From Sparse Data to Smart Decisions: Region-Specific Policy Evaluation via Simulation

Carson Dudley, Daniel Bergman, Harsh Jain, Kerri-Ann Norton, Erica Rutter, Marisa Eisenberg, Trachette Jackson

February 9, 2026

Abstract

Effective disease outbreak response requires actionable, region-specific guidance, but most modeling tools rely on detailed surveillance or strong assumptions, such as random mixing. Agent-based models (ABMs) allow us to capture key heterogeneity in contact patterns and intervention mechanisms, but linking these models with data is often computationally intractable, particularly at the larger scales needed for decision-making (e.g., county- or state-level). We present a simulation-based framework that combines ABMs with surrogate modeling to infer key transmission and severity parameters using only routine case and hospitalization data. This enables local health agencies to evaluate candidate interventions while explicitly accounting for uncertainty. Applied to COVID-19 in Michigan counties, our method recovers core parameters (transmissibility, latent period, asymptomatic transmissibility, underreporting, hospitalization risk, and duration) that align with empirical estimates, while demonstrating regional variation linked to age and comorbidity patterns. We find that intervention effectiveness cannot be reliably predicted from simple demographic predictors such as age structure, population density, or workforce participation. While school closures aligned with child population in some settings, other interventions showed weak or counterintuitive relationships with demographics. Traditional ODE models with random mixing assumptions cannot capture how interventions target specific contact networks, making it impossible to assess whether demographic proxies predict intervention success. Our framework addresses this gap by explicitly modeling intervention mechanisms within heterogeneous contact structures. Solely using routine case and hospitalization data, our method enables practical, uncertainty-aware decision support for local health agencies facing COVID-19, influenza, RSV, or future novel pathogens.

1 Significance Statement

During emerging outbreaks, policymakers must act before data are complete. We introduce a computational framework that bridges the gap between sparse local data and agent-based models, enabling evaluation of interventions under significant uncertainty. Our method captures how specific contact networks drive local transmission by moving beyond the random-mixing assumptions of compartmental models. We apply our framework to COVID-19 in diverse Michigan counties. Our method identifies which intervention policies are robust to the inherent uncertainty in early outbreaks for each specific community. This disease-agnostic approach empowers local agencies to make evidence-based decisions tailored to their unique population structures and contact patterns.

2 Introduction

During an emerging infectious disease outbreak, local health officials often face urgent policy decisions–whether to close schools, restrict workplaces, or advise isolation–but lack timely, region-specific modeling tools to guide those choices [1]. While national or global estimates of parameters related to disease transmission and severity may be available, they often fail to capture local variation in demographics, comorbidities, and healthcare capacity that drive real-world outcomes [2, 3]. In the early stages of outbreaks, data remain sparse, uncertainty is high, and public health agencies rarely have the time, expertise, or surveillance infrastructure to build detailed models [4]. Yet policy decisions cannot wait, underscoring the need for fast,

transparent, and locality-aware decision support. Without reliable regional estimates, resource allocation and intervention planning risk being poorly matched to actual needs [5].

A common approach to parameter estimation involves fitting simplified, differential equation (SIR-type) models, such as systems of ordinary differential equations (ODEs), to surveillance data [6, 7]. These models are computationally efficient and analytically tractable, making them useful for rapid assessment and forecasting. However, their ability to capture individual-level heterogeneity, stochasticity, and network or spatial structure is limited. Agent-based models (ABMs) offer a complementary approach by explicitly simulating the behavior of individuals within a population, allowing for more granular and mechanistic representations of transmission and disease progression [8, 9]. The primary drawback of ABMs is their computational cost, which makes parameter inference and uncertainty quantification difficult, particularly when fitting to real-world data [10]. Recent efforts have sought to address this challenge by borrowing optimization techniques from machine learning, allowing automated parameter tuning through repeated simulations [11]. These methods show promise for improving computational efficiency, but they often require specialized technical infrastructure that may be difficult to implement rapidly during emerging outbreaks when time and resources are constrained. In this study, we apply a surrogate-informed inference framework that reconciles the tractability of ODE models with the mechanistic fidelity of ABMs, enabling data-driven ABM parameter estimation and uncertainty quantification at speeds suitable for early-stage policy evaluation.

Our method leverages the recently introduced SMoRe ParS (Surrogate Modeling for Reconstructing Parameter Surfaces) approach [12, 13], which uses simplified differential equation models that preserve salient mechanisms to act as calibrated intermediaries between observed data and complex simulations. Specifically, we fit the same ODE surrogate model to both real-world surveillance data and to synthetic outputs generated by running the ABM across many parameter combinations. By comparing these two sets of ODE fits, we can identify which ABM configurations produce dynamics consistent with observed case and hospitalization trajectories. This approach enables rapid exploration of the full range of ABM parameter regimes consistent with local data, while preserving the mechanistic interpretability and high-resolution contact structure of the original agent-based simulations. We use SMoRe ParS because it has been successful in other domains where there exist both complex agent-based models that characterize system behavior as well as established, ODE-based equations that describe mean-field behavior well, similar to epidemiology [12, 13].

This approach enables uncertainty quantification that extends beyond the fixed structural assumptions of traditional ODE models. Because the ABM explicitly represents contact networks, age-structured mixing, and individual-level heterogeneity, the framework can capture trade-offs and compensatory dynamics (e.g., between contact patterns and disease transmissibility, or between age structure and hospitalization risk) that are embedded in the ABM structure but absent from simplified compartmental models. This allows decision-makers to evaluate interventions while accounting for uncertainty in both parameters and the underlying mechanisms of transmission.

We demonstrate the utility of this framework through a case study of early COVID-19 transmission in nine Michigan counties. Jointly fitting models to reported case and hospitalization time series, we infer a range of epidemiologically critical ABM parameters: symptomatic and asymptomatic transmissibility, latent periods, asymptomatic fractions, hospitalization probabilities and durations, and underreporting rates. These estimates align closely with independent empirical studies [14, 15], while showing region-specific variation driven by age structure and comorbidity prevalence. A schematic overview of the framework is shown in Figure 1, illustrating how county-level surveillance data is integrated with agent-based simulations and surrogate modeling to identify parameter regimes and evaluate interventions.

By embedding mechanistic ABMs into a tractable, data-driven inference pipeline, our framework empowers local policymakers to reason quantitatively about "known unknowns" and make context-specific decisions, even when data are sparse or rapidly evolving. The approach generalizes to other respiratory pathogens and broader domains requiring localized, uncertainty-aware policy planning.

3 Results

ABM Reproduces Observed Epidemic Dynamics

To simulate county-level outbreak dynamics under real-world conditions, we used an agent-based model (ABM) implemented in the Framework for Reconstructing Epidemiological Dynamics (FRED) simulation

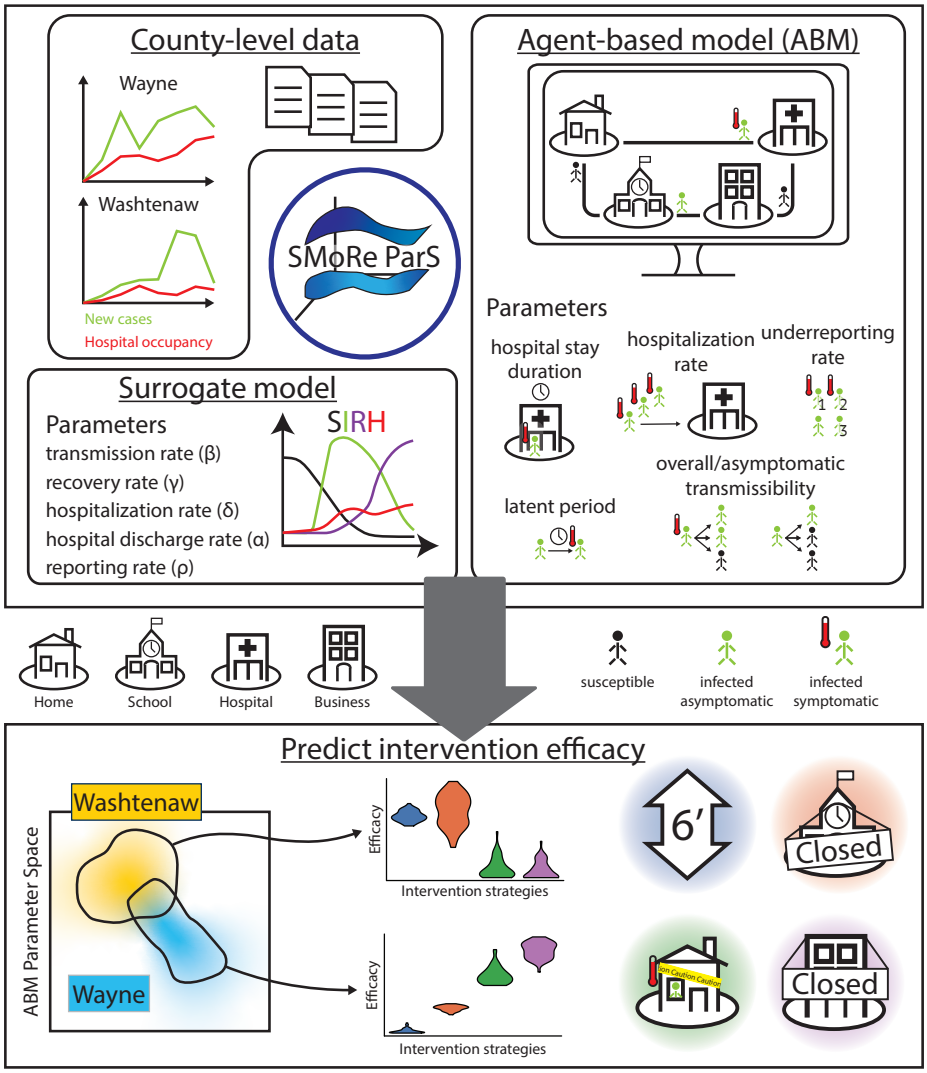

Figure 1: Schematic illustrating how SMoRe ParS integrates county-level data with an agent-based model and surrogate model to make predictions about specific interventions within counties. The four interventions are social distancing (blue), school closure (red), stay-home-if-symptomatic (green), and non-essential work-closure (purple).

platform (see Methods: Agent-Based Simulations) [16]. The model explicitly represents individuals and their interactions across home, school, work, and community settings, capturing demographic and behavioral heterogeneity relevant to local transmission dynamics. We focused our main analysis on three demographically distinct Michigan counties: Washtenaw County (younger, university-centered), Wayne County (densely populated urban center with high comorbidity burden), and Ontonagon County (rural with an aging population and limited local healthcare capacity). These counties were selected to span a wide range of demographic and infrastructural settings. To evaluate whether key conclusions, such as the relative effectiveness of specific interventions, held consistently within similar contexts, we also applied the framework to six additional counties approximately matched on structural features (e.g., urbanicity, age distribution, and healthcare access). Results for these are included in the Supplementary Information.

Using the SMoRe ParS framework, we first calibrated a simplified, ODE-based, SIR-type surrogate model to observed COVID-19 case and hospitalization data on a county-by-county basis [17]. We then pulled back these data-constrained surrogate fits to infer corresponding county-specific ABM parameter regions. This

surrogate-informed mapping enabled us to efficiently identify parameter sets which reproduce the observed epidemic trajectory without directly fitting the full ABM to data.

Figure 2 displays the results of calibrated ABM simulations in the three representative Michigan counties, obtained by averaging ten stochastic runs using parameter vectors sampled from the surrogate-constrained regions. The simulated epidemic time-courses (blue lines) closely track the reported case trajectories (black dots), accurately capturing both the rise and fall of the COVID-19 wave from late August to December 2020. This fit demonstrates that the surrogate-guided inference approach can successfully recover ABM configurations that explain observed dynamics without overfitting transient noise. We further validate our ABM fits by assessing the fit on a held-out data window in the appendix. Table 1 lists the inferred parameter ranges for each county, along with the specific parameter values used to generate the representative ABM simulations.

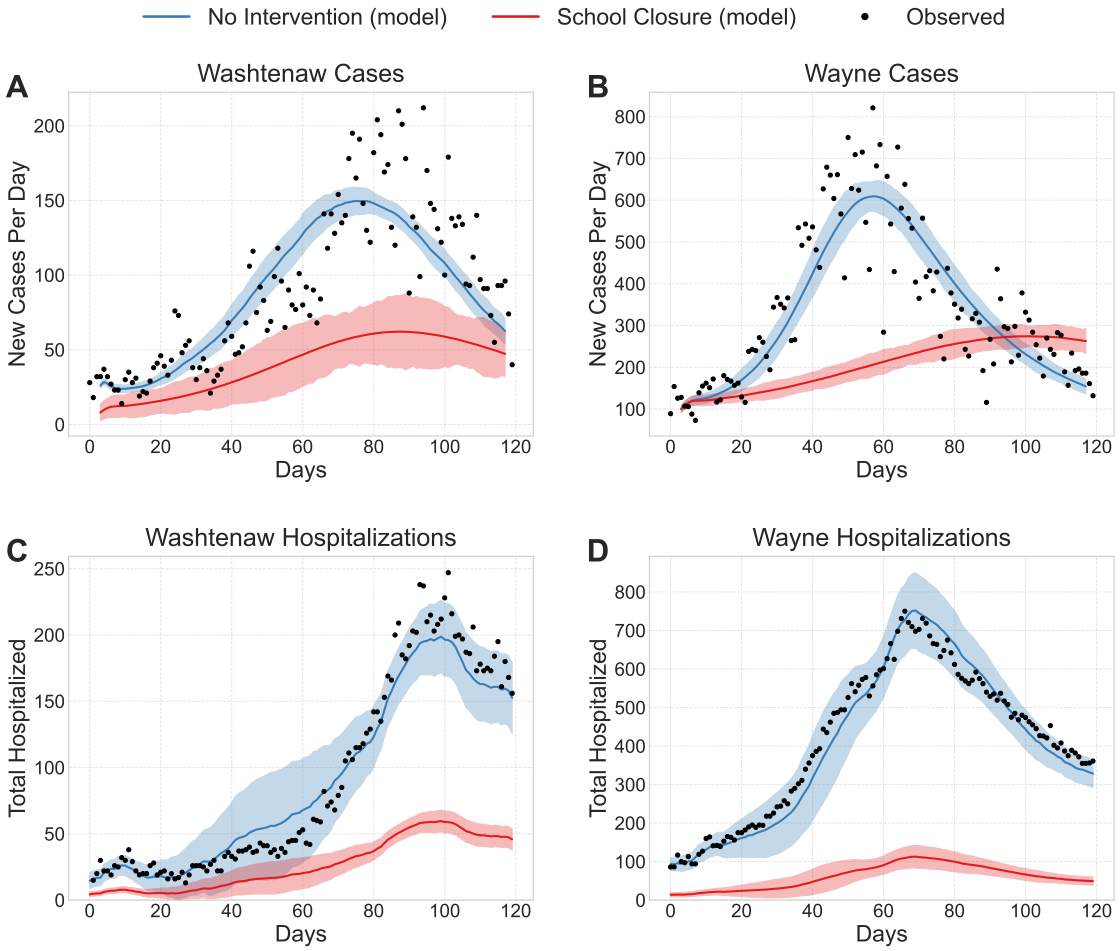

Figure 2: Calibration of the agent-based model (ABM) to COVID-19 cases (top row) and hospitalizations (bottom row) in Washtenaw (left) and Wayne (right) counties. Black dots represent observed data, with values fewer than 5 censored to protect patient confidentiality. Blue curves show ABM output averaged across ten simulation runs with no interventions, and red curves show ABM output under a school closure scenario. Shaded areas represent ± 1 SD across simulations with parameters varied within the confidence regions. Ontonagon County is not shown because case and hospitalization counts were so low that nearly all values were censored, particularly for hospitalizations.

ABM Parameter Estimates Align with Empirical Clinical Data and Demographics

For each county, we inferred the subset of ABM parameter combinations whose surrogate model projections matched the observed case and hospitalization time series. This calibration step yielded joint estimates of key epidemiological parameters, including transmissibility, latent period, asymptomatic fraction, hospitalization risk and duration, and reporting rates. Confidence intervals were derived by selecting ABM parameters whose surrogate projections fell within the 95% bounds of the fitted surrogate model (see Methods: Parameter Surface Reconstruction). By using these data-consistent parameter regions, rather than point estimates, we can evaluate interventions in a way that is more robust to the inherent non-identifiability of complex epidemic models.

The SMoRe ParS framework identified county-specific epidemiological parameter regions that reflect meaningful differences in reporting and severity risk (see Table 1). Notably, hospitalization probabilities aligned with expected demographic trends: Washtenaw County exhibited the lowest estimated hospitalization probability at 4.9%, consistent with its younger population. Wayne County's estimate was higher (7.0%), reflecting known disparities in chronic disease burden. Ontonagon County had the highest hospitalization probability at 9.2%, consistent with its older age distribution. These recovered risk profiles are consistent with age-stratified severity estimates established early in the pandemic [18].

These differences emerged despite limited and potentially incomplete surveillance data. In particular, Ontonagon County's small size and absence of a full-service hospital likely resulted in underreporting of hospitalizations, as patients were transported to neighboring counties. By jointly fitting both case and hospitalization time series and estimating reporting rates for each stream, the framework was able to recover plausible hospitalization risk estimates even in this sparse-data setting.

Other parameters, such as latent period duration, transmissibility, and hospitalization probability and duration, also varied by county. For example, Ontonagon County showed a longer hospitalization duration, potentially reflecting age effects, along with lower overall transmissibility, consistent with evidence that lower population density is associated with reduced transmission potential via reduced contact rates [19]. The recovered latent periods (3.7–5.85 days) align with clinical observations of the SARS-CoV-2 incubation period [20, 21]. Interestingly, asymptomatic transmissibility was consistent across all counties, despite large differences in age structure and comorbidity burden. This suggests that, unlike hospitalization risk or contact rates, asymptomatic infectiousness may be governed more by biological or pathogen-level factors than by local demographic composition.

Table 1: ABM Parameter estimates recovered by the SMoRe ParS framework for Washtenaw, Wayne, and Ontonagon Counties. Values represent point estimates with 95% confidence intervals derived from model calibration to case and hospitalization data.

ABM Parameter	**Washtenaw**	**Wayne**	**Ontonagon**
Hospitalization Probability	4.9% [3.3, 5.9]	7.0% [6.6, 9.4]	9.2% [3.8, 9.2]
Hospitalization Duration (days)	5.4 [3.5, 14.0]	8.6 [3.5, 12.1]	10.9 [10.5, 11.3]
Latent Period (days)	3.7 [3.6, 4.0]	5.85 [5.6, 6.02]	4.22 [3.87, 4.53]
Transmissibility	1.73 [1.64, 1.87]	1.81 [1.68, 1.98]	1.50 [1.43, 1.65]
Asymptomatic Transmissibility	38.0% [37.1, 42.3]	37.5% [36.9, 39.6]	38.5% [37.5, 39.8]
Case Reporting Rate	14.0% [10.1, 17.1]	9.1% [5.9, 13.9]	26.3% [21.2, 30.2]
Hospitalization Reporting Rate	82.0% [75.3, 84.2]	71.3% [65.3, 78.1]	10.0% [10.0, 11.3]

Translating Uncertainty into Intervention Analysis

For decision-making, the critical challenge is not just estimating parameters, but understanding how uncertainties and trade-offs between them shape policy outcomes. Our framework demonstrates these trade-offs by revealing the structure of parameter dependencies that arise from the calibration process. For example, the observed case trajectory can be reproduced by balancing force of infection against case reporting rate: higher transmission with lower detection yields similar observed cases versus lower transmission with higher

detection. By propagating this uncertainty into intervention evaluation, our framework reflects data-driven uncertainty.

To visualize these dependencies, we applied kernel density estimation to the top-fitting parameter sets from the surrogate model. Figure 3 displays contour plots for four surrogate model parameter relationships: (A) surrogate case reporting rate vs. force of infection, (B) hospitalization rate vs. discharge rate, (C) hospitalization rate vs. surrogate hospitalization reporting rate, and (D) basic reproduction number (a measure of transmissibility) vs. surrogate case reporting rate. The strength of parameter dependencies varies. Panels A and D have ridge-like structures indicating compensatory behavior: higher transmissibility can be offset by lower reporting rates to produce similar observed case trajectories. Panels B and C show weaker correlations, suggesting these parameter pairs are more independently constrained by the data.

The data strongly constrain some parameter combinations with weaker identifiability for others. In addition to data fitting, parameter tradeoffs can impact intervention efficacy analysis. Parameters that are not very compensatory for reproducing the observed dynamics may still interact in determining intervention effects. Visualizing the full structure of parameter dependencies underscores why point estimates are insufficient and motivates propagating complete parameter distributions into intervention simulations.

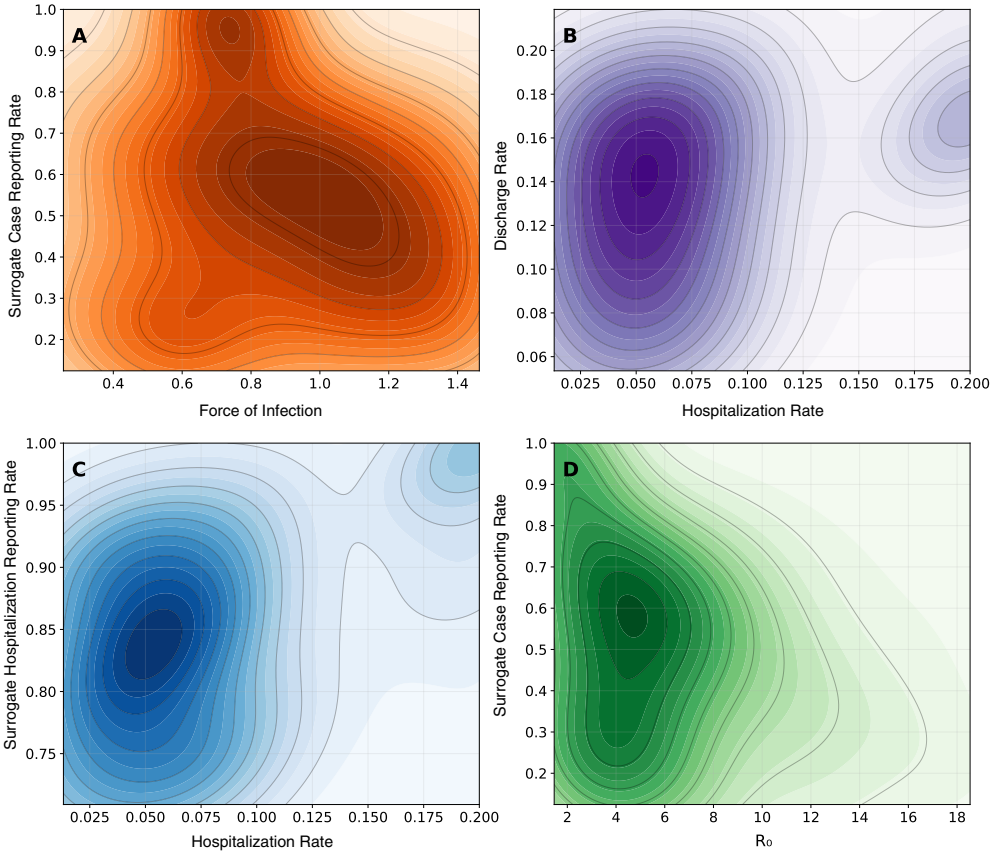

Figure 3: Parameter dependencies in the surrogate model. Contour plots show kernel density estimates over the top-fitting parameter sets for Washtenaw County. (A) Force of infection vs. surrogate case reporting rate shows a clear compensatory trade-off, where higher transmission can be offset by lower reporting. (B) Hospitalization probability vs. discharge rate (1/duration). (C) Surrogate hospitalization rate vs. surrogate hospitalization reporting rate. (D) Basic reproduction number (R_0) vs. surrogate case reporting rate also exhibits a negative relationship. The strength of these dependencies varies across parameter pairs, reflecting different levels of identifiability from the observed data. These parameter uncertainties and correlations are explicitly propagated in downstream intervention analyses.

Robust and Fragile Interventions under Uncertainty

Having calibrated the models for each county, we next evaluated the robustness of candidate non-pharmaceutical interventions (NPIs) by simulating each intervention across the full distribution of ABM parameter combinations identified as consistent with local surveillance data. The interventions considered included school closure, symptomatic isolation messaging ("stay home if sick"), non-essential workplace closure, and general social distancing.

To reflect the range of epidemic conditions a decision-maker might face, we sampled from the set of ABM parameter combinations whose surrogate projections fell within the 95% confidence region derived from the observed data. Then, we applied each intervention to these simulated outbreaks. This approach allowed us to capture variation in intervention effectiveness arising from uncertainty in transmission dynamics, disease severity, and underreporting.

Figure 4a-d summarizes the distribution of reductions in peak hospitalizations achieved by each intervention across counties and parameter regimes. Each violin represents the range of outcomes generated by varying underlying ABM parameters within the data-consistent region, illustrating both the median effectiveness and the uncertainty associated with each strategy. Figure 4f summarizes robustness, measured as the interquartile range (IQR) of peak reductions in hospitalizations. Interventions with smaller IQRs are more consistent across uncertainty, while larger IQRs reflect greater variability.

School Closure Impact Varies with Local Age Distribution

School closures consistently produced large and stable reductions in peak hospitalizations in Wayne and Washtenaw Counties (Figure 4a), with median reductions of approximately 85% and 70%, respectively. The effect was robust across parameter regimes and severity scenarios, reflecting the central role that school-based contact networks play in transmission in these counties, both of which have substantial student populations and higher population density overall.

In contrast, Ontonagon County saw much smaller and more variable reductions, with a median effect around 8% and wide stochastic spread. This result is consistent with expectations: Ontonagon has an aging population and relatively few school-aged children, so school closures remove a smaller share of transmission opportunities. In such low-density, older populations, the impact of school-based interventions is structurally constrained.

These patterns highlight the importance of tailoring intervention strategies to local demographic and contact structures. Where schools serve as major hubs of transmission, school closures can be a highly effective and reliable early intervention for slowing initial outbreak growth or flattening epidemic peaks. However, it is important to acknowledge that prolonged school closures may not be practical or sustainable throughout an extended outbreak or pandemic, given their impacts on education, childcare, and family economic stability. School closures also tend to be more socially and politically contentious than other interventions, requiring consideration of community values and trade-offs. In older or rural populations where schools contribute less to overall transmission, other strategies may be more impactful and face fewer implementation barriers.

Symptom-Based Interventions Are Vulnerable to Asymptomatic Spread

Stay-home-if-symptomatic policies reduced peak burden in many scenarios, but their effectiveness varied substantially with the proportion of asymptomatic infections (Figure 4b). In Washtenaw County, peak hospitalization reductions ranged from approximately 18% to 68% depending on asymptomatic spread assumptions, while Ontonagon County saw even wider variation and lower median effectiveness. The failure to capture asymptomatic transmitters limits this intervention's robustness under uncertainty. Policymakers should exercise caution when relying on symptom-based isolation alone, as performance can degrade sharply if asymptomatic transmission is more prevalent than anticipated (Figure 4e).

Workplace Closures Have Modest and Variable Effects Across Settings

Non-essential work closures led to modest reductions in peak hospitalizations across counties (Figure 4c). In Wayne County, the intervention produced consistent but small benefits, with peak reductions centered

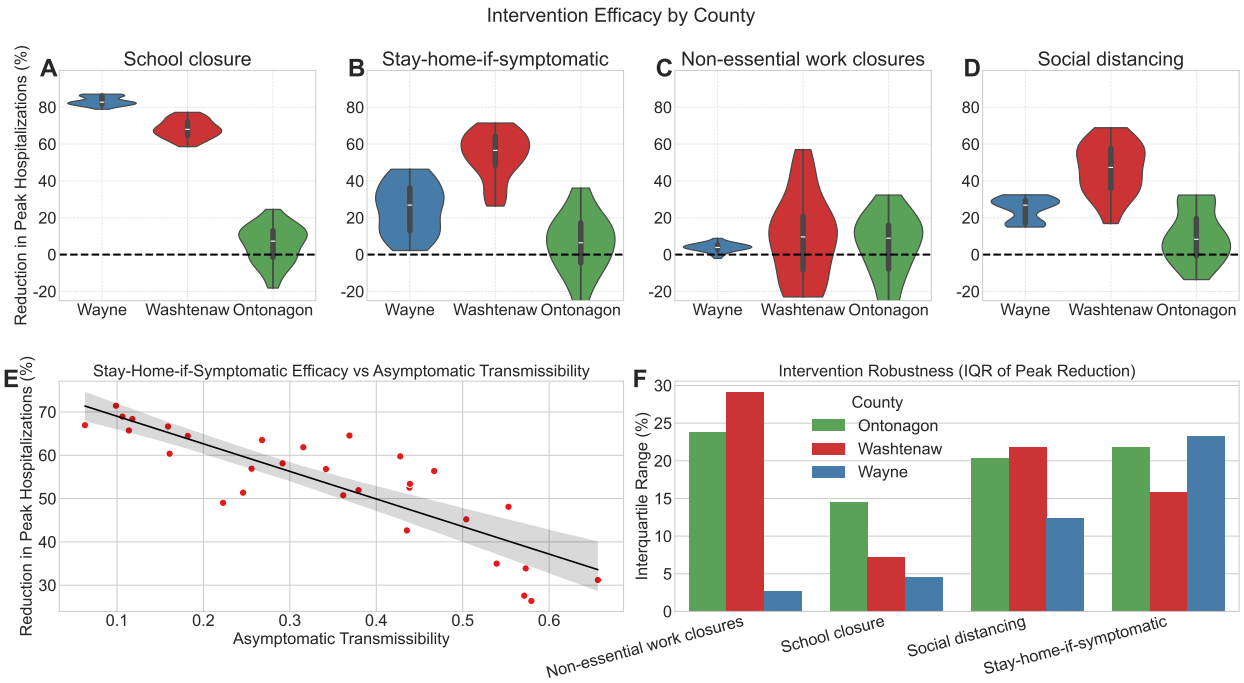

Figure 4: **Intervention effectiveness across counties under uncertainty.** Each panel shows the distribution of reductions in peak hospitalizations (%) for a single intervention, simulated across all data-consistent ABM parameter combinations in three Michigan counties. **(A)** K-12 school closures produced the largest and most consistent reductions in peak burden in Wayne and Washtenaw Counties (medians: ∼85% and ∼70%, respectively), with more modest and variable effects in Ontonagon (∼10%). **(B)** Stay-home-if-symptomatic policies were effective in Washtenaw (median ∼60%) but showed reduced and more variable impact, reflecting sensitivity to asymptomatic transmission. **(C)** Non-essential work closures yielded modest and inconsistent effects in all counties, with median reductions of 5-10% and high variance in Washtenaw and Ontonagon counties. **(D)** General social distancing produced moderate reductions in Washtenaw and Wayne, but limited and highly variable impact in Ontonagon. Shaded violins represent the distribution of outcomes across parameter regimes; negative values reflect stochastic variability in low-effect-size scenarios. **(E)** Effectiveness of stay-home-if-symptomatic interventions declines as the asymptomatic proportion increases, as shown for Washtenaw County. Each point represents a simulation under a different asymptomatic transmission assumption, and the trend highlights the structural vulnerability of symptom-based isolation to undetected spread. **(F)** Robustness of each intervention, measured as the interquartile range (IQR) of peak hospitalization reductions across retained parameter sets. Lower IQR indicates more stable performance across uncertainty. School closures consistently show low IQRs (high robustness), while symptom-based isolation and social distancing exhibit greater variance.

around 4%. In Washtenaw and Ontonagon Counties, outcomes were more variable, spanning from moderate reductions to small apparent increases in peak burden (negative y-values in Figure 4a-d).

These occasional negative estimates do not indicate that workplace closures were harmful; rather, they reflect the combination of small overall effect size and stochastic variability inherent in agent-based simulations. When effects are modest, random variation in outbreak timing or seeding can dominate the signal in individual runs.

The relative lack of variability in predicted hospitalization reductions in Wayne County likely reflects its large population size, which buffers against stochastic fluctuations and yields smoother aggregate dynamics. In smaller or more heterogeneous settings, such as Washtenaw and Ontonagon, workplace closures yield more variable outcomes, despite using the same intervention assumptions. Overall, workplace closures may contribute marginally to outbreak mitigation, but are unlikely to serve as reliable standalone interventions,

especially in settings where work-based contacts do not dominate transmission.

General Distancing Is Highly Sensitive to Transmission Dynamics

General social distancing produced variable reductions in peak hospitalizations across counties and parameter regimes (Figure 4d). Median effects ranged from 8–47%, with the highest reductions seen in urban counties like Wayne. In contrast, reductions were smaller and more variable in Ontonagon, where overall transmission is lower and social mixing is more limited.

Importantly, compliance assumptions were held constant across simulations. The variability in outcomes instead reflects sensitivity to underlying transmission characteristics—such as asymptomatic transmissibility and baseline contact rates. When asymptomatic spread is high, for example, social distancing among symptomatic individuals alone may fail to contain transmission.

While general distancing can meaningfully reduce burden under favorable conditions, its impact is highly parameter-dependent. It may be most effective when paired with other targeted interventions or in settings where broad behavioral change is achievable and well-aligned with transmission pathways.

County Demographics Do Not Reliably Predict NPI Effectiveness

To assess whether basic county-level demographic characteristics could explain the variation in intervention effectiveness observed across counties, we compared the median reduction in peak hospitalizations for each intervention to a demographic feature related to that intervention (Figure 5). For example, school closure effects were compared to the proportion of children under 18, workplace closure effects were compared to the labor force participation rate.

Among the four interventions examined, only school closures and workplace closures showed even moderate associations with their corresponding demographic predictors (child population share and labor force participation rate). School closures were more effective in counties with a higher proportion of children under 18 ($R^2 = 0.69$), a trend that is both intuitive and expected given the role of schools in facilitating youth transmission. Workplace closures, however, showed a surprising negative association with labor force participation rate ($R^2 = 0.61$), suggesting they were more effective in counties where fewer working-age adults were employed. This is unlikely to be a causal relationship; more plausibly, it reflects that workplace closure efficacy varies due to unmeasured structural or behavioral features not captured by basic demographics.

In contrast, the other interventions showed weak or no relationship with the chosen demographic predictors. Symptom-based isolation showed a slight negative trend with the proportion of residents over 65 ($R^2 = 0.22$), and social distancing showed no relationship at all with population density ($R^2 = 0.00$), despite expectations that higher-density areas might benefit more from reduced contacts.

Together, these findings underscore that county demographics alone are poor predictors of which interventions will be most effective in a given setting. Instead, they highlight the importance of a simulation-based approach that captures dynamic interactions between demographic structure, contact patterns, and intervention mechanisms under uncertainty, enabling more accurate and context-specific policy guidance.

4 Discussion

Infectious disease outbreaks often unfold under substantial uncertainty about key epidemiological parameters, particularly during the early stages when data are limited and rapidly evolving. This uncertainty complicates both scientific understanding and public health decision-making. Our results demonstrate that combining agent-based modeling with surrogate-informed parameter inference enables joint estimation of region-specific disease parameters and systematic evaluation of intervention strategies under uncertainty.

Applied to COVID-19 in demographically distinct counties, the framework successfully recovered plausible values for critical transmission and severity parameters, such as latent period, asymptomatic transmissibility, hospitalization probability, and hospitalization duration, using only routine case and hospitalization data. These inferred values aligned with the scientific literature while also exhibiting interpretable spatial variation. For instance, Ontonagon County's elevated hospitalization probability reflected its older population and limited local hospital capacity, while Wayne County's higher rate relative to Washtenaw County likely reflects a greater burden of chronic disease.

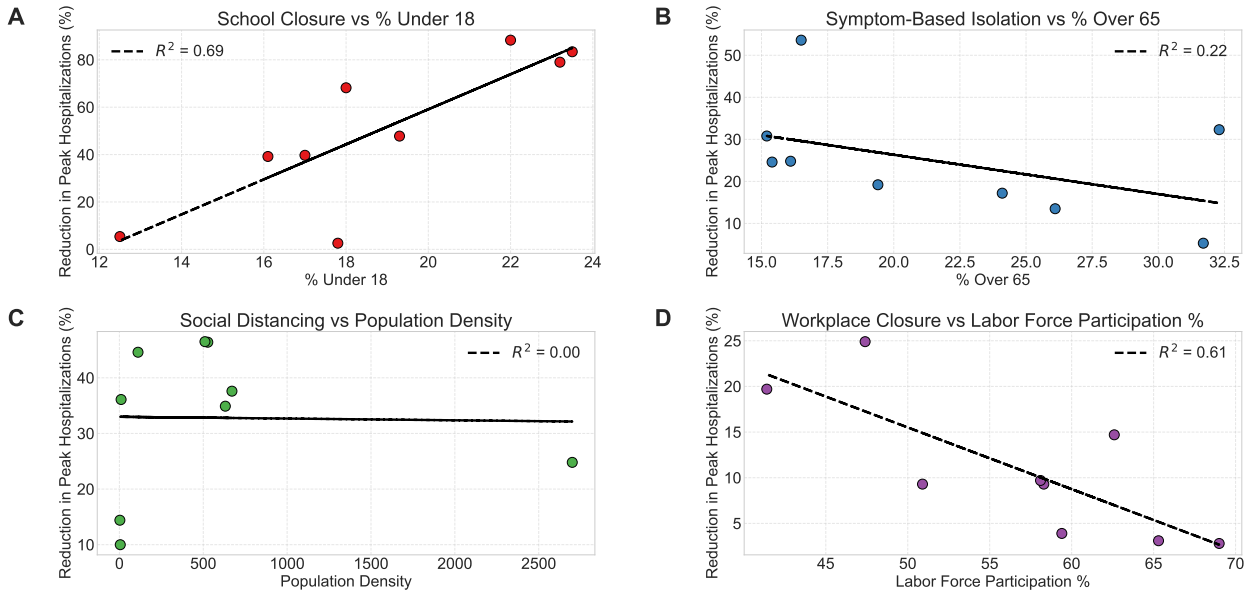

Figure 5: **County demographics do not reliably predict intervention effectiveness.** Each panel shows the relationship between the median effect of an intervention and a related demographic characteristic across Michigan counties. **(A)** School closure efficacy was positively associated with the percentage of children under 18 ($R^2 = 0.69$). **(B)** Symptom-based isolation showed a weak negative trend with the proportion of adults over 65 ($R^2 = 0.22$). **(C)** Social distancing effectiveness showed no relationship with population density ($R^2 = 0.00$). **(D)** Workplace closure effectiveness showed a surprising negative association with labor force participation rate ($R^2 = 0.61$), likely reflecting confounding structural factors rather than a causal relationship.

By providing distributions of possible parameter ranges, our framework aims to support robust intervention analysis under uncertainty and respect the inherent partial identifiability of infectious disease dynamics. We propagated full parameter distributions into downstream simulations, enabling decision-makers to evaluate not only expected outcomes but also the stability of those outcomes across plausible scenarios. This approach addresses the core challenge of early outbreak response: making policy decisions before key information is precisely known.

We found that school closures consistently produced the largest and most reliable reductions in peak hospitalizations when there was a large student population—reducing burden by 70–85% in Washtenaw and Wayne Counties—while offering limited benefit in Ontonagon County due to lower student density and reduced overall transmission potential. Symptom-based isolation yielded more variable effects, particularly sensitive to asymptomatic spread, with reductions ranging from 18–68% in Washtenaw depending on assumptions. General social distancing produced moderate but parameter-sensitive effects, while workplace closures had consistently modest impact across all settings. These patterns suggest that targeted interventions, such as school closures in student-heavy populations, offer greater robustness than broad behavioral strategies. However, policy makers also have to consider more than just public health implications when selecting interventions. School closures impact education, childcare, and family economic stability, so relying on them in a prolonged outbreak or pandemic may not be feasible.

Our results also demonstrate an important structural constraint: in very small or rural populations, even well-implemented community-level interventions may have limited effectiveness. In Ontonagon County, all interventions produced modest and highly variable reductions in peak burden. This likely reflects sparse contact networks and low baseline transmission, where stochastic case importations from other regions dominate and local interventions exert less control. In such contexts, regional coordination and enhanced surveillance for case introductions may be more impactful than traditional NPIs.

The surrogate-informed inference pipeline provides a scalable, data-efficient approach to local policy evaluation. Because it operates on standard surveillance data, it can be rapidly applied to a wide range

of pathogens. In any context where transmission can be mechanistically simulated and outcomes observed through routine surveillance, this framework enables uncertainty-aware intervention planning. In settings with limited analytic capacity, it allows local agencies to identify strategies that perform reliably across plausible scenarios, facilitating confident decision-making even under severe data constraints.

An important consideration for this framework is the potential for surrogate model misspecification. By design, the surrogate model is a simplification of the full ABM, otherwise the ABM would not be necessary. However, misspecification can also arise from leaving out key mechanisms that shape transmission dynamics. For example, if the surrogate model lacked underreporting parameters, which were critical for COVID-19, the framework might struggle to correctly map ABM parameter space to observed data [22]. The extent to which the high-dimensional ABM parameter space can compensate for structural omissions in the surrogate is an open question. In some cases, the flexibility of the ABM may allow it to recover appropriate dynamics despite surrogate misspecification. In others, missing mechanisms in the surrogate could systematically bias the identified ABM parameter region. Understanding the robustness of this approach to different forms and degrees of surrogate misspecification is an important direction for future work, and may inform best practices for surrogate model selection in different outbreak contexts.

The utility of this framework stems from the structural mapping between the complex agent-based system and the simplified surrogate. Mathematically, the surrogate ODE acts as a lower-dimensional manifold that captures the essential transmission geometry of the higher-dimensional agent-based model. By identifying the pre-image of data-consistent surrogate parameters, we can systematically pull back population-level uncertainty into the mechanistically rich ABM space. While this work provides a practical implementation for decision support, further theoretical development is needed to rigorously characterize the properties of this manifold mapping across different model classes. Formalizing the mathematical relationship between stochastic individual-level interactions and their deterministic mean-field approximations remains a critical area for future investigation to ensure robust inference in increasingly complex systems.

Another important limitation concerns compliance assumptions. Throughout our intervention simulations, we held compliance rates constant across counties and parameter regimes (see methods). In practice, compliance can vary substantially across communities due to differences in risk perception, trust in public health authorities, economic constraints, and cultural norms. Moreover, compliance is difficult to predict in advance and may evolve over the course of an outbreak as fatigue sets in or perceived risk changes. Our framework evaluates intervention effectiveness conditional on assumed compliance levels, but does not predict compliance itself. Future extensions could incorporate behavioral models or empirically derived compliance distributions to better capture this source of uncertainty.

Furthermore, while we assume epidemiological and behavioral parameters remain stationary throughout the simulation period, it is well-recognized that transmission dynamics evolve over time due to viral mutation, policy shifts, and behavioral fatigue. In our application to the 2020 wave in Michigan, assuming stationarity served as a regularization to maximize signal-to-noise ratio when comparing regional demographic structures across a narrow three month window. However, the SMoRe ParS framework is inherently modular; the surrogate model can be readily extended to include time-dependent parameters (e.g., $\beta(t)$ or $\rho(t)$) to capture longitudinal dynamics across multiple waves without requiring fundamental changes to the inference pipeline. Exploring the interplay between temporal shifts and spatial heterogeneity is a promising avenue for future research.

By combining mechanistic interpretability, uncertainty propagation, and regional specificity, this framework offers a practical method for real-time decision support in future outbreaks. It complements national-scale forecasting efforts by enabling grounded, localized policy planning tailored to the demographic and infrastructural realities of each community.

5 Methods

Agent-Based Simulations

We used the FRED (Framework for Reconstructing Epidemiological Dynamics) platform to simulate COVID-19 outbreaks under varying parameterizations [16]. Parameter sets were generated using Latin Hypercube Sampling (LHS) across ranges for latent period, transmissibility (overall and asymptomatic), hospitalization probability and duration, and case and hospitalization reporting rates. Each parameter set was run with

five stochastic replicates to capture variability. Model outputs included daily symptomatic cases and hospitalizations, which were aggregated and passed through the same surrogate fitting procedure as the observed surveillance data to ensure consistency.

Surrogate Model Selection and Fitting

To link ABM simulations with observed epidemic trajectories, we evaluated five candidate ODE-based surrogate models (SIRH, SEIRH, SEAIRH, $SEAI_1I_2RH$, SEIRHD), each of which had parameters governing reporting rates for both cases and hospitalizations. Models were fit to county-level COVID-19 case and hospitalization time series [17] using weighted least squares, and compared using the Akaike Information Criterion (AIC) [23]. The SIRH model provided the best overall fit in all three main representative counties (ΔAIC = 24.29 in Washtenaw, 2.99 in Wayne, and 31.91 in Ontonagon). Based on these results, SIRH was selected as the surrogate for all downstream analyses. Parameter confidence intervals were estimated using the profile likelihood method.

Parameter Surface Reconstruction

Within the SMoRe ParS framework, we mapped ABM parameters to surrogate model (SM) parameters to quantify uncertainty and constrain mechanistic space. For each ABM configuration, we ran simulations, fit the surrogate, and estimated confidence intervals using profile likelihood. These ABM–SM mappings were interpolated to form continuous response surfaces, which were then compared with surrogate fits to real data. ABM parameter sets consistent with the empirical 95% confidence regions were retained, defining county-specific parameter spaces for downstream intervention analyses.

Data Sources and Calibration

We obtained COVID-19 case and hospitalization data from public health surveillance systems at the county level in Michigan [17]. Case data consisted of daily incidence counts, while hospitalization data represented total hospital occupancy (i.e., prevalent, not incident, hospitalizations). We used census data for population [24]. No preprocessing or smoothing was applied to the data. However, due to known underreporting of COVID-19 cases and hospitalizations, we fit an underreporting rate as an additional parameter in the surrogate model calibration process.

Intervention Simulation and Robustness Evaluation

To assess how parameter uncertainty affects the projected outcomes of public health interventions, we simulated several non-pharmaceutical interventions within the FRED platform across the full range of parameter combinations consistent with observed data. The interventions evaluated included: (1) school closures (implemented by eliminating school and classroom contact rates), (2) stay-home-if-symptomatic policies (increasing probability of staying home to 60% [25], (3) non-essential work closures (implemented by reducing workplace and office contact rates by 66% [26]), and (4) general social distancing (implemented by reducing social contact rates by 60% [27]).

For each intervention, we evaluated outcomes using three primary metrics: (1) peak daily cases, (2) peak hospitalizations, and (3) cumulative burden of disease (total cases and total hospitalizations). Intervention robustness was assessed by evaluating performance at multiple locations within the identified ABM parameter region for each county. Interventions that performed consistently well across this space were classified as robust, while those whose effectiveness varied substantially were deemed sensitive to parameter uncertainty.

Implementation

All surrogate model fitting, profile likelihood computation, and intervention evaluation were implemented in Python. Agent-based simulations were run using the FRED simulation platform with custom scenario definitions. All code used in this study, including scripts for surrogate model fitting, ABM execution, parameter surface reconstruction, and intervention analysis, is available at `https://github.com/carsondudley1/smoreparscovid`.

6 Author Contributions

C.D. led the study and implemented the code. D.B., H.J., K.-A.N., and E.R. contributed equally to this work. D.B. created the conceptual figure. K.-A.N. performed the statistical significance analyses. M.E. provided guidance on relevant public health questions and framing. C.D., H.J., and T.J. planned the project. All authors shaped the direction of the project and reviewed and edited the manuscript.

References

[1] Natsuko Imai, Katy A. M. Gaythorpe, Sam Abbott, Sangeeta Bhatia, Sabine van Elsland, Kiesha Prem, Yang Liu, and Neil M. Ferguson. Adoption and impact of non-pharmaceutical interventions for covid-19. Wellcome Open Research, 5:59, 2020.

[2] Zichen Ma and Lior Rennert. An epidemiological modeling framework to inform institutional-level response to infectious disease outbreaks: a covid-19 case study. Scientific Reports, 14:7221, 2024.

[3] Jon Zelner, Nina B. Masters, Ramya Naraharisetti, Sanyu A. Mojola, Merlin Chowkwanyun, and Ryan Malosh. There are no equal opportunity infectors: Epidemiological modelers must rethink our approach to inequality in infection risk. PLOS Computational Biology, 0, 2022.

[4] Emma E. Glennon, Marjolein Bruijning, Justin Lessler, Ian F. Miller, Benjamin L. Rice, Robin N. Thompson, Konstans Wells, and C. Jessica E. Metcalf. Challenges in modeling the emergence of novel pathogens. Epidemics, 37:100516, December 2021.

[5] Dawn M. Bravata, Anthony J. Perkins, Laura J. Myers, Greg Arling, Ying Zhang, Alan J. Zillich, Lindsey Reese, Andrew Dysangco, Rajiv Agarwal, Jennifer Myers, Charles Austin, Ali Sexson, Samuel J. Leonard, Sharmistha Dev, and Salomeh Keyhani. Association of intensive care unit patient load and demand with mortality rates in US department of veterans affairs hospitals during the COVID-19 pandemic. JAMA Network Open, 4(1):e2034266, 2021.

[6] Marissa Renardy, Marisa Eisenberg, and Denise Kirschner. Predicting the second wave of covid-19 in washtenaw county, mi. Journal of Theoretical Biology, 507:110461, 2020.

[7] Ashleigh R. Tuite, Joseph Tien, Marisa Eisenberg, David J.D. Earn, Junling Ma, and David N. Fisman. Cholera epidemic in haiti, 2010: Using a transmission model to explain spatial spread of disease and identify optimal control interventions. Annals of Internal Medicine, 154(9):593–601, 2011.

[8] Olga Krivorotko, Maria Sosnovskaia, Irina Vashchenko, Cliff Kerr, and Daniel Lesnic. Agent-based modeling of covid-19 outbreaks for new york state and uk: Parameter identification algorithm. Infectious Disease Modelling, 7(1):30–44, March 2022. Epub 2021 Nov 27.

[9] Christian Nitzsche and Stefan Simm. Agent-based modeling to estimate the impact of lockdown scenarios and events on a pandemic exemplified on sars-cov-2. Scientific Reports, 14(13391), 2024.

[10] Vivek Srikrishnan and Klaus Keller. Small increases in agent-based model complexity can result in large increases in required calibration data. Environmental Modelling & Software, 138:104978, 2021.

[11] Ayush Chopra, Alexander Rodríguez, Jayakumar Subramanian, Arnau Quera-Bofarull, Balaji Krishnamurthy, B. Aditya Prakash, and Ramesh Raskar. Differentiable agent-based epidemiology, 2022.

[12] Harsh Vardhan Jain, Kerri-Ann Norton, Bernardo Bianco Prado, and Trachette L. Jackson. Smore pars: A novel methodology for bridging modeling modalities and experimental data applied to 3d vascular tumor growth. Frontiers in Molecular Biosciences, 9:1056461, 2022.

[13] Daniel R. Bergman, Kerri-Ann Norton, Harsh Vardhan Jain, and Trachette Jackson. Connecting agent-based models with high-dimensional parameter spaces to multidimensional data using smore pars: A surrogate modeling approach. Bulletin of Mathematical Biology, 86(11), 2024.

[14] Alexia Couture, A Danielle Iuliano, Howard H Chang, Neha N Patel, Matthew Gilmer, Molly Steele, Fiona P Havers, Michael Whitaker, and Carrie Reed. Estimating covid-19 hospitalizations in the united states with surveillance data using a bayesian hierarchical model: Modeling study. JMIR Public Health and Surveillance, 8(6):e34296, 2022.

[15] Hualei Xin, Yu Li, Peng Wu, Zhili Li, Eric HY Lau, Ying Qin, Liping Wang, Benjamin J Cowling, Tim K Tsang, and Zhongjie Li. Estimating the latent period of coronavirus disease 2019 (covid-19). Clinical Infectious Diseases, 74(9):1678–1681, 2022.

[16] John J. Grefenstette, Stephen T. Brown, Roni Rosenfeld, Justin Depasse, Nathaniel T. Stone, Patrick C. Cooley, William D. Wheaton, Alona Fyshe, David D. Galloway, Akshay Sriram, Hasan Guclu, Thomas Abraham, and Donald S. Burke. FRED (A Framework for Reconstructing Epidemic Dynamics): An open-source software system for modeling infectious diseases and control strategies using census-based populations. BMC Public Health, 13(1):940, Oct 2013.

[17] University of Michigan. Mi safe start map. Accessed April 15, 2024.

[18] Robert Verity et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases, 20(6), 2020.

[19] Karla Therese L Sy, Laura F White, and Brooke E Nichols. Population density and basic reproductive number of covid-19 across united states counties. PLOS ONE, 16(4), 2021.

[20] Qun Li et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. New England Journal of Medicine, 382(13):1199–1207, 2020.

[21] Stephen A Lauer et al. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. Annals of Internal Medicine, 172(9):577–582, 2020.

[22] Marisa C. Eisenberg, Joseph N. S. Eisenberg, Jeremy P. D'Silva, Eden V. Wells, Sarah Cherng, Yu-Han Kao, and Rafael Meza. Forecasting and uncertainty in modeling the 2014-2015 ebola epidemic in west africa, 2015.

[23] H. Akaike. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6):716–723, 1974.

[24] USAFacts. Our changing population: Michigan. `https://usafacts.org/data/topics/people-society/population-and-demographics/our-changing-population/state/michigan/county/washtenaw-county/`. Accessed: November 6, 2025.

[25] Centers for Disease Control and Prevention. Behind the model: Estimating the impact of updated isolation guidance on covid-19 transmission, April 2024. Accessed: 2025-08-04.

[26] Audrey Kearney and Cailey Muñana. Taking stock of essential workers. `https://www.kff.org/policy-watch/taking-stock-of-essential-workers/`, May 2020. Accessed: 2025-05-19.

[27] Dennis M. Feehan and Ayesha S. Mahmud. Quantifying population contact patterns in the united states during the covid-19 pandemic. Nature Communications, 12(1):893, February 2021.

[28] Sarabeth M. Mathis et al. Evaluation of FluSight influenza forecasting in the 2021–22 and 2022–23 seasons with a new target laboratory-confirmed influenza hospitalizations. Nature Communications, 15, July 2024.

[29] Alexander Rodriguez et al. Deepcovid: An operational deep learning-driven framework for explainable real-time covid-19 forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, 2021.

[30] Carson Dudley, Reiden Magdaleno, Christopher Harding, Ananya Sharma, and Marisa Eisenberg. Mantis: A foundation model for mechanistic disease forecasting. arXiv preprint arXiv:2508.12260, 2025.

[31] Dongxia Wu, Liyao Gao, Xinyue Xiong, Matteo Chinazzi, Alessandro Vespignani, Yi-An Ma, and Rose Yu. Deepgleam: A hybrid mechanistic and deep learning model for covid-19 forecasting, 2021.

Appendices

Agent-Based Model Parameter Sampling

We used Latin Hypercube Sampling (LHS) to generate parameter sets for agent-based simulations across the full range of plausible epidemiological values. The same parameter ranges and sampling strategy were applied to all three counties, with the same sampled parameters used across counties. Seven key parameters were varied, with the following continuous sampling ranges:

Table 2: Parameter ranges used for ABM sampling.

Parameter	**Range**
Latent period (days)	3.0 – 10.0
Asymptomatic transmissibility	0.00 – 0.75
Transmissibility	0.50 – 1.50
Hospitalization probability	0.00 – 0.20
Hospitalization duration (days)	3 – 21
Case Reporting Rate	10% – 100%
Hospitalization Reporting Rate	10% – 100%

A total of 250 unique parameter combinations were generated using Latin Hypercube Sampling, with each parameter sampled continuously and independently within its specified range. For each parameter set, 5 independent stochastic replicates were simulated to mitigate random variability arising from the agent-based dynamics. All simulations were initialized identically across replicates to ensure comparability.

Intervention Implementation Details

All interventions were simulated under consistent compliance assumptions across counties. Because the study period captured an ongoing epidemic wave rather than epidemic initiation, simulations were initialized with seeded infections and hospitalizations based on observed surveillance data for each county as of August 28, 2020. Interventions were applied from the beginning of each simulation and remained in effect throughout the full simulation period. Specific implementation assumptions were as follows: school closures were modeled by fully eliminating school and classroom contact rates; stay-home-if-symptomatic policies increased the daily probability of isolation for symptomatic individuals to 60% [25]; non-essential work closures were modeled by reducing workplace and office contact rates by 66% [26]; and general social distancing was modeled by reducing community and social contact rates by 60% [27].

Surrogate Model Fitting

To connect agent-based model (ABM) simulations to observed epidemic trajectories, we evaluated several compartmental ordinary differential equation (ODE) models as candidate surrogates. The models considered were:

- **SIRH (Susceptible–Infectious–Recovered–Hospitalized):**

$$\begin{aligned}
\frac{dS}{dt} &= -\beta SI \\
\frac{dI}{dt} &= \beta SI - \gamma I - \delta I \\
\frac{dH}{dt} &= \delta I - \alpha H \\
\frac{dR}{dt} &= \gamma I + \alpha H
\end{aligned}$$

- **SEIRH (Susceptible–Exposed–Infectious–Recovered–Hospitalized):**

$$\begin{aligned}
\frac{dS}{dt} &= -\beta SI \\
\frac{dE}{dt} &= \beta SI - \kappa E \\
\frac{dI}{dt} &= \kappa E - \gamma I - \delta I \\
\frac{dH}{dt} &= \delta I - \alpha H \\
\frac{dR}{dt} &= \gamma I + \alpha H
\end{aligned}$$

- **SEAIRH (Susceptible–Exposed–Asymptomatic–Infectious–Recovered–Hospitalized):**

$$\begin{aligned}
\frac{dS}{dt} &= -\beta S(I + \eta A) \\
\frac{dE}{dt} &= \beta S(I + \eta A) - \kappa E \\
\frac{dA}{dt} &= \kappa E \cdot p_A - \gamma_A A \\
\frac{dI}{dt} &= \kappa E \cdot (1 - p_A) - \gamma I - \delta I \\
\frac{dH}{dt} &= \delta I - \alpha H \\
\frac{dR}{dt} &= \gamma I + \gamma_A A + \alpha H
\end{aligned}$$

- **SEAI$_1$I$_2$RH (with separate compartments for mild and severe infection):**

$$\begin{aligned}
\frac{dS}{dt} &= -\beta S(I_1 + I_2 + \eta A) \\
\frac{dE}{dt} &= \beta S(I_1 + I_2 + \eta A) - \kappa E \\
\frac{dA}{dt} &= \kappa E \cdot p_A - \gamma_A A \\
\frac{dI_1}{dt} &= \kappa E \cdot (1 - p_A) \cdot p_M - \gamma_1 I_1 \\
\frac{dI_2}{dt} &= \kappa E \cdot (1 - p_A) \cdot (1 - p_M) - \gamma_2 I_2 - \delta I_2 \\
\frac{dH}{dt} &= \delta I_2 - \alpha H \\
\frac{dR}{dt} &= \gamma_A A + \gamma_1 I_1 + \gamma_2 I_2 + \alpha H
\end{aligned}$$

- **SEIRHD (Susceptible–Exposed–Infectious–Recovered–Hospitalized–Deceased):**

$$\begin{aligned}
\frac{dS}{dt} &= -\beta SI \\
\frac{dE}{dt} &= \beta SI - \kappa E \\
\frac{dI}{dt} &= \kappa E - \gamma I - \delta I \\
\frac{dH}{dt} &= \delta I - \alpha H - \mu H \\
\frac{dR}{dt} &= \gamma I + \alpha H \\
\frac{dD}{dt} &= \mu H
\end{aligned}$$

Each model was fit to observed COVID-19 case and hospitalization time series using a weighted least squares approach. Model selection was performed using the Akaike Information Criterion (AIC) [23]. The SIRH model provided the best fit across all counties and was therefore selected as the final surrogate model for all downstream analysis.

Model fitting was performed via weighted least squares under an observation model incorporating both underreporting and measurement error. Specifically, we estimated eight parameters jointly:

- β: transmission rate
- γ: recovery rate
- δ: hospitalization rate
- α: hospital discharge rate
- ρ_{cases}: reporting rate for cases
- ρ_{hosp}: reporting rate for hospitalizations
- σ_{cases}: observation noise for cases
- σ_{hosp}: observation noise for hospitalizations

Model fitting was performed via weighted least squares. The objective function minimized was:

$$\text{RSS} = \sum_t (C_t - \rho_{\text{cases}} \cdot \hat{C}_t)^2 + w_H \sum_t (H_t - \rho_{\text{hosp}} \cdot \hat{H}_t)^2$$

where C_t and H_t are the observed daily cases and hospitalizations at time t, $\hat{C}_t$ and $\hat{H}_t$ are model-predicted quantities, and w_H is a weighting factor applied to hospitalization residuals to ensure both data streams contribute meaningfully to the fit despite differences in scale. We used $w_H = 5$.

Optimization was performed using the L-BFGS-B algorithm with multiple random initializations to avoid local minima. Parameter bounds were specified for each parameter (see Table 3). The full epidemic time series for each county was fit jointly to both case and hospitalization data.

Parameter Surface Reconstruction

Following the SMoRe ParS framework, we constructed explicit mappings between ABM parameters and surrogate model (SM) parameters to enable systematic uncertainty quantification and data-informed constraint of mechanistic model space. For each ABM parameter combination sampled via Latin Hypercube Sampling, we simulated the agent-based model, extracted summary trajectories (e.g., hospitalization time series), and fit the SIRH surrogate model to these trajectories using maximum likelihood estimation.

Table 3: Parameter bounds used in surrogate model optimization.

Parameter	Lower Bound	Upper Bound
β	0.01	2.0
γ	0.01	0.5
δ	0.001	0.2
α	0.01	0.5
ρ_{cases}	0.1	1.0
ρ_{hosp}	0.1	1.0
σ_{cases}	1.0	100.0
σ_{hosp}	1.0	100.0

This procedure yielded surrogate parameter estimates for each sampled ABM parameter combination. To quantify uncertainty in these estimates, we applied the profile likelihood method to each ABM–SM fit, calculating 95% confidence intervals for each surrogate parameter by profiling the objective function with respect to each parameter while optimizing over the others. This yielded, at each ABM parameter combination, confidence intervals on the surrogate parameters conditional on that ABM setting.

The resulting discrete mappings—from ABM parameter vectors $\vec{p}_{\text{ABM}}$ to surrogate parameter estimates and their confidence intervals—were then used to reconstruct continuous response surfaces via multivariate interpolation. Specifically, we used `scipy`'s `LinearNDInterpolator` to construct interpolated functions of the form:

$$p_{\text{SM},i} = f_i(\vec{p}_{\text{ABM}}) \pm C.B._i(\vec{p}_{\text{ABM}}) \tag{1}$$

where $p_{\text{SM},i}$ denotes the ith surrogate parameter (e.g., $\beta, \gamma, \delta, \alpha$), f_i is the interpolated best-fit value, and $C.B._i$ represents the corresponding 95% confidence bounds, both obtained via interpolation from the set of profile-likelihood-derived estimates.

To determine which ABM parameter combinations are consistent with the real-world data, we first fit the surrogate model directly to the observed data using weighted least squares and profile likelihood, generating a data-derived confidence region for each surrogate parameter. We then used the reconstructed response surfaces to project every ABM parameter vector $\vec{p}_{\text{ABM}}$ into surrogate space, and retained only those ABM parameter combinations whose projected surrogate values fell within the empirical 95% confidence region inferred from data. The union of these retained points defines an implicit confidence region over ABM parameters, incorporating both data-derived surrogate uncertainty and interpolation-based ABM–surrogate relationships.

This reconstruction process was performed independently for each county, enabling region-specific parameter space characterization while maintaining a consistent calibration pipeline across all study sites. The resulting confidence regions over ABM parameters support downstream analyses such as policy ranking, robustness checks, and posterior predictive simulations within data-consistent mechanistic regimes.

Results for Additional Counties

To evaluate whether the intervention patterns observed in Washtenaw, Wayne, and Ontonagon Counties held across structurally similar settings, we analyzed six additional Michigan counties grouped by demographic and infrastructural similarity. Specifically, we selected:

- **Like Washtenaw (younger, university-centered):** Isabella County and Ingham County (also contains state capital)
- **Like Wayne (urban, high comorbidity burden):** Genesee County and Kent County
- **Like Ontonagon (rural, aging population):** Luce County and Iron County

Table 4: ABM Parameter estimates recovered by the SMoRe ParS framework for the additional counties. Values represent point estimates with 95% confidence intervals derived from model calibration to case and hospitalization data.

ABM Parameter	**Isabella**	**Ingham**	**Genesee**	**Kent**	**Luce**	**Iron**
Hospitalization Probability	5.9% [3.3, 7.9]	6.2% [5.1, 7.6]	8.2% [6.8, 12.5]	7.2% [5.8, 10.3]	8.2% [6.9, 10.1]	9.2% [8.6, 10.1]
Hospitalization Duration (days)	6.4 [4.5, 14.0]	8.6 [3.5, 12.1]	11.9 [11.2, 12.4]	9.9 [8.5, 14.0]	7.9 [6.5, 9.0]	10.1 [8.8, 14.0]
Latent Period (days)	5.2 [5.02, 5.40]	4.65 [4.4, 4.89]	5.16 [4.84, 5.63]	4.76 [4.54, 5.01]	3.76 [3.54, 3.98]	5.98 [5.39, 7.01]
Transmissibility	1.63 [1.54, 1.77]	1.81 [1.68, 1.98]	1.90 [1.83, 1.98]	1.78 [1.73, 1.82]	1.76 [1.71, 1.80]	1.49 [1.43, 1.60]
Asymptomatic Transmissibility	36.0% [35.1, 38.5]	37.3% [36.7, 39.6]	38.3% [37.1, 39.8]	37.3% [37.1, 37.8]	38.3% [37.6, 38.9]	37.6% [37.3, 38.0]
Case Reporting Rate	14.0% [11.3, 16.4]	16.1% [12.4, 19.9]	26.3% [21.2, 30.2]	22.0% [18.2, 26.2]	6.0% [6.0, 12.3]	24.2% [18.0, 27.3]
Hospitalization Reporting Rate	93.0% [86.5, 100.0]	100.0% [93.2, 100.0]	97% [91.1, 100.0]	100.0% [95.1, 100.0]	6.0% [6.0, 10.1]	11.0% [10.0, 19.1]

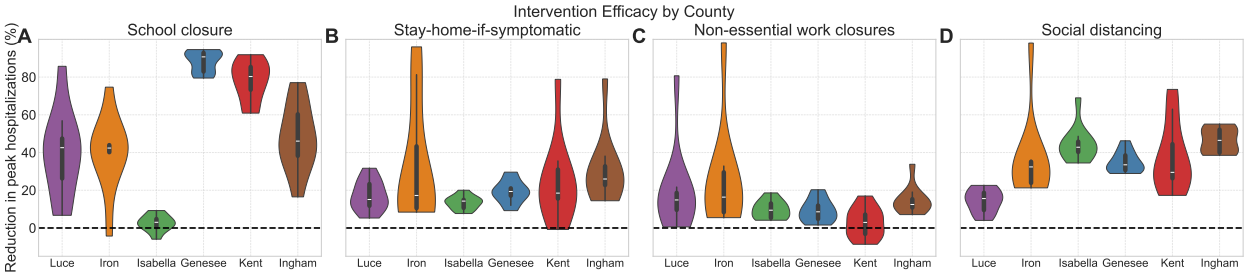

Figure 6: **Intervention effectiveness across additional counties.** Violin plots show the distribution of reductions in peak hospitalizations (%) under four non-pharmaceutical interventions, simulated across data-consistent ABM parameter combinations.

(A) School closures remained highly effective in counties structurally similar to Washtenaw and Wayne. Ingham, Kent, and Genesee exhibited the largest reductions in peak hospitalizations. Ingham showed strong but slightly more variable reductions. An exception was Isabella County, where school closures produced minimal effect. This likely reflects the unique population structure of Isabella, which has a disproportionately high university student population and relatively fewer school-aged children. Because our intervention model targets K–12 settings and does not modify university transmission, the impact of school closures is possibly muted in counties where universities dominate the educational landscape. In contrast, rural Ontonagon-like counties (Luce and Iron) also showed impactful but highly variable school closure effects.

(B) Stay-home-if-symptomatic policies again showed the strongest performance in university-oriented counties. Ingham and Isabella mirrored Washtenaw's high median reductions, while Wayne-like counties (Genesee, Kent) had more moderate effects. Ontonagon-like counties (Luce, Iron) showed wide variance and relatively weak performance, reinforcing the policy's vulnerability to asymptomatic transmission and limited

control in sparse rural networks.

(C) Non-essential work closures continued to produce modest and inconsistent reductions across all counties, with little structural variation. Even in more urban counties like Genesee and Kent, median effects were small, and wide spreads were sometimes observed, particularly in Iron and Luce. These findings support the main conclusion that work closures are a low-impact strategy in most contexts.

(D) Social distancing remained moderately effective in urban and university-adjacent counties. Isabella, Genesee, and Ingham saw stable, substantial reductions. In Kent and Iron, the results were also positive but more variable. In Luce, the results were much less substantial, possibly reflecting lower baseline connectivity due to rurality. Isabella and Ingham had the highest average effect, possibly reflecting highest baseline connectivity due to the university settings.

Overall, these results reinforce the main text conclusions:

- **School closures** are consistently robust in populations with many school-aged children.
- **Symptom-based isolation** is highly context-dependent and weakened by asymptomatic spread.
- **Work closures** produce only modest effects.
- **Social distancing** can produce strong effects particularly in high connectivity settings.

These additional counties support the conclusion that intervention performance is tightly coupled to local demographic and structural features, and that effective outbreak policy design must account for these context-specific factors.

Out-of-Sample Validation

To assess the reliability and stability of our framework, we performed a temporal hold-out validation. We calibrated the agent-based model (ABM) using only the first 80 days of the epidemic wave for Wayne and Washtenaw counties—a period primarily characterized by exponential growth—and evaluated the resulting simulations against the full 120-day observed time series.

In Wayne County, where the 80-day training window included the initial peak, the resulting ABM projections remained highly consistent with the held-out data (figure 7). The parameters identified using the partial dataset did not deviate significantly from those found using the full 120-day window, yielding a stable fit across both periods.

For Washtenaw County, the validation provided a more rigorous test as the outbreak had not yet peaked by day 80 and was still in the exponential growth phase. Calibrating solely on this growth period resulted in a higher transmissibility estimate compared to the full-data fit (point estimate 1.92 vs. 1.73). Interestingly, while the surrogate model had an accurate fit to the 80-day training period (which you would expect, because it was fit on that), the resulting ABM projection exhibited a slight overshoot during the initial training period rather than the held-out window (figure 7). For the surrogate model, the higher estimated transmissbility led it to peak later, with a good fit initially and then an overshoot. For the ABM, the peak timing was more accurate, but the initial growth phase was overshot. These results indicate that, while fitting to a narrow window of exponential growth can introduce bias in parameter fitting, especially for our simplified surrogate models, our method's projection into the higher-dimensional ABM space potentially provides a corrective mechanism. By connecting the data-consistent parameter sets to a realistic population architecture, the framework mitigates the impact of incomplete surveillance signals, yielding fits that are more representative of the full epidemic wave and more biologically plausible than would be expected from the limited training window.

To be clear, our goal is validation of the structural stability of our approach, not proposing use for real-time epidemic forecasting. While forecasting models optimize predictive accuracy (sometimes with mechanistic models similar to what we use but often with statistical or deep learning approaches [28, 29, 30, 31]), the goal of our framework is to identify data-consistent parameter regimes that allow for robust policy evaluation. Our hold-out analysis demonstrates that even when calibrated on incomplete epidemic phases, the mapping to the higher-dimensional ABM structure provides some regularization that prevents the model from drifting into potentially biologically implausible regimes (like extreme transmission).

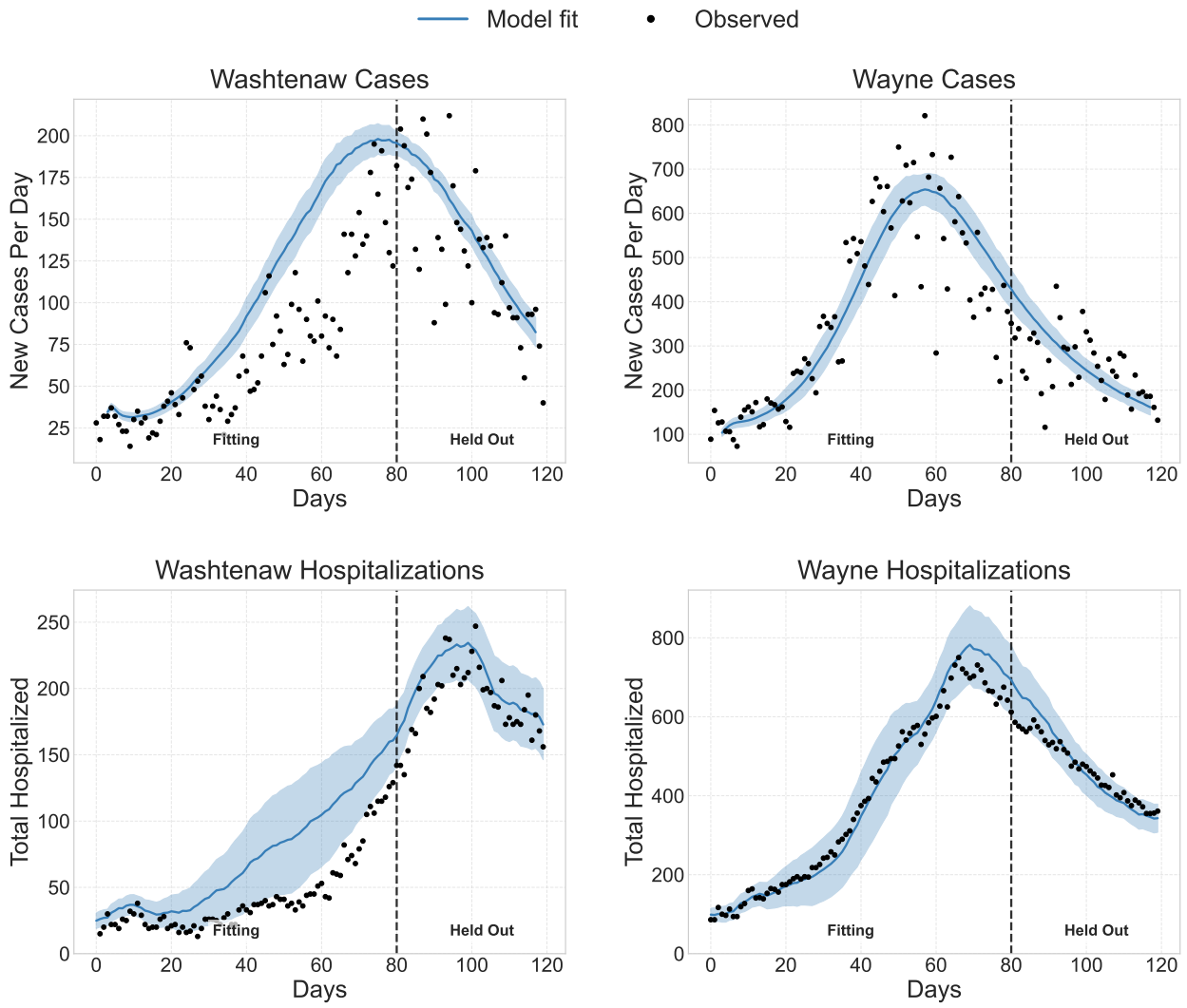

Figure 7: **Out-of-Sample Validation in Wayne and Washtenaw Counties.** The framework was calibrated using surveillance data from the first 80 days (fitting) and projected forward for the remaining 40 days (which were held out from fitting). Solid lines represent the ABM mean across 10 simulations, with shaded regions indicating ± 1 SD across stochastic replicates. In Wayne County, parameters remained stable, yielding an accurate fit across both windows. In Washtenaw County, calibrating solely on the exponential growth phase (pre-peak) resulted in higher transmissibility estimates, leading to a slight overshoot during the initial period, though the model relatively successfully anticipated the peak timing and subsequent decay rate.

Statistical Analysis

Outcome. All tests use the ratio of peak hospitalizations under an intervention to the baseline (no intervention) peak for the same county and parameter setting. Ratios < 1 indicate reductions; ratios > 1 indicate increases.

Design and error control. For each comparison family we ran a one-way ANOVA and, when significant, Tukey's HSD for all pairwise contrasts (familywise error control within the family). We report p-values; distributional summaries are shown in the main-text figures.

Across counties (per intervention)

For each intervention, we compared distributions across Washtenaw, Wayne, and Ontonagon (each $n = 30$).

- **School closure.** ANOVA $p = 7.04 \times 10^{-58}$; all pairwise contrasts significant (Tukey $p < 10^{-10}$).
- **Social distancing.** ANOVA $p = 1.17 \times 10^{-18}$; all pairwise contrasts significant (Tukey $p \leq 2.6 \times 10^{-5}$).
- **Workplace closure.** ANOVA $p = 0.312$; no pairwise differences significant (all Tukey $p \geq 0.338$).
- **Stay-home-if-symptomatic.** ANOVA $p = 6.50 \times 10^{-26}$; all pairwise contrasts significant (Tukey $p \leq 7.5 \times 10^{-6}$).

Within counties (across interventions)

For each county, we compared school closure, social distancing, workplace closure, and stay-home-if-symptomatic (each $n = 30$).

- **Wayne County.** ANOVA $p = 2.49 \times 10^{-104}$; all pairwise contrasts significant (smallest $p = 0.003$ for social vs. symptom).
- **Washtenaw County.** ANOVA $p = 4.64 \times 10^{-29}$; all pairwise contrasts significant except social vs. symptom (Tukey $p = 0.247$, n.s.).
- **Ontonagon County.** ANOVA $p = 0.387$; no pairwise differences significant (all Tukey $p \geq 0.323$).

Interpretation. Across counties, school closure, social distancing, and symptom-based isolation show statistically detectable differences in effect, whereas workplace closure does not. Within counties, intervention effects differ in Wayne and Washtenaw but are not statistically distinguishable in Ontonagon, consistent with greater stochastic variability in small, sparse populations.

Assumptions and robustness. ANOVA and Tukey's HSD assume approximate normality and homoscedasticity. Given bounded ratios and potential heteroskedasticity, we emphasize robust summaries (medians/IQRs) in the figures. As a sensitivity check, results were qualitatively unchanged under nonparametric testing (Kruskal–Wallis with Dunn's post hoc and Holm correction), and under percentile bootstrap confidence intervals for median differences (not shown).