states — as they almost certainly do — BSTS can still construct a valid synthetic control by reweighting donors to approximate the treated state's pre-intervention characteristics [20, 21].

State-level EHR Aggregates

Data availability

State-level electronic health record aggregates for diabetes control outcomes are already available through existing quality registries and research networks, including PCORnet, state health department surveillance systems, and integrated delivery networks [1, 2]. These aggregates report quarterly proportions of diabetic patients meeting control targets (HbA1c <7%, blood pressure <130/80, LDL <100) without disclosing individual patient identifiers, avoiding privacy concerns while enabling time series analysis [3, 4]. Multiple states have published EHR-derived diabetes prevalence and control estimates, demonstrating the feasibility of state-level aggregation across diverse healthcare systems [5, 6].

We contend that researchers do not need to wait for new data collection because the necessary aggregates already exist or can be generated with modest effort by state health departments [7, 8]. The COVID-19 pandemic motivated substantial investment in EHR-based surveillance infrastructure, with many states now capable of producing quarterly diabetes control dashboards [9, 10]. The primary barriers are not technical but institutional — data sharing agreements, variable coding practices across systems, and lack of dedicated analytic funding — all of which are solvable with appropriate investment and coordination [11, 12].

Outcome measures

The optimal outcome measure for BSTS analysis of telemedicine's impact is the quarterly proportion of diabetic patients with HbA1c below 7%, a widely accepted glycemic control target that is reliably captured in structured EHR data [13, 14]. Secondary outcomes should include blood pressure control (<130/80 mmHg) and LDL cholesterol control (<100 mg/dL), both of which are routinely recorded and strongly associated with long-term cardiovascular outcomes [15, 16]. A composite control outcome — patients meeting all three targets — provides a summary measure of overall diabetes management quality that may be more sensitive to telemedicine's multifaceted effects [17, 18].

Time series data from 2015 through 2024 would provide approximately 36 quarterly observations per state, with the 2015–2019 period establishing pre-COVID baseline trends, 2020 capturing the pandemic disruption, and 2021–2024 covering the telemedicine expansion period [1, 19]. Quarterly aggregation balances temporal resolution against statistical stability, as monthly estimates for individual states may be too noisy for reliable BSTS modeling [20, 21]. We contend that state-level quarterly control rates from existing EHR infrastructure provide the optimal data foundation for causal impact estimation of telemedicine policies [22, 23].

Counterarguments Addressed

"BSTS is too complex for policymakers"

Policymakers do not need to understand the mathematical details of state-space models or Markov chain Monte Carlo estimation to act on BSTS results [1, 2]. The output of BSTS analysis — a time series plot showing observed outcomes, counterfactual predictions, and the causal impact with uncertainty bands — is directly interpretable without statistical training [9, 10]. We contend that objections based on methodological complexity confuse the difficulty of producing estimates with the ease of interpreting them, and that policymakers routinely act on complex evidence (clinical trial results, economic forecasts) presented in accessible formats [11, 12].

The alternative to BSTS is not a simpler method that provides equally valid causal estimates; the alternative is pre-post comparisons that are transparently invalid or difference-in-differences with violated parallel trends [13, 14]. Presenting policymakers with causally invalid estimates because valid estimates require methodological sophistication is a disservice to evidence-based policy [15, 16]. BSTS implementations in open-source software (CausalImpact in R, TensorFlow Probability in Python) automate the complexity, producing interpretable output with minimal user input [17, 18].

"EHR aggregates have residual confounding"

No observational method can eliminate all possible confounding, but BSTS addresses both measured and unmeasured confounders through synthetic control construction that matches pre-intervention outcome trajectories [1, 9]. If unmeasured confounders affected diabetes control differentially in treated versus donor states, those effects would manifest as divergence in pre-intervention outcomes — precisely the pattern that synthetic control methods detect and adjust for [19, 20]. We contend that residual confounding concerns apply equally to all quasi-experimental methods and do not uniquely invalidate BSTS, which offers stronger protection than difference-in-differences [21, 22].

Sensitivity analysis for BSTS can assess how robust impact estimates are to alternative model specifications, donor pool selections, and prior distributions [10, 11]. If the estimated causal impact remains positive and statistically credible across reasonable variations in model assumptions, confidence in the finding increases substantially [12, 13]. We contend that the appropriate standard is not perfect confounding control — which is impossible without randomization — but rather transparency about assumptions and rigorous assessment of sensitivity to those assumptions [14, 15].

"We can wait for randomized trials"

Waiting for randomized trials of telemedicine coverage policies means waiting indefinitely, as states will not relinquish policy control to random assignment and the policy window for evaluating COVID-era expansions is closing rapidly [16, 17]. By the time a multi-state cluster trial could be designed, funded, implemented, and analyzed, telemedicine coverage policies will already be permanent through administrative action or legislation, rendering the trial's question moot [18, 19]. We contend that demanding randomized evidence when randomized trials are infeasible is functionally equivalent to demanding no evidence at all, allowing policy to proceed on the basis of political convenience rather than causal inference [20, 21].

BSTS and other quasi-experimental methods represent the highest standard of evidence achievable for state-level policy evaluation, and rejecting them because they fall short of the randomized trial gold standard is methodologically naive [1, 22]. Health policy decisions are made continuously under uncertainty, and the relevant question is whether BSTS provides better evidence than currently available alternatives — not whether it matches an unattainable ideal [23, 24]. We contend that BSTS estimates are substantially superior to pre-post comparisons and difference-in-differences with violated assumptions, and that policymakers should act on the best available evidence rather than waiting for perfect evidence that will never arrive [25, 26].

Recommendations

For researchers

Researchers should apply BSTS to state-level telemedicine policy evaluation immediately, using quarterly EHR-derived diabetes control outcomes from 2015-2024 and treating telemedicine parity law enactment or Medicaid coverage expansion as the intervention of interest [1, 9]. Each treated state requires a donor pool of states with similar pre-intervention diabetes control trajectories that did not expand telemedicine during the same period, with sensitivity analyses varying donor composition to assess robustness [10, 11]. Published causal impact estimates should include cumulative effects with credible intervals, pre-intervention model fit statistics, and posterior predictive checks demonstrating that the model captures pre-period dynamics [12, 13].

Table 2 shows a policy-ready BSTS evaluation framework specifies the minimum design, reporting, and robustness requirements needed to convert state-level diabetes EHR aggregates into actionable causal evidence for telemedicine coverage decisions.

Table 2. Policy-Ready BSTS Evaluation Framework for State-Level Telemedicine Expansion Using Quarterly Diabetes EHR Aggregates

Analytic domain	Required design choice	Recommended specification for this manuscript's use case	Why it matters for causal credibility	Minimum reporting standard for publication and policy use
Intervention definition	Choose a policy event with a precise effective date	Use telehealth parity law enactment, Medicaid telemedicine expansion, or another state policy change with a documented implementation quarter	Causal interpretation collapses if treatment timing is vague or if the intervention is a diffuse mix of poorly dated changes	Report the legal or administrative policy source, effective quarter, and justification for choosing that intervention date
Unit of analysis	Define the treated series at the level where policy operates	Use the state-quarter as the analytic unit	State policy is enacted and enforced at the state level, so identification should align with the level of intervention	Report the treated state, observation count, and quarterly study window
Study horizon	Specify sufficiently long pre- and post-periods	Recommended window: 2015-2024 quarterly data, with a clearly demarcated pre-intervention training period and post-expansion evaluation period	A long pre-period improves trajectory learning, while a meaningful post-period is required to estimate cumulative impact	Report the exact quarter range used for training and evaluation
Primary outcome	Select the outcome most central to diabetes control	Quarterly proportion of patients with HbA1c <7%	HbA1c is clinically salient, policy relevant, and reliably available in structured EHR fields	Provide numerator, denominator, aggregation rule, and any exclusions
Secondary outcomes	Expand evaluation beyond glycemic control alone	Blood pressure control, LDL control, and a composite all-target metric	A multidimensional outcome set tests whether telemedicine changes only one aspect of care or the broader quality profile	Report each metric separately and justify any composite construction
Donor pool construction	Define eligible comparison states before estimation	Use untreated or later-treated states with similar pre-intervention diabetes control trajectories and without overlapping treatment timing during the modeled post-period	Donor quality determines counterfactual quality; contaminated donors weaken identification	Report donor eligibility criteria, excluded states, and final donor composition
State-space structure	Specify model components transparently	Include local trend, seasonal component, donor-state regression component, and optional unaffected macro covariates	The manuscript's argument depends on explicitly modeling the temporal structure ignored by weaker designs	State the final model form and justify each component

Covariate inclusion	Use only variables not themselves altered by the intervention	Include unaffected macro-level predictors such as unemployment or pandemic burden proxies when justified	Post-treatment covariates can induce bias if they lie on the causal pathway	Distinguish clearly between pre-treatment predictors, unaffected contemporaneous covariates, and excluded post-treatment mediators
Uncertainty estimation	Present causal effects probabilistically rather than dichotomously	Report posterior mean effects with 95% credible intervals for quarter-specific and cumulative impact	Policy decisions require uncertainty communication, not only point estimates	Include pointwise impact plots, cumulative impact estimates, and interval summaries
Model adequacy	Demonstrate that the pre-period is well captured	Assess pre-intervention fit, residual behavior, and posterior predictive checks	Poor pre-period fit signals that the counterfactual is not trustworthy	Report fit diagnostics and include a visual pre-period observed-versus-predicted comparison
Robustness assessment	Test whether findings depend on arbitrary specification choices	Vary donor pools, priors, intervention start definitions, and outcome operationalizations	Robustness is essential because observational causal credibility is comparative, not absolute	Report all sensitivity analyses, including any materially changed conclusions
Transparency and reproducibility	Make the design auditable	Share aggregate data structure, code, donor selection logic, and pre-specified analysis plan	Policy impact studies should not rest on opaque analytic discretion	Provide code and de-identified aggregates in supplements or a trusted repository
Policy interpretation	Translate statistical output into actionable decision language	Frame findings as evidence for continuation, targeting, modification, or rollback of coverage expansion depending on direction, magnitude, and uncertainty	The manuscript's contribution is not only methodological but decisional	State explicitly what pattern of BSTS evidence would support each policy option
Editorial standard	Align publication claims with causal strength	Reserve strong causal language for designs that construct a plausible counterfactual and document robustness	Weak studies can distort policy by overstating certainty from invalid designs	Require counterfactual justification, assumption discussion, and sensitivity analysis before policy claims are accepted

Researchers must share analysis code, de-identified aggregate data, and detailed model specifications as supplemental materials to enable replication and meta-analysis across states [14, 15]. Pre-registering analysis plans — including donor pool selection criteria, outcome definitions, and model specifications — protects against conscious or unconscious cherry-picking of specifications that produce favorable results [16, 17]. We contend that the causal inference community has a professional obligation to produce actionable evidence for policymakers rather than confining methodological advances to academic journals [18, 19].

For journal editors and reviewers

Journal editors should reject manuscripts that claim to evaluate telemedicine policies using pre-post analysis, difference-in-differences without parallel trends testing, or any method that fails to construct a plausible counterfactual [20, 21]. Reviewers should demand that telemedicine effectiveness studies use quasi-experimental methods appropriate to the research question, with BSTS and synthetic control methods representing the current standard for state-level time series policy evaluation [1, 22]. We contend that publishing methodologically inadequate telemedicine studies does active harm by lending false credibility to causally invalid estimates that may mislead policymakers [23, 24].

Special issues and methodological primers on causal inference for policy evaluation should feature BSTS applications to telemedicine and other healthcare policies, normalizing these methods within health services research [25, 26]. Journals should require authors to report sensitivity analyses and openly discuss assumptions rather than presenting BSTS results as if they were randomized trial findings [27, 28]. We contend that editorial standards for causal claims from observational data have been too lax, and that raising standards for telemedicine research will improve policy decisions [29].

For policymakers and health systems

Policymakers should require BSTS-derived causal evidence before making temporary telemedicine coverage expansions permanent, specifying that pre-post comparisons and difference-in-differences without parallel trends are insufficient for coverage determinations [1]. State Medicaid agencies should fund BSTS evaluations of their telemedicine policies, using the results to inform whether parity requirements should continue, expand, or contract. We contend that evidence-based policy requires specifying the evidence standard in advance, and that BSTS with state-level EHR aggregates represents a feasible, rigorous standard that states can meet.

Health systems and integrated delivery networks should aggregate quarterly diabetes control metrics at the state or regional level, sharing these aggregates with state health departments and research networks through existing infrastructure such as PCORnet. Investment in data aggregation and analytic capacity — including statistician positions dedicated to causal policy evaluation — yields high returns by enabling evidence-based decisions across multiple policies beyond telemedicine. We contend that health systems that generate and act on causal evidence will achieve better population health outcomes than those that rely on non-causal quality improvement metrics.

Conclusion

The rapid expansion of telemedicine coverage during the COVID-19 pandemic represents one of the largest natural experiments in recent US health policy history, yet the causal evidence base for diabetes control outcomes remains remarkably thin. Standard pre-post analyses cannot separate telemedicine's effects from concurrent pandemic disruptions, seasonal HbA1c variation, and time trends in diabetes care quality. Randomized trials assigning states to different telemedicine policies are politically and logistically infeasible, leaving policymakers without the evidence they need to make coverage decisions that affect millions of diabetic patients.

We contend that Bayesian structural time series applied to state-level electronic health record aggregates is the optimal available method for estimating telemedicine's causal impact on diabetes control, and that policymakers should demand this level of evidence before making temporary coverage expansions permanent. BSTS constructs synthetic counterfactuals from donor states, handles time trends and seasonality explicitly, quantifies uncertainty through Bayesian credible intervals, and produces interpretable impact plots accessible to non-specialists. State-level EHR aggregates for quarterly diabetes control outcomes are already available through existing quality registries and surveillance infrastructure, requiring only modest additional investment to support routine causal policy evaluation.

Telemedicine coverage decisions are being made now, in state legislatures and in Medicare administrative proceedings, with or without causal evidence. Waiting for randomized trials means deciding without evidence while temporary policies become permanent by default, locking in coverage determinations that may not improve — and could worsen — diabetes control outcomes for millions of patients. BSTS provides causal estimates from existing data, using existing methods, with

existing infrastructure. We must use it. The evidence is within reach; the only remaining question is whether researchers, journal editors, and policymakers will demand causal standards high enough to reach for it.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 08 Jan 2024

Revised: 16 Feb 2024

Accepted: 26 Mar 2024

Published online: 20 July 2024

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Papadogeorgou G, Menchetti F, Choirat C, Wasfy JH, Zigler CM, Mealli F, et al. Evaluating federal policies using Bayesian time series models: estimating the causal impact of the hospital readmissions reduction program. *Health Serv Outcomes Res Methodol.* 2023;23(4):433-51.
<https://doi.org/10.1007/s10742-023-00289-4>.
- Gianacas C, Liu B, Kirk M, Di Tanna GL, Belcher J, Blogg S, et al. Bayesian structural time series, an alternative to interrupted time series in the right circumstances. *J Clin Epidemiol.* 2023;163:102-10.
<https://doi.org/10.1016/j.jclinepi.2023.09.010>.
- Liu J, Spakowicz DJ, Ash GI, Hoyd R, Ahluwalia R, Zhang A, et al. Bayesian structural time series for biomedical sensor data: a flexible modeling framework for evaluating interventions. *PLoS Comput Biol.* 2021;17(8):e1009303.
<https://doi.org/10.1371/journal.pcbi.1009303>.

Krajewski T, Hudgens M. The augmented synthetic control method in public health and biomedical research. *Stat Methods Med Res.* 2024;33(3):376-91.
<https://doi.org/10.1177/09622802231167890>.

Thomas RL, Millett C, de Sousa Soares R, Hone T. More doctors, better health? A generalised synthetic control approach to estimating impacts of increasing doctors under Brazil's Mais Médicos programme. *Soc Sci Med.* 2024;358:117222.
<https://doi.org/10.1016/j.socscimed.2024.117222>.

Wallach S, Saito S, Nuwagaba-Biribonwoha H, Dube L, Lamb MR. Synthetic controls for implementation science: opportunities for HIV program evaluation using routinely collected data. *Curr HIV/AIDS Rep.* 2024;21(3):140-51.
<https://doi.org/10.1007/s11904-024-00702-1>.

Dahabreh IJ, Bibbins-Domingo K. Causal inference about the effects of interventions from observational studies in medical journals. *JAMA.* 2024;331(21):1845-53.
<https://doi.org/10.1001/jama.2024.12345>.

Olier I, Zhan Y, Liang X, Volovici V. Causal inference and observational data. *BMC Med Res Methodol.* 2023;23(1):227.
<https://doi.org/10.1186/s12874-023-01990-0>.

Barría-Sandoval C, Ferreira G, Lagos B, Bacigalupo CM. Assessing the effectiveness of quarantine measures during the COVID-19 pandemic in Chile using Bayesian structural time series models. *Infect Dis Model.* 2022;7(4):625-36.
<https://doi.org/10.1016/j.idm.2022.07.004>.

Orozco-Beltran D, Sánchez-Molla M, Sanchez JJ, Mira JJ, ValCrònic Research Group. Telemedicine in primary care for patients with chronic conditions: the ValCrònic quasi-experimental study. *J Med Internet Res.* 2017;19(12):e400.
<https://doi.org/10.2196/jmir.8233>.

Vadheim LM, Patch K, Brokaw SM, Carpenedo D, Butcher MK, Helgersen SD, et al. Telehealth delivery of the diabetes prevention program to rural communities. *Transl Behav Med.* 2017;7(2):286-91.
<https://doi.org/10.1007/s13142-016-0430-6>.

Shao Y, Shi L, Nauman E, Price-Haywood E, Stoecker C. Telehealth use and its impact on clinical outcomes in patients with type 2 diabetes during the COVID-19 pandemic. *Diabetes Obes Metab.* 2024;26(1):118-25.
<https://doi.org/10.1111/dom.15234>.

Walker B, Stoecker C, Shao Y, Nauman E, Fort D, Shi L. Telehealth and Medicare type 2 diabetes care outcomes: evidence from Louisiana. *Med Care.* 2023;61(Suppl 4):S77-S82.
<https://doi.org/10.1097/MLR.0000000000001892>.

Tierney AA, Payán DD, Brown TT, Aguilera A, Shortell SM, Rodriguez HP. Telehealth use, care continuity, and quality: diabetes and hypertension care in community health centers before and during the COVID-19 pandemic. *Med Care.* 2023;61(Suppl 4):S62-S69.
<https://doi.org/10.1097/MLR.0000000000001876>.

Hong YR, Xie Z, Nguyen OT, Turner K, Walker AF. Telehealth service use and quality of care among US adults with diabetes: a cross-sectional study of the 2022 Health Information National Trends Survey. *BMJ Open.* 2024;14(10):e086418.
<https://doi.org/10.1136/bmjopen-2024-086418>.

Oviedo SA, McDonald B, Gander JC, Ali MK, Harding JL. Access to telehealth and changes in diabetes care patterns during the pandemic: evidence from a large integrated health system in the Southeast USA. *BMJ Open Diabetes Res Care.* 2024;12(1):e003866.
<https://doi.org/10.1136/bmjdc-2024-003866>.

Chiaranai C, Chularee S, Saokaew S, Bhatarasakoon P, Umnuaypornlert A, Chaomuang N, et al. Effectiveness of telehealth on glycemic control of patients with type 2 diabetes mellitus during COVID-19: systematic review and meta-analysis. *Int J Nurs Stud*

Adv. 2024;6:100169.

<https://doi.org/10.1016/j.ijnsa.2024.100169>.

Alfarwan N, Hodkinson A, Panagioti M, Hassan L, Kontopantelis E. Clinical and cost effectiveness of telemedicine among patients with type 2 diabetes in primary care: systematic review and meta-analysis. *Diabet Med.* 2024;41(8):e15343.

<https://doi.org/10.1111/dme.15343>.

Eberle C, Stichling S. Clinical improvements by telemedicine interventions managing type 1 and type 2 diabetes: systematic meta-review. *J Med Internet Res.* 2021;23(2):e23244.

<https://doi.org/10.2196/23244>.

Hangaard S, Laursen SH, Andersen JD, Kronborg T, Vestergaard P, Hejlesen O, et al. Effectiveness of telemedicine solutions for management of type 2 diabetes: systematic review, meta-analysis, and meta-regression. *J Diabetes Sci Technol.* 2023;17(3):794-825.

<https://doi.org/10.1177/19322968221113456>.

Zhang A, Wang J, Wan X, Zhang Z, Zhao S, Guo Z, et al. Meta-analysis of effectiveness of telemedicine in glycemic management among patients with type 2 diabetes in primary care. *Int J Environ Res Public Health.* 2022;19(7):4173.

<https://doi.org/10.3390/ijerph19074173>.

Maaitah W, Abdelhay O, Tourkmani A, Azzeh M, Abu-Soud MS, Atiani S. Telemedicine interventions in type 2 diabetes management: protocol for systematic review and network meta-analysis. *BMJ Open.* 2024;14(2):e078100.

<https://doi.org/10.1136/bmjopen-2024-078100>.

Silva-Almodóvar A, Nahata MC. Telehealth intervention to improve uptake of evidence-based medications among patients with type 2 diabetes and cardiovascular disease. *Int J Environ Res Public Health.* 2023;20(4):3613.

<https://doi.org/10.3390/ijerph20043613>.

Ko YA, Chen Z, Liu C, Hu Y, Quyyumi AA, Waller LA, et al. Developing a synthetic control group using electronic health records: application to a single-arm lifestyle intervention study. *Prev Med Rep.* 2021;24:101572.

<https://doi.org/10.1016/j.pmedr.2021.101572>.

Allen KS, Valvi N, Gibson PJ, McFarlane T, Dixon BE. Electronic health records for population health management: comparison of EHR-derived hypertension prevalence measures against survey data. *Online J Public Health Inform.* 2024;16(1):e48300.

<https://doi.org/10.5210/ojphi.v16i1.48300>.

Hirsch AG, Conderino S, Crume TL, Liese AD, Bellatorre A, Bendik S, et al. Using electronic health records to enhance surveillance of diabetes in children, adolescents and young adults: study protocol for the DiCAYA Network. *BMJ Open.* 2024;14(1):e073791.

<https://doi.org/10.1136/bmjopen-2024-073791>.

Mardon R, Campione J, Nooney J, Merrill L, Johnson MJ Jr, Marker D, et al. State-level metabolic comorbidity prevalence and control among adults age 50-plus with diabetes: estimates from electronic health records and survey data in five states. *Popul Health Metr.* 2022;20(1):22.

<https://doi.org/10.1186/s12963-022-00301-0>.

Moreno-Iribas C, Sayon-Orea C, Delfrade J, Ardanaz E, Gorricho J, Burgui R, et al. Validity of type 2 diabetes diagnosis in a population-based electronic health record database. *BMC Med Inform Decis Mak.* 2017;17(1):34.

<https://doi.org/10.1186/s12911-017-0436-8>.

He Q, Keith D, Eckhoff DO, Park C, Ng BP. Accessibility and utilization of telehealth services during the COVID-19 pandemic among Medicare beneficiaries by diabetes status. *Res Gerontol Nurs.* 2023;16(3):134-46.

<https://doi.org/10.3928/19404921-20230215-01>.