

# Bracing for the Blow: Weekly Primary-Care Billing Before and After Danish Workplace Closures, 1991–2023

A Stage-1 Registered Report Protocol

Kellogg Center for Science of Science and Innovation (CSSI), Northwestern University  
Copenhagen Center for Social Data Science (SODAS), University of Copenhagen  
Statistics Denmark  
The ROCKWOOL Foundation

*This protocol was drafted by a multi-agent AI research harness under human direction.*

August 31, 2026 — Protocol frozen at git commit `fc59c71f6ac3` (stamped by the registration driver after the final audit)

## Abstract

A workplace closure is an event in administrative records but a process for the people who live through it. Workers usually know, or fear, that a workplace is dying long before the doors close. Modern register-based displacement studies, however, measure health outcomes after separation and in annual event time; even the finest published utilization clock around a mass-layoff event is monthly, drawn from one US firm’s insurance claims. Danish health-insurance billing offers a weekly clock — the HONUUGE fee week — that has not yet been used to study establishment death. We will trace general-practice billing quarterly, and descriptively weekly, around workplace closures for cohorts 2009–2023. Using the worker-flow canon, we will distinguish true closures from spurious, en-bloc absorptions and riskset-match every exposed worker to up to three stable-workplace controls based on a full year of pre-baseline billing history. A Holm family contains three confirmatory quantities, each verified at 0.95 power or better for floor effects below published magnitudes: the pre-boundary contact excess in the last two terminal-year quarters; the true-closure excess in the post-year, with the spurious-closure arm serving as a registered falsification gate that can only subtract support; and the out-of-hours versus daytime composition of the peri-event excess. Before any outcome data are touched, a realized-N gate demotes any underpowered member to exploratory. All claims are associational, and anticipation is worded strictly as pre-boundary excess relative to an annually dated boundary. Nulls will be bounded, never accepted. This remediated second registration follows a first Stage-0 power simulation that emptied the confirmatory family. The protocol therefore publishes the complete records of both power runs (parameters, per-test results, and stdout; the round-1 script survives as its recorded parameterization), together with the demotion trail and remediation record.

**PROTOCOL REGISTRATION.** This document is a Stage-1 Registered Report protocol prepared *before any access to the underlying microdata*. It is written against the Nature Human Behaviour Registered Report Stage-1 review criteria (importance of the question; capacity of the design to answer it; logic and plausibility of the hypotheses; soundness and feasibility of the pipeline; replication-grade methodological detail; outcome-neutral conditions). Every hypothesis, estimand, threshold, contingent decision rule, forecast, and interpretation template in this document is carried in structured form by the machine-readable registration accompanying this protocol, frozen at the same git commit; the seeded power-simulation script and the records of both Stage-0 power runs — the voided first round and the remediated re-run that this protocol registers — accompany them in the deposit bundle. Data reside within Statistics Denmark’s research scheme (Forskningsservice), which acts as independent gatekeeper; the authors certify no prior access to the microdata. Secondary analyses of existing datasets are eligible for Registered Reports at Nature and Nature Human Behaviour under exactly this certification. Hypothesis statements are frozen: they carry their initial wording for later outcome scoring and cannot be amended after in-principle acceptance.

## 1 Introduction

A workplace can be dying while every worker still appears employed. Workers usually know, or fear, that closure is coming long before annual records show a separation. Job displacement is among the most-studied economic shocks (Sullivan and von Wachter 2009), but its health footprint does not begin on the day the doors close. A 55-year lineage of closure research shows measurable traces during this period. Researchers have observed plant-gate physiology in the months before shutdown (Kasl and Cobb 1970), general-practice consultations around a single UK factory closure including its insecurity phase (Beale and Nethercott 1985), hospital admissions before and after a Danish shipyard closure (Iversen, Sabroe and Damsgaard 1989), self-reported morbidity under anticipated job change (Ferrie et al. 1995), and formal lead effects of plant closures on survey outcomes (Wunder and Zeydanli 2021). Modern register-based displacement-health studies, however, measure outcomes after separation and at much coarser resolution. The canonical Danish, Swedish, and Austrian studies observe mortality, hospitalization, and health costs in annual event time (Browning and Heinesen 2012; Eliason and Storrie 2009; Kuhn, Lalive and Zweimüller 2009), while the finest published utilization clock around a mass-layoff event is monthly and comes from one US firm’s insurance claims (Hamad, Modrek and Cullen 2015). Danish health-insurance billing offers a weekly clock — the HONUFE fee week, populated 1990–2025 — that has not yet been used to study establishment death.

Two questions remain unresolved: whether primary-care contact already rises while a workplace is dying, and whether any post-closure excess belongs to genuinely destroyed jobs rather than to workplaces that disappear only on paper. Neither question has been studied at register scale with matched controls and sub-annual resolution. The recorded 62-query novelty-boundary search deposited with the registration verified each of three specific gaps as open. No study estimates population-scale trajectories of administrative primary-care utilization before establishment closure; the construct-level ancestors follow one factory and one shipyard, and the finest modern evidence follows one firm monthly. Displacement research uses the worker-flow classification separating true establishment deaths from en-bloc absorptions (Benedetto et al. 2007; Hethey-Maier and Schmieder 2013) to exclude cases, but no published study turns the absorbed pseudo-closures into an explicit placebo arm for a health outcome. Nor has an event-based study tested whether peri-event excess care tilts toward out-of-hours, unplanned services.

The expected direction is genuinely contested. US evidence finds that utilization falls under layoff-threat insecurity and after job loss (Hamad, Modrek and Cullen 2015; Schaller and Stevens 2015), and the Danish displacement literature contains a published null on stress-related hospitalizations (Browning, Møller Danø and Heinesen 2006). By contrast, the small-scale closure

literature finds that consultations rise. We therefore pre-register both the null and the reversal as informative outcomes.

The timing matters beyond this immediate literature. If care-seeking moves before the closure-year boundary, every closure-based displacement-health design that treats the pre-closure year as a clean baseline mis-times the shock and understates its size. That methodological problem reaches beyond primary care. For Danish health planning, anticipatory demand around mass closures is directly actionable, while its timing composition — daytime versus out-of-hours — distinguishes deferred routine care from unplanned, distress-driven contact. Those demand responses call for different provision. An inert spurious-closure arm would give closure-based designs a reusable severity device; a non-inert arm would instead caution that closure exposure partly measures selection into dying workplace identities. Because the design is powered to 0.95 at floor effects of 2% of baseline, a null is also informative: it bounds anticipatory volume responses below magnitudes that would matter for the literature’s pre-trends or for provision planning.

The study gains its leverage by linking records that no other setting supplies at this scale. The IDA workplace register (IDAS) dates every Danish establishment’s terminal event and records the final-closure, forward-identity, and absorption-share information required by the worker-flow canon. The IDA employment register (IDAN) attaches workers to workplaces each November and follows the workforce into successor units. The health-insurance billing registers — SSSY, with SYSI for the legacy era — date every publicly reimbursed general-practice contact by fee week.

For closure cohorts 2009–2023, we begin with workers whose primary November-status employment in year  $c - 1$  was at a workplace that disappears between years  $c$  and  $c + 1$ . We classify the event as a true closure when no successor absorbs the workforce and the maximum successor share is below 0.30. We classify it as a spurious closure when an en-bloc successor exists and the share is at or above 0.70; we exclude and monitor the ambiguous mid-zone. Freezing cohorts in November of  $c - 1$ , before the terminal year, prevents the sinking-ship attrition channel (Schwerdt 2011) from conditioning the sample. We then riskset-match each exposed worker to up to three stable-workplace controls within exact strata, using Mahalanobis distance over a full year of pre-baseline billing history together with tenure and wage decile. A stacked matched-set difference-in-differences event study estimates quarterly contrasts, accompanied by weekly descriptive trajectories.

This is a contribution in measurement, not in discovering the phenomenon. Researchers have measured health during the anticipation phase of closure for 55 years, and a Danish closure study already covers the pre-closure window (Iversen, Sabroe and Damsgaard 1989); we therefore make neither a phenomenon-discovery nor a Danish-first claim. Danish job insecurity has also been tied to medication use (Rugulies et al. 2009), and lead effects of plant closures are already an estimand (Wunder and Zeydanli 2021). The true-versus-spurious distinction is measurement canon, so H2’s novelty is confined to applying that placebo to health billing.

The study instead claims the first register-scale, matched, weekly-resolution estimate of primary-care billing trajectories around workplace closures. Its falsification architecture combines a spurious-closure gate, a pre-baseline placebo panel, and a pre-outcome realized-N demotion gate so that every registered outcome, including a null, carries interpretable weight. Annual November-status dating means that all pre-event quantities are described as pre-boundary excess relative to the closure-year boundary, never as announcement-dated anticipation. The four registers contain no diagnoses, hospital contacts, or prescriptions, so the outcomes are utilization volume, fees, and timing composition — never health states.

The registration history also belongs in the opening. This is a remediated second registration because the first Stage-0 power simulation — deposited, seeded, and independently re-run — emptied the confirmatory family under the registered fixed-point demotion rule. The program’s audit voided the resulting all-exploratory registration. The remediation changed which quantities

carry confirmatory weight without touching the frozen hypothesis wording, recalibrated one power basis to a published Danish quantity, and enlarged the sampling plan only through levers with cited published bases; the enlargement is licensed jointly by a demotion-only realized-N gate that runs before any outcome data are accessed. No floor effect was raised. This protocol publishes the records of both power runs — the round-1 run as archived stdout and parameterization, the remediated run as the deposited script — along with the diagnosis and the full remediation record. Section 7 gives the details; we regard the registration’s history as part of its evidence.

**Claim-class ceiling:** Associational — the canonical ceiling token recorded in the machine-readable registration, fixed before writing and binding on all prose. In full: associational for the matched event-study contrasts, the contrastive falsification gate, and the composition contrast; descriptive for trajectory panels and the bounded priority claim. Binding vocabulary rules: “is associated with” / “consistent with” / “suggests” only; “effect”, “impact”, “causes” prohibited throughout; anticipation always worded as pre-boundary excess relative to the annually dated closure-year boundary, never as announcement-dated anticipation; no phenomenon-discovery or Danish-first framing; H2 novelty confined to the placebo’s application to health billing. H2 verdict semantics are fixed in advance: “consistent with H2” requires *both* the confirmatory core ( $\theta_{2a}$  Holm-significant, predicted direction) *and* a passing falsification gate F2; the strict comparative magnitude is only ever reported as an exploratory estimate with its CI. Exploratory results (H1\_trend, H2\_diff, E1, E3, E4, and any gate-demoted member) carry the word “exploratory” in the heading and in every sentence stating them. The prose label never exceeds the evidence label. The prose in this protocol and any resulting Stage-2 report will not exceed this evidence class.

## Research questions and hypotheses

**RQ1.** Does primary-care contact already rise in the final quarters before a workplace’s closure-year boundary — before any separation is visible in annual employment records?

**RQ2.** Does the post-closure billing excess attach to genuinely destroyed jobs (true closures) rather than to workplaces that die only on paper (spurious, en-bloc absorptions)?

**RQ3.** Is the peri-event excess compositionally tilted toward out-of-hours, unplanned services rather than daytime, scheduled care?

### RQ1, H1 — confirmatory ( $\theta_1$ ).

H1 (anticipation): Compared to riskset-matched workers at stable workplaces with parallel 52-week baseline billing histories, workers employed at workplaces in their final year before permanent closure (IDAS ENDNEDL) will show elevated weekly primary-care contact rates (SYSI/SSSY contacts dated by HONUKE; YDLANT, KONTAKT) already in the two calendar quarters preceding the closure-year boundary, with the excess growing as the boundary approaches — before any separation is visible in annual IDAN records. *Registered tiering:* the statement’s core assertion (elevated contact in the two pre-boundary quarters) is the confirmatory estimand  $\theta_1$ ; the strict growth rider (“with the excess growing as the boundary approaches”) is tiered exploratory as the gated contrast H1\_trend, because its only confirmatory operationalization is verified structurally underpowered (Section 7). The wording itself is untouched. (*Direction: positive;  $\theta_1 > 0$ .*)

### RQ1, H0.

Substantive H0: Primary-care contact among workers at closing workplaces does not move before the closure-year boundary: any anticipatory response at or above 2% of baseline is ruled out at the design’s error rates, locating whatever billing response exists at or after separation. A Holm-significant *decrease* is not this null but a pre-registered informative reversal (deferred or forgone care under insecurity, per Hamad, Modrek and Cullen 2015). (*Direction: null.*)

**RQ2, H2 — confirmatory ( $\theta_{2a}$ , with falsification gate F2).**

H2 (true vs spurious closure): Compared to workers whose workplace disappears only administratively — absorbed into a successor unit with a high absorbed-worker share (IDAS ANDELOPS) and preserved forward identity (IDFREM, REGELFR), so employment continues — workers at true permanent closures (ENDNEDL with low ANDELOPS) will show a larger and more persistent post-event excess in weekly contacts and gross fees (BRUHON) through the first post-closure year. *Registered tiering*: the statement’s asserted excess is tested confirmatorily as  $\theta_{2a}$ , the true-closure arm’s post-year excess over its own matched controls; the comparative quantifier is enforced by falsification gate F2 outside the Holm family (the spurious arm’s excess  $\theta_{2b}$ , bounded, with a pre-committed disconfirmation trigger that can only subtract support), and the strict differential is reported as the exploratory H2\_diff. The wording itself is untouched. (*Direction: positive;  $\theta_{2a} > 0$ , with gate F2 required not to fire for the comparative reading.*)

**RQ2, H0.**

Substantive H0: Workers at true closures show no post-boundary excess in primary-care billing through the first post-closure year relative to matched controls (post-year excess at or above 2% of baseline ruled out); together with  $\theta_{2b}$  near zero this locates any billing response, if  $\theta_1$  found one, strictly before the boundary. (*Direction: null.*)

**RQ3, H3 — confirmatory ( $\theta_3$ ).**

H3 (distress composition): Compared to the out-of-hours share of contemporaneous contacts among their matched controls, closure-exposed workers’ peri-event excess contacts will contain a disproportionately higher share of out-of-hours and off-schedule services (YDLTID time-of-day codes; VAGTOMR in the SYSI era), consistent with unplanned, distress-driven care-seeking rather than deferred routine care. (*Direction: positive;  $\theta_3 > 0$ .*)

**RQ3, H0.**

Substantive H0: The peri-event excess is compositionally unremarkable: the out-of-hours share of excess contacts does not differ from the daytime share by 5 points or more (bootstrap CI converted to ruled-out gaps against the realized MDE), consistent with scheduled as much as distress-driven care. (*Direction: null.*)

**Exploratory tier.**

Five heading-segregated exploratory quantities, never promoted under any contingency — H2\_diff, E1, E3, and E4 Holm-corrected within family as a severity aid, and H1\_trend gated on the H1 verdict and reported at unadjusted  $\alpha = 0.05$ : H1\_trend, the gated within-terminal-year growth contrast excerpted verbatim from H1’s rider; H2\_diff, the true-minus-spurious differential excerpted verbatim from H2’s comparative quantifier; E1, a SYSI-era descriptive replication of the H1 contrast for closure cohorts 1992–2003; E3, gross-fee analogues and heterogeneity by sex, age band, and baseline wage decile; and E4, an arm-comparative attrition accounting panel quantifying the sinking-ship channel, including a post-treatment-conditioned exploratory split of the  $\theta_1$  contrast by whether the exposed worker retains November-year- $c$  attachment to the event workplace. (*Direction: exploratory; directions pre-stated only where indicated.*)

## 2 Data: registers, variables, and linkage

The analysis server contains only the IDA workplace register (IDAS), the IDA employment register (IDAN), and two health-insurance billing registers: the current SSSY series and the legacy SYSI series. Because no diagnosis, hospital, prescription, or population-spine register is in scope, we study utilization volume, fees, and timing composition — never health states. Before accessing any microdata, we checked every variable and its validity years against the

parsed public TIMES/HKV schema documentation. The Statistics Denmark Forskningservice inventory for the four-register analysis server verifies PNR on all person registers, KOM in IDAS, and VAGTOMR and SIKREAMT in SYSI, although the public parses omit all four variables. ARB\_SEKTORKODE in IDAN presents the reverse case. This private-sector eligibility filter is documented publicly in TIMES *ida-ansættelser*, but it does not appear in the recorded Forskningservice inventory extract, so its claimed 2008+ validity rests only on the public parse. Stage 1 validates its presence and coverage in the analysis environment before classification begins; Section 5 registers a frozen fallback based on ARB\_HOVED\_BRA\_DB07. Every expected magnitude comes from published Statistics Denmark series or the schema documentation; we have not accessed the microdata.

## Registers and variables

- **IDA workplaces (IDAS).** The *IDA arbejdssteder* series is documented in TIMES/HKV as valid from 1980, and the panel spans 1980–2024 for event cohorts  $c$  in 2009–2023. LBNR supplies the workplace serial number, while IDFREM/IDTILB trace identity chains; ARBNR and CVRNR are the 2007+–vintage workplace and firm keys. ENDNEDL marks a workplace’s final closure, and its validity through 2023 caps the last cohort. IDFREM/IDTILB trace workplace identity forward and backward in time, REGELFR records the rule governing whether a workplace is retained in the following year, and FRELFR identifies closure through absorption. ANDELOPS and ANDELUDS record absorbed and split-off employee shares, which cross-check the worker-flow classification. To define stable controls and workplace size, we use AFGRATE, the subsequent-period exit rate, requiring  $< 30\%$  for stable controls; ANTALXFR, the following year’s change in employees, requiring  $|\text{ANTALXFR}| \leq 30\%$ ; and ANTNOV, November employment, with eligibility set at 5–1000 and event presence requiring  $\text{ANTNOV} \geq 5$ . ARB\_HOVED\_BRA\_DB07 supplies 2008+ industry strata, and inventory-verified KOM supplies workplace municipality, which we map to the five post-2007 regions. The pre-2008 legacy keys ARBGNR/ARBSTK serve only the exploratory SYSI-era module.
- **IDA employments (IDAN).** The *IDA ansættelser* series is documented in TIMES/HKV as valid from 1980, and the panel spans 1980–2024. Inventory-verified PNR forms the person spine. LBNR, together with ARBNR/CVRNR, links an employment to its workplace in the 2008+ vintage. ANSAAR records the year employment began at the workplace, allowing us to measure tenure and apply the inclusion floor  $\text{ANSAAR} \leq c - 2$ . BREDT\_LOEN\_BELOEB records broad salary amount; its largest value in  $c - 1$  defines the primary job and the baseline wage decile, with ties broken deterministically by larger LOENTIMER and then lower ARBNR. ARB\_SEKTORKODE records workplace sector from 2008+ and applies the private-sector filter through the modal sector among year- $c - 1$  employees. Because only the public parse verifies this variable and the recorded analysis-server inventory extract omits it, Stage 1 validates it and Section 5 provides the frozen ARB\_HOVED\_BRA\_DB07 fallback. ANSXFREM, the change of employment through the following year, feeds the exploratory attrition panel E4. The legacy job characteristics STILL, TILKNYT, ANSDAGE, and TIMELON, together with the legacy keys, serve only the exploratory era.
- **Health insurance statistics, current series (SSSY).** The *Sygesikringsstatistik* series contains service rows for 2005–2025; extraction over 2006–2025 guarantees the complete  $c - 3$  lookback for the first cohort,  $c = 2009$ . Inventory-verified PNR provides the person spine, and HONUGE provides the weekly clock by recording the fee week. KONTAKT is the row-level contact marker that we sum into quarterly GP contact counts, subject to a frozen fallback definition if validation in the analysis environment fails. YDLANT counts services, while BRUHON records the gross fee to the doctor and is compared nominally only within cohort  $\times$  calendar-quarter cells. SPEC2 is the 2-digit specialty code, valid from

2005. Frozen code lists assign 70–80 to daytime GP and 81–84 and 89 to out-of-hours GP. Although parsed value labels describe codes 71, 73, and 75 as “evening etc.” GP services, the registered primary classifier keeps them on the daytime side, as disclosed and addressed in Section 5. YDLTID records service time for the robustness out-of-hours classifier; missing or “not relevant” values are classified as daytime, conservatively against H3. AFRPER characterizes settlement lag descriptively, while BEHANDLINGSDATO, valid from 2021+, calibrates the HONUUGE settlement lag but never replaces HONUUGE. We exclude modal group-2 insured using SIKGRUP. ALDERIMP and KOENIMP provide imputed age and sex for exact strata, including explicit unknown categories, and YDERNR supplies the provider number used by the frozen fallback contact definition.

- **Health insurance statistics, legacy series (SYSI).** The series spans 1990–2005 and is used *only* for the exploratory SYSI-era replication E1 in cohorts 1992–2003. PNR and HONUUGE identify people and fee weeks. Because SPEC2 does not exist pre-2005, the frozen era code list uses the 6-digit SPECIALE specialty. YDLANT is the era’s contact proxy because KONTAKT also does not exist pre-2005. The remaining measures are BRUHON, YDLTID, inventory-verified VAGTOMR for the out-of-hours service area in 1997–2005, AFRPER, and SIKGRUP. Era patient geography comes from SIKREKOM/SIKREAMT for E1 only; SSSY contains no patient geography, and SIKREAMT is inventory-verified.

## Linkage

PNR links people across IDAN, SSSY, and SYSI. For the 2008+ vintage that contains the entire confirmatory tier, LBNR links workplaces together with ARBNR/CVRNR; in the pre-2008 vintage used only for the exploratory era module, it links them with ARBGNR/ARBSTK. IDFREM/IDTILB connect workplace identities over time. We never pool the two eras across the 2007/08 IDA vintage break. An outcome-neutral linkage check requires at least 80% of cohort workers, counting both exposed workers and the control pool, to have at least one SSSY row of any kind during the five-year window from  $c - 3$  through  $c + 1$ . The published Danish benchmark for annual GP contact prevalence is roughly 85%; if linkage falls below the registered floor, the pipeline halts at Stage 3.

## Register clocks and their vintage caveats

The design combines two administrative clocks, each of which measures a different kind of time and carries statutory or vintage caveats. The protocol addresses those caveats explicitly.

*The weekly billing clock.* HONUUGE records the *fee* week, which is the week of settlement rather than necessarily the week of service. Because billing lag can blur the closure boundary, the confirmatory estimands use quarters and remain robust to lags up to four weeks. Before estimating any event study, we calibrate the lag with BEHANDLINGSDATO for 2021+ rows. A median lag above four weeks activates the registered guard-band contingency in Section 6, while a week-zero step diagnostic in the weekly descriptive series bounds any claim about timing shape. The classification vocabulary also changes at the 2005 vintage. SPEC2 and KONTAKT exist only in SSSY (2005+), which confines the confirmatory tier to SSSY-era cohorts. The SYSI era, with SPECIALE codes, YDLANT-weighted service rows, and VAGTOMR for 1997–2005, remains exploratory only.

*The annual closure clock.* Annual November status dates an establishment’s terminal event: its workplace identity is present in year  $c$  with  $\text{ANTNOV} \geq 5$  and absent in year  $c + 1$ . Because no announcement date is in scope, we infer anticipation windows rather than date them to an announcement, and we describe every pre-event quantity as pre-boundary excess. ENDNEDL ends in 2023 and therefore caps the last cohort. IDAS and IDAN extend through 2024, providing the  $c + 1$  identity-absence and worker-flow checks for cohort 2023, while SSSY extends through 2025 to provide that cohort’s post-year outcomes. The IDA keys change at the 2007/08 vintage,

from ARBGNR/ARBSTK before the break to ARBNR/CVRNR after it. ARB\_HOVED\_BRA\_DB07 and ARB\_SEKTORKODE begin in 2008, and workplace municipality maps to the five regions created in the 2007 structural reform. These are three further reasons to confine the confirmatory tier to cohorts 2009–2023. If the ENDNEDL flag-year convention in the analysis environment differs from the assumed “flagged in the workplace’s last active year,” a registered contingency in Section 6 mechanically re-anchors  $c$  to the last IDAS year with  $\text{ANTNOV} \geq 5$  and reports the recalibration verbatim.

## Population

The registered SSSY-era cohorts begin with people whose primary November-status employment in year  $c - 1$  was either at an event workplace or at a stable control workplace. For closure years  $c$  in 2009–2023, an event workplace has an IDAS identity present in year  $c$  with  $\text{ANTNOV} \geq 5$  and absent in year  $c + 1$ . A stable control retains the same identity from  $c - 2$  through  $c + 1$ , has no ENDNEDL, has  $\text{AFGRATE} < 30\%$ , and satisfies  $|\text{ANTALXFR}| \leq 30\%$  in years  $c - 1$  and  $c$ . We freeze workplace attachment in November of  $c - 1$ .

The unit of analysis is the person  $\times$  event-quarter, aggregated from person  $\times$  HONUGE fee-week records and stacked by workplace-event cohort in 1:3 matched sets. At baseline, included workers are age 25–59, using ALDERIMP from the most recent SSSY row between  $c - 3$  and  $c - 1$ ; people with no such row enter explicit unknown age and sex strata rather than being dropped. They have at least one full year of tenure, defined by  $\text{ANSAAR} \leq c - 2$ , and work at a private-sector workplace according to modal ARB\_SEKTORKODE with November size 5–1000. We classify event workers as exposed to a true closure when ENDNEDL records final closure and the maximum worker-flow successor share is  $s(w) < 0.30$ . We classify them as exposed to a spurious closure when IDFREM/REGELFR/FRELFREM preserves a forward identity and  $s(w) \geq 0.70$ .

Every exclusion is reported with its count. We exclude an ambiguous successor share when  $0.30 \leq s(w) < 0.70$  and monitor whether this group remains under 35% of terminal events. We also exclude modal SIKGRUP group-2 insured, expected to comprise 1–2%; workers appearing in more than one event cohort, whom we assign to the earliest; control-pool workers at workplaces with a terminal event or mass separation from  $c - 2$  through  $c + 1$ ; and workers with missing ANSAAR or missing or zero annual pay, expected to comprise under 2%. Mass-separation events at surviving workplaces remain entirely outside the confirmatory tier as a matter of scope discipline.

Schema documentation and published sources, rather than realized data, provide the expected  $N$ . Across 15 cohorts, we expect 10,000–20,000 eligible exposed workers per cohort-year, or 150,000–300,000 exposed true-closure workers in total. At the registered 90% match-rate floor, this yields at least 135,000 matched sets; the spurious-closure group is expected to contain 30–60% as many workers as the true-closure group. The denominators behind this sizing are roughly 227,000 IDAS workplace rows per year, 3.2M IDAN rows per year, and 88M SSSY rows per year in the Forskningsservice inventory. The remaining bases are Statistics Denmark business-demography establishment death rates of roughly 2–5% among employer workplaces, the demonstrated Danish register closure-cohort scale of Browning and Heinesen (2012), and published GP contact rates of roughly 85% annual prevalence and 4–5 contacts per person-year. Sections 6 and 7 license this sizing only in conjunction with the pre-outcome realized- $N$  demotion gate. If a confirmatory member falls below its binding  $N^*$  or its blind re-simulated power falls below 0.95 at the floor, it is demoted to exploratory before any post-baseline outcome is accessed; floor inflation can never resolve the shortfall.

### 3 Design and identification

#### The design in brief

At the center of the design is a stacked 1:3 riskset-matched difference-in-differences event study of quarterly GP billing around true and spurious Danish workplace closures in SSSY-era cohorts 2009–2023. H2’s confirmatory core is the true-closure excess in the post-year. A spurious-arm falsification gate sits outside the Holm family, the out-of-hours power calibration rests on published evidence, and a realized-N gate can demote tests before any outcome is accessed.

We begin with workers whose primary November-status employment in year  $c - 1$  was at a workplace that disappears between years  $c$  and  $c + 1$ . The worker-flow canon (Benedetto et al. 2007; Hethcote-Maier and Schmieder 2013) classifies an event as a true closure when no successor absorbs the workforce and as a spurious closure when an en-bloc successor exists; the design excludes the ambiguous mid-zone and monitors its mass with a pre-registered bimodality check. Within exact strata for cohort, sex, age band, NACE section, size class, and region, we riskset-match each exposed worker to up to three stable-workplace controls by Mahalanobis distance over a full year of pre-baseline billing history, tenure, and wage decile. The stacked matched-set estimator includes matched-set and cohort-by-calendar-quarter fixed effects, clusters at the event workplace, tests a  $c - 2$  placebo panel against a numeric failure threshold, and calibrates the settlement lag in the fee-week clock.

The Holm family contains three confirmatory quantities, each an honest operationalization of a frozen bank hypothesis. For H1,  $\theta_1$  measures the pre-boundary excess in the last two quarters of the terminal year, described strictly relative to the annually dated boundary. The frozen statement’s growth rider remains a gated exploratory contrast because its ramp-implied floor is verified as undetectable at any honest N.

For H2,  $\theta_{2a}$  measures the true-closure arm’s post-year excess over its own matched controls, which is the excess that the frozen statement asserts through the first post-closure year. A registered falsification gate enforces the comparative quantifier. It estimates and bounds the spurious-closure arm’s excess,  $\theta_{2b}$ , and applies a pre-committed disconfirmation trigger: a spurious excess that is flagged positive and reaches at least half the true-arm excess disconfirms H2’s comparative reading regardless of  $\theta_{2a}$ . The strict differential remains exploratory. This structure follows the verified Stage-0 diagnosis. When the spurious arm is one-third the size of the true arm, its variance dominates the differential, which then requires roughly 5.5 times any honest variance budget. The spurious arm is therefore a falsification device, not an estimation target.

For H3,  $\theta_3$  retains the out-of-hours versus daytime relative-excess gap unchanged. Only its power basis changed, from a self-declared over-conservative 3% out-of-hours base rate to the conservative end of the published Danish range. The simulation uses 8.0% of daytime GP volume, or 7.4% of all GP contacts, against a published contact share of 8–12% (Christensen and Olesen 1998; Moth et al. 2012). Before the confirmatory run, an outcome-neutral check verifies that the realized share lies within the published range.

#### The identifying variation, stated without causal language

The identifying variation comes from terminal events recorded in IDAS across 15 closure cohorts. IDAN worker-flow successor shares distinguish true from spurious events, and riskset-matched workers at stable workplaces provide the comparison within exact strata after matching on 52 weeks of pre-baseline billing history. Annual November-status dating provides no sharp discontinuity within scope. The claim ceiling is therefore associational, every pre-event quantity is described as pre-boundary excess, and the spurious-closure placebo together with the  $c - 2$  placebo panel carries the severity burden. We use no causal machinery — no instrument, discontinuity, or synthetic control. The estimator is the simplest one that makes the registered contrasts exact.

## Alternatives considered

- *The round-1 conservative design (voided)*. A single-era stacked matched-pair (1:1) difference-in-differences, cohorts 2009–2022, sampling self-handicapped to half the lower bound of its own derived expectation (75,000 pairs), H2 as the true-minus-spurious differential, H3 simulated at a 3% out-of-hours base rate. Identical contrast logic; it differed only in matching ratio, sampling plan, H2’s confirmatory construct, and H3’s power basis. Its deposited Stage-0 power simulation emptied the confirmatory family via the registered fixed-point rule (H2 0.358, H1\_trend 0.380, H3 0.614, H1 0.840 alone), and the program’s audit voided the resulting all-exploratory registration on 2026-08-31. The remediation used authorized honest levers only — no floor was raised — and both power runs’ records are published with the protocol.
- *Design B, the ambitious three-arm weekly event study 1991–2023*. Weekly-resolution confirmatory estimands, both eras confirmatory, and a mass-separation arm at surviving workplaces. Weekly confirmatory estimands are hostage to the HONUUGE settlement lag; the SYSI era cannot match on age or sex and sits across the 2007/08 IDA vintage break; the mass-separation arm breaks one-study scope discipline. Its honest elements — the ANTNOV cap of 1000, pre-committed reversed-sign interpretations, and the absorption placebo — were absorbed into the registered design; the SYSI era survives as exploratory E1 and weekly resolution as descriptive trajectories under the week-zero step diagnostic.

## Threats, each paired with the analysis that addresses it

- *Anticipatory attrition — healthier or more mobile workers leave sinking workplaces before closure (Schwerdt 2011), biasing late-window composition*. Addressed by freezing the cohort at November-status attachment in year  $c - 1$ , before the terminal year; intention-to-follow regardless of subsequent separation; and the E4 arm-comparative attrition panel, labeled post-treatment-conditioned descriptives.
- *Billing responses to separations already realized within the terminal year — under intention-to-follow, workers who separate from the closing workplace early in year  $c$  (the wave E4 describes) contribute post-job-loss utilization to the pre-boundary  $Q3(c)/Q4(c)$  window, so a Holm-significant  $\theta_1$  could be generated wholly by realized early job loss with zero anticipation. Direction: inflates  $\theta_1$  relative to a pure-anticipation reading. The frozen clause “before any separation is visible in annual IDAN records” holds at the cohort-anchoring level (attachment is frozen at November of  $c - 1$ , before the terminal year) but not for individual early leavers, whose separation is visible in the year- $c$  IDAN record*. Addressed by the pre-committed H1 interpretation, which names both channels with the E4 year- $c$  separation share bounding the realized-separation channel; the E4 arm-comparative separation shares themselves; and a registered post-treatment-conditioned exploratory split of  $\theta_1$  by November-year- $c$  attachment (E4), which localizes the channel without touching the confirmatory estimand.
- *Selection of sicker workers into dying workplaces masquerading as anticipation or post-year excess*. Addressed by exact-plus-Mahalanobis matching on 52 weeks of pre-baseline billing (contacts, fees, out-of-hours contacts, contact-weeks) in year  $c - 1$ ; the  $c - 2$  placebo panel with a numeric failure threshold; and falsification gate F2, which separates identity-death from economic job loss with a pre-committed disconfirmation trigger that can only take support away from H2.
- *HONUUGE is the fee week, not the service week — billing lag blurs the boundary*. Addressed by quarter-level primary estimands (robust to lags  $\leq 4$  weeks); the BEHANDLINGSDATO calibration (2021+) with a guard-band contingency; AFRPER used descriptively; and the week-zero step diagnostic bounding any timing-shape claim.

- *Annual closure dating — the design cannot see announcement dates, so “anticipation” could be mislabeled.* Addressed by defining and wording the estimand exclusively as pre-boundary excess; the claim ceiling forbids announcement-locked language; the gated H1\_trend partial pattern pre-commits the flat-elevation reading when no within-year growth is seen.
- *Local-economy co-shocks hitting exposed and control workers simultaneously.* Addressed by exact matching on region and NACE section plus cohort  $\times$  calendar-quarter fixed effects; residual local shocks affect controls too and attenuate the contrast (bias toward null, stated).
- *Control contamination by quietly dying or shrinking workplaces.* Addressed by the stable-control definition (unchanged identity  $c - 2$  to  $c + 1$ , no ENDNEDL, AFGRATE  $< 30\%$ ,  $|\text{ANTALXFR}| \leq 30\%$ ); the control-drop rule on data revision; and the forward-window  $(c + 2/c + 3)$  contamination sensitivity with a registered direction note.
- *Reverse causality at small firms — workforce or owner illness causes the closure.* Addressed by the size floor  $\text{ANTNOV} \geq 5$ ; the associational claim ceiling with the reverse channel named in the Limitations; exploratory size heterogeneity.
- *The 2007/08 IDA vintage break contaminating any pooled estimate.* Addressed by confining the confirmatory tier entirely to post-break cohorts ( $c \geq 2009$ ); the SYSI era is exploratory replication and never pooled.
- *Unobservable death and emigration (no population spine in scope) deflating post-event billing.* Addressed by naming the scope limitation and its direction (deflates the post window — conservative for  $\theta_1$ , attenuating for  $\theta_{2a}$ ); a pre-specified sensitivity excluding possible-death/emigration sets; and a 2-percentage-point later-presence trigger forcing best/worst-case bounds on all post-boundary conclusions.
- *Spurious-arm thinness — the structural cause of the first round’s collapse.* Addressed by letting no confirmatory quantity touch the spurious arm:  $\theta_{2a}$  is a true-arm-only contrast; the spurious arm serves as falsification gate F2 (whose job needs bounding, not detection, and whose operating characteristics are verified at Stage 0) and as the exploratory H2\_diff; spurious-arm death caps the H2 verdict rather than silently weakening a test.
- *Residual underpowering if realized  $N$ , realized match ratio, or the realized out-of-hours share falls short of the sampling plan.* Addressed by the pre-outcome Stage-2 realized- $N$  gate with binding  $N^*$  thresholds from the deposited simulation; a blind re-simulation on realized set counts, realized  $k$  distribution, baseline moments, and the realized out-of-hours share; demotion-only resolution before any outcome access; and publication of the realized-power/MDE table regardless.

## Falsification architecture

1. *Falsification gate F2 (spurious-closure placebo arm).*  $\theta_{2b}$  with the pre-committed disconfirmation trigger; outside the Holm family; verified operating characteristics registered (pass probability 0.886 under the hypothesis-true DGP, 0.978 under a strictly inert placebo; trigger probability 0.79 at a floor-sized non-inert placebo, at 45,000 spurious sets — the pre-matching worker count at the 30% band lower bound; applying the registered 90% match-rate floor implies  $\sim 40,500$  sets, so the stated characteristics are computed at a  $\sim 10\%$  flattered lower-bound  $N$ ); the Stage-2 blind re-verification at realized spurious  $N$  is the binding value.
2. *The  $c - 2$  pre-baseline placebo panel.* Exposed-minus-control contact difference across the four quarters of year  $c - 2$  with a numeric failure threshold — the panel fails if and only if the contrast is significant at  $\alpha = 0.05$  and  $|\text{estimate}| > 1\%$  of the control baseline mean,

the single registered condition, identical wherever this gate appears (Sections 5–7); one pre-specified re-match, then kill of the registered event-study tier.

3. *Week-zero boundary-step diagnostic* on the weekly descriptive series (settlement-lag contamination bound on any timing-shape claim).
4. *Successor-share threshold sensitivities* (20/80 and 40/60 alongside the primary 30/70) for  $\theta_{2b}$ /H2\_diff sign stability.
5. *Successor-share bimodality check* at Stage 1: the ambiguous mid-zone must hold under 35% of terminal events.
6. *Later-presence differential trigger* (2 percentage points), forcing best/worst-case bounds on all post-boundary conclusions.
7. *Control forward-window ( $c + 2/c + 3$ ) contamination sensitivity* with a registered direction note: exclusion should weakly increase magnitudes; a decrease is flagged as diagnostic.
8. *Classification cross-validation*: worker-flow classes versus the IDAS-native flags (ENDNEDL, IDFREEM, REGELFR, FRELFREM, ANDELUDS), with a 75% agreement kill floor and the full cross-tabulation released.

## Claim–evidence matrix

Each claim the Stage-2 report may make, with its evidence class fixed in advance.

- *Pre-boundary excess* (associational, confirmatory):  $\theta_1$  from the stacked 1:3 matched-set DiD event study, Holm-adjusted within the three-test family, licensed by the passed  $c - 2$  placebo panel, the matching-balance table, and the pre-outcome realized-N gate certifying at least 0.95 power at the 2% floor.
- *The excess grows toward the boundary* (associational, exploratory, gated): the H1\_trend contrast computed only if  $\theta_1$  Holm-rejects; the week-zero step diagnostic bounds settlement-lag contamination; disclosed as verified structurally underpowered for confirmatory status.
- *True-closure post-year excess whose spurious-arm counterpart is bounded below the registered fraction of the true-arm excess (one-sided 95% upper bound)* (associational; confirmatory core plus registered falsification gate — a passing gate bounds the spurious-arm excess, it cannot license an absence claim):  $\theta_{2a}$  Holm-adjusted, plus gate F2 with its pre-committed trigger, the threshold-sensitivity table, and the classification cross-validation; the strict differential magnitude is exploratory H2\_diff with its CI.
- *Out-of-hours composition tilt* (associational, confirmatory):  $\theta_3$  with a 999-draw cluster-bootstrap CI; power basis recalibrated to the published Danish out-of-hours share and re-verified against the realized share before the confirmatory run; robustness under the YDLTID-augmented classifier.
- *Bounded priority claim* (descriptive): the first register-scale, matched, weekly-resolution estimate of primary-care billing trajectories around workplace closures — positioned against the recorded 62-query search; expressly *not* a phenomenon-discovery or Danish-first claim (Iversen, Sabroe and Damsgaard 1989 and Rugulies et al. 2009 precede).
- *Weekly trajectory figures* (descriptive): pooled exposed-minus-matched-control weekly series across at least 100 events per figure, disclosure-compliant, with the step diagnostic governing permissible timing-shape language; no test attached.

- *Methodological disclosure* (binding): this protocol is a remediated second registration; both power runs' records, the fixed-point demotion log, and the remediation note are published with it; no floor effect was raised, and every N lever is cited to a published basis with a pre-outcome realized-N kill floor.
- *Scope statement* (binding): outcomes are utilization volume, fees, and timing composition — never health states; only IDAN, IDAS, SSSY, and SYSI exist on the analysis server.

## 4 Study design table

Hypotheses appear verbatim as frozen; interpretation sentences are the pre-committed conclusions.

| Question                                                                                                                                                                             | Hypothesis (verbatim)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Sampling plan (power)                                                                                                                                                                                                                                                                                                                                                                                                                           | Analysis plan                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | Interpretation given different outcomes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>RQ1.</b> Does primary-care contact already rise in the final quarters before a workplace's closure-year boundary — before any separation is visible in annual employment records? | <p>H1 (anticipation): Compared to riskset-matched workers at stable workplaces with parallel 52-week baseline billing histories, workers employed at workplaces in their final year before permanent closure (IDAS ENDNEDL) will show elevated weekly primary-care contact rates (SYSI/SSSY contacts dated by HONUKE; YDLANT, KONTAKT) already in the two calendar quarters preceding the closure-year boundary, with the excess growing as the boundary approaches — before any separation is visible in annual IDAN records. <i>H0</i>: Primary-care contact among workers at closing workplaces does not move before the closure-year boundary: any anticipatory response at or above 2% of baseline is ruled out at the design's error rates, locating whatever billing response exists at or after separation. A Holm-significant <i>decrease</i> is not this null but a pre-registered informative reversal (deferred or forgone care under insecurity, per Hamad, Modrek and Cullen 2015). <i>Tiering</i>: confirmatory core <math>\theta_1</math>; the growth rider is the gated exploratory H1_trend.</p> | <p>Closure cohorts 2009–2023 (15 cohorts); expected 150,000–300,000 exposed true-closure workers, <math>\geq</math> 135,000 1:3 matched sets at the 90% match-rate floor. Verified a priori power <b>0.980</b> at the 2% floor under full-family Holm — the least-powered confirmatory test. Binding realized-N gate <math>N_{H1}^* = 118,000</math> sets (supersedes the provisional 100,000); Stage-2 blind re-simulation, demotion-only.</p> | <p><math>\theta_1 = (\beta_{Q3(e)} + \beta_{Q4(e)})/2</math> from the stacked matched-set DiD event study (set FE + cohort <math>\times</math> calendar-quarter FE; CR1 SEs clustered on event workplace; controls weighted <math>1/k</math>); winsorization at the 99.9th percentile within era <math>\times</math> calendar quarter; zero utilization is a true zero, never missing; Poisson QML robustness; <math>c - 2</math> placebo panel with numeric threshold and a single pre-specified re-match; contingencies: ENDNEDL re-anchoring, coarsening ladder, settlement-lag guard band, KONTAKT fallback (Section 6).</p> | <p>If <math>\theta_1 &gt; 0</math>, Holm-significant: we will conclude that workers at closing workplaces show excess primary-care contact in the final two pre-boundary quarters, consistent with anticipatory care-seeking and/or billing responses to separations already realized early within the terminal year, with the E4 year-<i>c</i> separation share bounding the latter channel — as a pre-boundary association relative to an annually dated boundary, expressly not an announcement-dated anticipation effect (first register-scale, matched, weekly-resolution estimate in primary care; not a phenomenon discovery). The gated exploratory H1_trend then determines which timing-shape language is permitted. If not Holm-significant (null): we will conclude that we have no evidence that primary-care contact is elevated in the two pre-boundary quarters among workers at closing workplaces relative to matched stable-workplace controls, and we will report the excess magnitudes the 95% CI rules out, quoted against the realized MDE; the null is informative at the registered <math>\geq 0.95</math> power. If <math>\theta_1 &lt; 0</math>, Holm-significant (reversed): we will conclude — contrary to our hypothesis — that pre-boundary contact falls, a pattern consistent with forgone or deferred routine care under job insecurity (Hamad, Modrek and Cullen 2015); pre-registered as informative, disconfirming, never reframed as support.</p> |

| Question                                                                                                                              | Hypothesis (verbatim)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Sampling plan (power)                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Analysis plan                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Interpretation given different outcomes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|---------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>RQ2.</b> Does the post-closure billing excess attach to genuinely destroyed jobs rather than to workplaces that die only on paper? | H2 (true vs spurious closure): Compared to workers whose workplace disappears only administratively — absorbed into a successor unit with a high absorbed-worker share (IDAS ANDELOPS) and preserved forward identity (IDFREM, REGELFR), so employment continues — workers at true permanent closures (ENDNEDL with low ANDELOPS) will show a larger and more persistent post-event excess in weekly contacts and gross fees (BRUHON) through the first post-closure year. <i>H0</i> : Workers at true closures show no post-boundary excess in primary-care billing through the first post-closure year relative to matched controls (post-year excess at or above 2% of baseline ruled out); together with $\theta_{2b}$ near zero this locates any billing response, if $\theta_1$ found one, strictly before the boundary. <i>Tiering</i> : confirmatory core $\theta_{2a}$ (true-arm post-year excess); comparative quantifier enforced by gate F2 outside the Holm family; strict differential is exploratory H2_diff. | Same true-arm plan ( $\geq 135,000$ sets); spurious arm expected at 30–60% of the true count, $\geq 15,000$ spurious sets required for gate F2. Verified a priori power <b>0.998</b> at the 2% floor under full-family Holm. Verified F2 operating characteristics at 45,000 spurious sets: pass 0.886 (hypothesis-true), 0.978 (inert), trigger 0.79 at a floor-sized non-inert placebo. Binding gates: $N_{H2a}^* = 95,000$ true sets; F2 re-verified blind at realized spurious N. | $\theta_{2a} = \text{mean of } \beta_{\text{true},k} \text{ over } Q1(c+1) - Q4(c+1)$ from the true-arm stacked model; $\theta_{2b}$ and exploratory H2_diff from the pooled model with event-class $\times$ event-quarter interactions; gate F2 fires iff $\theta_{2b}$ 's 95% CI excludes zero positively <i>and</i> $\theta_{2b} \geq 50\%$ of $\hat{\theta}_{2a}$ ; BRUHON fee secondary hierarchically gated on $\theta_{2a}$ ; successor-share threshold sensitivities 20/80, 40/60; later-presence trigger forces bounds; possible-death/emigration sensitivity; same FE, clustering, outlier and missing rules as RQ1. | If $\theta_{2a} > 0$ , Holm-significant, F2 passes, sign stable across thresholds: we will conclude that workers at true closures show a post-boundary billing excess through the first post-closure year whose spurious-arm counterpart is bounded below $X\%$ of the true-arm excess, where $X$ is the realized one-sided 95% upper bound on $\theta_{2b}$ as a share of $\hat{\theta}_{2a}$ (the registered gate quantity; a passing gate bounds the spurious-arm excess — it cannot license an absence claim) — consistent with job-loss exposure rather than dying-firm selection; comparative magnitude reported only from exploratory H2_diff with its CI. If Holm-significant but gate F2 fires: we will conclude that H2's comparative reading is disconfirmed — the placebo is not inert, and the shared excess is consistent with selection into dying workplace identities or reorganization stress — regardless of $\theta_{2a}$ ; the $\theta_1$ interpretation is weakened. If not Holm-significant and $\theta_{2b}$ near zero (null): we will conclude that we have no evidence of post-boundary excess primary-care billing after terminal events of either kind, report the CI-ruled-out magnitudes for both arms, and note this locates any billing response (per $\theta_1$ ) before the boundary. If not Holm-significant and $\theta_{2b}$ flagged positive (95% CI excludes zero): we will conclude the placebo arm is not inert — the billing excess appears where employment continued under a new workplace identity — read the pattern as evidence against closure-specific job-loss exposure driving billing, and report the CI-ruled-out magnitudes for both arms; the H2 |

| Question                                                                                             | Hypothesis (verbatim)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Sampling plan (power)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | Analysis plan                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Interpretation given different outcomes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>RQ3.</b> Is the peri-event excess compositionally tilted toward out-of-hours, unplanned services? | H3 (distress composition): Compared to the out-of-hours share of contemporaneous contacts among their matched controls, closure-exposed workers' peri-event excess contacts will contain a disproportionately higher share of out-of-hours and off-schedule services (YDLTID time-of-day codes; VAGTOMR in the SYSI era), consistent with unplanned, distress-driven care-seeking rather than deferred routine care. <i>H0</i> : The peri-event excess is compositionally unremarkable: the out-of-hours share of excess contacts does not differ from the daytime share by 5 points or more (bootstrap CI converted to ruled-out gaps against the realized MDE), consistent with scheduled as much as distress-driven care. | Same true-arm plan ( $\geq 135,000$ sets); out-of-hours base rate simulated at 8.0% of daytime GP volume (7.4% of all GP contacts), the conservative end of the published Danish 8–12% contact-share range, with an outcome-neutral check verifying the realized share before the confirmatory run. Verified a priori power <b>1.000</b> at the 5-point floor under full-family Holm. Binding gate $N_{H3}^* = 90,000$ sets (provisional value retained as the stricter bound over the re-run's 66,000). | $\theta_3 = (\theta_{OOH}/\mu_{OOH,ctrl}) - (\theta_{day}/\mu_{day,ctrl})$ over the peri-event window $Q3(c) - Q2(c+1)$ ; cluster-bootstrap percentile CI (999 draws, seed 20260831, clusters = event workplaces); frozen code lists (out-of-hours = SPEC2 81–84, 89; robustness classifier adds YDLTID codes 2–5, 7–9; missing YDLTID classified daytime, conservative; SPEC2 71/73/75 “evening etc.” GP rows stay daytime in the primary classifier by registered choice — disclosure and validation in Section 5 — with both classifiers reported side by side in Table T7); same stacks, outlier and missing rules as RQ1. | If $\theta_3 > 0$ , Holm-significant, with daytime excess $\geq 0$ : we will conclude that the peri-event excess is disproportionately out-of-hours, consistent with unplanned, distress-driven care-seeking rather than deferred routine care (composition claim only; no health-state claim is possible in scope). If $\theta_3 > 0$ , Holm-significant, with daytime excess $< 0$ (total down, out-of-hours up): we will conclude that closure exposure is associated with a compositional shift matching the forgone-routine-care-plus-acute-distress pattern (Hamad, Modrek and Cullen 2015); this supports H3's composition claim while disconfirming any volume-increase reading of $\theta_1$ . If not Holm-significant (null): we will conclude that we have no evidence that the peri-event excess is compositionally tilted toward out-of-hours care, and report the relative-excess differentials the 95% bootstrap CI rules out against the realized MDE; any $\theta_1$ excess is then compositionally unremarkable — quoted with the registered caveat that the primary classifier counts SPEC2 71/73/75 evening-GP services as daytime, which attenuates $\theta_3$ and correspondingly weakens the null bound (Section 5). If $\theta_3 < 0$ , Holm-significant (reversed): we will conclude — contrary to our hypothesis — that the excess is disproportionately daytime/scheduled, consistent with planned or deferred routine care; H3 is disconfirmed. |

| Question                                                                                                         | Hypothesis (verbatim)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Sampling plan (power)                                                                                                                                                                                                                                                                                                                                                                                                                | Analysis plan                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Interpretation given different outcomes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Exploratory tier.</b> Where does the excess localize — in time shape, across arms, eras, fees, and subgroups? | <b>H1_trend</b> (rider excerpted verbatim from H1): "...with the excess growing as the boundary approaches ..." — gated on $\theta_1$ Holm-rejecting in the predicted direction.<br><b>H2_diff</b> (quantifier excerpted verbatim from H2): "...a larger and more persistent post-event excess in weekly contacts and gross fees (BRUHON) through the first post-closure year ..." <b>E1</b> : SYSI-era descriptive replication of the H1 contrast, cohorts 1992–2003. <b>E3</b> : gross-fee analogues and heterogeneity by sex, age band, wage decile. <b>E4</b> : arm-comparative attrition accounting panel, plus a post-treatment-conditioned $\theta_1$ split by November-year- $c$ attachment. | Same cohorts, sets, and disclosure floors as the confirmatory tier; no 0.95 power target applies. Stage-0 reference values: gated joint power 0.772 for H1_trend at the ramp-implied 1.714% floor; 0.460 for H2_diff at the hypothesis-true +1.5-point differential — both verified infeasible as confirmatory members, which is why they are tiered here. E1 skippable (under 300 era closures or a failed era coverage benchmark). | H1_trend: $\beta_{Q4(c)} - \beta_{Q1(c)}$ with 95% CI, computed only if the gate is met; otherwise the quarterly trajectory is shown descriptively with no contrast statistic.<br>H2_diff: pooled-model differential with 95% CI and the quarter-by-quarter persistence profile.<br>E1: YDLANT-weighted service-row proxy, legacy keys, matching without age/sex. E3: within-family Holm as a severity aid. E4: class-specific year- $c$ separation shares, descriptive only, no test; plus the $\theta_1$ split by November-year- $c$ attachment (post-treatment-conditioned, selection caveat mandatory). Every row labeled exploratory. | Exploratory results never gate, upgrade, or revise any confirmatory verdict. H1_trend gate met and positive with CI excluding zero: reported, as a gated exploratory estimate, as consistent with — but not probative of — intensifying anticipatory care-seeking; annual dating still forbids announcement-locked claims. Gate met but CI includes zero: we will conclude the excess is a flat elevation across the terminal year, lean on gate F2 and the $c - 2$ placebo to bound the selection component, and score H1's growth rider not supported at the available resolution. Gate not met: no trend statistic is computed or foregrounded. H2_diff in the non-predicted direction (spurious > true): we will report — contrary to the frozen comparative expectation — that the placebo is not inert, reinforcing whatever gate F2 showed and capping the interpretive weight of the true/spurious contrast. |

The machine-readable registration fixes the scoring rule in advance: at Stage 2 an outcome attaches to each confirmatory hypothesis's frozen statement as confirmed, partially confirmed, or reversed, adjudicated strictly against the pre-committed interpretation table. H1 scores confirmed only if  $\theta_1$  Holm-rejects in the predicted direction *and* the gated exploratory H1\_trend supports the frozen growth rider; partially confirmed when  $\theta_1$  rejects but the rider is unsupported (the pre-committed flat-elevation reading) or ungated; reversed if Holm-significant negative. H2 scores confirmed only if  $\theta_{2a}$  Holm-rejects in the predicted direction *and* gate F2 does not fire *and* the  $\theta_{2b}$  sign pattern is stable across the successor-share thresholds; partially confirmed when the comparative quantifier is untestable (spurious-arm kill); reversed if  $\theta_{2a}$  is Holm-significant negative *or* gate F2 fires. H3 scores confirmed only if  $\theta_3$  Holm-rejects in the predicted direction (the total-down/out-of-hours-up pattern is pre-registered as scoreable support); reversed if significant negative. A Holm-nonsignificant result attaches no confirmation — it is a bounded informative null, never "accepted." A member demoted by the realized-N gate or killed by its criterion attaches no score; the trail is disclosed.

## 5 Analysis plan

### Preregistered analyses: estimators, named exactly

We estimate a stacked matched-set difference-in-differences event study by OLS in R  $\geq 4.3$  using `fixest::feols`, with versions frozen in the registered script. The outcome  $y_{iq}$  is the winsorized quarterly GP contact count. We regress it on treated  $\times$  event-quarter indicators for the eight quarters from  $Q1(c)$  through  $Q4(c+1)$ , together with matched-set and cohort  $\times$  calendar-quarter fixed effects. Each set contains one exposed worker and up to three controls, with controls weighted  $1/k_s$  within the set. The four quarters of year  $c - 1$  form the omitted baseline; quarters in year  $c - 2$  enter only the placebo panel. CR1 cluster-robust standard errors are clustered at the event workplace, and each control inherits its set’s event cluster.

The true-arm stacked model yields  $\theta_{2a}$ . A pooled model with event-class  $\times$  event-quarter interactions yields  $\theta_{2b}$  for gate F2 and the exploratory H2\_diff. One table reports successor-share sensitivities at 20/80 and 40/60 alongside the primary 30/70 thresholds. The ratio contrast  $\theta_3$ , defined in the design table, receives percentile CIs from a cluster bootstrap over event workplaces with 999 draws and RNG seed 20260831.

A pre-specified sensitivity re-estimates  $\theta_1/\theta_{2a}$  after excluding sets whose control workplace acquires a terminal event in  $c+2$  or  $c+3$ . The registered direction note states that contamination attenuates toward the null, so exclusion should weakly increase magnitudes; a decrease is diagnostic, and truncation of the window for the last two cohorts is disclosed. The pre-specified count-model robustness check uses Poisson QML through `fixest::fepois`.

We match by 1:3 nearest-neighbor Mahalanobis distance within exact strata, without replacing controls across sets. The target is  $k = 3$  and the minimum is  $k = 1$ . We report the realized  $k$  distribution, feed it into the Stage-2 blind re-simulation, and process exposed workers in seeded random order using seed 20260831. Exact strata cross cohort  $c$  with sex from KOENIMP, including unknown; a 5-year age band from ALDERIMP, including unknown; NACE section from ARB\_HOVED\_BRA\_DB07; workplace size class from ANTNOV (5–9/10–19/20–49/50–99/100–499/500–1000); and region, obtained by mapping workplace KOM to the five post-2007 regions.

All Mahalanobis components come from year  $c - 1$ : total GP contacts,  $\log(1 + \text{BRUHON})$ , out-of-hours contacts, weeks with at least one contact, tenure measured as  $c - 1$  minus ANSAAR, and wage decile from BREDT\_LOEN\_BELOEB. Calipers require a baseline-contact difference  $\leq 3$  and a  $\log(1 + \text{BRUHON})$  difference  $\leq 0.5$ . By design, the model includes no post-matching regression covariates. The descriptive contrast rests entirely on exact strata, Mahalanobis matching, matched-set fixed effects, and cohort  $\times$  calendar-quarter fixed effects.

Confirmation requires a Holm-adjusted two-sided  $p < .05$  within the final family and a point estimate in the predicted direction; a Holm-significant estimate in the non-predicted direction is disconfirming for the confirmatory test, never supportive. Each confirmatory hypothesis closes, at Stage 2, with a boldface Summary verdict paragraph opening “Consistent with” or “Contrary to our hypothesis.”

### Preprocessing, in order

1. Freeze code lists before data access: daytime GP = SPEC2 codes 70–80; out-of-hours GP = SPEC2 codes 81–84 and 89; the robustness out-of-hours classifier adds YDLTID codes 2–5 and 7–9. Code labels are quoted verbatim from the parsed value labels and re-validated in-environment before estimation. Registered disclosure on the evening GP codes: the parsed value labels mark SPEC2 71 (“General practitioner, consultation, evening etc.”), 73 (“General practitioner, telephone consultation, evening etc.”), and 75 (“General practitioner, home visit, evening etc.”) as explicitly off-schedule GP services; they nonetheless remain on the daytime side of the *primary* classifier because H3’s out-of-hours construct and its published power basis

are keyed to the lægevagt out-of-hours service (SPEC2 81–84, 89), the quantity the published Danish 8–12% share describes. The attenuation direction is stated for both readings: counting evening GP as daytime attenuates  $\theta_3$  toward zero, which is conservative against the positive H3 conclusion but anti-conservative for the null-bounding “compositionally unremarkable” reading, whose Stage-2 statement therefore carries a mandatory caveat. In-environment validation must verify that the YDLTID-augmented robustness classifier captures SPEC2 71/73/75 rows (their evening time codes 2–5); if YDLTID proves unpopulated on those rows, the robustness classifier — and only the robustness classifier — is amended mechanically and with disclosure to include SPEC2 71/73/75 directly. The SPEC2 71/73/75 volume share is reported, and both classifiers appear side by side in Table T7.

2. Build the IDAS event panel per era key system (ARBNR/CVRNR/LBNR, 2007+): workplace identity chains via IDFREM/IDTILB; identity-disappearance events between  $c$  and  $c + 1$  flagged; ENDNEDL, REGELFR, FRELFR, ANDELOPS, ANDELUDS attached.
3. Compute worker-flow successor shares  $s(w)$  from IDAN: among the event workplace’s year- $c$  November primary employees, the largest share whose year- $c + 1$  primary workplace is a single successor identity (computed on the year- $c$  workforce, independent of the frozen  $c - 1$  cohort).
4. Classify events: true closure (ENDNEDL final and  $s(w) < 0.30$ ), spurious closure (forward identity and  $s(w) \geq 0.70$ ), else ambiguous-excluded; cross-tabulate against IDAS flags; run the successor-share bimodality check at Stage 1.
5. Extract frozen cohorts at November of  $c - 1$ ; build control pools from stable workplaces; run 1:3 matching with realized- $k$  accounting.
6. Extract SSSY 2006–2025: person  $\times$  HONUGE-week aggregation of KONTAKT (contacts), YDLANT (services), and BRUHON (fees) within the frozen code lists; map weeks to quarters by ISO week of HONUGE (1–13/14–26/27–39/40–52; ISO week 53 to Q4).
7. Winsorize person-quarter outcomes at the 99.9th percentile within era  $\times$  calendar quarter (thresholds computed pre-estimation and frozen).
8. Assemble the panel: 16 event-quarters per set,  $Q1(c-2)$ – $Q4(c+1)$ ; estimation panel  $Q1(c-1)$ – $Q4(c+1)$ ; placebo panel  $Q1(c-2)$ – $Q4(c-2)$ .
9. Report the BEHANDLINGSDATO (2021+) settlement-lag calibration table before any event-study estimation.
10. Do not deflate BRUHON: nominal fees are compared only within cohort  $\times$  calendar-quarter cells.
11. Run the registered sensitivity suite once and report it together: successor-share thresholds; the control forward-window exclusion; un-winsorized outcomes; Poisson QML; and the possible-death/emigration set exclusion.

### **Multiplicity: the registered rule**

The registered convention (Holm 1979), reproduced from the machine-readable registration with notation typeset and otherwise verbatim: Holm–Bonferroni across the three confirmatory tests ( $\theta_1$ ,  $\theta_{2a}$ ,  $\theta_3$ ) at family-wise  $\alpha = 0.05$ , two-sided  $p$ -values with pre-registered directions; a Holm-significant estimate in the non-predicted direction is disconfirming, never supportive (exemplar convention). Hierarchical gates: the BRUHON fee secondary is tested only if  $\theta_{2a}$  Holm-rejects; the exploratory H1\_trend contrast is computed only if  $\theta_1$  Holm-rejects in the

predicted direction. Falsification gate F2 sits outside the Holm family as a registered severity device with a pre-committed disconfirmation trigger (it consumes no family alpha and can only take support away, never add it). Exploratory families (H2\_diff; E1; E3; E4) receive Holm correction within family as a severity aid and are segregated by heading. If the Stage-2 realized-N gate demotes members, the family shrinks by the registered fixed-point iteration; the surviving family’s alphas are those of the reduced family. Null results are bounded, not accepted: each null is converted into the effect magnitudes the 95% CI rules out, quoted against realized MDEs from the Stage-2 blind re-simulation. No other multiplicity convention applies anywhere in this protocol.

### Contingent decision rules (pre-committed; nothing left to choose)

1. If the ENDNEDL flag-year convention differs in-environment from the assumed “flagged in the workplace’s last active year”: re-anchor mechanically so that  $c$  always denotes the last IDAS year with  $\text{ANTNOV} \geq 5$  for the identity; the estimand is unchanged; the recalibration is reported verbatim.
2. If ARB\_SEKTORKODE is absent from the analysis-server IDAN extract (it is publicly documented on TIMES *ida-ansattelser* but unconfirmed in the recorded Forskningservice inventory; presence and 2008+ coverage are validated at Stage 1 before any classification work): apply the frozen fallback — the private-sector filter is defined from ARB\_HOVED\_-BRA\_DB07 instead, a workplace counting as private if and only if its NACE DB07 section is not O (public administration and defence), P (education), or Q (human health and social work) — with the substitution and its effect on eligible-workplace and worker counts reported verbatim; no other filter change is permitted.
3. If the match rate is below 95% of eligible exposed workers at target  $k = 3$  under full exact strata: apply the coarsening ladder in order, stopping at the first rung reaching 95% matched with  $k \geq 1$  — (1) NACE section to 10 aggregate industry groups; (2) drop the region stratum; (3) age bands 5-year to 10-year. Never drop sex, cohort, size class, or the baseline-billing calipers.
4. If the median HONUGE settlement lag (BEHANDLINGSDATO calibration, 2021+ rows) exceeds 4 weeks: apply the guard band — exclude  $Q1(c+1)$  from the  $\theta_{2a}/\theta_{2b}$  windows (means over  $Q2(c+1)$ – $Q4(c+1)$ ) and report  $\theta_1$  both as registered and with a 4-week-shifted boundary; the registered  $\theta_1$  window remains primary.
5. If KONTAKT does not behave as a row-level contact marker on in-environment validation: switch to the frozen fallback contact definition (count of distinct person  $\times$  HONUGE-week  $\times$  YDERNR  $\times$  code-group cells with at least one positive-fee row); report the switch and level effect; re-run all estimands under both definitions with the fallback as primary.
6. If OLS and Poisson QML disagree in sign on any primary quantity: foreground the pre-committed OLS estimate, report both, flag the discordance in the Summary paragraph; no post hoc model selection.
7. The  $c - 2$  placebo panel *fails* if and only if the placebo contrast is significant at  $\alpha = 0.05$  and  $|\text{estimate}| > 1\%$  of the control baseline mean (50% of the  $\theta_1$  floor effect) — the single registered condition, stated identically in the outcome-neutral check (Section 6), this rule, and the H1 placebo-kill criterion (Section 7). On failure: execute the single pre-specified re-match adding year- $c - 2$  baseline contacts to the Mahalanobis set; if the placebo fails again under the same condition, the entire registered event-study tier is killed and the study reports descriptive trajectories only. The two non-complementary cells are pre-committed passes with mandatory disclosure, never discretion: (a) significant at  $\alpha = 0.05$  with  $|\text{estimate}| \leq$

- 1% — no re-match, no kill; the Table T7 placebo row flags the significance and states that a sub-floor-scale imbalance was detected; (b) not significant with  $|\text{estimate}| > 1\%$  — no re-match, no kill; a mandatory caution paragraph reports the placebo magnitude against the  $\theta_1$  floor and the realized placebo MDE, and the H1/H2 Summary paragraphs must cite it.
8. If a matched control’s own workplace acquires a terminal event within  $c - 2$  to  $c + 1$  after matching (data revision): drop that control from the set; drop the whole set only if no control remains; report dropped counts; no re-matching.
  9. If YDLTID is missing or coded “not relevant” on a GP row: classify as daytime (conservative against the H3 prediction); report the missingness share; if it exceeds 20% of peri-event GP rows, H3’s kill criterion triggers.
  10. If a Mahalanobis distance ties exactly: break deterministically by smallest absolute baseline-contact difference, then by seeded RNG (seed 20260831).
  11. If IDAS-flag versus worker-flow classification agreement is below 90% but at least 75%: proceed with the worker-flow classification as registered (the canon); publish the full disagreement cross-tabulation; below 75%, H2’s kill criterion governs.
  12. If the sign of  $\theta_{2b}$  or H2\_diff is not stable across the three successor-share threshold pairs: flag classification fragility in the pre-committed H2 Summary paragraph; the 30/70 estimate remains primary; no threshold shopping.
  13. If the exposed-versus-control difference in later-presence rates (any IDAN record in year  $c+1$  or any SSSY row in  $c+1$  or later) exceeds 2 percentage points: a mandatory bounding paragraph accompanies all post-boundary conclusions, converting the differential into best/worst-case bounds; the trigger and threshold are registered, not discretionary.
  14. If the control forward-window sensitivity yields *smaller* magnitudes than the primary estimates: flag as diagnostic of something other than contamination (registered direction note); report verbatim; no re-specification.
  15. Stage-2 realized-N gate: if realized matched sets (or spurious sets for F2) fall below a member’s binding  $N^*$ , or the Stage-2 blind re-simulation (realized Ns, realized  $k$  distribution, baseline-year moments only) shows power below 0.95 for a member at its floor: that member is demoted to exploratory *before* any post-baseline outcome is accessed; the Holm family shrinks and alphas recompute by the registered fixed-point iteration; demotion-only — no member, threshold, or floor may move in the other direction; the realized-power/MDE table is published either way.
  16. If the realized control-group out-of-hours share of GP contacts is below 5% (below the published 8–12% range): re-verify H3’s power basis in the Stage-2 blind re-simulation at the realized share; if below 0.95 at the 5-point floor, H3 demotes to exploratory per the realized-N gate; if the realized share is below 2%, H3’s kill criterion governs instead.
  17. If the spurious arm is killed at Stage 1 or 2 (events below 150, sets below 3,000, classification agreement below 75%, or bimodality unresolved): gate F2 and  $\theta_{2b}$ /H2\_diff are reported unavailable (drop disclosed before outcome access);  $\theta_{2a}$  remains in the Holm family if its own  $N^*$  is met, but the H2 verdict is capped at “partially consistent — comparative quantifier untestable,” and the Discussion states that the excess cannot then be distinguished from selection into disappearing workplace identities.

## Outlier and missing-data rules

Person-quarter GP contact counts and BRUHON sums are winsorized at the 99.9th percentile within era  $\times$  calendar quarter (thresholds computed pre-estimation and frozen); un-winsorized estimates are reported as sensitivity; no observation is deleted for outlyingness. Absence of SSSY rows is zero utilization by construction of the claims universe, never missing data — person-quarters are retained with  $y = 0$ . Persons with no SSSY row in  $c - 3$  to  $c - 1$  enter explicit unknown age/sex matching strata. Workers with missing ANSAAR or missing/zero annual pay in the  $c - 1$  IDAN record are excluded pre-matching with counts reported. Missing YDLTID maps to daytime (rule above). Post-event disappearance from both IDAN and SSSY — possible death or emigration, neither observable in scope — is retained under intention-to-follow; a pre-specified sensitivity re-estimates excluding sets where the exposed member (or all controls) has zero SSSY rows across  $c$  to  $c + 1$  and no IDAN record in  $c + 1$ . The 2-percentage-point later-presence trigger applies as registered.

## 6 Power and pre-analysis simulation

### Method and data-generating process

Power is certified by a deposited Monte Carlo simulation (Python 3, numpy; root seed 20260831; 500 draws for the primary scenario, 250 per sensitivity scenario), using exact set-level-difference equivalents of the registered `fixest::feols` specification with CR1 event-workplace clustering. This is the *remediated re-run* of the same deposited Stage-0 script whose first round emptied the confirmatory family; the round-1 record survives verbatim in the program’s design file and is summarized below.

The data-generating process follows person-quarter GP contact counts across a 12-event-quarter panel: four baseline quarters in  $c - 1$ ,  $Q1$ – $Q4$  of the terminal year  $c$ , and  $Q1$ – $Q4$  of  $c + 1$ . Counts are conditionally Poisson, with a daytime-GP control mean of 1.1 contacts per person-quarter based on Statistics Denmark. Persistent person gamma frailty has variance 1.25 and cancels in within-set DiD; transient person-quarter lognormal shocks have log-variance 0.25 and do not cancel. We retained this overdispersion split *unchanged* from the failed round. Together, the components yield 53.5% zero person-quarters and cross-sectional variance 3.49.

Event-workplace  $\times$  quarter shocks are lognormal and follow AR(1) with  $\rho = 0.5$  and log-variance 0.0375. The realized event-level ICC is approximately 0.014, in the lower half of the design’s stated 0.01–0.05 plausibility band. The 0.95 certification is therefore conditional on that half. The Stage-2 blind re-simulation at the realized variance components is the binding re-verification, and registered demotion gates apply if it falls short.

The remediation used five audit-authorized, honest levers and raised no floor. (1) *Sampling*. We sample at the *lower* bound of the full derived expectation; the voided round had self-handicapped to half. This gives 135,000 true-closure 1:3 matched sets across 12,600 event workplaces and 45,000 spurious sets across 4,200 events, the 30% lower bound of the 30–60% band. (2) *Matching*. We use 1:3 riskset matching. The simulation represents each set’s three controls exactly as one Poisson draw from their summed conditional rates, weighted 1/3 and assigned the set’s event cluster. (3) *H2*. The confirmatory quantity is re-operationalized as  $\theta_{2a}$ , the true-arm-only four-quarter post-year excess at the unchanged 2% floor; the spurious arm serves only gate F2 and the exploratory  $H2\_diff$ . (4) *Out-of-hours volume*. We recalibrated the base rate from the self-declared over-conservative 3% to 8.0% of daytime GP volume, or 7.4% of all GP contacts, which is the conservative end of the published Danish range (Christensen and Olesen 1998; Moth et al. 2012). (5) *Serial correlation*. We build the registered discount into the DGP instead of assuming it away. Transient person shocks follow AR(1) with  $\rho = 0.5$ , while marginals, zero share, variance, and ICC remain identical to round 1. This reduces the averaging gain in every multi-quarter window, including  $\theta_{2a}$ ’s, and weakly raises every SE relative to round

1.

Effects multiply the exposed means. In year  $c$ , a linear ramp has  $\text{mean}(Q3, Q4) = 2\%$ , the  $\theta_1$  floor, and implies a  $Q4 - Q1$  trend of 1.714% for the exploratory H1\_trend. The true post-year excess is 2.0%, the  $\theta_{2a}$  floor; the voided round’s DGP used 2.5%, so simulation at the floor is conservative. In the hypothesis-true world, the spurious residual is 0.5%; we retained it rather than tune it to an idealized inert zero, which is reported as a reference. The out-of-hours-versus-daytime relative gap is +5 points, the  $\theta_3$  floor. The simulation omits winsorization, conservatively. Holm–Bonferroni operates over the confirmatory family at family-wise  $\alpha = 0.05$ , uses two-sided tests, counts rejection only in the predicted direction, and applies the demotion-only fixed-point rule at the final family’s Holm alphas.

## Lowest-effect basis

The registered floor effects are unchanged from the voided round — no floor was raised to buy power. H1: +2% relative pre-boundary excess of the control baseline mean (0.022 contacts per person-quarter on a 1.1 base), set below the layoff-insecurity utilization shifts of roughly 5–10% in Hamad, Modrek and Cullen (2015). H2: +2% relative true-arm post-year excess, set below Hamad, Modrek and Cullen (2015) and below the post-displacement utilization and expenditure magnitudes in Kuhn, Lalive and Zweimüller (2009). H3: a 5-point gap in relative peri-event excess between out-of-hours and daytime contacts, below the published composition shifts (emergency up, outpatient down) in the same insecurity literature; what changed for H3 is only the power basis (the out-of-hours base rate), not the floor. Exploratory reference values: H1\_trend at the ramp-implied 1.714% of baseline; H2\_diff at the hypothesis-true +1.5-point differential. These are smallest effects of interest per the NHB registered-report convention and the design’s own commitments.

## Results

The fixed-point iteration converged at step 1 with zero demotions: the final confirmatory family is {H1, H2, H3} at its own Holm alphas, all at or above 0.95. The least-powered confirmatory test is H1 ( $\theta_1$ ) at **0.980** under full-family Holm.

| Quantity           | Tier                | Floor effect                         | Assumed N                    | Power               |
|--------------------|---------------------|--------------------------------------|------------------------------|---------------------|
| $\theta_1$ (H1)    | confirmatory        | +2% of baseline                      | 135,000 sets / 12,600 events | 0.980               |
| $\theta_{2a}$ (H2) | confirmatory        | +2% of baseline                      | 135,000 sets / 12,600 events | 0.998               |
| $\theta_3$ (H3)    | confirmatory        | +5-point gap                         | 135,000 sets / 12,600 events | 1.000               |
| H1_trend           | exploratory (gated) | 1.714% of baseline                   | same                         | 0.772 (joint)       |
| H2_diff            | exploratory         | +1.5-point differential              | 135,000 vs. 45,000 sets      | 0.460               |
| F2 pass            | gate characteristic | spurious residual +0.5%              | 45,000 spurious sets         | 0.886 (0.978 inert) |
| F2 trigger         | gate characteristic | $\theta_{2b} = \theta_{2a}$ at floor | 45,000 spurious sets         | 0.790               |

Estimator recovery is unbiased (mean estimates 0.0220/0.0219/0.0501 against targets 0.022/0.022/0.050), and the remediation’s own scaling arithmetic is confirmed: H1’s mean  $z = 4.32$ , H2’s 4.84, H3’s 5.79, against the 4.04 needed for 0.95 at Holm’s worst-case  $\alpha/3$ . The binding pre-outcome demotion thresholds fixed by this run (demotion-only: where the re-run implies a looser threshold than the provisional value, the stricter provisional value stands):  $N_{H1}^* = 118,000$  matched sets (tightened from the provisional 100,000);  $N_{H2a}^* = 95,000$  true sets (tightened from 60,000), plus at least 15,000 spurious sets for gate F2’s registered operating characteristics;  $N_{H3}^* = 90,000$  sets (the provisional value retained as the stricter bound over the re-run’s 66,000). The 135,000-set plan covers all three, but H1’s margin is only about 13%: realized N below roughly 87% of the lower-bound expectation demotes H1 first.

## Sensitivity and honest reading

Two registered stress scenarios discipline the verdict. At half the planned sets (67,500 — the neighborhood of the provisional gates), full-family Holm powers fall to  $H1 = 0.840$ ,  $H2 = 0.904$ ,  $H3 = 0.980$ : two members fall below 0.95, directly validating that the pre-outcome realized-N demotion gate is load-bearing, not decorative. At halved floors, powers are 0.512/0.596/0.800: the design has essentially no confirmatory resolution below its floors, which is why every null must be reported as CI-ruled-out magnitudes against realized MDEs.

The falsification gate’s verified operating characteristics are registered at their verified values, not at aspirational targets. The trigger probability when  $\theta_{2b}$  equals  $\theta_{2a}$  at floor is 0.790 at 45,000 spurious sets — below the aspirational 0.95; scaling implies roughly 75,000–85,000 spurious sets would be needed for 0.95 trigger sensitivity, within the upper half of the 30–60% spurious-arm band but above the lower-bound sizing (at the 15,000-set viability floor the trigger falls to about 0.5, and F2 then reports bounds only, per the registered fallback). The pass probability under the hypothesis-true DGP (spurious residual 0.5%) is 0.886, marginally below an aspirational 0.90 and within one Monte Carlo standard error; under a strictly inert placebo it is 0.978. The gate remains valid as a severity device — it can only subtract support, and its false-fire risk is low — but its sensitivity to a floor-sized non-inert placebo is materially below 0.95, which this protocol states rather than hides. The program’s driver reconciled the resulting *powered-with-caveats* verdict deterministically, and both actions are carried in this protocol: (1) F2’s operating characteristics are restated from aspirational targets to the verified values, with the Stage-2 blind re-simulation at realized spurious N registered as the binding re-verification; (2) the binding  $N_{H1}^*$  is recorded as 118,000 sets, superseding the provisional 100,000. The frozen bank hypothesis wording is untouched. Finally, the exploratory reference powers confirm the tiering decisions:  $H1\_trend$  at 0.772 (gated, joint) and  $H2\_diff$  at 0.460 would both fail as confirmatory members, exactly as the round-1 diagnosis found (0.380 and 0.358); their exploratory tiering is validated, not arbitrary.

## Null bounding

No null will be “accepted.” Every Holm-nonsignificant confirmatory result is converted into the effect magnitudes its 95% CI rules out — for example, “the data are inconsistent with a pre-boundary excess larger than  $X\%$  of baseline” — quoted against the realized MDE from the Stage-2 blind re-simulation, with a caution paragraph. At the registered power, a full-family null bounds anticipatory and post-closure volume responses below 2% of baseline and the composition tilt below 5 points: magnitudes below anything the cited literature reports.

## Verification

The deposited script’s re-run was verified by the harness (exit 0, seed 20260831; every registered number above matched exactly; calibration in band: zero share 0.535, out-of-hours share 8.0% of daytime volume, event ICC 0.014), and independently reproduced by the program’s driver on 2026-08-31: of 45 stdout lines, 3 differed — all wall-clock stamps — and every numeric result was identical. Both Stage-0 runs (the voided round and the remediated re-run), the fixed-point demotion log, and the remediation record are published with the protocol.

## 7 Outcome-neutral quality checks

Eleven checks stand between every stage and the next — ten orthogonal to the confirmatory hypotheses, plus the week-zero step diagnostic (check 10), whose threshold is scaled by the realized Q3–Q4(c) excess and is therefore scheduled in Stage 4 with the weekly descriptive series, after the confirmatory contrasts are computed. None examines any registered contrast in the hypothesis direction.

1. *Coverage benchmark*: annual GP contacts per 1,000 population in the extract (SPEC2 70–80), 2008–2024, against the published Statistics Denmark health-insurance series. Threshold: within  $\pm 10\%$  every year; any year outside triggers documented reconciliation before estimation.
2. *Linkage/observability floor*: share of cohort workers with at least one SSSY row in  $c - 3$  to  $c + 1$ . Threshold:  $\geq 80\%$  (the published annual GP contact prevalence benchmark is roughly 85%); below floor halts at Stage 3.
3. *Closure-rate plausibility*: annual true-closure event rate among private size-5+ workplaces, and the spurious/true ratio. Threshold: rate in  $[1\%, 8\%]$  every cohort year, consistent with the published business-demography basis used for sizing; ratio in  $[0.1, 1.5]$ ; violations trigger a classification audit before any outcome data access.
4. *Classification cross-validation*: agreement between worker-flow classes and IDAS-native flags. Threshold:  $\geq 75\%$  agreement (the kill floor for H2’s comparative machinery); full cross-tabulation released.
5. *Successor-share bimodality (Stage 1)*: the distribution of  $s(w)$  across terminal events must be visibly bimodal with the ambiguous mid-zone ( $0.30 \leq s(w) < 0.70$ ) containing under 35% of terminal events, per the worker-flow classification literature; violation triggers the classification audit and, if unresolved, the spurious-arm kill.
6. *Matching balance and match rate*. Threshold: standardized mean difference  $\leq 0.10$  on every matching covariate and on unmatched auxiliaries after matching (control aggregates weighted  $1/k$ ); match rate  $\geq 90\%$  of eligible exposed ( $k \geq 1$ ) after the coarsening ladder; realized mean  $k$  reported.
7. *Pre-baseline placebo (design validity, orthogonal to hypotheses)*: exposed-minus-control contact difference across the four quarters of year  $c - 2$ . Threshold (the single registered condition, stated identically in contingent rule 7 of Section 5 and the H1 placebo-kill criterion of Section 7): the check *fails* if and only if the placebo contrast is significant at  $\alpha = 0.05$  and  $|\text{estimate}| > 1\%$  of the control baseline mean; failure triggers the single pre-specified re-match, then kill of the registered event-study tier. The two non-complementary cells pass with mandatory disclosure and no discretion: significant with  $|\text{estimate}| \leq 1\%$  passes, with the significance flagged in the Table T7 placebo row; non-significant with  $|\text{estimate}| > 1\%$  passes, with a mandatory caution paragraph quoting the magnitude against the  $\theta_1$  floor and the realized placebo MDE.
8. *HONUGE settlement-lag calibration* on 2021–2024 GP rows via BEHANDLINGSDATO. Threshold:  $\geq 75\%$  of rows billed within 4 weeks of the treatment week; a median lag above 4 weeks triggers the guard-band contingency.
9. *No floor or ceiling in outcomes*: distribution of control-group person-quarter GP contacts. Threshold: zero-share in  $[30\%, 85\%]$ ; 99th percentile  $\leq 30$  contacts per quarter; the control weekly series shows the expected summer trough (ISO weeks 28–31 mean at least 10% below the annual weekly mean — a HONUGE parsing sanity check).
10. *Week-zero boundary-step diagnostic* on the weekly descriptive series: the discontinuity in the exposed-minus-control series at the first HONUGE week of year  $c + 1$ , relative to the total pre-boundary excess. Threshold: step  $\leq 50\%$  of the total  $Q3(c) + Q4(c)$  excess; if exceeded, quarter-level estimates are foregrounded as the only interpretable resolution and a caution paragraph converts the step size into an explicit bound on any timing-shape claim.

11. *Published-range out-of-hours share benchmark* (licenses H3’s remediated power basis): the realized control-group out-of-hours share of baseline-year GP contacts against the published Danish 8–12% range. Threshold: in [5%, 15%], proceed as registered; in [2%, 5%), H3’s realized-power re-verification governs; below 2%, H3’s kill criterion; above 15%, a code-list re-audit before estimation.

## 8 Threats, falsification, and kill criteria

The threat-by-analysis pairing and the falsification architecture are registered in Section 3 and in the machine-readable registration. The kill criteria are verbatim commitments:

- **H1.** *Project-viability kill*: stop the project if Stage 1 yields fewer than 500 true-closure events or Stage 2 yields fewer than 10,000 matched sets. *Confirmatory demotion gate (pre-outcome)*: if realized matched sets fall below  $N_{H1}^*$  (binding value 118,000 from the Stage-0 re-run) or the Stage-2 blind re-simulation shows power below 0.95 at the 2% floor,  $\theta_1$  demotes to exploratory before any post-baseline outcome is accessed. *Placebo kill*: if the  $c - 2$  placebo fails after the single pre-specified re-match — failure defined by the single registered condition, identical in Sections 5 and 6: significant at  $\alpha = 0.05$  and  $|\text{estimate}| > 1\%$  of the control baseline mean — the entire registered event-study tier is killed and the study reports descriptive trajectories only.
- **H2.** *Spurious-arm kill (comparative machinery)*: if fewer than 150 spurious-closure events or fewer than 3,000 spurious matched sets survive Stage 2, or worker-flow versus IDAS-flag agreement is below 75%, or the bimodality check fails unresolved — gate F2,  $\theta_{2b}$ , and H2\_diff are dropped (disclosed before outcome access) and the H2 verdict is capped at “partially consistent — comparative quantifier untestable.” *Confirmatory demotion gate for  $\theta_{2a}$* : realized true sets below  $N_{H2a}^*$  (binding value 95,000) or blind re-simulated power below 0.95 at the 2% floor demotes  $\theta_{2a}$  pre-outcome. Gate F2 additionally requires at least 15,000 spurious sets for its registered operating characteristics; below that, F2 reports bounds only and the verdict cap applies.
- **H3.** *Kill*: if out-of-hours contacts are below 2% of all control-group GP contacts in the baseline year (composition not estimable), or the out-of-hours classifier is unresolvable (YDLTID missing or “not relevant” on more than 20% of peri-event GP rows and SPEC2 out-of-hours codes absent). *Confirmatory demotion gate*: realized control out-of-hours share below 5% with blind re-simulated power below 0.95 at the 5-point floor, or realized matched sets below  $N_{H3}^*$  (90,000), demotes  $\theta_3$  pre-outcome.

A killed hypothesis is reported as killed, never silently dropped; a demoted member is reported under the exploratory heading with its demotion trail; nulls are bounded per Section 7; and any deviation from this protocol is disclosed in running prose with its reason and approver.

## 9 Limitations

The design rests on explicit assumptions and has explicit failure modes. All remain inside the associational claim ceiling, and none is defended beyond what the registers can support.

*What we assume.* SSSY (2005–2025) and SYSI (1990–2005) are assumed to capture the universe of publicly reimbursed primary-care billing for insured residents. An absence of rows therefore means no billed care, not missingness. We assume that KONTAKT is a row-level contact marker whose sum yields contact counts, with a frozen fallback definition behind the validation in the analysis environment.

We take the meanings of ENDNEDL, IDFREM/IDTILB, REGELFR, and FRELFR from Statistics Denmark’s IDA workplace-continuity documentation. That source is register documentation rather than DOI-indexed literature. Because the primary classification uses the worker-flow measure  $s(w)$  computed from IDAN, misreading an IDAS flag would weaken only the cross-check.

November-status attachment observes workplace membership annually, leaving intra-year changes unobservable ex ante. The design addresses this with the frozen  $c - 1$  anchor. As Section 3 notes, however, early year- $c$  separations then contribute realized post-job-loss billing to the pre-boundary window under intention-to-follow. The E4 separation shares and the registered  $\theta_1$  attachment split bound this channel but cannot remove it. We assume that DST-imputed age and sex are accurate enough for stratified matching and otherwise use explicit unknown strata.

We assume that HONUGE settlement lags are predominantly within four weeks, but calibrate rather than presume that condition and register a guard-band contingency. We also assume that lag structure does not differ between exposed and matched controls in the same calendar quarters. Death and emigration are unobservable within scope. Their post-event incidence is assumed small; their deflating direction is conservative for the pre-boundary claim and attenuating for the post-year claim, and the later-presence trigger bounds the post-boundary side. Nominal-fee comparisons require no deflator because they occur within cohort  $\times$  calendar-quarter cells. We assume that GP billing categories remain institutionally stable at the 2-digit level across 2007–2024.

Published evidence underpins the sizing assumptions, and each has a registered failure path. Closure rates must lie in the business-demography-consistent band; otherwise, the realized-N gate demotes rather than stretches. The out-of-hours share is assumed to lie in the published 8–12% band; otherwise the failure path runs through the published-range check and the realized-share re-simulation, with demotion or kill only if the re-simulated power falls short. The control pool is assumed to support 1:3 matching at a 90% match rate; the coarsening ladder, realized- $k$  accounting, and gate govern a shortfall. We assume that  $\theta_{2a}$ ’s four-quarter averaging gain survives serial correlation, but build the registered discount into the simulation DGP rather than relying on that assumption. Finally, we retain the DGP’s overdispersion split unchanged from the failed first round. We deliberately did not revise it in the direction that would raise power, so the remediation cannot be accused of tuning.

*How this can fail, and what happens then.* Realized event or worker counts may land far below the schema-derived expectations — this is a documentation-only corpus, and no realized Ns exist anywhere in this protocol. The design now bears that risk openly: the realized-N gate demotes underpowered members loudly and pre-outcome instead of letting an underpowered confirmatory test run, and a sparse-closure Denmark kills or demotes cleanly rather than quietly.

KONTAKT or ENDNEDL semantics may differ from assumption; both carry frozen fallback rules, but level shifts in outcomes or event dating would force reported recalibration and consume credibility. The settlement lag may be materially longer than four weeks; the quarter-level estimands survive, but the weekly-resolution headline weakens to quarterly resolution in practice, and the week-zero diagnostic makes that explicit and pre-committed. The absorption-share semantics of ANDELOPS may differ from the idea’s gloss; classification routes through IDAN worker flows, so only the cross-check weakens. Spurious-closure events may be too heterogeneous to serve as a clean placebo; a non-inert placebo now shows up as gate F2 firing or an H2\_—diff reversal, which caps H2’s verdict by pre-commitment instead of silently contaminating a differential test.

The realized out-of-hours share may fall below the published range; the benchmark check catches this before estimation, and H3 demotes or dies pre-outcome instead of running underpowered. Realized N may land below the full-expectation lower bound; the gate demotes affected members, and the study survives with a smaller confirmatory family — or none, in which case it reports as registered exploratory estimation with the demotion trail published, the first round’s

outcome now a contingency rather than the design.

Unobservable death and emigration deflate post-boundary billing; the direction is stated, the sensitivity is pre-specified, the trigger forces bounds, and no in-scope fix exists. Sector misclassification via modal codes at multi-establishment firms attenuates contrasts; its share is bounded in the balance tables; and if ARB\_SEKTORKODE itself proves absent in-environment, the frozen ARB\_HOVED\_BRA\_DB07 fallback of Section 5 substitutes, with counts reported. The deterministic primary-job rule may misattribute exposure for multi-job workers, biasing toward null; the share is reported. Matching without replacement may exhaust thin strata; realized  $k < 3$  is accounted honestly in the blind re-simulation rather than assumed away.

The SYSI-era replication may prove infeasible; it is bounded as skippable exploratory work that cannot contaminate the confirmatory tier. And readers may over-read gate F2 as proof of causal job-loss effects; the associational ceiling, the consistency-vocabulary cap, and the pre-committed F2-fired interpretation guard the prose.

Two scope limits are intrinsic and bear repeating. Closure timing is annual, so anticipation windows are inferred, never announcement-dated, and the design measures pre-boundary excess only. And the four registers contain no health states: a rise in billing is a rise in billed contact — demand, access, and distress are interpretations to be worded as “consistent with,” never as findings.

## 10 Planned exhibits

The exhibits below reserve the complete map of Stage-2 results; their numbering follows the committed execution plan exactly, and Stage 2 will add no ad hoc exhibit. Disclosure rules bind every released object: no statistic based on fewer than 5 persons or fewer than 3 workplaces; event-count tables rounded to the nearest 10; weekly trajectory figures pooling at least 100 events and 5,000 person-weeks per plotted point; fee aggregates passing a 50% single-workplace dominance check; suppression applied before Holm reporting.

### Figures

- **Figure F1 — Quarterly event-study panel.** Treated  $\times$  event-quarter coefficients with 95% CIs across  $Q1(c-2)$ – $Q4(c+1)$ , with placebo quarters, the omitted-baseline marker, and terminal-year and post-year quarters; panel (a) the true arm, panel (b) the spurious arm from the pooled model. Fed by the registered  $\theta_1/\theta_{2a}$  event-study estimation.
- **Figure F2 — Weekly descriptive trajectory.** The pooled exposed-minus-matched-control weekly GP contact series around the closure-year boundary, with the week-zero step diagnostic annotated; if the step threshold is exceeded, the figure carries the registered caution and quarter-level estimates are foregrounded. Fed by the weekly descriptive aggregation.
- **Figure F3 — Successor-share distribution.** Histogram of worker-flow successor shares  $s(w)$  across terminal events with the 0.30/0.70 thresholds and the mid-zone mass — the bimodality-check exhibit. Fed by the Stage-1 classification pipeline.
- **Figure F4 — Out-of-hours versus daytime relative excess ( $\theta_3$ ).** The two relative-excess series over the peri-event window with the bootstrap CI of their gap; primary and YDLTID-augmented classifiers side by side. Fed by the registered H3 analysis.
- **Figure F5 — Matching balance (love plot).** Standardized mean differences before and after matching for all matching covariates and unmatched auxiliaries, with the 0.10 line marked. Fed by the Stage-2 matching output.

- **Figure F6 — SYSI-era replication (exploratory; skippable).** The quarterly SYSI-era analogue of Figure F1(a) for cohorts 1992–2003, labeled exploratory, on the YDLANT-weighted proxy outcome. Fed by the E1 module.

## Tables

- **Table T1 — Event extraction and classification by cohort year.** Terminal events; true, spurious, and ambiguous counts (rounded to 10); the annual true-closure rate among private size-5+ workplaces; the spurious/true ratio. Fed by Stage-1 event extraction.
- **Table T2 — Classification cross-validation.** Worker-flow classes versus IDAS-native flags: full cross-tabulation and agreement rate, plus classification counts under the 20/80 and 40/60 thresholds. Fed by Stage-1 classification.
- **Table T3 — Cohort flow accounting.** Inclusion and exclusion counts per rule, eligible exposed per cohort, match rate, the coarsening-ladder rung used, and the realized  $k$  distribution. Fed by Stage-2 cohort construction and matching.
- **Table T4 — Matching balance.** Per-covariate standardized mean differences (weighted  $1/k$ ), caliper compliance, realized mean  $k$ . Fed by Stage-2 matching.
- **Table T5 — Power and demotion trail.** Stage-0 power results from *both* rounds — the voided first run and the remediated re-run — and the Stage-2 blind realized-power/MDE table with binding  $N^*$ , realized sets, and any demotions with the fixed-point recomputation. Fed by the deposited simulation and the Stage-2 gate; published regardless of outcome.
- **Table T6 — Outcome-neutral check panel.** All eleven registered checks with thresholds and pass/fail status. Fed by the Stage-1 through Stage-3 check battery.
- **Table T7 — Confirmatory results.**  $\theta_1$ ,  $\theta_{2a}$ ,  $\theta_3$ : point estimates (contacts per person-quarter and % of baseline), 95% CIs, two-sided  $p$ , Holm-adjusted verdicts; the gate F2 row ( $\theta_{2b}$ , its one-sided 95% upper bound, trigger evaluation);  $\theta_3$  under both out-of-hours classifiers side by side (primary SPEC2 list and YDLTID-augmented, per Section 5); the gated BRUHON fee secondary (if computed); the  $c - 2$  placebo row (with the non-complementary-cell flags of Section 6); realized MDEs beside every null with the CI-ruled-out magnitudes. Fed by the registered Stage-4 estimation.
- **Table T8 — Sensitivity suite.** Successor-share thresholds; the control forward-window exclusion (with truncation note); un-winsorized outcomes; Poisson QML; the possible-death/emigration exclusion — one table, run once. Fed by the registered sensitivity suite.
- **Table T9 — Exploratory tier.** H1\_trend (or a row stating the gate was not met), H2\_diff with the quarter-by-quarter differential profile, E3 fee and subgroup analyses (within-family Holm), and the E1 summary; every row labeled exploratory. Fed by the Stage-5 analyses.
- **Table T10 — Attrition and later-presence accounting.** The E4 year- $c$  separation shares by event class (post-treatment-conditioned label and selection caveat), the post-treatment-conditioned exploratory  $\theta_1$  split by November-year- $c$  attachment, and the later-presence rates with the 2-percentage-point trigger status. Fed by the E4 module and the bounding machinery.

## 11 Disclosure control and ethics

All analysis occurs inside Statistics Denmark’s Forskningservice; only aggregate outputs leave. The rules applied to every released object: (1) no statistic based on fewer than 5 persons or fewer

than 3 workplaces; (2) event-count tables rounded to the nearest 10; (3) weekly trajectory figures pool at least 100 events and 5,000 person-weeks per plotted point; (4) fee (BRUHON) statistics released only as cell means or coefficients over minimum-cell aggregates, with a dominance check that no single workplace contributes more than 50% of any released fee aggregate; (5) released objects are limited to model coefficients with SEs and CIs, balance tables, calibration and benchmark tables, figure aggregates, rounded event counts, the Stage-0 (both rounds) and Stage-2 power/MDE tables with the demotion trail, and the analysis and power scripts — never microdata, workplace identifiers (ARBNR/CVRNR/LBNR), or person-level values; (6) minimum-cell suppression is applied *before* Holm reporting, so suppression can never be outcome-selective; (7) the  $c - 2$  placebo, the gate F2 outputs, the step and bimodality diagnostics, the later-presence rates, the E4 attrition panel, and all outcome-neutral check outputs are released under the same rules, ensuring the full audit trail exits the environment.

This study is a secondary analysis of pseudonymized administrative registers under Denmark’s research-statistics scheme. No participant is contacted, and no output informs an individual-level decision. Processing relies on the research and statistics provisions of the GDPR and the Danish Data Protection Act, implemented through the Forskningservice authorization for this collaboration. Forskningservice’s gatekeeping certifies that no team member had prior access to the microdata and enforces output disclosure control — nothing else. The staged data-access discipline (outcome registers untouched until Stage 4; the blind Stage-2 gate on pre-boundary years only) is self-enforced by the team, not gatekeeper-enforced, and is made auditable by logged, deposited extract scripts whose execution order and data scopes are recorded in the run manifest.

## 12 Timeline

The staged plan is binding, and its stop rules — not the calendar — govern. We anticipate completion approximately 24 months after Forskningservice project approval: Stage 0, months 0–3; Stage 1, months 3–8; Stage 2, months 8–12; Stage 3, months 12–15; Stage 4, months 15–18; Stage 5, months 18–21; the Stage-2 report and disclosure clearance, months 21–24. These durations are estimates, nothing more. The project application is prepared, and the gatekeeper certifies no prior data access by any team member.

### Stage 0 — Remediated registration freeze and re-run of the deposited power simulation (months 0–3; no data access).

Freeze code lists, thresholds, seeds (20260831), and the full analysis script; re-run the deposited Stage-0 simulation under the remediated parameters (1:3 matched sets; the sampling plan at the full derived expectation’s lower bound on the published closure-rate basis; the out-of-hours base rate at the published 8%; the confirmatory family at unchanged floors; the F2 gate operating characteristics; the registered serial-correlation discount). The output fixes the binding  $N^*$  demotion thresholds and the realized-MDE machinery. Publish both the first-round (failed) and remediated power runs with the protocol. *Stop rule:* stop — do not register this candidate; the program substitutes the runner-up — if any confirmatory member fails 0.95 at its floor at the registered sampling plan and the registered fixed-point demotion would empty the family again. Floors may not be raised and  $N$  may not be asserted beyond the published-basis expectation to pass this gate.

### Stage 1 — Event extraction and classification (months 3–8; IDAS and IDAN only; no outcome register touched).

Build the workplace event panel for cohorts 2009–2023, compute worker-flow successor shares, classify true/spurious/ambiguous, validate ARB\_SEKTORKODE presence and 2008+ coverage in the IDAN extract (frozen fallback per Section 5), and run the closure-rate plausibility check, the classification cross-validation, and the bimodality check. *Stop rule:* stop the project if true-closure events number fewer than 500. Drop the spurious-arm machinery

(gate F2,  $\theta_{2b}$ , H2\_diff — disclosed before any outcome access, H2 verdict capped) if spurious events number fewer than 150, classification agreement falls below 75%, or bimodality fails unresolved. Report projected exposed N against the sampling plan; a projection below the smallest  $N^*$  already forecasts (but does not yet execute) demotion.

**Stage 2 — Cohort construction, 1:3 matching, and the pre-outcome realized-N gate (months 8–12; IDAN covariates and SSSY pre-boundary years  $c - 3..c - 1$  only (no year- $c$  or later rows)).**

Extract frozen November  $c - 1$  cohorts, build control pools, execute 1:3 matching with the coarsening ladder, and produce balance tables and the realized  $k$  distribution; then run the *blind* realized-N gate: re-run the deposited simulation on realized set counts, realized  $k$ , baseline-year moments, and the realized control out-of-hours share — no post-baseline outcome touched. Publish the realized-power/MDE table. *Stop rule*: stop if the match rate is below 90% after the full ladder or any matching-covariate imbalance above 0.10 persists; matched sets below 10,000 stop the project; spurious sets below 3,000 drop the spurious-arm machinery. Gate (demotion-only, binding): any confirmatory member whose realized sets fall below its  $N^*$  or whose blind re-simulated power falls below 0.95 at its floor is demoted to exploratory *now*, before any outcome access; the Holm family shrinks by the registered fixed-point iteration; no promotion of anything, ever.

**Stage 3 — Outcome-neutral validation (months 12–15; no registered estimand computed).**

The coverage benchmark against published series, the linkage floor, floor/ceiling checks, the HONUGE lag calibration via BEHANDLINGSDATO, seasonality sanity, the published-range out-of-hours benchmark, and the  $c - 2$  placebo panel. *Stop rule*: stop if the coverage benchmark misses  $\pm 10\%$  without documented reconciliation, the 80% linkage floor fails, or the  $c - 2$  placebo fails after the single pre-specified re-match (which kills the registered event-study tier per the kill criteria, downgrading to descriptive trajectories only).

**Stage 4 — Confirmatory estimation (months 15–18; frozen code, single run).**

Estimate the surviving Holm family as gated at Stage 2, with Holm correction; evaluate gate F2 and its pre-committed trigger; compute the gated BRUHON fee secondary if  $\theta_{2a}$  rejects and the gated exploratory H1\_trend if  $\theta_1$  rejects; run the registered sensitivity suite and the later-presence trigger; produce the pre-committed boldface Summary verdict paragraph per hypothesis; convert every null into CI-ruled-out magnitudes against realized MDEs. *Stop rule*: no discretionary stop — all surviving registered tests are run and reported exactly once regardless of results (the Stage-2 Registered Report principle: decisions never based on outcomes); demoted members are reported under the exploratory heading with their demotion trail.

**Stage 5 — Exploratory tier (months 18–21; segregated by heading).**

The H2\_diff differential and its quarter-by-quarter persistence profile; the E1 SYSI-era replication (1992–2003, legacy keys, no age/sex matching, with the inventory-only variables disclosed); weekly-resolution trajectory figures under the step diagnostic; E3 fees and heterogeneity (Holm within family as a severity aid); the E4 attrition accounting panel. *Stop rule*: skip E1 entirely (reported as skipped, not failed) if SYSI-era extraction yields fewer than 300 true closures or the era’s coverage benchmark fails; no exploratory result can revise any confirmatory verdict or Stage-4 language.

## References

- Beale, N., Nethercott, S. (1985). Job-loss and family morbidity: a study of a factory closure. *Journal of the Royal College of General Practitioners* 35:510–514.

- Benedetto, G., Haltiwanger, J., Lane, J., McKinney, K. (2007). Using worker flows to measure firm dynamics. *Journal of Business & Economic Statistics* 25(3):299–313.
- Browning, M., Heinesen, E. (2012). Effect of job loss due to plant closure on mortality and hospitalization. *Journal of Health Economics* 31(4):599–616.
- Browning, M., Møller Danø, A., Heinesen, E. (2006). Job displacement and stress-related health outcomes. *Health Economics* 15(10):1061–1075.
- Christensen, M.B., Olesen, F. (1998). Out of hours service in Denmark: evaluation five years after reform. *BMJ* 316(7143):1502–1505.
- Eliason, M., Storrie, D. (2009). Does job loss shorten life? *Journal of Human Resources* 44(2):277–302.
- Ferrie, J.E., Shipley, M.J., Marmot, M.G., Stansfeld, S., Davey Smith, G. (1995). Health effects of anticipation of job change and non-employment: longitudinal data from the Whitehall II study. *BMJ* 311(7015):1264–1269.
- Hamad, R., Modrek, S., Cullen, M.R. (2015). The effects of job insecurity on health care utilization: findings from a panel of U.S. workers. *Health Services Research*, published online 29 September 2015 (print 2016;51(3):1052–1073).
- Hethey-Maier, T., Schmieder, J.F. (2013). Does the use of worker flows improve the analysis of establishment turnover? Evidence from German administrative data. *Schmollers Jahrbuch* 133(4):477–510.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6(2):65–70.
- Iversen, L., Sabroe, S., Damsgaard, M.T. (1989). Hospital admissions before and after shipyard closure. *BMJ* 299(6707):1073–1076.
- Kasl, S.V., Cobb, S. (1970). Blood pressure changes in men undergoing job loss: a preliminary report. *Psychosomatic Medicine* 32(1):19–38.
- Kuhn, A., Lalive, R., Zweimüller, J. (2009). The public health costs of job loss. *Journal of Health Economics* 28(6):1099–1115.
- Moth, G., Flarup, L., Christensen, M.B., Olesen, F., Vedsted, P. (2012). *Kontakt- og sygdomsmønsteret i lægevagten: LV-KOS 2011*. Aarhus: Research Unit for General Practice, Aarhus University.
- Rugulies, R., Thielen, K., Nygaard, E., Diderichsen, F. (2009). Job insecurity and the use of antidepressant medication among Danish employees with and without a history of prolonged unemployment. *Journal of Epidemiology and Community Health*.
- Schaller, J., Stevens, A.H. (2015). Short-run effects of job loss on health conditions, health insurance, and health care utilization. *Journal of Health Economics* 43:190–203.
- Schwerdt, G. (2011). Labor turnover before plant closure: “Leaving the sinking ship” vs. “Captain throwing ballast overboard.” *Labour Economics* 18(1):93–101.
- Sullivan, D., von Wachter, T. (2009). Job displacement and mortality: an analysis using administrative data. *Quarterly Journal of Economics* 124(3):1265–1306.
- Wunder, C., Zeydanli, T. (2021). The early costs of plant closures: evidence on lead effects on workers’ subjective and objective outcomes. *Journal of Economic Behavior & Organization*.