Hospital Closures and Their Local Labor Market Fallout

Estimating the Causal Impact on Employment in German Municipalities

---

Submitting Student: Niklas Pawelzik

Thesis Supervisor: Prof. Christian Traxler
Second Reader: Prof. Asya Magazinnik
Study Programme: Data Science for Public Policy (MSc)

Year of Graduation: 2025
Word Count: 7700

Hertie School

CONTENTS

EXECUTIVE SUMMARY

As Germany undertakes its most significant hospital reform in decades, hundreds of closures are expected in the coming years. While debates often center on efficiency and care quality, broader economic implications, particularly for local labor markets, remain underexplored.

This study investigates how hospital closures between 2010 and 2020 affected employment in German municipalities. Drawing on a newly compiled dataset of confirmed closures and panel data on municipal employment, it applies a Difference-in-Differences approach to estimate causal effects. Municipalities with closures are compared to structurally similar ones that remained unaffected, allowing the analysis to isolate the employment impact of closures from broader economic trends.

Affected municipalities saw significant and persistent declines in employment as a result of the closure, growing steadily over the eight-year post-closure window. These findings are consistent across model specifications and supported by multiple robustness checks. Depending on the estimation strategy, average employment declines range from –0.5 to –0.55 percentage points, rising to –1.2 to –1.36 points by the end of the observation period. No signs of employment recovery are observed, even several years after closure, suggesting durable negative effects.

These findings suggest that hospital closures can trigger broader economic disruptions beyond the loss of hospital jobs alone. Hospitals often function as local anchor institutions, supporting service sectors, stabilizing labor demand, and shaping regional economic structures. Their withdrawal can have ripple effects, particularly in areas with limited employment alternatives.

Prior studies have suggested that hospital closures might have especially strong effects in rural areas or for women. This study does not find consistent evidence of such variation. However, the estimation strategy does not support formal statistical comparisons between subgroups, and very large cities as well as remote rural areas are underrepresented in the sample of municipalities that experienced closure. As a result, the findings should not be interpreted as confirming uniform effects, but rather as describing outcomes in the types of municipalities where closures have actually occurred.

In the context of ongoing reforms, this study provides timely empirical evidence of the labor market risks associated with hospital closures. It highlights the importance of considering local labor market impacts when evaluating healthcare restructuring. Closure decisions should be judged not only in terms of service provision, but also by their long-term economic effects on affected communities. Policymakers should weigh these labor market risks carefully when planning restructuring, and consider mitigation strategies for affected municipalities.

1 INTRODUCTION

Germany is undergoing a major hospital reform that is expected to reshape the national healthcare landscape. As part of this restructuring, the federal government has indicated that hundreds of hospitals may close (Bundesministerium für Gesundheit, 2025; Politico, 2024; WirtschaftsWoche, 2024). Critics warn the actual number may be even higher, with potentially far-reaching consequences for healthcare access and local communities (Deutscher Bundestag, Parlamentsnachrichten, 2024; Deutschlandfunk, 2024). Yet one aspect remains largely overlooked in public debate: the labor market impact of such closures (Berger et al., 2024). A growing international literature has emphasized hospitals' broader economic role in this regard; in Germany, however, the employment effects of hospital closures have received limited empirical attention (Fischer, 2024, p. 62). This study addresses this gap by examining a decade of hospital consolidation prior to the current reform wave, using panel data from 2010 to 2020. Against this backdrop, it poses the central question: What employment effects have hospital closures had at the municipal level in Germany?

To answer this question, the study constructs and draws on a unique municipal panel dataset. Hospital closures are identified through a multi-stage validation process that triangulates structural data, registry entries, and local media reports.[1] Employment outcomes are derived from the Bertelsmann foundation's *Wegweiser Kommune* dataset, and municipalities are classified using the *RegioStaR* typologies developed by the Federal Ministry for Digital and Transport. Based on this dataset, the primary estimation strategy applies the Difference-in-Differences (DiD) framework by Callaway and Sant'Anna (2021b), which accommodates key challenges posed by hospital closures, including staggered treatment timing and dynamic effects. The analysis estimates the Average Treatment Effect on the Treated (ATT), providing a causal estimate of employment impacts in municipalities that experienced confirmed closures.

This study makes three contributions to the literature. *First*, it adds new descriptive insights into the structure, geography, and timing of hospital closures in Germany during the 2010s. *Second*, it introduces an ATT-compatible strategy to improve estimator reliability in contexts with strong treatment–control imbalance, a common feature when policy shocks affect a limited number of units.[2] A pre-matching step enhances covariate balance without altering the estimand, offering a conceptually grounded and practical alternative to trimming or regression-based adjustments. *Third*, the study offers the first systematic estimate of the employment effects of hospital closures in Germany, extending a literature that has so far been largely U.S.-focused. Results indicate persistent employment declines in affected municipalities, with average treatment effects ranging from –0.5 to –0.55 percentage points across models, growing to between –1.2 and –1.36 points toward the end of the observation window. No clear recovery is observed, and effects vary little by gender or rural–urban classification. These findings highlight hospitals' role in employment and underscore the wider socioeconomic consequences of healthcare restructuring. While the findings are robust across estimators and supported by

[1] In this study, "hospitals" denote general acute care hospitals (*allgemeine Krankenhäuser*) offering comprehensive somatic services. Specialized institutions in psychiatry or neurology are excluded.

[2] The term "imbalance" here refers to the treated-to-control unit ratio, as used in the machine learning literature. This contrasts with imbalance from incomplete panel data, often discussed in DiD contexts.

multiple sensitivity checks, they are limited to municipalities with confirmed closures and adequate data coverage. The results should therefore be interpreted as internally valid but not nationally generalizable. Nonetheless, the study offers timely evidence on the labor market consequences of hospital consolidation, particularly relevant in light of Germany's ongoing healthcare reforms.

The paper develops its argument in the following steps. It first situates the study at the intersection of health policy, local labor markets, and causal inference. A contextual chapter introduces the German case, focusing on the institutional framework shaping hospital planning and closure dynamics. The data section outlines the construction of a municipal panel dataset, presents key descriptive patterns, and explains how they inform the empirical strategy. The methods section introduces the estimator and discusses identification. The results section presents the estimated employment effects, including dynamic patterns, subgroup differences, and robustness checks. The discussion chapter reflects on the findings' broader implications, and the conclusion summarizes the study's contributions and relevance for research and policy.

2 LITERATURE REVIEW

This chapter situates the study within two key strands of research. First, it reviews work on hospitals as local economic anchors, highlighting their role beyond healthcare provision. Second, it discusses recent advances in Difference-in-Differences estimation, with emphasis on challenges relevant to hospital closures. Together, these literatures inform the study's empirical and identification strategy.

2.1 Hospitals as Cross-Sectoral Economic Drivers

The interconnectedness of health and broader societal structures has long been recognized in public health research. The *Health in All Policies* (HiAP) framework emerged to emphasize that health outcomes are shaped by decisions across sectors, encouraging an integrated lens (Baum et al., 2014). More recently, the *Health for All Policies* (HfAP) perspective has extended this logic in the opposite direction, pointing out how health policy in turn generates effects beyond the healthcare domain (Simon, 2016, p. 141; Greer et al., 2022). One contribution in a recent volume systematizing the HfAP-approach shows how health infrastructure shapes labor market dynamics notably by creating direct employment, influencing wages, and sustaining service-sector demand (Williams et al., 2024). While these mechanisms apply broadly to healthcare, they are especially salient for hospitals, which often serve as major employers and anchor institutions within their communities (Simon, 2016, p. 367; Dubb & Howard, 2012; Schildt & Rubin, 2015). Consequently, hospital *closures* can constitute localized economic shocks, with ripple effects well beyond the healthcare sector (Mandich & Dorfman, 2017). The HfAP framework therefore offers a coherent and directional model for understanding the socioeconomic reach of health policy -— and for analyzing hospital closures not just as healthcare events, but structural disruptions in local labor markets.

A growing body of empirical evidence supports underscores the economic significance of hospitals (Mills et al., 2024). Health and social work activities account for more than 8% of total employment in the EU, with nearly two thirds of them employed in hospitals

(Williams et al., 2024, pp. 130–131; European Commission, 2015; Eurostat, 2022). US studies show that hospital closures eliminate a wide range of jobs across skill levels. The effects are often particularly acute for lower-skilled workers, who face greater barriers to re-employment, and for women, who make up a disproportionately large share of the hospital workforce. (Williams et al., 2024, p. 132; Vogler, 2020). In addition, local businesses reliant on hospital staff and patient traffic — such as pharmacies, restaurants, and transport services — may experience reduced demand and secondary employment losses (Mandich & Dorfman, 2017). These employment declines are often found to persist over time (Alexander & Richards, 2023) and tend to hit rural municipalities harder, as their economy often relies even more on their hospital.

However, while this literature provides manifold indication of the labor market effects of hospital closures, it remains overwhelmingly US-focused. European research on hospital consolidation has largely centered on issues of financial sustainability, system efficiency, and access to care (see, for example, Ghislandi et al., 2025; Nyholt, 2024; Timofeyev et al., 2024). Labor market outcomes, by contrast, have received little systematic attention. This is an important gap, not least because European healthcare systems differ significantly from their U.S. counterpart in terms of financing, workforce regulation, and regional planning authority (Wendt et al., 2009). These differences may reshape the employment effects of hospital closures. To what extent U.S. findings extend to European systems remains an open question. This study addresses this gap by providing a causal estimate of the employment effects of hospital closures at the municipal level in Germany, a European setting with both a dense hospital landscape and a history of structural consolidation (Fischer, 2024, p. 50; Brunn et al., 2023). While the existing literature highlights the disruptive potential of hospital closures, rigorous assessment of their labor market impacts requires modern causal inference techniques. The next section reviews recent methodological innovations that inform the study's empirical strategy.

2.2 Causal Estimation in Non-Experimental Settings

To estimate the causal impact of hospital closures on local employment, the expanding Difference-in-Differences literature provides a wide range of flexible tools well suited to applied policy evaluation (Baker et al., 2025). At its core, DiD compares changes in outcomes over time between units exposed to an intervention and those that are not (Imbens & Wooldridge, 2009). In the context of hospital closures, municipalities experiencing closures can be compared to structurally similar municipalities without closures, a setup widely applied in previous U.S. studies. However, applying DiD methods to such settings presents several challenges that have prompted substantial developments in the recent econometric literature. Roth et al. (2023) and Baker et al. (2025) provide instructive guidance on the choice and implementation of appropriate designs, which also informs the present paper.

A first challenge arises when treatment timing varies across units (*staggered treatment*) and treatment effects evolve over time (*dynamic treatment effects*) (Athey & Imbens, 2022).So-called two-way fixed effects (TWFE) models, traditionally the default DiD estimator in applied policy research, rely on strong assumptions and can yield biased results under such conditions (De Chaisemartin & d'Haultfoeuille, 2023; Goodman-Bacon, 2021). These conditions have motivated the development of new estimators among which the estimator by Callaway and Sant'Anna (2021b) is particularly influential. It compares each cohort of treated units to an appropriate control group at each point in time, and then aggregates those estimates to obtain both dynamic and overall effects. This structure is well-suited to hospital closures, that appear at different moments in time and where employment impacts may unfold gradually.

A second complication concerns spatial interdependence. In hospital closure settings, the assumption that untreated units remain unaffected by treatment (a foundational requirement of conventional DiD designs) may be violated: Closures could plausibly affect neighboring areas via commuting patterns, shifting healthcare demand, or labor market adjustments (Alexander & Richards, 2023). If such spillovers exist, estimates that treat neighboring units as unaffected controls may be biased (Clarke, 2017). While some adjustments for spillovers have been proposed (Butts, 2023), they typically require additional assumptions. A more conservative alternative, adopted in this study, is to exclude municipalities adjacent to treated units.

Another complication arises from imbalance in the number of treated and untreated units. Several authors have pointed out that, when only a small fraction of the sample is treated, weighting estimators (including Callaway and Sant'Anna's doubly robust estimator) may place disproportionate weight on a small set of controls, inflating variance and reducing robustness (Crump et al., 2009; Huber et al., 2013; Khan & Tamer, 2010; Yang & Ding, 2018). Common responses like trimming (excluding poorly comparable units) or switching to pure outcome regression come with trade-offs (Callaway & Sant'Anna, 2021b, p. 21), so recent work has proposed to instead match treated units to structurally similar controls before estimation (Ballinari, 2024; Imai et al., 2023). The present study follows this logic, adapting a matching-based procedure to preserve the ATT estimand while improving balance.

Finally, the question of heterogeneous treatment effects has re-emerged as a key topic in the DiD literature. Labor market impacts of hospital closures may plausibly vary by gender or degree of rurality. While the estimator by Callaway and Sant'Anna allows for covariate adjustment, it does not directly recover distinct subgroup-specific effects. More flexible tools for modeling heterogeneity have shown promise in other designs (Chernozhukov et al., 2018; Wager & Athey, 2018), but their applicability to DiD settings remains an open question (Roth et al., 2023) that would go beyond the scope of this project. Exploratory subgroup analysis, though limited in terms of inference, remains a pragmatic strategy to descriptively assess whether treatment effects appear consistent across key dimensions and complements the main analysis in this study.

In sum, the reviewed literature provides both the substantive and methodological foundations for this study. Theoretical and empirical work on hospitals as local anchor institutions highlights the potential for closures to generate lasting labor market disruptions, while recent advances in DiD estimation offer causal tools to identify such effects. Yet most existing studies are rooted in the U.S. context, leaving open the question of how these dynamics unfold under different institutional conditions. The following chapter therefore introduces the German case, providing the necessary background on hospital structures, planning regimes, and closure patterns that shape the empirical analysis.

3 CASE STUDY: GERMAN *PLANKRANKENHÄUSER*

This chapter introduces the empirical setting of the study, outlining how hospital closures occur in Germany and why they matter economically. It builds on the conceptual foundations established in the literature review by providing background on the German hospital landscape, with particular attention to the role of hospitals in local employment and the institutional context in which closure decisions are made. This includes the planning framework that governs hospital operations as well as the recent reform debate that has brought hospital consolidation to the forefront of policy discussions.

With more than 1.4 million employees, hospitals are among the largest employers in Germany (Simon, 2016, p. 381; Statistisches Bundesamt, 2025c). Their particular relevance for female employment becomes clear when considering that women constitute roughly three-quarters of the hospital workforce (Statistisches Bundesamt, 2025a). Beyond those directly employed by hospitals, two additional tiers of employment amplify their economic importance. First, hospitals support a wider local service ecosystem. Businesses such as pharmacies, restaurants, transport services, and retail establishments often depend on hospital staff and patient traffic. Second, many facilities have outsourced non-clinical services such as cleaning, catering, or technical support. Although these workers continue to operate on hospital premises, they are excluded from official hospital employment statistics, creating a sizable ”hidden workforce” whose vulnerability becomes visible when closures occur (Simon, 2016, pp. 381—384). The closure of a hospital in Germany therefore affects the local labor market often in at least three ways: a) through its direct employment links, b) through its outsourced workforce operating on hospital premises; and c) through external actors whose economic activity is indirectly tied to the hospital’s presence. The shutdown of a facility can therefore be expected to trigger ripple effects throughout the municipality’s economy (Simon, 2016, pp. 141, 150).

Understanding the dynamics of hospital closures in Germany requires attention to the institutional planning structures that determine where hospitals operate and where they are at risk of closure. The German hospital landscape is deeply shaped by its federal planning system. Each state is required to develop a hospital plan (*Krankenhausplan*) that serves two core goals: (1) ensuring sufficient regional healthcare provision and (2) regulating costs and capacity by managing the number and type of hospitals (Van den Berg et al., 2019). Hospitals included in these plans are designated as *Plankrankenhäuser* (plan hospitals). As of 2010, roughly 98% of beds in general hospitals were located in *Plankrankenhäuser* (university hospitals included), underlining their central role in both healthcare delivery and regional employment (Simon, 2016, pp. 377–378). Their status grants them a formal care mandate, eligibility for state investment funding, and is the main path to reimbursement under the statutory health insurance system (Fischer, 2024, pp. 52–56; Porter & Guth, 2012). Inclusion in or removal from the state plan plays a decisive role in determining hospitals’ long-term financial viability. Hospital closures in Germany are therefore not merely market-driven events but in reality often reflect state-level planning decisions, distinguishing them clearly from closure dynamics in the US. Plankrankenhäuser are the most directly affected by planning reforms and therefore central to debates on hospital consolidation.

Against this institutional backdrop, the current national hospital reform has reignited long-standing debates - framed either as much needed efficiency gains through consolidation or as severe “Krankenhaussterben” (hospital die-off), depending on one’s perspective (Berger et al., 2024). However even before this reform, Germany had seen a steady decline in hospital numbers: Over the course of the past 30 years, official records indicate a 20% decline in total hospitals numbers (Statistisches Bundesamt, 2025b). While these counts must be interpreted with caution (as discussed in the data section), they nonetheless indicate a clear and ongoing contraction of the hospital landscape. These institutional and policy dynamics position Germany as a valuable case for examining the labor market consequences of hospital restructuring. Its combination of a dense hospital network, a formalized planning regime, and a sustained record of state-induced closures offers a unique empirical setting for understanding how public health policy impacts local labor markets.

4 DATA

With the institutional context established, this chapter describes the construction of the dataset used to estimate the employment effects of hospital closures in Germany. It outlines the main data sources, the closure validation strategy, and the variables used to assign treatment and measure employment outcomes. To ensure analytical precision and policy relevance, the analysis focuses on confirmed closures of hospitals classified as *Plankrankenhäuser* prior to closure, ensuring that treated hospitals had public service relevance and were directly affected by hospital planning. The study period for identifying hospital closures spans from 2010 to 2020 and captures a full decade of consolidation prior to the COVID-19 pandemic, avoiding confounding effects from pandemic-related labor market disruptions and regional policy responses. Employment outcomes draw on municipal panel data from 2006–2020, allowing for stable pre-treatment baselines and dynamic post-treatment estimation.

4.1 Data on Hospitals and Hospital Closures

Hospital closures are identified using two primary sources. The *Annual Hospital Quality Reports* (Gemeinsamer Bundesausschuss, 2023) provide institutional identifiers for longitudinal tracking, while the *hospital register* (Statistische Ämter des Bundes und der Länder, n.d.) is used to verify whether a hospital was a *Plankrankenhaus* prior to closure. Identifying closures can be challenging due to inconsistencies in official data: Changes in institutional IDs and hospital classification, administrative mergers, and differing definitions of what constitutes a hospital contribute to discrepancies across sources (Simon, 2016, pp. 144, 151). Previous research suggests that only a fraction of hospitals disappearing from administrative datasets reflect actual closures (Preusker et al., 2014).

A multi-step validation strategy was used to ensure accuracy. Institutional IDs were tracked over time, and facilities that reappeared under different IDs were excluded. Google Maps and local news reports were used to verify closure status that is typically well-documented given the political and economic relevance of such events. Where uncertainty remained, direct inquiries were sent to local and state-level health authorities. Finally, hospital names and addresses were cross-checked against the hospital register to confirm classification and closure.

This rigorous process resulted in a validated set of confirmed closures of *Plankrankenhäuser*. Given the treated-to-control unit imbalance, prioritizing false negatives over false positives is a deliberate trade-off that follows established research practice and strengthens internal validity.

4.2 Employment Outcomes and Covariate Structure

Municipality-level data from the *Wegweiser Kommune* platform, provided by the Bertelsmann Stiftung form the basis for estimating labor market effects (Bertelsmann Stiftung, 2023). The main outcome is the employment rate, defined as the share of employed individuals in the working-age population (18-64). Gender-disaggregated data enable analysis of potential differences given the gendered structure of hospital employment. Data coverage is incomplete and limited to municipalities with more than 5000 inhabitants. Missing values vary across municipalities and indicators, reducing the usable sample and shaping treatment–control composition. Nonetheless, it remains one of Germany's most consistent municipal labor market sources.

To adjust for structural differences between municipalities, the study applies the *RegioStaR* typology developed by the Federal Ministry for Digital and Transport (Bundesministerium für Digitales und Verkehr, 2022). The primary specification uses a modified version yielding four analytically tractable types: (1) Urban-City, (2) Urban-Surrounding, (3) Rural-City, and (4) Rural-Village;[3] as well as an aggregated version for binary rural–urban comparisons. During the DiD pre-matching stage (see Methodology chapter), a more granular 6-type-classification is used to increase structural similarity between treated and control. While for estimation coarser groupings are chosen to preserve statistical power, matching does not involve statistical estimation, and thus allows finer stratification. This staged approach balances structural nuance and comparability while ensuring reliable inference.

4.3 Descriptive Patterns and Estimation Design

The validation procedure identified 56 confirmed closures.[4] A comparison with official statistics shows that only 31% of reported net exits represent confirmed closures, a figure highly consistent with prior studies (see Appendix A.1). This highlights both the validity of the identification strategy used here, as well as the need for caution when interpreting official counts.

This dataset offers new descriptive insights into the regional closure patterns and the types of municipality most affected, insights that would remain obscured in official statistics alone. Figure 1 presents the spatial distribution of confirmed closures highlighting a clear concentration in the Western German states of Lower Saxony, North Rhine-Westphalia, and Baden-Württemberg; a regional pattern that matches subsequent trajectories (Bündnis Klinikrettung, 2024). Table 1 shows that while rural closures dominate public discourse, over

[3]Throughout the paper, the "Metropolis" and "Large City" categories from the RegioStaR typologies are merged due to the small number of treated units in each. This improves covariate balance and is theoretically justified, as no systematic differences in closure effects are expected: both types likely share key institutional and labor market characteristics relevant to the analysis.

[4]Table 1 includes additional confirmed closures for which the exact year of closure could not be determined. These cases were excluded from subsequent steps, see Appendix A.2 for details.

half of sample closures occurred in medium-sized urban and rural cities, and notably, more than 15% in large cities.

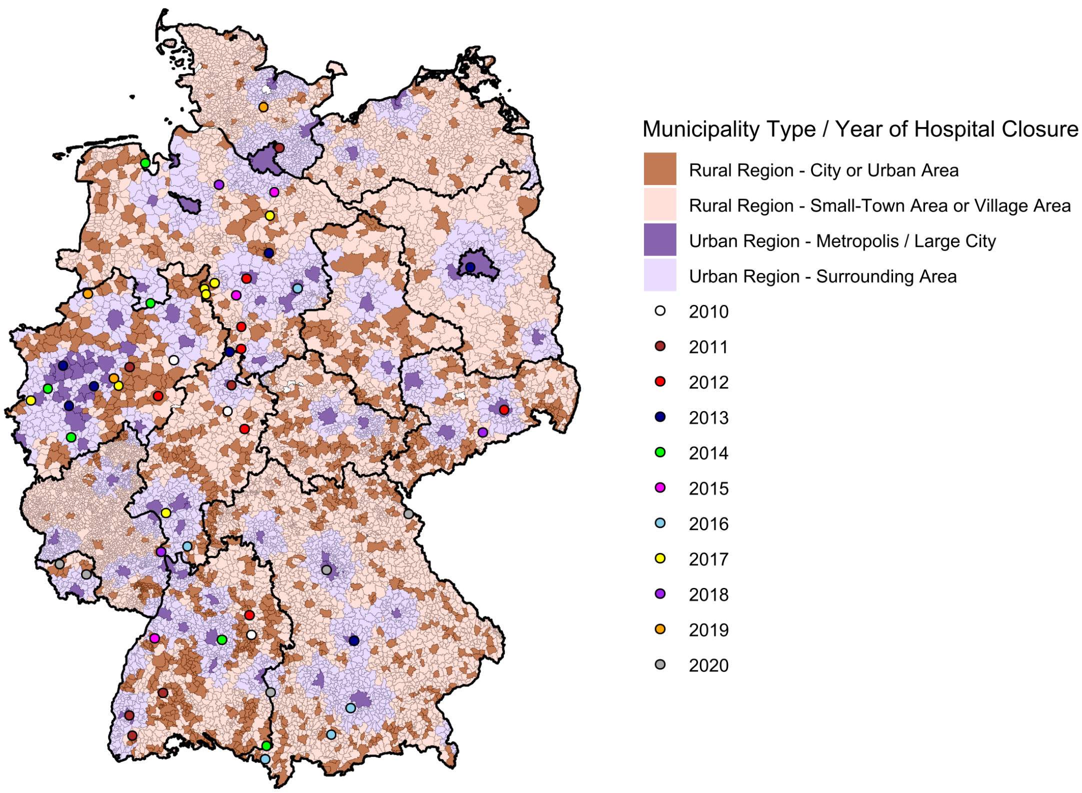

Figure 1. Hospital Closures in Germany, 2010–2020. *Points represent confirmed hospital closures, identified through multi-source validation. Background shading indicates municipality type, based on the RegioStaR typology.*

Taken together, these patterns highlight the contribution of the dataset beyond causal identification: By validating and mapping hospital closures, the study adds new descriptive insight into the geography of healthcare retrenchment in Germany during the 2010s. The next chapter outlines the study's empirical strategy in detail, including how closures are matched to control units, how dynamic effects are estimated over time, and how subgroup variation is accounted for in light of the structural patterns described above.

5 METHODS

This chapter presents the methodological framework for estimating the causal impact of hospital closures on local employment in Germany. The analysis builds on the estimator developed by Callaway and Sant'Anna, whose properties make it well-suited for evaluating policy interventions like hospital closures. As outlined in the literature review, its design aligns

Region Type	Municipality Type	Without Closure	With Closure	Share with Closure (% of type)	Share of All Closures (% of total)
Urban	Metropolis / Large City	70	12	14.6%	17.4%
	Surrounding Area	1374	18	1.3%	26.1%
	Small-Town / Village Area	1895	4	0.2%	5.8%
Rural	Central City	99	9	8.3%	13.0%
	Urban Area	1136	15	1.3%	21.7%
	Small-Town / Village Area	6335	11	0.2%	15.9%
Total		**10909**	**69**	—	**100%**

Table 1. Distribution of Confirmed Hospital Closures by Municipality Type (2010–20).

with the empirical structure of this study, though adaptations are made to address specific contextual challenges.

The estimator computes group-time average treatment effects ($\text{ATT}_{g,t}$), representing the difference between observed employment outcomes and the counterfactual trajectory in the absence of treatment, for each cohort g (defined by year of closure) and time period t (relative to the year of closure). These effects are estimated *on the treated* (municipalities that experienced a hospital closure) and capture dynamic post-treatment trajectories over time. Following Callaway and Sant'Anna (2021b), and given the large control pool, the control group consists of municipalities that remain untreated throughout the observation window (rather than including those that had not yet experienced a closure by the time a treated unit was exposed). $\text{ATT}_{g,t}$ estimates are aggregated across cohorts and time periods using a weighted average, with weights reflecting the number of treated units in each cohort. This yields both an overall ATT and a dynamic treatment profile by year relative to treatment, capturing how employment evolves after closure. Estimation is implemented using the `did` package by Callaway and Sant'Anna (2021a).

5.1 Key Assumptions

The Callaway and Sant'Anna estimator relies on a set of identifying assumptions, some slightly differing from those common in other DiD-settings. This section outlines the most relevant assumptions and how they are addressed in the present study, both conceptually and empirically.

First, as with all DiD estimators, causal identification requires that, absent treatment, treated and control units would have followed similar outcome trajectories. In the Callaway and Sant'Anna framework, this *parallel trends assumption* applies within treatment cohorts and may be relaxed to hold only *conditionally* on observed covariates. This assumption is assessed by comparing pre-treatment employment trends, disaggregated by municipality type. The results of this diagnostic are presented in the main results section.

Second, the *no treatment revocation assumption* requires that once a unit receives treatment, it remains treated. Since the dataset compilation defines hospital closures as permanent events, this assumption holds here by design.

Third, the estimator assumes that treatment does not affect outcomes before implementation. Given the possibility of workforce reductions, service wind-downs, or staffing adjustments due to uncertainty, this *no anticipation assumption* might be challenged in the context of hospital closures. To assess this, an alternative specification is estimated that allows for anticipation effects in the year prior to closure from the model.

Lastly, the estimator assumes that outcomes in untreated municipalities are unaffected by treatment in nearby areas. In the given context, the *no spillover assumption* raises a legitimate concern: given hospitals' role as regional employers and service hubs, closures may indirectly affect neighboring labor markets through commuting ties, market linkages, or shifts in healthcare demand. Recent methodological work has proposed formal DiD frameworks for estimating such spillovers (Butts, 2023). However, applying such models would require strong assumptions about spillover direction, intensity, and consistency across municipality types; or alternatively necessitate estimating heterogeneous effects across many treated–control constellations. As this effort not only lies beyond the scope of this study but, more importantly, exceeds what the available treatment pool reasonably allows, this study adopts a conservative exclusion strategy: all municipalities that share a direct border with a treated unit are excluded from the control group, ensuring that remaining controls serve as unaffected counterfactuals. This also aligns with the study's focus on local employment effects within treated municipalities, rather than broader regional consequences. Although reducing the number of available controls, the large initial pool maintains sufficient power. This trade-off prioritizes internal validity and minimizes contamination risk.

5.2 Estimation Under Treatment-Control Imbalance

Despite sound identification assumptions, treatment imbalance presents practical challenges. This section explains how these are addressed within the estimation strategy, beginning with an overview of the three counterfactual construction approaches proposed by Callaway and Sant'Anna. *Outcome Regression (OR)* models the untreated outcome as a function of observed covariates. This method is straightforward but relies on correct model specification; misspecification can lead to bias. *Inverse Probability Weighting (IPW)* reweights control observations based on their estimated probability of treatment, aligning the covariate distribution of the control group with that of the treated group. However, misspecification of the propensity score model can lead to distorted weights and biased estimates. *Doubly Robust (DR)* estimation combines *OR* and *IPW*, offering increased robustness to misspecification in either component. This is the recommended default in most empirical applications (Callaway & Sant'Anna, 2021b, pp. 205–206, 212).

However, while the DR estimator is more robust than either *OR* or *IPW* alone, its performance may still be undermined by strong treatment–control imbalance. As discussed in the literature review, such imbalances are more than technical nuisances: they can lead to inflated weights for a small number of control units, increasing variance, reducing estimator stability, and potentially introducing bias. Callaway and Sant'Anna acknowledge this issue but set it aside after briefly noting two potential responses: excluding units with extreme estimated probabilities of treatment (trimming) or switching to a pure OR specification. Yet, the authors note that both approaches involve trade-offs: trimming can improve weighting performance but changes the estimand by excluding part of the treated sample, while OR eliminates weighting altogether but again increases sensitivity to functional form assumptions.

To address this issue without compromising estimator validity, this study adapts the pre-weighting matching-based strategy proposed by Ballinari (2024). While originally developed in the context of DML-based ATE estimation, its core idea of pre-matching treated and control units to improve balance *before* applying a weighted estimator translates well to ATT estimation: the key modification is that the calibration step is omitted, as the goal is not population representativeness but credible inference for treated units. Because the matched sample is fixed by design, this approach does not support resampling-based inference, requiring a shift to symmetric confidence intervals instead (see Appendix B.1). Although limiting the use of bootstrap diagnostics, it allows the estimator to address instability from extreme weights without altering the estimand. Building on this setup, all treated units are matched to a subset of structurally similar controls based on key covariates, in this case, the municipality type. DR estimation is then applied to the matched subsample. This approach reduces the influence of extreme weights, improves covariate overlap, and preserves the original ATT target parameter, as all treated units are retained. To assess the reliability of this approach, the matching and weighting steps are repeated across multiple subsamples. If ATT estimates remain stable across replications, this suggests estimator robustness. Conversely, large variation would indicate sensitivity to the choice of control units, a potential source of concern.

In sum, the analysis compares three estimation strategies under treatment–control imbalance: a default DR estimator (following Callaway and Sant'Anna), a baseline Outcome Regression (OR), and a pre-matched DR estimation (following Ballinari).. Comparing these approaches ensures that results are not driven by sample imbalance and that the methodology fits the empirical context.

5.3 Subgroup and Robustness Analysis Design

To assess robustness beyond assumption diagnostics, the study adds theoretically motivated subgroup models and placebo tests. First, exploratory *subgroup analyses* are used to probe potential heterogeneity in treatment effects suggested in prior studies. These take two complementary forms: models are estimated separately for municipality subgroups (e.g., urban vs. rural), and outcome variables are disaggregated to explore differential effects across subpopulations within treated units (e.g., by gender). Both approaches target different dimensions: variation at the unit level and variation within treated populations. These analyses are descriptive and do not involve formal statistical tests of between-group differences.

Secondly, two *placebo tests* are conducted. In the first, placebo treatment timing is assigned to actual treated municipalities at an earlier date. In the second, a set of untreated municipalities is randomly assigned a placebo treatment. In both cases, finding systematic effects would raise concerns about the design's internal validity.

Taken together, these tests strengthen the credibility of the study's findings by addressing potential violations of identification assumptions, weighting instability, and contextual outliers. They complement the triangulation across estimators that helps ensuring that results are not overly influenced by model specification or treatment–control imbalance. This design combines recent causal inference advances with adaptations to the study's structure and context, providing a robust foundation for the analysis.

6 RESULTS

With the estimation framework established, this chapter presents the empirical findings: it first assesses core assumptions, then reports main effect estimates across the three main models, and finally explores robustness and potential subgroup variation.

6.1 Assumption Checks

This section evaluates the key assumptions for causal identification: conditional parallel trends, no anticipation effects, and no spillovers. Evidence presented below suggests that these assumptions plausibly hold in the context of this study. Where appropriate, additional sensitivity analyses are provided in the appendix to complement the findings reported.

To assess the parallel trends assumption, Figure 2 displays pre-treatment employment trajectories for treated and control municipalities, disaggregated by municipality type. Trends appear closely aligned, with only minor divergences, mainly driven by municipalities exiting the sample after their hospital closure. This supports the assumption that, conditional on structural characteristics, treated and control municipalities followed similar pre-treatment paths.

The main models include the full pre-treatment period. To test robustness, an alternative specification reclassifies the final pre-treatment year as part of the treatment period. As shown in Appendix C.1, the results remain broadly consistent: while estimates shift slightly and post-treatment confidence intervals widen, the overall effect trajectory is stable. This suggests that anticipatory behavior, and thus violation of the no-anticipation assumption, did not meaningfully distort the main findings.

To minimize bias from potential spillover effects, the main estimation sample excludes all municipalities that share a direct border with a treated unit, as well as units with unclear treatment status and their immediate neighbors. Figure 3 visualizes the resulting classification, taking into account only municipalities with employment data available. The estimation sample includes 42 directly treated municipalities, 371 excluded due to spillover risk, and 11 removed for unclear treatment status, leaving 2,303 as eligible controls. A more detailed comparison of the final and original sample compositions and an alternative model that reintroduces the excluded spillover-prone municipalities that yields slightly smaller treatment effects are provided in Appendix C.2 and Appendix A.2. This suggests that the exclusion strategy might provide a conservative bias correction. Moving beyond these key assumptions, the next section presents the core estimates and compares identification strategies.

6.2 Core Estimates of Employment Effects

This section presents the core findings of the three estimation strategies: (1) the default DR estimator, (2) a basic OR model, and (3) a DR estimator applied to a pre-matched subsample. All models share the same treatment definition and outcome variable. The analysis focuses on the eight years after treatment to ensure comparability and avoid inflated standard errors due to declining sample size in longer post-treatment windows.

The *default DR model* is estimated on the full sample DiD-sample, using inverse probability weighting and regression adjustment, controlling for a four-category urban–rural typology to balance regional heterogeneity. The *baseline OR model* applies the same covariates but

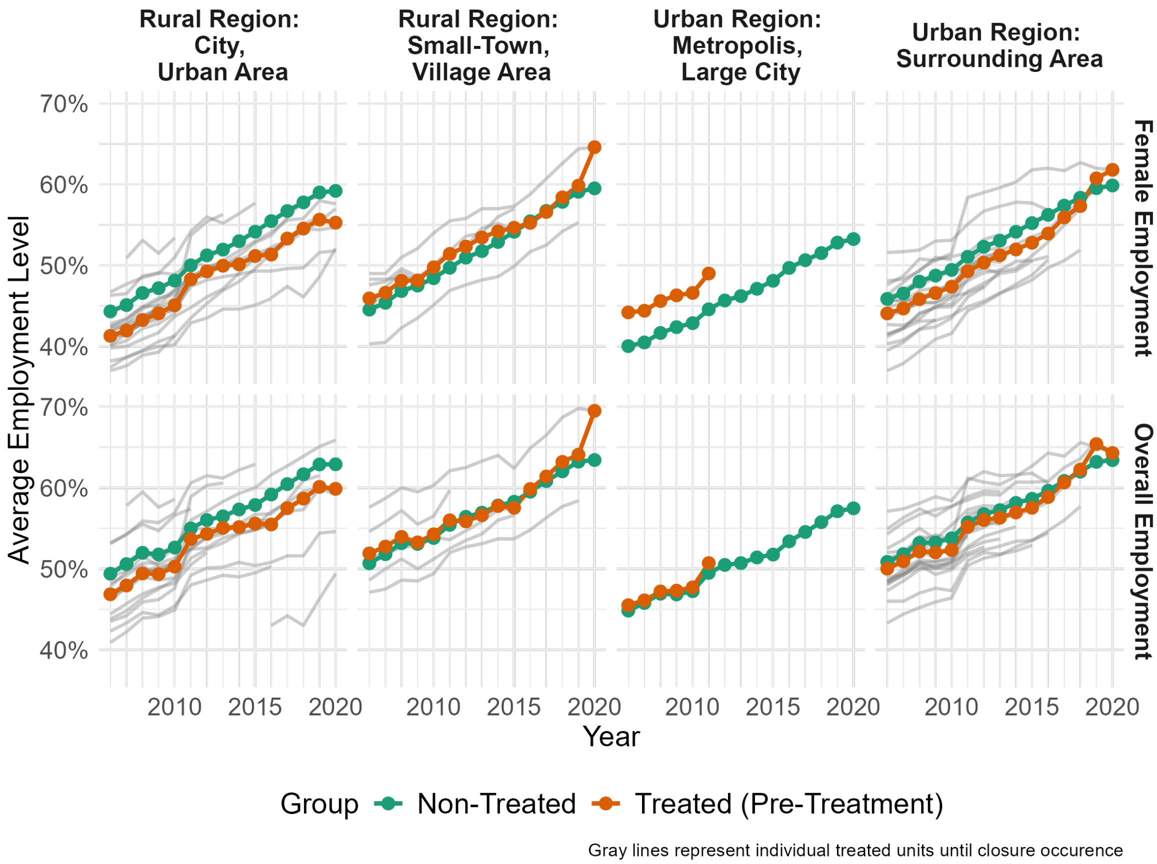

Figure 2. Parallel Trends in Employment Prior to Closure. *Average employment levels over time for treated municipalities (prior to closure) and control municipalities, conditioned on municipality type.*

omits the weighting component. The *pre-matched DR model* follows the default DR approach but only after each treated municipality is matched to five controls based on a six-category municipality typology. Matching is repeated 100 times, with DR estimation applied to each resulting sample and the reported results being based on the mean ATT across these runs (see Appendix B.3 for more detailed results).

Table 2 presents the ATTs across the three specifications. All models indicate a statistically significant negative effect of closure on employment, with ATT estimates ranging from -0.55 to -0.5. The default *DR* and *OR* estimators yield virtually identical point estimates; suggesting that the weighting component in the *DR model* adds little to the estimation. In contrast, the *pre-matched DR model* yields a slightly stronger effect estimate (–0.55). Nonetheless, the effect estimate remains substantively in line with the *DR* and *OR* models.

All three models show consistent dynamic patterns: pre-treatment coefficients fluctuate around zero, while post-treatment effects become increasingly negative, reaching magnitudes that well exceed one percentage point in the final years of observation. The dynamic estimates from the full-sample models are nearly identical throughout, reinforcing that inverse probability

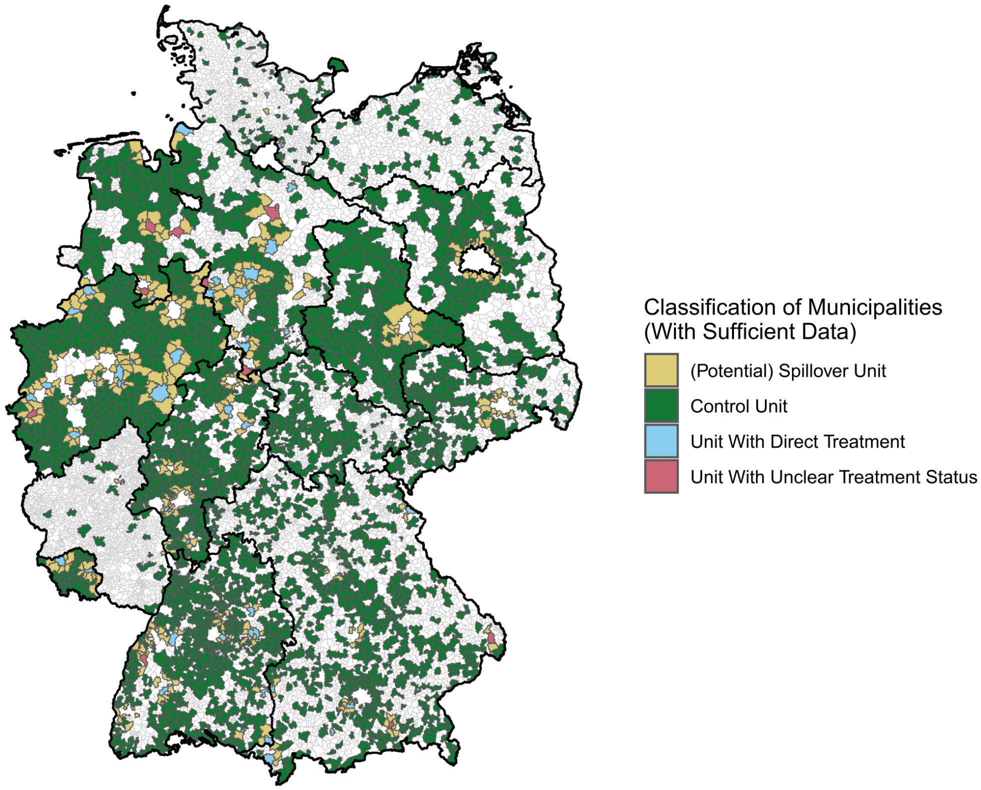

Figure 3. Treatment Status of German Municipalities, 2010–20. *Unfilled areas lack necessary data for analysis.*

weighting adds little in this context. The *pre-matched DR model* produces more pronounced effects, reaching a maximum of –1.36 percentage points (compared to –1.2 with the other models). This stronger effect is also relatively stable across iterations, which may indicate that estimates based on the full sample slightly understate the true effect size. Aside from this difference, the temporal patterns remain highly consistent across all models in terms of direction, timing, and scale, which strengthens confidence in the robustness of the results.

Given the consistent results across all three estimator strategies, the default DR model is used as the primary specification for subsequent robustness checks and extensions. While the pre-matched DR model provides hints that the full-sample models may somewhat understate the treatment impact, the computational demands of iterative matching make it impractical to extend to the full set of analyses. More importantly, the default DR model retains bootstrap-based confidence intervals, which offer valuable diagnostic insights into irregularities such as skewness or estimator instability. In contrast, the pre-matched model necessitates the use of symmetric confidence intervals throughout, which may obscure such issues (see Appendix B.1 for a discussion of this limitation). Since the default bootstrap inference is recommended by Callaway and Sant'Anna (2021b) and performs well in the non-matched setting, it provides a

Model Specification	ATT Estimate	Standard Error	95% Confidence Interval
Default DR	–0.50%	0.18	[–0.85%, –0.15%]
Basic OR	–0.50%	0.18	[–0.86%, –0.14%]
Pre-Matched DR (Average)	–0.55%	0.18	[–0.90%, –0.19%]

Table 2. ATT Estimates of Employment Effects Across Estimation Models. *Values show average percentage point changes in municipal employment following hospital closures, based on three model types: doubly robust (DR), outcome regression (OR), and pre-matched DR. All confidence intervals are symmetric; see Appendix B.1 for details.*

principled and efficient basis for further testing.

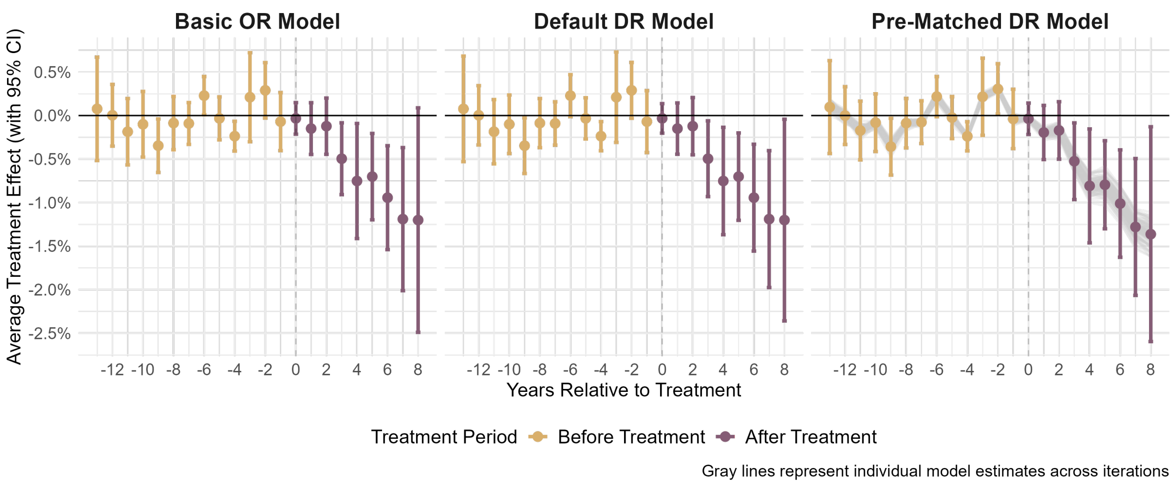

Figure 4. Dynamic Treatment Effects Across Three Model Specifications. *Each panel shows employment effects over time with symmetric 95% confidence intervals. Gray lines in the matched panel indicate results from 100 iterations.*

6.3 Subgroup Effects and Robustness Findings

This section reports additional robustness checks, including placebo tests and subgroup analyses. Two strategies are used: placebo tests to rule out spurious correlations, and subgroup analyses to explore potential effect heterogeneity. The latter includes both differences across subgroups of treated municipalities (e.g., urban vs. rural contexts) and across subpopulations within treated units (e.g., gendered employment outcomes). These analyses follow the robustness design outlined in the methods chapter and complement the main estimator results presented above.

Two placebo tests are used: one assigns treatment randomly to untreated municipalities, the other assigns placebo timing to actual treated municipalities, several years before the real closure. In both cases, the same estimators are applied. Neither placebo design produces statistically significant effects in the post-treatment period (see Appendix C.3). This provides

Figure 5. Dynamic Employment Effects by Subgroup. *Each panel shows estimated treatment effects over time with 95% bootstrap confidence intervals, based on the default DR estimator. Results are disaggregated by urban–rural classification and gender.*

reassurance that the observed treatment effects in the main analysis are not artifacts of underlying trends, spurious correlations, or sample structure.

To explore whether hospital closures have differential effects depending on the structural context of a municipality as suggested by some US-studies, separate models are estimated for urban and rural areas.[5] Figure 5 shows the time-varying ATT estimates for each group, and Table 3 reports the corresponding average effects. Point estimates are nearly identical between urban and rural municipalities (–0.63 and –0.62, respectively), with wider confidence intervals for rural areas. While previous literature has suggested that closures may have more pronounced effects in rural contexts, the results from this sample offer no clear evidence in favor of that hypothesis. However, sample composition must be considered: most treated municipalities in the final estimation sample are medium-sized urban and rural towns, whereas very large cities and remote village areas are underrepresented. The resulting pattern therefore does not fully reflect the distribution of hospital closures in Germany described in the Data chapter (see Figure 3 and Appendix A.2), and the subgroup analysis therefore effectively compares a narrower band of urban and rural municipalities with broadly similar structural characteristics, limiting generalizability, especially for municipalities at the extremes of the urban–rural spectrum.

A second set of subgroup analyses explores whether the observed employment effects differ by gender. Since hospital closures may disproportionately affect sectors with high female employment, a separate model is estimated for female employment. The resulting estimate is somewhat more variable, showing a decline of –0.53 percentage points compared to –0.50 for overall employment, indicating broadly similar effects with minor differences

[5]The binary rural–urban typology is used for subgroup analysis, as the DiD sample is heavily concentrated in two of the four municipality types. Subgroup estimation for the remaining categories is limited by sample size (see Appendix A.2)

	Overall ATT	Overall Standard Error	95% Confidence Interval
Effect on Employment Level by Municipality Type			
Urban Municipality	–0.63%	0.18	[–0.99%, –0.28%]
Rural Municipality	–0.62%	0.37	[–1.35%, 0.10%]
Effect on Employment Level by Gender			
Female Employment	–0.53%	0.20	[–0.93%, –0.13%]
Overall Employment	–0.5%	0.18	[–0.85%, –0.15%]

Table 3. Average Treatment Effects by Gender and Municipality Type. *Estimates are based on the default doubly robust model. All values refer to percentage point changes in municipal employment following hospital closures.*

in magnitude. However, these results do not provide conclusive evidence of gender-specific heterogeneity. Because *all* subgroup models (regarding gender *and* rurality) are estimated independently, no formal statistical tests of between-group differences are conducted. These analyses serve as suggestive robustness checks, not definitive tests of heterogeneity.

Taken together, these results nevertheless reinforce the credibility of the main findings. The absence of significant placebo effects supports the design's internal validity, while subgroup results are in line with the main effect estimates. At the same time, the results underscore the importance of considering sample composition and context when interpreting estimated effects. These considerations set the stage for a critical assessment of how the findings align with previous research and what they may suggest for policy and future inquiry.

7 DISCUSSION

Building on the preceding analysis, this chapter examines the results in relation to the existing literature and reflects on their implications. Rather than generalizing beyond the sample, it aims to contextualize what the estimated effects reveal about local labor market impacts in the observed settings. Broader implications are taken up in the concluding chapter.

The findings provide robust evidence that hospital closures are associated with persistent and statistically significant declines in municipal employment, with average effects estimated between –0.5 and –0.55 percentage points across models, and peak declines reaching –1.2 to –1.36 percentage points toward the end of the observation period. These results hold across all three estimators and a wide range of robustness checks, reinforcing the internal validity of the design. Notably, the dynamic results suggest that employment losses do not occur abruptly but accumulate over time, with no signs of recovery or convergence within the eight-year post-treatment window. This sustained decline is consistent with theoretical expectations of hospitals as anchor institutions: their closure disrupts not only direct healthcare employment but also broader local economic structures. The observed declines and patterns by which thy unfold are plausibly interpreted as the cumulative result of a localized economic shock with sectoral spillover effects within the treated municipality.

Contrary to findings from some U.S. studies, the analysis does not reveal systematically stronger effects in rural municipalities. However, this finding must be interpreted with caution for two reasons. First, the rural and urban subgroups in the estimation sample are dominated by medium-sized municipalities. More remote rural areas and very large cities are underrepresented among treated units compared to the initial overall sample. Second, the analysis compares independently estimated subgroup models, rather than directly testing for differences in treatment effects. This means that the absence of significant divergence in point estimates should not be read as evidence of equivalence and reflects both the limited variation within the treated sample and the descriptive, non-inferential nature of these subgroup comparisons.

A similar logic applies to the gender-disaggregated models. While one might expect disproportionate impacts on female employment given the healthcare sector's demographic profile, the data do not provide strong support for this hypothesis. The results for female employment mirror those for the overall sample in both magnitude and trajectory, though with somewhat wider confidence intervals. Again, this does not indicate that men and women are equally affected, but only that the available data do not offer strong empirical basis to conclude otherwise.

It is furthermore important to underscore that the results do not estimate a population-wide average treatment effect; the findings reflect the specific municipalities in which hospital closures actually occurred instead of a representative cross-section of all German municipalities. The estimation sample is also non-random in other important ways. Municipalities were selected based on confirmed closure events between 2010 and 2020 and further narrowed through the exclusion of units with incomplete data or those at risk of spillover contamination. These exclusions were necessary to ensure credible identification, but they come at the cost of external validity.

Importantly, the estimation sample underrepresents very large cities, where data limitations were more common, and also remote rural municipalities, where closures may have occurred but were harder to confirm retrospectively or lacked complete employment data. If closures in large urban areas lead to smaller employment shocks due to more diversified labor markets, the ATT estimates may slightly overstate national effects. Conversely, if more vulnerable rural areas are underrepresented, results may understate the true impact.

A further boundary of the design is geographic: by construction, the estimated effects reflect the consequences of closure for the municipality in which the hospital was located. They do not capture downstream effects on neighboring areas, such as shifting employment patterns, commuting spillovers, or changes in regional healthcare access. The exclusion of spillover-prone municipalities ensures internal validity but limits the scope of the findings to within-municipality effects.

The observed distribution of closures being clustered in medium-sized municipalities primarily in western Germany, reflects broader institutional and policy dynamics. Closures were more likely where consolidation was administratively feasible and politically acceptable. Extremely rural hospitals may have been protected from closure due to local resistance or policy exemptions. This helps explain the absence of many remote rural units in the final sample and also underscores that closures during this period were shaped by a combination of economic, institutional, and political constraints. This, in turn, shapes how the results should be interpreted. The findings do not speak to a general effect of hospital closures under all conditions, but to the observed consequences in the subset of municipalities where closures actually occurred, a group likely shaped by both structural vulnerability and institutional feasibility. A federal-level intervention like the current reform may shift this balance, potentially leading to different closure patterns emerging in the future.

While descriptive and retrospective in nature, the study's findings carry clear relevance for current policy debates. Hospital consolidation remains a central feature of German healthcare reform, with major structural changes currently underway. The evidence presented here provides a pre-reform benchmark for understanding one key dimension of closure effects: their labor market consequences in affected municipalities. These effects should be considered in future policymaking, especially when assessing trade-offs between efficiency, quality, accessibility, and local economic stability. The findings offer timely insight into an emerging policy challenge. The final chapter summarizes the study's contributions and outlines directions for future research and policymaking.

8 CONCLUSION

This study has examined the local employment effects of hospital closures in Germany between 2010 and 2020. Using a tailored adaptation of the DiD estimator by Callaway and Sant'Anna, it estimates the Average Treatment Effect for municipalities with confirmed closures. Results indicate a clear and persistent decline in municipal employment: average treatment effects range from –0.5 to –0.55 percentage points across models, increasing to between –1.2 and –1.36 points by the end of the eight-year post-closure window. These effects are robust across estimation strategies and placebo tests, and no recovery is observed. While prior literature suggests potential heterogeneity by gender or rurality, this study finds no strong evidence of systematic differences. However, the sample is composed mostly of medium-sized municipalities, with few very large cities or remote rural areas included. Subgroup models are estimated independently and do not support formal comparisons. The absence of variation should therefore be interpreted as a lack of evidence, not confirmation of uniform effects. Estimates reflect local impacts in treated municipalities and should not be generalized to national averages or wider regions.

Beyond these results, the study offers two broader contributions. First, it introduces a practical estimator adaptation for settings with strong treatment–control imbalance, providing an alternative to trimming or outcome regression. Second, it compiles a validated dataset of confirmed hospital closures across a full decade, addressing known gaps in official data and illustrating how aggregate hospital counts can obscure meaningful variation in closure dynamics.

Several directions for future research emerge from the study's findings and its limitations. First, to better assess potential heterogeneity, future research could examine closures in very large cities or more remote rural areas. While theory suggests contextual variation, this study's sample does not support inference on that dimension. Methods such as the Augmented Synthetic Control Method (Ben-Michael et al., 2021) may enable estimation in smaller subgroups if longer pre-treatment trajectories are available. Second, the range of labor market outcomes should be expanded. While this study focuses on overall employment levels, important aspects like wages, job stability, and full- vs. part-time composition remain unexplored. Third, the dataset developed here could support comparative work across European health systems. Denmark, for instance, offers compatible data on closure patterns (Nyholt, 2024). Finally, forward-looking ex-ante evaluations would add policy-relevant insights, especially in light of recent reform efforts in Germany, where political actors have criticized the lack of anticipatory impact assessments (Deutsche Krankenhausgesellschaft, 2024).

Together, the findings bring empirical clarity to a rarely quantified aspect of healthcare restructuring. They support accounts of hospitals as local anchor institutions and offer timely evidence for ongoing policy debates. While the study takes no normative stance on closures, it makes one point clear: they carry lasting labor market consequences.

APPENDIX

A DATA SOURCES AND SAMPLE COMPOSITION

A.1 Validating Hospital Closure Identification Against Official and External Sources

Official statistics on hospital numbers in Germany differ substantially across sources and years. These reflect differing legal definitions and reporting frameworks across institutions, rather than errors. Even within a single source, year-to-year counts may reflect mergers, reclassifications, or billing status updates rather than actual closures (Reimbursement Institute, 2019). While useful for tracking system-level dynamics, they do not reliably capture local closures, which is what the present study aims to identify.

To assess the reliability of the present study´s approach, three measures of hospital contraction are compared in Figure 6: net changes in official counts from the Federal Statistical Office, confirmed closures compiled in this study, and those reported by Preusker et al. (2014). Despite minor definitional differences, both use largely aligned validation methods, making Preusker et al.'s dataset a valuable benchmark.

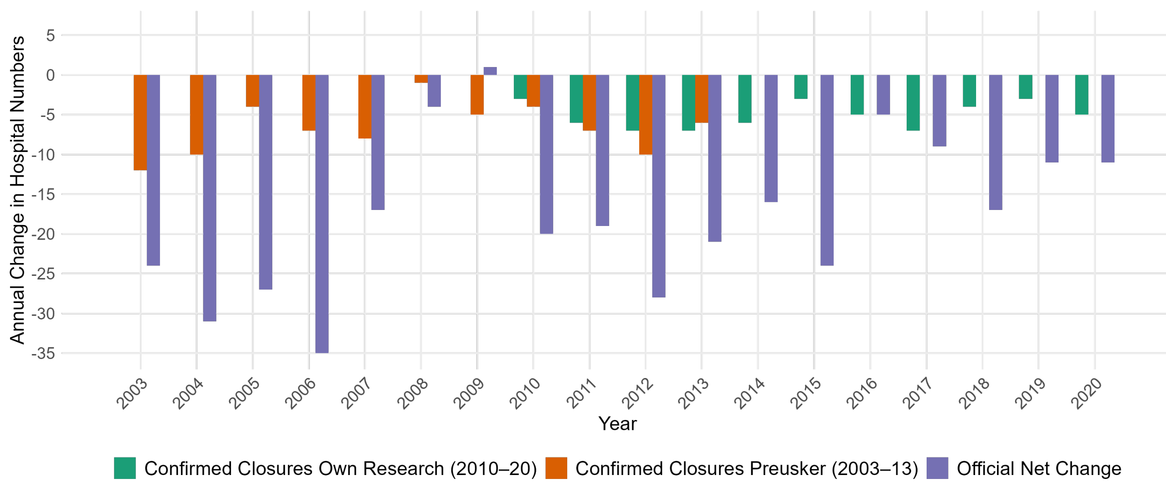

Figure 6. Annual Hospital Closure Counts by Source (2003–20). *The figure compares official net changes in total hospital numbers with independently confirmed closures reported by Preusker et al. (2014) and those compiled in this study.*

The similarity between the two validated datasets is striking: Preusker confirms 33% of net exits between 2003 and 2013; this study finds a near-identical 31% confirmation rate for 2010 to 2020. Compiled independently and covering different periods, both arrive at almost identical rates, which strongly supports the robustness of the method applied here.

This convergence highlights two key points: First, that official net change figures should not be interpreted as direct evidence of closures; and second, that a registry-based confirmation strategy offers a more reliable basis for evaluating the local employment effects of hospital closures.

A.2 Tracking Municipality Coverage from All Closures to DiD Sample

Table 4 shows confirmed closures by municipality type for the full dataset and the subset used for DiD estimation. The rightmost columns offer two complementary perspectives: the share of municipalities within each type that experienced a closure, and the share of all confirmed closures that occurred in each type.

Region Type	Municipality Type	Without Closure	With Closure	Share with Closure (% of type)	Share of All Closures (% of total)
Urban	Metropolis / Large City	70 (10)	12 (1)	14.6% (9.1%)	17.4% (2.3%)
	Surrounding Area	1374 (903)	18 (17)	1.3% (1.9%)	26.1% (39.5%)
	Small-Town / Village Area	1895 (240)	4 (2)	0.2% (0.8%)	5.8% (4.7%)
Rural	Central City	99 (53)	9 (5)	8.3% (8.6%)	13.0% (11.6%)
	Urban Area	1136 (599)	15 (13)	1.3% (2.1%)	21.7% (30.2%)
	Small-Town / Village Area	6335 (490)	11 (5)	0.2% (1.0%)	15.9% (11.6%)
Total		**10909 (2295)**	**69 (43)**	—	**100% (100%)**

Table 4. Municipality Types by Treatment Status in Full and DiD Samples (in parentheses). *Percentages show closure rates within each type and the type's share of all confirmed closures.*

In the full sample, confirmed closures are most common in medium-sized cities and urban areas in urban or rural regions. Combined, they represent roughly half of all confirmed closures. The highest closure rates occur in large cities, where roughly 15% experienced a closure during the study period. This is expected, as large urban areas typically host multiple hospitals, making individual closures more likely without necessarily reducing access to care. While comprehensive hospital counts by municipality type are unavailable, a substantial share of all hospitals is likely located in large cities, further increasing the likelihood of observed closures in this group.

In contrast, closures in rural small-town and village areas are rare in both absolute and relative terms. Closures in these areas may be harder to detect due to limited digital media coverage; however, there is no evidence of underreporting skewed toward the early part of the observation period. This weakens the concern that low rural closure rates stem solely from detection bias. More plausibly, many of the over 6,000 municipalities in this category never hosted a hospital to begin with. Moreover, when hospitals do exist in such areas, institutional safeguards apply: under Germany's state-level planning regulations, hospitals in rural regions cannot easily be removed from the hospital plan unless sufficient coverage is maintained.

Turning to the DiD estimation sample, which includes only municipalities with complete employment panel data and excludes those with potential spillover exposure, medium-sized municipalities remain central. Large urban areas are often dropped due to inconsistent employment data,and the *Wegweiser Kommune* dataset excludes very small municipalities with fewer than 5,000 inhabitants from the outset. This exclusion criterion also lends indirect support to the assumption that many untreated municipalities in the most rural category likely never hosted a hospital: while nearly half of the treated municipalities in this group remain in the DiD sample, only a small fraction of untreated units do. Since small population size is a key reason for exclusion, this suggests that many of the untreated rural municipalities are particularly small and thus unlikely to have supported a hospital in the first place. As a result, the DiD sample is heavily concentrated in two municipality types, with very few treated units from the most rural and most urban categories. Given this skew, a binary rural–urban distinction is more appropriate for subgroup analysis, as finer classifications offer little additional insight. This framing aligns the analysis with the set of municipalities for which reliable treatment and outcome data are available.

B ESTIMATION MODELS AND SPECIFICATION VARIANTS

B.1 Assessing Confidence Intervals Across Estimation Strategies

This section explains the methodological deviation from the default confidence interval construction provided by Callaway and Sant'Anna (2021a). This adjustment applies only to the model comparison involving the pre-matched DR specification in Subchapter 6.2. There, bootstrap-based confidence intervals produce systematic irregularities. To ensure comparability within that particular model comparison, symmetric confidence intervals ($\pm 1.96 \times$ SE) are reported across all three specifications discussed (default DR, basic outcome regression, and pre-matched DR). Outside of this comparison, all models that do not involve matching rely on the default bootstrap inference provided by the `did` package to evaluate robustness checks and alternative specifications of the default DR estimator.

Across repeated runs of the pre-matched DR estimator, a systematic anomaly in bootstrap-based inference emerged: the bootstrap percentile CIs consistently failed to contain the point estimate of the simple ATT. The fact that this happened despite the point estimates being highly stable across runs and closely aligned with those from the default DR and outcome regression models seems to suggest that the problem is likely a systematic bias in the bootstrap distribution used for inference. This reflects a known methodological issue (Austin & Small, 2014; Hill & Scott, 2010): the regularity conditions that justify the bootstrap can break down in the presence of matching (Roth et al., 2023, p. 26; Abadie & Imbens, 2008).

In the current context, control units are selected through stratified random matching based on covariates, forming a fixed, reduced dataset with improved covariate balance that is then passed into the difference-in-differences estimator. However, when bootstrap inference is applied *after* matching, the resampling does *not* reapply the matching procedure. Instead, it draws repeated samples from already matched data. It treats this as a random sample from the population, which it is not. This violates the structure of the matched design, where treated units were intentionally paired with control units based on municipality type. Consequently, the `did` bootstrap resamples may no longer preserve the balance or representativeness that matching was meant to ensure (Imbens, 2004: 5-21). In the given case, this seems to manifest in percentile CIs that systematically failed to cover the point estimate.

In contrast, symmetric confidence intervals operate on the fixed, matched dataset without resampling, relying solely on local variability around the estimate. For this reason, the paper reports symmetric 95% CIs for all ATT estimates instead of bootstrap-based intervals. This approach ensures comparability across all models and ATT types applied, allowing to judge differences as reflections of the underlying estimators and data instead of diverging inference methods. These benefits arguably justify in the given case the deviation from the Callaway and Sant'Anna estimator's default support of bootstrap inference.

B.2 Tabulating Dynamic Treatment Effects Across Estimation Models

As a numerical complement to Figure 4 in the main text, Table 5 provides the dynamic ATT estimates for all three models used in the main analysis. For each year relative to treatment, the table reports the point estimate and confidence interval, allowing for a direct comparison of effect trajectories across model types. Confidence intervals are calculated symmetrically ($\pm 1.96 \times$ standard error) to maintain consistency across models and ensure comparability (see Appendix B.1 for details on confidence interval construction). Across all three specifications, post-treatment effects grow in magnitude over time and reach substantial levels in the later years, while pre-treatment coefficients remain close to zero.

Year	Default DR Model		Basic OR Model		Pre-Matched DR Model	
	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval
-13	0.08%	[–0.53%, 0.68%]	0.08%	[–0.52%, 0.67%]	0.10%	[–0.44%, 0.63%]
-12	0.00%	[–0.34%, 0.34%]	0.00%	[–0.35%, 0.36%]	0.00%	[–0.33%, 0.33%]
-11	–0.19%	[–0.56%, 0.18%]	–0.17%	[–0.57%, 0.20%]	–0.17%	[–0.51%, 0.16%]
-10	–0.10%	[–0.44%, 0.24%]	–0.10%	[–0.48%, 0.28%]	–0.08%	[–0.41%, 0.25%]
-9	–0.35%	[–0.67%, –0.03%]	–0.35%	[–0.65%, –0.04%]	–0.36%	[–0.68%, –0.03%]
-8	–0.09%	[–0.37%, 0.19%]	–0.09%	[–0.39%, 0.22%]	–0.09%	[–0.37%, 0.20%]
-7	–0.09%	[–0.34%, 0.16%]	–0.09%	[–0.33%, 0.15%]	–0.08%	[–0.32%, 0.17%]
-6	0.23%	[–0.01%, 0.47%]	0.23%	[0.01%, 0.45%]	0.22%	[–0.01%, 0.45%]
-5	–0.03%	[–0.27%, 0.20%]	–0.03%	[–0.28%, 0.21%]	–0.02%	[–0.27%, 0.22%]
-4	–0.24%	[–0.41%, –0.07%]	–0.24%	[–0.41%, –0.07%]	–0.24%	[–0.41%, –0.07%]
-3	0.21%	[–0.31%, 0.73%]	0.21%	[–0.30%, 0.72%]	0.22%	[–0.23%, 0.66%]
-2	0.29%	[–0.03%, 0.61%]	0.29%	[–0.03%, 0.61%]	0.31%	[0.02%, 0.60%]
-1	–0.07%	[–0.43%, 0.29%]	–0.07%	[–0.40%, 0.26%]	–0.04%	[–0.38%, 0.30%]
0	–0.03%	[–0.20%, 0.14%]	–0.03%	[–0.21%, 0.15%]	–0.04%	[–0.22%, 0.14%]
1	–0.15%	[–0.45%, 0.14%]	–0.15%	[–0.45%, 0.15%]	–0.20%	[–0.51%, 0.12%]
2	–0.12%	[–0.45%, 0.21%]	–0.12%	[–0.45%, 0.20%]	–0.17%	[–0.50%, 0.16%]
3	–0.50%	[–0.93%, –0.06%]	–0.50%	[–0.91%, –0.08%]	–0.53%	[–0.97%, –0.09%]
4	–0.75%	[–1.37%, –0.14%]	–0.75%	[–1.41%, –0.09%]	–0.81%	[–1.46%, –0.16%]
5	–0.70%	[–1.20%, –0.20%]	–0.70%	[–1.20%, –0.21%]	–0.80%	[–1.30%, –0.29%]
6	–0.94%	[–1.56%, –0.33%]	–0.94%	[–1.54%, –0.35%]	–1.01%	[–1.63%, –0.40%]
7	–1.19%	[–1.98%, –0.40%]	–1.19%	[–2.01%, –0.37%]	–1.28%	[–2.07%, –0.50%]
8	–1.20%	[–2.36%, –0.04%]	–1.20%	[–2.49%, 0.09%]	–1.36%	[–2.60%, –0.13%]

Table 5. Dynamic Treatment Effects by Year Relative to Closure. *Values represent percentage point changes in overall municipal employment. Confidence intervals are symmetric.*

B.3 Exploring Estimate Variation Across Matched Iterations

As described in Section 6.2, the pre-matched DR model applies a matching procedure in which each treated municipality is paired with five structurally similar control units. This process is repeated 100 times using random draws from the available donor pool. The final estimate reported in the main text represents the mean ATT across these 100 iterations.

Statistic	ATT Estimate	Standard Error	95% Confidence Interval
Minimum (Min)	–0.69%	0.17	[–1.05%, –0.33%]
First Quartile (Q1)	–0.58%	0.18	[–0.93%, –0.23%]
Mean	–0.55%	0.18	[–0.90%, –0.19%]
Third Quartile (Q3)	–0.51%	0.18	[–0.87%, –0.16%]
Maximum (Max)	–0.43%	0.19	[–0.76%, –0.08%]

Table 6. Summary Statistics for ATT Estimates Across 100 Pre-Matched DR Runs. *Values report the distribution of point estimates, standard errors, and symmetric confidence intervals.*

To assess the stability and dispersion of these estimates, Figure 7 presents the distribution of overall ATT estimates across all model runs. The histogram reveals a reasonably symmetric and unimodal distribution centered near the reported mean of –0.55, suggesting consistent performance of the estimator across matched samples.

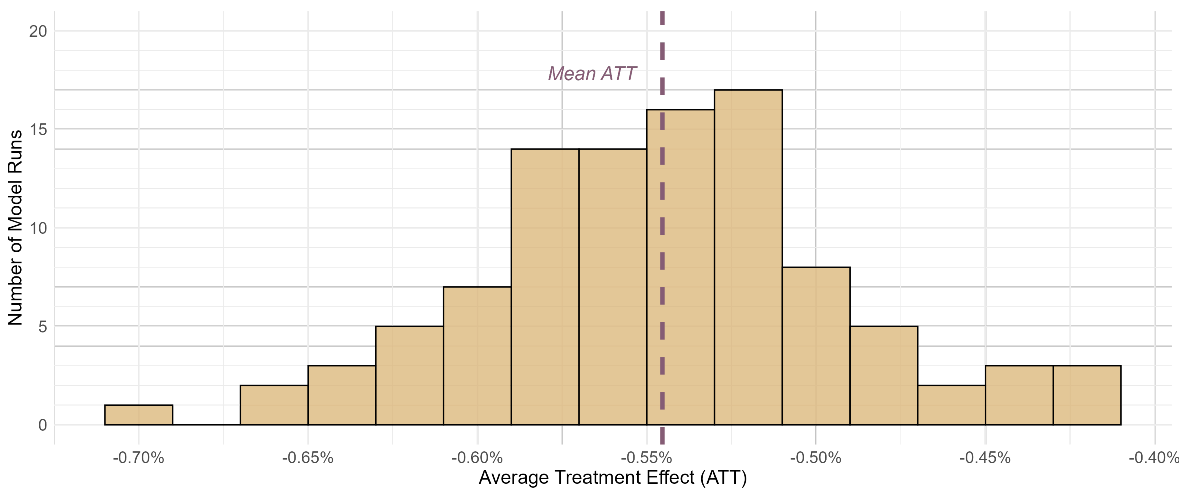

Figure 7. Histogram of ATT Estimates Across 100 Pre-Matched DR Iterations. *The plot shows the distribution of average treatment effects on employment. Each bar reflects the number of model runs in a given bin, with the vertical line indicating the overall mean.*

Table 6 complements this visual overview by providing a numeric summary of the distribution of the ATT estimates, standard errors, and confidence bounds across the iterations. The variation is modest across all metrics, reinforcing the robustness of the findings discussed in the main results section.

The narrow range of estimates and the stable distribution of standard errors across iterations support the reliability of the matching procedure. While the matched sample size is smaller than in the full-sample estimations, the consistency of the ATT (including the dynamic ATTs depicted in Figure 4 in the main text) suggests that the effect of hospital closures is not highly sensitive to the specific donor pool chosen. This stability strengthens the interpretation that the slightly larger effect size in the pre-matched DR model is not an artifact of sampling variability, but may reflect a more refined comparison between structurally comparable municipalities.

C ROBUSTNESS CHECKS AND IDENTIFICATION ASSUMPTIONS

C.1 Testing for Anticipation Effects

To assess the potential influence of anticipatory behavior on the estimated treatment effects, a robustness check is conducted using an alternative model specification that allows for a one-year anticipation window. This approach accounts for the possibility that municipalities or affected actors might begin adjusting prior to the official closure date (for instance in response to announcements or rumors about the potential closure, staffing changes, or early service reductions).

To compare both specifications, Figure 8 displays the dynamic average treatment effects of the baseline (no anticipation) and the one-year anticipation model, plotted side by side. The results reveal a high degree of overlap in the estimated effects, indicating no evidence of systematic pre-treatment divergence or pronounced anticipation effects. Point estimates in the one-year anticipation model are marginally smaller overall, and confidence intervals are wider across most time periods. This increased uncertainty leads to slightly reduced statistical significance in the post-treatment years, despite the general shape of the effect trajectory remaining consistent.

The early post-treatment period tends to show slightly stronger negative effects under the no-anticipation model, while effects in the later years are marginally more pronounced in the anticipation specification. However, these differences are modest and not systematically directional. In line with guidance from the estimator's authors, and given the comparability of results across specifications, the main analysis relies on the simpler no-anticipation model.

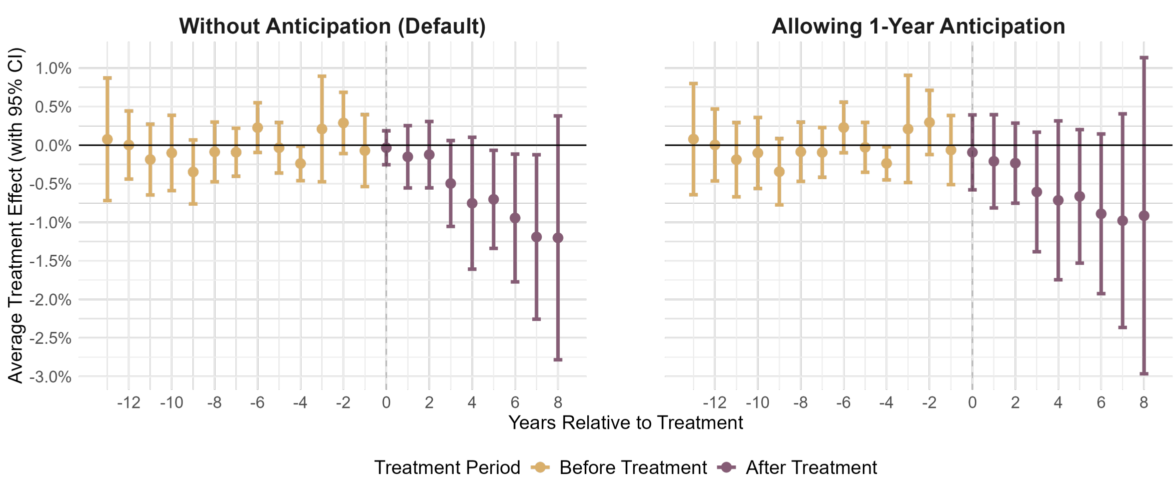

Figure 8. Dynamic ATT Estimates With and Without Anticipation Window. *Compares employment effects from a model that includes the year before closure as treatment.*

C.2 Testing for Spillover Contamination

To examine the sensitivity of the estimated dynamic treatment effects to the specification of the control group, this robustness check assesses the sensitivity of the results to the inclusion of control units that may have been indirectly affected by treatment due to spatial proximity. Specifically, the comparison contrasts estimates derived from the full sample (including potential spillover municipalities in the control group) with those obtained after excluding municipalities within a predefined spatial radius of treated units (as used in the main body).

The estimator and outcome variable remain unchanged, relying on the default DR specification and measuring overall employment levels. The two specifications produce virtually identical dynamic treatment effect trajectories. Figure9 shows that including potential spillover units leads to marginally smaller point estimates in some post-treatment periods. This slight attenuation is accompanied by a modest narrowing of confidence intervals, likely due to the increased number of control observations when spillover units are retained.

Although no formal test of statistical equivalence is conducted, the near-identical trajectories suggest that spatial spillovers are minimal, if present at all. Confidence intervals narrow slightly when potential spillover units are included, likely due to the larger control group, while the overall pattern and magnitude of effects remain stable. Given the already substantial size of the control group, the version excluding potentially affected municipalities is used in the main analysis to minimize any risk of contamination. The robustness check indicates that this decision is conservative, as the results would likely remain unchanged regardless.

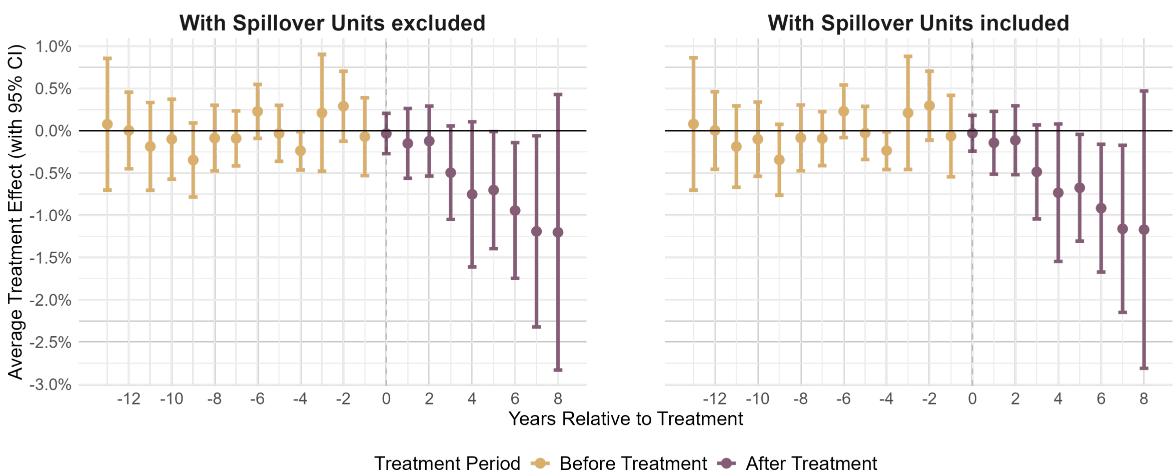

Figure 9. Dynamic ATT Estimates With and Without Exclusion of Spillover Units. *Compares employment effects from models that do and do not exclude potentially contaminated municipalities.*

C.3 Evaluating Effects Under Placebo Assignment

To evaluate whether the observed treatment effects in the main analysis could be driven by spurious correlations or structural sample features, two placebo tests were conducted using alternative treatment assignments. In the first test, a subset of never-treated municipalities was randomly selected and assigned pseudo-treatment years between 2010 and 2020. In the second test, the actual set of treated municipalities was retained, but each unit was assigned a placebo treatment year in 2011. To ensure that the placebo post-treatment period remained uncontaminated by actual closures, only units with real treatment dates occurring after 2015 were included. This allowed for an artificial treatment window from 2011 to 2014, unaffected by real-world interventions. Both placebo models were estimated using the default doubly robust estimator.

Figure 10 visualizes the resulting dynamic treatment effects. In both placebo settings, post-treatment trajectories do not resemble the effects found with the real treatments exposure, and none of the post-treatment coefficients are statistically significant at conventional levels. This includes both the artificially assigned pre-closure period in the second test and the full post-period in the random-treatment design. If anything, estimates appear to fluctuate randomly around zero, with confidence intervals consistently spanning the null.

These findings provide further reassurance that the effects identified in the main analysis are not artifacts of data structure or modeling choices. The placebo specifications behave as expected under the null, reinforcing confidence in the causal interpretation of the primary results.

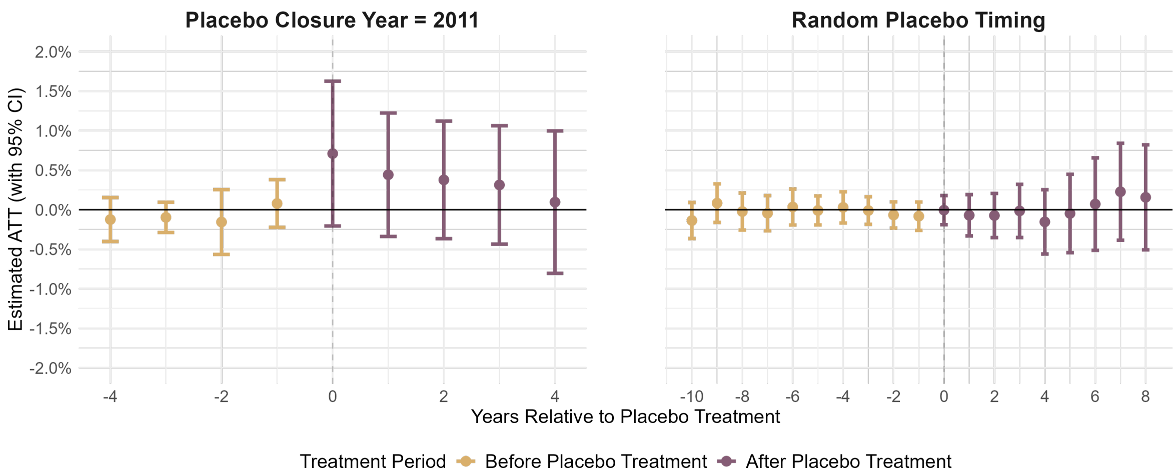

Figure 10. Dynamic ATT Estimates From Placebo Tests. *Compares effects from models using randomly reassigned or artificially shifted treatment years as placebo treatments.*

REFERENCES

Abadie, A., & Imbens, G. W. (2008). On the failure of the bootstrap for matching estimators. *Econometrica*, *76*(6), 1537–1557. https://doi.org/10.3982/ECTA6474

Alexander, D., & Richards, M. R. (2023). Economic consequences of hospital closures. *Journal of Public Economics*, *221*, Article 104821. https://doi.org/10.1016/j.jpubeco.2023.104821

Athey, S., & Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, *226*(1), 62–79. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.10.012

Austin, P. C., & Small, D. S. (2014). The use of bootstrapping when using propensity-score matching without replacement: A simulation study. *Statistics in Medicine*, *33*(24), 4306–4319. https://doi.org/10.1002/sim.6276

Baker, A., Callaway, B., Cunningham, S., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2025). Difference-in-differences designs: A practitioner's guide. https://arxiv.org/abs/2503.13323

Ballinari, D. (2024). Calibrating doubly-robust estimators with unbalanced treatment assignment. *Economics Letters*, *241*, Article 111838. https://doi.org/https://doi.org/10.1016/j.econlet.2024.111838

Baum, F., Lawless, A., Delany, T., Macdougall, C., Williams, C., Broderick, D., Wildgoose, D., Harris, E., Mcdermott, D., Kickbusch, I., et al. (2014). Evaluation of health in all policies: Concept, theory and application. *Health Promotion International*, *29*(Supplement 1), i130–i142. http://www.jstor.org/stable/45153211

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, *116*(536), 1789–1803.

Berger, E., Eriksen, A., Wiedicke, A., Ortlepp, L., Caglar, R., & Busse, R. (2024). The media's coverage and framing of hospital reforms: The case of germany. *Journal of Public Health*, 1–13. https://doi.org/10.1007/s10389-024-02271-6

Bertelsmann Stiftung. (2023). Wegweiser Kommune [Data licensed under Data licence Germany 2.0. Retrieved April 20, 2025]. https://www.wegweiser-kommune.de

Brunn, M., Kratz, T., Padget, M., Clément, M.-C., & Smyrl, M. (2023). Why are there so many hospital beds in germany? *Health Services Management Research*, *36*(1), 75–81. https://doi.org/10.1177/09514848221080691

Bundesministerium für Digitales und Verkehr. (2022). RegioStaR – Regionalstatistische Raumtypisierung [Spatial typology for regional statistics] [Retrieved April 20, 2025]. https://doi.org/10.57806/KY0FOA1B

Bundesministerium für Gesundheit. (2025). *Krankenhausreform: Warum die Reform notwendig ist [Hospital reform: Why the reform is necessary]* [Retrieved February 13, 2025]. https://www.bundesgesundheitsministerium.de/themen/krankenhaus/krankenhausreform.html

Bündnis Klinikrettung. (2024). *Gemeingut Krankenhaus retten [Saving public hospitals as a public good]* [Retrieved April 18, 2025]. https://www.gemeingut.org/krankenhausschliessungen/#1683637052964-11c159ba-64ad

Butts, K. (2023). *Difference-in-differences estimation with spatial spillovers* [arXiv preprint]. https://doi.org/10.48550/arXiv.2105.03737
Callaway, B., & Sant'Anna, P. H. (2021a). Did: Difference in differences [R package version 2.1.2]. https://bcallaway11.github.io/did/
Callaway, B., & Sant'Anna, P. H. (2021b). Difference-in-differences with multiple time periods. *Journal of econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001
Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2018). *Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india* (NBER Working Paper). National Bureau of Economic Research. https://doi.org/10.3386/w24678
Clarke, D. (2017). *Estimating difference-in-differences in the presence of spillovers* (MPRA Paper No. 81604). University Library of Munich. https://EconPapers.repec.org/RePEc:pra:mprapa:81604
Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, *96*(1), 187–199. https://doi.org/10.1093/biomet/asn055
De Chaisemartin, C., & d'Haultfoeuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The econometrics journal*, *26*(3), C1–C30. https://doi.org/10.1093/ectj/utac017
Deutsche Krankenhausgesellschaft. (2024). Gesetzgebungsprozess ohne Folgenabschätzung gefährdet massiv die Patientenversorgung [Legislative process without impact assessment massively jeopardizes patient care] [Retrieved February 13, 2025]. https://www.dkgev.de/dkg/presse/details/gesetzgebungsprozess-ohne-folgenabschaetzung-gefaehrdet-massiv-die-patientenversorgung/
Deutscher Bundestag, Parlamentsnachrichten. (2024). Experten-Kritik an der Krankenhausreform [Expert criticism of hospital reform] [Gesundheit — Anhörung — hib 633/2024. Published on 25 September 2024. Editorial team: Dr. Stephan Balling et al.]. Retrieved April 15, 2025, from https://www.bundestag.de/presse/hib/kurzmeldungen-1020706
Deutschlandfunk. (2024). Hospital reform and the end of case-based payments in germany [Retrieved February 13, 2025]. https://www.deutschlandfunk.de/krankenhaeuser-reform-lauterbach-fallpauschale-100.html
Dubb, S., & Howard, T. (2012). Leveraging anchor institutions for local job creation and wealth building. *Big ideas for job creation*, 01–31.
European Commission. (2015). Hospitals and healthcare – sectoral social dialogue [Retrieved April 17, 2025]. https://employment-social-affairs.ec.europa.eu/policies-and-activities/eu-employment-policies/social-dialogue/cross-industry-and-sectoral-social-dialogue/hospitals-and-healthcare-sectoral-social-dialogue_en#:~:text=Social%20partners,high%2Dquality%20health%20care%20services
Eurostat. (2022). Businesses in human health and social work activities sector [Retrieved April 17, 2025]. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Businesses_in_human_health_and_social_work_activities_sector&oldid=631972#:~:text=update:%20March%202026-,Highlights,added%20(%E2%82%AC590%

20billion).&text=This%20article%20presents%20an%20overview,collections%20such%20as%20health%20statistics.

Fischer, G. (2024). *Das deutsche Gesundheitswesen kompakt 2024: Akteure, Strukturen, Finanzierung [The German healthcare system compact 2024: players, structures, financing]*. MWV.

Gemeinsamer Bundesausschuss. (2023). Qualitätsberichte der Krankenhäuser [Quality reports of hospitals] [Retrieved June 29, 2024]. https://qb-datenportal.g-ba.de/

Ghislandi, S., Renner, A.-T., & Varghese, N. E. (2025). The impact of budget cuts on individual patient health: Causal evidence from hospital closures. *Journal of Health Economics*, *101*, Article 102975. https://doi.org/10.1016/j.jhealeco.2025.102975

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Greer, S. L., Falkenbach, M., Siciliani, L., McKee, M., Wismar, M., & Figueras, J. (2022). From health in all policies to health for all policies. *The lancet public health*, *7*(8), e718–e720. https://doi.org/10.1016/S2468-2667(22)00155-4

Hill, J. L., & Scott, M. (2010). The difficulties of estimating treatment effects in matching designs. *Journal of Econometrics*, *158*(2), 286–293. https://doi.org/10.1016/j.jeconom.2010.03.032

Huber, M., Lechner, M., & Wunsch, C. (2013). The performance of estimators based on the propensity score. *Journal of Econometrics*, *175*(1), 1–21. https://doi.org/10.1016/j.jeconom.2012.11.006

Imai, K., Kim, I. S., & Wang, E. H. (2023). Matching methods for causal inference with time-series cross-sectional data. *American Journal of Political Science*, *67*(3), 587–605. https://doi.org/10.1111/ajps.12685

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics*, *86*(1), 4–29.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, *47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5

Khan, S., & Tamer, E. (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica*, *78*(6), 2021–2042. https://doi.org/10.3982/ECTA7372

Mandich, A. M., & Dorfman, J. H. (2017). The wage and job impacts of hospitals on local labor markets. *Economic Development Quarterly*, *31*(2), 139–148. https://doi.org/10.1177/0891242417691609

Mills, C. A., Yeager, V. A., Unroe, K. T., Holmes, A., & Blackburn, J. (2024). The impact of rural general hospital closures on communities—a systematic review of the literature. *The Journal of Rural Health*, *40*(2), 238–248. https://doi.org/10.1111/jrh.12810

Nyholt, N. (2024). Left behind: Voters' reactions to local school and hospital closures. *European Journal of Political Research*, *63*(3), 884–905. https://doi.org/10.1111/1475-6765.12622

Politico. (2024). Germany's biggest hospital reform: What it means for healthcare and hospital closures [Retrieved February 13, 2025]. https://www.politico.eu/article/germany-biggest-reform-hospital-system-health-care-karl-lauterbach/

Porter, M. E., & Guth, C. (2012). *Redefining german health care: Moving to a value-based system*. Springer Science & Business Media.

Preusker, U. K., Müschenich, M., & Preusker, S. (2014). *Darstellung und Typologie der Marktaustritte von Krankenhäusern Deutschland 2003–2013 [Presentation and typology of market exits of hospitals in Germany 2003–2013]* (Commissioned report). GKV-Spitzenverband. https://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/krankenhaeuser/gutachten_marktaustritt_krankenhaeuser/Gutachten_Marktaustritte_KH_07-2014_gesamt.pdf

Reimbursement Institute. (2019). *Anzahl der Krankenhäuser in Deutschland [Number of hospitals in Germany]*. Retrieved April 22, 2025, from https://reimbursement.info/blog/anzahl-der-krankenhaeuser-in-deutschland/

Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, *235*(2), 2218–2244. https://doi.org/10.1016/j.jeconom.2023.03.008

Schildt, C., & Rubin, V. (2015). *Leveraging anchor institutions for economic inclusion* (Policy report). PolicyLink. https://www.policylink.org/resources-tools/leveraging-anchor-institutions

Simon, M. (2016). *Das Gesundheitssystem in Deutschland: Eine Einführung in Struktur und Funktionsweise [The healthcare system in Germany: An introduction to its structure and functioning]*. Hogrefe AG.

Statistische Ämter des Bundes und der Länder. (n.d.). Verzeichnis der Krankenhäuser und Vorsorge- oder Rehabilitationseinrichtungen in Deutschland [Directory of Hospitals and Preventive or Rehabilitation Facilities in Germany]: Krankenhausverzeichnis [Hospital directory], *Wiesbaden*. https://www.statistischebibliothek.de/mir/receive/DESerie_mods_00002778

Statistisches Bundesamt. (2025a). Gesundheitspersonal: Deutschland, Jahre, Einrichtungen, Geschlecht (Tabelle 23621-0001)[Healthcare personnel: Germany, years, facilities, gender (Table 23621-0001)] [Data set. Retrieved April 11, 2025]. https://www-genesis.destatis.de/datenbank/online/table/23621-0001

Statistisches Bundesamt. (2025b). Krankenhäuser, Betten, Patienten: Deutschland, Jahre (Tabelle 23111-0001) [Hospitals, beds, patients: Germany, years (Table 23111-0001)] [Data set. Retrieved April 11, 2025]. https://www-genesis.destatis.de/datenbank/online/table/23111-0001

Statistisches Bundesamt. (2025c). Personal in Krankenhäusern: Deutschland, Stichtag (Tabelle 23111-0002) [Personnel in hospitals: Germany, reference date (Table 23111-0002)] [Data set. Retrieved April 11, 2025]. https://www-genesis.destatis.de/datenbank/online/table/23111-0002

Timofeyev, Y., Goldenova, V., Mantaeva, E., & Jakovljevic, M. (2024). The impact of hospital competition on the quality of care in europe: A systematic review. *Healthcare (Basel, Switzerland)*, *12*(22), Article 2218. https://doi.org/10.3390/healthcare12222218

Van den Berg, N., Radicke, F., Stentzel, U., Hoffmann, W., & Flessa, S. (2019). Economic efficiency versus accessibility: Planning of the hospital landscape in rural regions using a linear model on the example of paediatric and obstetric wards in the northeast of germany. *BMC health services research*, *19*, 1–13. https://doi.org/10.1186/s12913-019-4016-2

Vogler, J. (2020). Rural hospital closures and local economic decline. *Available at SSRN 3750200*. https://doi.org/dx.doi.org/10.2139/ssrn.3750200

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Wendt, C., Frisina, L., & Rothgang, H. (2009). Healthcare system types: A conceptual framework for comparison. *Social Policy & Administration*, *43*(1), 70–90. https://doi.org/10.1111/j.1467-9515.2008.00647.x

Williams, G. A., Rockwell, O. A., & Greer, S. L. (2024). Sdg8, promoting decent work and economic growth: Health policies for good jobs. In S. L. Greer, M. Falkenbach, J. Figueras, & M. Wismar (Eds.), *Health for all policies: The co-benefits of intersectoral action* (pp. 130–148). Cambridge University Press. https://doi.org/10.1017/9781009467766.008

WirtschaftsWoche. (2024). Lauterbach announces hundreds of hospital closures as part of reform [Retrieved February 13, 2025]. https://www.wiwo.de/politik/deutschland/krankenhausreform-lauterbach-kuendigt-an-hunderte-krankenhaeuser-werden-schliessen/30046948.html

Yang, S., & Ding, P. (2018). Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores. *Biometrika*, *105*(2), 487–493. https://doi.org/10.1093/biomet/asy008

REPLICATION MATERIALS

The full replication package for this thesis, including code, processed datasets, and documentation, is available on GitHub: https://github.com/nikpaw/MDS_MT_clean

The replication package includes:

- The full reproducible codebase, organized into scripts for data cleaning, merging, modeling, and robustness testing.
- Processed datasets used for estimation, adhering to relevant data protection guidelines.
- Output files such as figures, reports, and tables.
- Comprehensive documentation describing the project structure and instructions for reproducing the results.

To reproduce the results:

1. Clone or download the GitHub repository.
2. Open the RStudio project file `MDS_MT_clean.Rproj`.
3. Execute the master script `R/00_master/00_master_run_reproducible.R` to run the full analysis pipeline on preprocessed data.

Software requirements and package dependencies are detailed within the repository documentation.