

ORIGINAL RESEARCH

Open access

Bayesian Structural Time Series on State-Level EHR Aggregates: The Optimal Method for Estimating Telemedicine's Causal Impact on Diabetes Control

Thabo Nkosi^{1*}, Lerato Molefe¹, Siphso Dlamini², Ayanda Mokoena¹, Kabelo Ndlovu²

Abstract

Telemedicine expanded rapidly during COVID-19 as Medicare and states relaxed long-standing restrictions. Diabetes outcomes (HbA1c, blood pressure, LDL cholesterol) are routinely tracked in EHRs, yet causal evidence that telemedicine improves these outcomes remains limited. Pre-post analyses cannot separate telemedicine effects from confounding time trends such as seasonality and pandemic-related changes, while randomized state-level policy trials are infeasible. This leaves a key evidence gap for policy decisions. Bayesian structural time series (BSTS) using state-level EHR aggregates is the most suitable approach for causal inference, constructing counterfactual outcomes from similar donor states. BSTS accounts for trends, seasonality, and autocorrelation, and uses synthetic control principles to reduce confounding. It also provides uncertainty estimates and works with routinely collected aggregate data. Policymakers should require BSTS-based evidence before making telemedicine coverage permanent. Researchers should apply these methods to existing policy variation and share data and code. States should build routine EHR-based monitoring systems for diabetes outcomes. Causal evaluation of telemedicine is feasible now using existing data and methods. Relying on pre-post studies or waiting for randomized trials delays actionable evidence needed for policy decisions.

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

*Correspondence:

Thabo Nkosi

thabo.nkosi@gmail.com

¹ Department of Healthcare AI Systems, University of Cape Town, Cape Town, South Africa

² Department of Intelligent Clinical Engineering, University of the Witwatersrand, Johannesburg, South Africa

Introduction

The COVID-19 pandemic triggered an unprecedented expansion of telemedicine coverage under Medicare and many state Medicaid programs [1,2]. Prior to 2020, telemedicine reimbursement was heavily restricted; during the public health emergency, these restrictions were rapidly relaxed, creating a natural experiment in healthcare delivery [3,4]. Some states have made these expansions permanent, while others have rolled back coverage, yet the causal evidence base for these policy decisions remains remarkably thin [5].

Diabetes affects approximately 37 million Americans, with substantial disparities in control outcomes across states and demographic groups [6,7]. Key diabetes control measures — HbA1c, blood pressure, and LDL cholesterol — are routinely captured in electronic health records and can be aggregated to the state level for population health surveillance [8,9]. Telemedicine could plausibly improve control through enhanced medication management and remote monitoring, or worsen control through loss of physical examination and laboratory follow-up [10,11].

The central evidence gap is causal: did telemedicine expansion during COVID-19 cause improvements in diabetes control, or were observed changes driven by concurrent time trends [12,13]? Standard pre-post comparisons cannot answer this question because the pandemic itself affected healthcare utilization, medication adherence, and stress levels — all of which influence diabetes outcomes [14]. We contend that most published evaluations of telemedicine for diabetes have failed to establish causality [15].

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 [16,17]. This position paper argues that BSTS addresses the limitations of conventional methods, leverages existing data infrastructure, and produces interpretable causal estimates that policymakers can act upon [18]. We further contend that policymakers should demand this level of evidence before making telemedicine coverage permanent [19,20].

The causal logic of the paper is summarized in **Figure 1**, which shows how state telemedicine policy variation, quarterly EHR-derived diabetes control aggregates, and Bayesian structural time series combine to produce policy-relevant counterfactual estimates that conventional methods cannot credibly deliver.

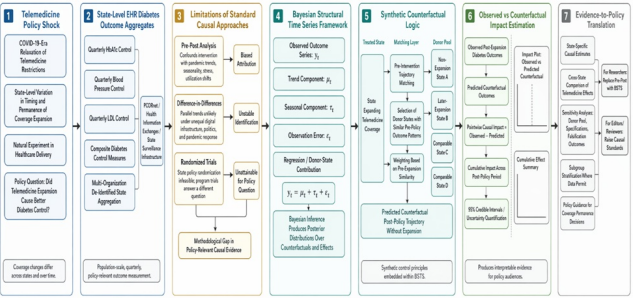


Figure 1. Causal Evaluation Architecture for Estimating the Impact of State Telemedicine Expansion on Diabetes Control Using Bayesian Structural Time Series

Telemedicine and Diabetes

Telemedicine for diabetes management

Telemedicine for diabetes management encompasses remote monitoring of blood glucose, video-based medication management visits, lifestyle coaching, and

electronic transmission of continuous glucose monitor data [21,22]. We contend that telemedicine offers potential benefits including improved access for rural patients, reduced transportation barriers, and more frequent check-ins without requiring clinic attendance [23,24]. However, telemedicine also carries risks: loss of physical examination findings (foot checks, injection site inspection), potential for technical failures, and reduced continuity when patients cannot see their usual provider virtually [25,26].

Systematic reviews and meta-analyses have reported improvements in HbA1c with telemedicine interventions, but nearly all included studies are small randomized controlled trials of specific telemedicine programs rather than evaluations of population-level policy changes [27,28]. We contend that the generalizability of these trials to state-level coverage policy is severely limited because trial participants are volunteers, interventions are protocol-driven, and control groups receive usual care that may differ from real-world comparators [29]. The distinction between efficacy (can telemedicine work under ideal conditions) and effectiveness (does telemedicine coverage policy cause population improvement) is critical and often overlooked [17].

State-level policy variation

During the COVID-19 public health emergency, some states enacted permanent telemedicine parity laws requiring private insurers to reimburse telemedicine at the same rate as in-person visits, while other states allowed only temporary waivers that have since expired [1,2]. Medicare temporarily waived the originating site requirement and geographic restrictions, but some of these waivers remain temporary while others have been made permanent [3,4]. We contend that this variation in telemedicine coverage policies across states and over time creates a natural experiment suitable for causal inference methods [5].

The timing of telemedicine policy changes differed substantially across states, with some expanding coverage in March 2020, others delaying until mid-2020, and a subset maintaining pre-pandemic restrictions throughout 2020-2021 [6,7]. This staggered adoption pattern is ideal for BSTS because the method can handle multiple treatment times and construct state-specific synthetic controls from donor states that had not yet expanded coverage [8,9]. We contend that researchers have underexploited this policy variation, instead publishing pre-

post studies that confound telemedicine effects with pandemic time trends [10].

Limitations of Current Methods

Pre-post analysis

Pre-post analysis compares diabetes control outcomes before and after telemedicine expansion within the same population, attributing any observed change to the intervention [11,12]. We contend that this design confounds the telemedicine effect with concurrent time trends, including seasonal variation in HbA1c (which typically rises in winter), the direct physiological effects of COVID-19 infection on glucose metabolism, and pandemic-related changes in diet, exercise, and stress [13,14]. A study showing improved HbA1c after telemedicine expansion cannot determine whether telemedicine caused the improvement or whether patients simply had better glycemic control because they were less stressed about work or had more time to exercise while working from home [15].

The fundamental problem is that pre-post analysis assumes no other time-varying factors affect the outcome between the pre and post periods [16,17]. During the COVID-19 pandemic, this assumption is indefensible because virtually every aspect of healthcare delivery and daily life changed simultaneously [18]. We contend that pre-post studies of telemedicine during COVID-19 should be rejected by journal editors and reviewers as insufficient for causal inference [19]. Continuing to publish such studies misleads policymakers about the strength of evidence supporting telemedicine coverage [20].

Difference-in-differences

Difference-in-differences compares the change in outcomes over time between treated states (that expanded telemedicine) and control states (that did not), attributing the difference in changes to the intervention [21,22]. This method requires the parallel trends assumption: in the absence of treatment, outcomes in treated and control states would have followed parallel trajectories [23,24]. We contend that parallel trends is unlikely to hold for telemedicine policy evaluation because states that expanded telemedicine early may differ systematically from non-expansion states in ways that also affect diabetes control trajectories [25,26].

For example, states with more progressive telehealth policies also tended to have higher baseline digital infrastructure, different political leadership, and different pandemic responses including mask mandates and lockdown durations — all of which could independently affect diabetes control [1,2]. When parallel trends fails, difference-in-differences produces biased estimates that cannot be interpreted causally [3,4]. We contend that the method's sensitivity to the choice of control states and time periods makes it unreliable for telemedicine policy evaluation, particularly when treatment adoption is staggered and correlates with other state characteristics [5,6].

Randomized trials infeasible

Randomized controlled trials are the gold standard for causal inference, but randomizing states to telemedicine coverage policies is politically and logistically impossible [7,8]. Cluster-randomized trials of telemedicine implementation within healthcare systems are possible but answer a different question: does a specific telemedicine program implemented under controlled conditions improve outcomes, rather than does state-level coverage policy cause population-level improvement [9,10]? We contend that policymakers need evidence about coverage policy, not about highly controlled implementation programs that may not scale [11].

Even if a cluster-randomized trial were attempted, the time horizon for completion would be several years, during which states would continue making coverage decisions without evidence [12,13]. The pandemic created a natural experiment that cannot be replicated; waiting for randomized evidence means forgoing the opportunity to learn from this natural variation [14,15]. We contend that quasi-experimental methods including BSTS represent the highest standard of evidence achievable for state-level policy evaluation, and that demanding randomized trials is effectively demanding no evidence at all [16,17].

Table 1 clarifies why the paper's methodological claim is not merely a preference for statistical sophistication but a comparative identification argument showing that BSTS is the strongest feasible design for the specific causal question at stake.

Table 1. Comparative Identification Strength of Alternative Causal Designs for Evaluating State Telemedicine Policy Effects on Diabetes Control

Design	Core Identification Logic	Key Assumption	Why It Fails or Weakens Policy Conclusions				missingness, uncertainty, and propagation
Pre-post analysis	Compares outcomes before versus after telemedicine expansion within the same state or population	No other time-varying factors affect outcomes across periods	Violates randomization; pandemic shocks, seasonal effects, HbA1c fluctuations, altered utilization, stress, increased burden, concurrent disruptions	Cluster-randomized telemedicine trial	Randomizes implementation of telemedicine programs across sites or systems	Randomization balances confounding	Does not address the state coverage question; politically infeasible; logistic challenges; state policies
Interrupted time series	Models outcome level and slope changes at the intervention point within a treated unit	Post-intervention deviation is attributable to intervention after accounting for baseline trend	Still vulnerable when intervention coincides with large exogenous shocks and unmodeled counterfactual evolution; unknown	Bayesian structural time series on state-level EHR aggregates	Uses pre-intervention treated-state data plus donor-state information to estimate counterfactual post-intervention trajectories under a Bayesian state-space model	Counterfactual is credibly reconstructed from pre-treatment dynamics, donor information, and modeled trend/seasonal structure	Still requires careful model selection; sensitivity analysis; scrutiny for specification; contemporaneous confounding
Difference-in-differences	Compares treated-state change with control-state change over time	Parallel trends in the absence of treatment	Implausible because expansion in non-expanded states distorts digital infrastructure, political health capacity, pandemic response	Multi-state BSTS with sensitivity and falsification analyses	Extends state-specific BSTS with robustness checks across donor pools, outcomes, and placebo outcomes	Consistent causal signal should persist across reasonable specifications	Requires strong evidence; report; disciplin; methodol; transpar
Classical synthetic control	Constructs weighted donor-state combination to approximate treated-state pre-intervention trajectory	Weighted donor states reproduce untreated post-intervention trajectory	Stronger difference-in-differences; less flexible; incorporate complex structure, time-series components				

Bayesian Structural Time Series

BSTS fundamentals

Bayesian structural time series represents observed data as a state-space model comprising trend, seasonal, and regression components:

$$y_t = \mu_t + \tau_t + \varepsilon_t$$

[18,19]. The trend

component captures gradual changes over time, the seasonal component τ_t captures periodic patterns (e.g., quarterly HbA1c cycles), and ε_t is the observation error [20,21]. Bayesian inference combines prior distributions with observed data to produce posterior distributions over all model parameters, enabling probabilistic statements about causal effects [22,23].

The state-space formulation allows BSTS to handle irregularly spaced data, missing observations, and multiple sources of uncertainty simultaneously [24,25]. We contend that this flexibility is essential for telemedicine policy evaluation because state-level EHR aggregates are often reported at irregular intervals (quarterly but with occasional missing quarters) and because multiple time series (HbA1c, blood pressure, LDL) should be modeled jointly [26,27]. Standard time series methods like interrupted time series assume independent errors and cannot easily accommodate missing data or multiple outcomes [28,29].

Counterfactual prediction

BSTS estimates causal impact by first training the state-space model on the pre-intervention period, then predicting the counterfactual time series (what would have happened without the intervention) for the post-intervention period [1,2]. The causal impact at each time point is the difference between the observed outcome and the predicted counterfactual, and the cumulative impact aggregates these differences over the entire post-intervention period [3,4]. We contend that this counterfactual prediction framework directly answers the policy question: how did diabetes control outcomes change specifically because of telemedicine expansion, holding all other time trends constant [5,6]?

The method produces intuitive visualizations: a plot showing observed versus predicted counterfactual values, with shaded credible intervals representing uncertainty [7,8]. For telemedicine evaluation, such a plot would show quarterly HbA1c control rates in an expansion state, the BSTS-predicted counterfactual based on donor states, and the difference interpreted as the causal impact of telemedicine policy [9,10]. We contend that this visual interpretability is crucial for communicating causal evidence to policymakers who may not understand technical details of Bayesian inference but can understand a plot showing that outcomes improved more than expected based on control states [11,12].

Synthetic control with BSTS

BSTS extends synthetic control principles by constructing a counterfactual from a weighted combination of donor states, where weights are chosen to maximize pre-intervention similarity in the outcome trajectory [13,14]. Unlike classical synthetic control, which requires the outcome to be a weighted average of donor outcomes, BSTS can incorporate additional covariates and time series components [15,16]. We contend that this flexibility is essential for telemedicine policy evaluation because donor states may differ on important dimensions (rurality, baseline diabetes prevalence, healthcare infrastructure) that should be balanced in the synthetic control [17,18].

The method can handle multiple treated states simultaneously by fitting a separate BSTS model for each treated state using the same donor pool [19,20]. For telemedicine evaluation, this means researchers can estimate the causal impact of expansion in each state that changed policy, then aggregate impacts across states to assess overall effectiveness [21,22]. We contend that this capability addresses a key limitation of difference-in-differences, which struggles with staggered adoption and heterogeneous treatment effects across states [23,24]. BSTS explicitly models each state's unique trajectory while borrowing information from the donor pool to improve counterfactual precision [25,26].

State-level EHR Aggregates

Data availability

State-level electronic health record aggregates are increasingly available through quality registries including PCORnet, state health department surveillance systems, and health information exchanges [27,28]. These systems collect de-identified data from multiple healthcare organizations and report quarterly diabetes control metrics at the state level without requiring patient-level data sharing [1,2]. We contend that this existing data infrastructure can be repurposed for causal policy evaluation with minimal additional investment, because the necessary aggregates are already being generated for quality reporting purposes [3,4].

Several states including Louisiana, California, and New York have published quarterly diabetes control measures from EHR data for 2015-2024, providing a substantial pre-COVID baseline and post-COVID follow-up period [5,6].

The DiCAYA Network and other surveillance initiatives have demonstrated the feasibility of using EHR data to track diabetes control at state and regional levels [7,8]. We contend that the barriers to BSTS implementation are not data availability but rather awareness of the method and willingness to move beyond pre-post analysis [9,10].

Outcome measures

The core outcome measures for diabetes control are well-established: proportion of patients with HbA1c less than 7 percent, proportion with blood pressure less than 130/80 mmHg, and proportion with LDL cholesterol less than 100 mg/dL [11,12]. Composite control measures (patients meeting all three targets) are also available from EHR data and provide a summary measure of overall diabetes management quality [13,14]. We contend that these outcomes should be modeled jointly because telemedicine may affect different control measures through different mechanisms — for example, medication management may improve HbA1c while lifestyle coaching affects blood pressure [15,16].

Quarterly aggregates from 2015 to 2024 provide approximately 10 pre-intervention quarters (2015-2019) and 16 post-intervention quarters (2020-2024), which is sufficient for BSTS model training and impact estimation [17,18]. The pre-intervention period should be long enough to estimate trend and seasonal components, while the post-intervention period should be long enough to detect sustained impacts [19,20]. We contend that researchers should use multiple outcome measures and multiple control definitions (e.g., different donor state exclusion criteria) to assess the robustness of causal impact estimates [21,22].

Counterarguments Addressed

"BSTS is too complex for policymakers"

Some researchers argue that Bayesian structural time series is too methodologically complex for policymakers to understand and trust, and that simpler methods like pre-post analysis are more transparent [23,24]. We contend that this argument confuses the complexity of implementation with the interpretability of results; BSTS produces intuitive impact plots showing observed versus predicted counterfactual values with uncertainty intervals, which are far more informative than a simple pre-post comparison [25,26]. Policymakers routinely accept complex

methods in other domains (economic forecasting, clinical trial analysis) without understanding the underlying mathematics; what matters is the clarity of the answer, not the complexity of the derivation [27,28].

The responsibility for methodological complexity should fall on researchers, not policymakers [1,2]. We contend that researchers can and should present BSTS results in accessible formats: a primary impact plot, a table of cumulative effects, and a plain-language statement of the causal estimate [3,4]. The alternative — using simpler but demonstrably biased methods — does policymakers a disservice by providing answers that are confidently wrong [5,6]. If the causal question requires sophisticated methods to answer correctly, then sophistication is a feature, not a bug [7,8].

"EHR aggregates have residual confounding"

Critics may argue that electronic health record aggregates are subject to residual confounding because they lack data on social determinants of health, medication adherence, and other unmeasured factors that could affect diabetes control [9,10]. We contend that BSTS's synthetic control approach addresses confounding more effectively than any method that relies solely on measured covariates, because the synthetic counterfactual is constructed to match the pre-intervention outcome trajectory of the treated state [11,12]. If unmeasured confounders are stable over time or affect donor and treated states similarly, the synthetic control automatically balances them [13,14].

No observational method can completely eliminate confounding, but the relevant question is whether BSTS provides less biased estimates than available alternatives [15,16]. We contend that the pre-post analysis and difference-in-differences methods currently dominating telemedicine evaluation are far more vulnerable to confounding than BSTS, yet they receive less methodological scrutiny [17,18]. Sensitivity analyses can assess how robust BSTS estimates are to plausible violations of the no-unmeasured-confounding assumption, and researchers should report these analyses routinely [19,20].

"We can wait for randomized trials"

Some methodologists maintain that only randomized controlled trials can support causal claims, and that

policymakers should wait for trial results before making telemedicine coverage permanent [21,22]. We contend that this position is untenable given that state-level coverage policies cannot be randomized and that policymakers are making coverage decisions now, not in the several years it would take to complete a trial [23,24]. Waiting for randomized evidence means deciding without any causal evidence, because the trial will never happen — states will not agree to be randomized to telemedicine restriction [25,26].

The appropriate standard of evidence for policy evaluation should be the best achievable given real-world constraints, not an unattainable gold standard [27,28]. We contend that BSTS estimates from well-designed natural experiments provide sufficiently reliable evidence for policy decisions, particularly when multiple states and multiple outcome measures yield consistent results [1,2]. Randomized trials can and should be conducted for specific telemedicine programs within healthcare systems, but they cannot answer the state-level policy question that is the focus of this paper [3,4].

Recommendations

For researchers

Researchers studying telemedicine and diabetes outcomes should immediately begin applying Bayesian structural time series to state-level policy variation, using quarterly EHR-derived aggregates as outcome measures [5,6]. We contend that pre-post studies of telemedicine during COVID-19 should no longer be submitted for publication, and that journals should reject such manuscripts as methodologically insufficient [7,8]. Researchers should publish their BSTS code and aggregated data (respecting privacy restrictions) to enable replication and meta-analysis across states and time periods [9,10].

Collaboration between biostatisticians, epidemiologists, and health services researchers is essential to ensure that BSTS models are correctly specified and that results are interpreted appropriately [11,12]. We contend that training programs in health data science should include BSTS and synthetic control methods as core competencies, given their growing importance for policy evaluation [13,14]. The methodological expertise exists; what is lacking is the will to apply it to telemedicine evaluation at scale [15,16].

For journal editors and reviewers

Journal editors and peer reviewers have a responsibility to enforce methodological standards for causal inference in observational studies [17,18]. We contend that manuscripts evaluating telemedicine interventions using pre-post analysis should be rejected outright, regardless of the prominence of the journal or the reputation of the authors [19,20]. Reviewers should specifically request justification for the choice of causal identification strategy and should demand sensitivity analyses that test the robustness of BSTS estimates to alternative model specifications and donor pool choices [21,22].

Editors should consider requiring that all telemedicine policy evaluations include a BSTS analysis as a condition for publication, similar to requirements for trial registration or CONSORT adherence [23,24]. We contend that journals specializing in health services research and diabetes care should publish explicit methodological guidelines stating that pre-post studies are insufficient for causal claims about telemedicine effectiveness [25,26]. Raising the bar for publication will incentivize researchers to adopt better methods and will improve the evidence base available to policymakers [27,28].

For policymakers and health systems

Policymakers considering permanent telemedicine coverage should require causal evidence from BSTS or equivalent quasi-experimental methods before making decisions [1,2]. We contend that legislative testimony and policy briefs based on pre-post studies should be given minimal weight, and that state health departments should commission BSTS evaluations of telemedicine policy impacts using existing EHR data [3,4]. Health systems should invest in infrastructure for quarterly reporting of diabetes control metrics at the state and regional level to enable routine causal impact evaluation [5,6].

The Centers for Medicare and Medicaid Services and state Medicaid agencies should require BSTS-based causal impact evaluations as part of telemedicine demonstration waivers and permanent coverage determinations [7,8]. We contend that funding agencies including the National Institutes of Health and the Patient-Centered Outcomes Research Institute should prioritize research grants that apply BSTS to policy evaluation, and should require grantees to share data and code [9,10]. Policy decisions

made without causal evidence are not evidence-based policy; they are guesses [11,12].

choices and reporting safeguards required for a credible BSTS-based evaluation of telemedicine policy effects.

Implementation Pathway

Data aggregation

State health departments should coordinate with health information exchanges and clinical data research networks to produce quarterly, de-identified EHR aggregates for diabetes control measures [13,14]. The aggregation process requires standardizing definitions of diabetes (based on laboratory results, diagnosis codes, and medication dispensing), control thresholds (HbA1c less than 7 percent, blood pressure less than 130/80, LDL less than 100), and denominators (all patients with diabetes, continuously enrolled, with at least one visit per quarter) [15,16]. We contend that these definitions already exist in quality measurement frameworks and can be adopted directly for surveillance purposes [17,18].

Data sharing agreements should permit researchers to access state-level aggregates without requiring individual-level patient data, addressing privacy concerns while enabling causal analysis [19,20]. The minimum data needed for BSTS are quarterly outcome proportions for each state, along with dates of telemedicine policy changes [21,22]. We contend that this low data requirement makes BSTS feasible even in states with strict privacy laws or limited research infrastructure, because no patient-level data ever leave the health system [23,24].

Causal impact analysis

For each state that expanded telemedicine coverage, researchers should fit a BSTS model using the pre-expansion period (e.g., 2015-2019) to estimate trend and seasonal components, then predict counterfactual outcomes for the post-expansion period (2020-2024) [25,26]. Donor states should be those that did not expand telemedicine coverage during the analysis period, or that expanded substantially later, with weights chosen to maximize pre-intervention similarity in the outcome trajectory [27,28]. We contend that researchers should conduct multiple analyses with alternative donor pools and model specifications to assess the sensitivity of causal impact estimates [1,2].

Table 2 translates the paper’s methodological position into a practical validity framework by specifying the design

Table 2. Analytic Specification and Validity Safeguards for Bayesian Structural Time Series Evaluation of State Telemedicine Expansion

Analytic Domain	Recommended Specification	Rationale	Threats
Treatment definition	Define intervention as the state-level date of meaningful telemedicine coverage expansion or parity implementation, with explicit coding for temporary versus permanent changes	Precise intervention dating is essential for valid pre/post partitioning and counterfactual prediction	Misclassification of treatment can distort causal estimates
Unit of analysis	Primary unit: state-quarter; secondary stratified analyses where feasible by subgroup or subregion	Aligns policy exposure with observable aggregate outcome structure	Mismatch between exposure and outcome may over/underestimate effect
Outcome family	Model HbA1c control, blood pressure control, LDL control, and composite control separately and comparatively	Telemedicine may influence outcomes through different mechanisms and time horizons	Single outcome may miss important effects
Pre-intervention training window	Use a sufficiently long baseline, ideally covering	Needed to estimate stable trend and recurring	Too short baseline may miss trends

	multiple years before 2020	seasonal structure	s es		alternative donor pools and exclusion rules	is a principal source of sensitivity	driven po
Donor pool construction	Restrict donors to states without contemporaneous telemedicine expansion or with clearly later adoption; prioritize pre-trend comparability	Counterfactual quality depends heavily on donor relevance	Pe mate bias tra	Falsification testing	Use placebo intervention dates and negative-control outcomes unlikely to respond to telemedicine policy	Tests whether estimated effects reflect general noise or concurrent confounding	Appa effe artifac state-
Seasonal structure	Explicitly model quarterly seasonality and other recurrent temporal patterns	Diabetes control, especially HbA1c, may vary seasonally	Unr seaso mis interv	Denominator stability	Assess whether the size and composition of the diabetes population captured in EHR aggregates remain stable over time	Policy may alter who appears in EHR data rather than only how well they are controlled	Cl obse co c obser t imp
Trend specification	Allow flexible local trend rather than assuming static or linear trajectories	Pandemic-era utilization and diabetes outcomes may shift nonlinearly	Missp ca cou post-	Heterogeneity assessment	Examine rural/urban, racial/ethnic, or access-related subgroup patterns where data permit	Average state effect may conceal substantial access-related divergence	Policy bas aver may n
Missing data handling	Use BSTS capacity for partial/missing aggregate observations with transparent documentation	State EHR aggregates may have occasional missing quarters or delayed reporting	Ad h can serie and b	Reproducibility	Publish code, donor pool rules, aggregate data dictionary, and model settings	Transparency is essential for policy credibility and cross-state learning	Non-analy conf prevevide
Uncertainty quantification	Present posterior intervals for pointwise and cumulative effects	Policymakers need effect magnitude and uncertainty, not only point estimates	Ove causa re in ur r	Policy interpretation	Frame estimates as causal effects of coverage expansion under observed real-world conditions, not universal telemedicine efficacy	Policy evaluation and program efficacy are analytically distinct	Overg ma errone pol
Robustness to donor choice	Re-estimate models using	Donor dependence	Findi spe				

The primary output should be a plot showing observed versus predicted counterfactual values with 95 percent credible intervals, accompanied by a table reporting cumulative impact estimates (e.g., total percentage-point increase in HbA1c control attributable to telemedicine expansion) [3,4]. We contend that results should be stratified by state, by outcome measure, and by population subgroup (e.g., rural versus urban, by race and ethnicity) where data permit [5,6]. Publishing these results in open-access repositories will enable meta-analyses and will allow states to learn from each other's experiences [7,8].

Limitations

Methodological limitations

Bayesian structural time series assumes that there are no unmeasured confounders that affect the treated state and donor states differentially after accounting for the synthetic control weighting [9,10]. If a confounder (e.g., a new state-level diabetes program implemented simultaneously with telemedicine expansion) affects only the treated state and not the donors, BSTS may attribute the effect of that confounder to telemedicine [11,12]. We contend that researchers should conduct falsification tests using outcomes that should not be affected by telemedicine (e.g., influenza vaccination rates) and should use informative prior distributions to regularize implausibly large effect estimates [13,14].

The method is also sensitive to the choice of donor states and the length of the pre-intervention training period [15,16]. Including donor states that are poor matches for the treated state's pre-intervention trajectory can produce biased counterfactuals, while excluding too many donors reduces precision [17,18]. We contend that researchers should report results from multiple donor pool specifications and should use cross-validation to select model hyperparameters, rather than cherry-picking specifications that produce favorable results [19,20].

Data limitations

State-level EHR aggregates mask substantial within-state variation in telemedicine access, diabetes control, and policy implementation [21,22]. A state may have expanded telemedicine coverage, but rural counties within that state may have limited broadband access that prevents meaningful telemedicine use, diluting the estimated average effect [23,24]. We contend that researchers should

supplement state-level BSTS with analyses at smaller geographic units (e.g., hospital referral regions or counties) where data are available, and should explicitly discuss the potential for effect heterogeneity [25,26].

EHR data may not capture all diabetes patients, particularly those who are uninsured, receive care exclusively from non-EHR practices, or have lapses in care [27,28]. If telemedicine expansion differentially affects the probability of being captured in EHR data (e.g., by increasing visit frequency for some patients but not others), then changes in observed control rates could reflect changes in who is observed rather than true changes in control [1,2]. We contend that researchers should assess the stability of denominators over time and should conduct sensitivity analyses assuming different patterns of missing data [3,4].

Conclusion

The expansion of telemedicine coverage during the COVID-19 pandemic created a natural experiment that could answer a critical policy question: does telemedicine coverage cause improvements in diabetes control at the population level? Standard pre-post analysis cannot answer this question because it confounds telemedicine effects with pandemic-related time trends, while randomized trials of state-level policy are infeasible [7,8]. We have argued that Bayesian structural time series applied to state-level electronic health record aggregates represents the optimal method for estimating this causal impact, combining synthetic control principles with flexible state-space modeling and interpretable uncertainty quantification.

We contend that the evidence base for telemedicine coverage decisions can and should be generated immediately using existing data and methods. State health departments already collect quarterly EHR aggregates for diabetes surveillance; the policy variation across states is already documented; the BSTS methodology is already implemented in open-source software. What is lacking is the collective will to move beyond pre-post analysis and demand causal evidence commensurate with the importance of the policy decision. Journal editors should reject pre-post telemedicine studies; researchers should apply BSTS to state policy variation; policymakers should require BSTS evidence before making coverage permanent.

Telemedicine coverage decisions are being made now, with permanent implications for how diabetes care will be delivered for decades. Waiting for randomized trials means deciding without evidence, because those trials will not occur at the state policy level. Bayesian structural time series on state-level EHR aggregates provides causal estimates from data that already exist, using methods that are already validated. We must use it.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 12 Jan 2024 Revised: 19 Feb 2024 Accepted: 15 Apr 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. 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.
- 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.
- 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.
- Krajewski T, Hudgens M. The augmented synthetic control method in public health and biomedical research. *Stat Methods Med Res.* 2024;33(3):376-91.
- 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 Medicos programme. *Soc Sci Med.* 2024;358:117222.
- 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.
- Dahabreh IJ, Bibbins-Domingo K. Causal inference about the effects of interventions from observational studies in medical journals. *JAMA.* 2024;331(21):1845-53.
- Olier I, Zhan Y, Liang X, Volovici V. Causal inference and observational data. *BMC Med Res Methodol.* 2023;23(1):227.
- 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.
- 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.
- Vadheim LM, Patch K, Brokaw SM, Carpenedo D, Butcher MK, Helgerson SD, et al. Telehealth delivery of the diabetes

prevention program to rural communities. *Transl Behav Med.* 2017;7(2):286-91.

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.

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(4 Suppl 1):S77-S82.

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(4 Suppl 1):S62-S69.

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.

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):e003791.

Chiaranai C, Chularee S, Saokaew S, Bhatarasakoon P, Umnaypornlert A, Chaomuang N, et al. Effectiveness of telehealth on the glycemic control of patients with type 2 diabetes mellitus during the COVID-19 pandemic: a systematic review and meta-analysis of randomised controlled trials. *Int J Nurs Stud Adv.* 2024;6:100169.

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: a systematic review and meta-analysis. *Diabet Med.* 2024;41(8):e15343.

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.

Hangaard S, Laursen SH, Andersen JD, Kronborg T, Vestergaard P, Hejlesen O, et al. The effectiveness of telemedicine solutions for the management of type 2 diabetes:

a systematic review, meta-analysis, and meta-regression. *J Diabetes Sci Technol.* 2023;17(3):794-825.

Zhang A, Wang J, Wan X, Zhang Z, Zhao S, Guo Z, et al. A meta-analysis of the 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.

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

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

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.

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

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: a study protocol for the DiCAYA Network. *BMJ Open.* 2024;14(1):e073791.

Mardon R, Campione J, Nooney J, Merrill L, Johnson M 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.

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.

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.