
title: "First emergency accidental injury and subsequent employment and wage income" **subtitle:** "Outcome-blind preregistration report" **date:** "2026-09-04T09:30:00-05:00"

Registration boundary

- Research ID: `restricted42_acute_injury`
- Mode: `schema_first`
- Version: 1
- Knowledge cutoff: 2026-09-04T09:04:07-05:00
- Outcome data accessed: **no**

Research scope

Question. Among labor-attached people aged 25-54, how does a first hospital-ascertained emergency accidental injury, versus remaining free of a qualifying injury on the same index date, change unconditional employment and unconditional annual wage income in the first three complete post-index calendar years?

Primary estimand. Two separately registered ATT objects: the treated-person average change in the probability of positive annual wage income and the treated-person average change in $\ln$ annual wage income, each at event years +1,+2,+3 relative to -1 and versus same-date not-yet-treated risk-set comparators.

Population and setting. People aged 25-54 at a qualifying LPR contact, directly measured by `V_ALDER`, with valid person linkage, five complete prior LPR years, and `IND` wage-income records in both complete pre-index years; comparators satisfy the same criteria on the treated person's date and are not yet qualifying-injured. Unit: Person within treated-person risk-set stack and complete calendar year.. Setting: Danish register data within the exclusive readable `_42_table_schema` boundary.. Period: Maximal set of complete index calendar years within 2005-2018 that supports five prior LPR years, complete pre-index years -2 and -1, and complete post-index years +1 through +3; the index calendar year is excluded from all outcome contrasts..

Inferential ceiling. Conditional causal ATT only if every exposure, support, linkage, observability, no-anticipation, and pretrend gate passes; otherwise association, description, or feasibility-only reporting as explicitly routed. Diagnostics cannot establish exchangeability, latent frailty and physical job demands remain open, and death and emigration are not identifiable.

Theory and contribution

Focal mechanism. Acute injury can reduce functional capacity and impose treatment and recovery time, causing workforce exit or reduced earning capacity; employment is therefore analyzed unconditionally and never used to select the wage-income sample.

Rival mechanisms.

- Frailty, risk behavior, and deteriorating health can jointly cause injury, care seeking, and labor decline.

- Physical job demands and occupational sorting can jointly determine accident exposure and counterfactual earnings.
- Planned contacts or coding intensity can create apparent event timing and ascertainment differences.
- Calendar and seasonal shocks can change injury hazards and labor outcomes together.
- Differential loss of register observability can mimic labor-market effects.

Contribution. An outcome-blind, explicitly gated comparison of a broadly defined first emergency accidental injury with same-date risk sets, separating unconditional employment from unconditional wage income and limiting causal language when timing diagnostics fail.

Question-to-output map

Question

Can the restricted schema support the frozen first-emergency-injury risk set before any post-index outcome value
 How does first emergency accidental injury change unconditional positive-wage employment in complete post-ind
 How does first emergency accidental injury change unconditional annual wage income in complete post-index year

Execution governance

Authority. This YAML and its bound upstream digests govern execution; no observed result may alter construction, estimators, thresholds, or branches.

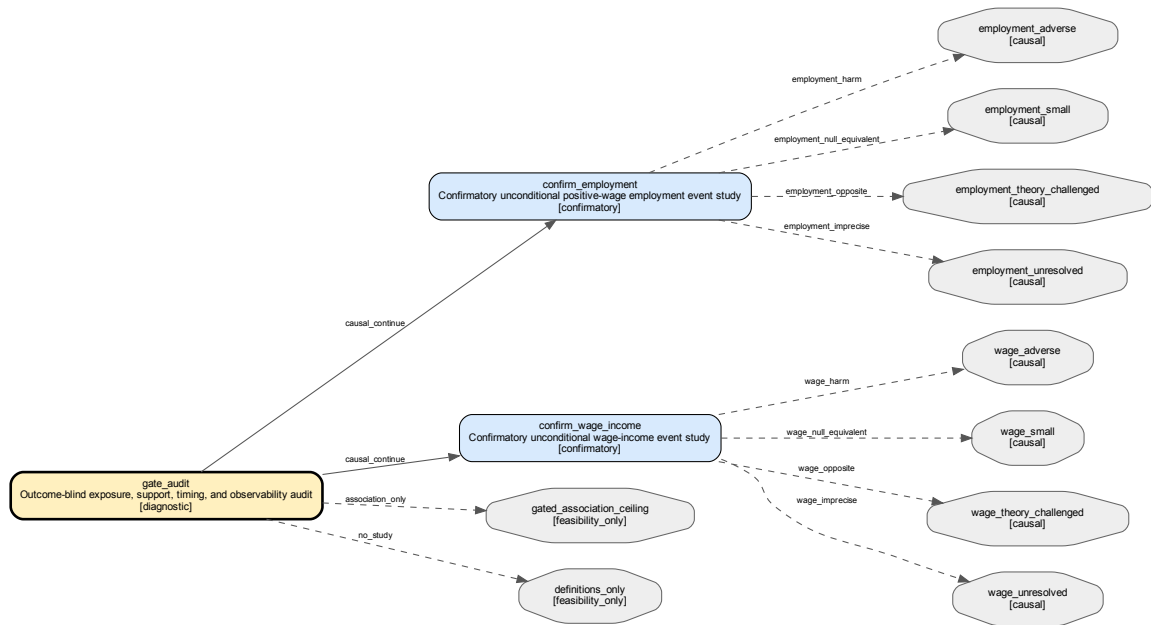
Execution order. gate_audit -> confirm_employment -> confirm_wage_income

Environment. Python 3.12; pandas==2.2.3, numpy==2.1.3, scipy==1.14.1, statsmodels==0.14.4, pyarrow==18.1.0. Capture: Write python --version and pip freeze with SHA-256 hashes to preregistration/run/environment.txt. Seed: Set NumPy PCG64 seed 420904 for comparator tie-breaking and 421904 for the synthetic dry run; no other randomness is permitted. Time zone: Europe/Copenhagen. Locale: C.UTF-8.

Data-access sequence. Resolve and hash official dictionaries; read only keys, treatment fields, dates, age, and IND row-presence plus pre-index LOENMV_13; run all gates; only after gate_decision equals causal_continue may read post-index LOENMV_13 values for confirmatory fits.

Failure policy. Failed assertions stop the named experiment and follow only its registered decision or failure route; no substitute field, horizon, code list, estimator, or sample is allowed.

Experiment graph



gate_audit: Outcome-blind exposure, support, timing, and observability audit

Epistemic role: **diagnostic**. Determine whether the confirmatory design is executable and its maximum claim ceiling without reading post-index outcome values.

Activation. Run when Always first; skip when Never.

Hypothesis. All registered feasibility and conditional-identification gates pass.

Design and estimand. descriptive; The vector of pre-outcome feasibility indicators for codebook validity, complete years, event counts, emergency completeness, washout, linkage, risk-set support, pretrends, and follow-up record presence.

Identifying variation. No effect is estimated; this audit uses treatment-defining fields and pre-index outcomes only.

Counterfactual. Same-date eligible not-yet-treated people define support and pretrend benchmarks, not an outcome effect in this experiment.

Analysis. Deterministic counts, proportions, standardized differences, and OLS lead regressions.: Two separate weighted lead regressions, one for positive wage and one for asinh wage, at event years -2 with -1 reference after risk-set weighting.

Pseudocode execution plan

Closed identity. Can the restricted schema support the frozen first-emergency-injury risk set before any post-index outcome value is inspected? -> The vector of pre-outcome feasibility indicators for codebook validity, complete years, event counts, emergency completeness,

washout, linkage, risk-set support, pretrends, and follow-up record presence. -> `gate_vector` -> `gate_vector_and_claim_ceiling` -> `gate_decision`.

Sources

ID	Resource and tables	Fields
<code>gate_lpr_adm</code>	<code>readable_42_table_schema: LPR_ADM</code>	PNR, RECNUM, CPRTJEK, CPRTYPE, C_A
<code>gate_lpr_diag</code>	<code>readable_42_table_schema: LPR_DIAG</code>	RECNUM, C_DIAG, C_DIAGTYPE
<code>gate_lpr_accident</code>	<code>readable_42_table_schema: LPR_ULYK</code>	RECNUM, C_ULYK, C_ART, C_TILULYK
<code>gate_ind</code>	<code>readable_42_table_schema: IND</code>	PNR, LOENMV_13, CPRTJEK, CPRTYPE

Variable construction `gate_vector` (statistic, One row per registered gate.).

Sources: PNR, RECNUM, C_ADIAG, C_DIAG, C_INDM, C_ULYK, D_INDDTO, V_ALDER, LOENMV_13. Time: Pre-index leads use -2 versus -1; follow-up observability reads only row presence for +1 through +3. Missingness: Missing code labels fail codebook resolution; missing keys enter linkage failures; missing post-year rows enter observability denominators.

1. Resolve official version-specific labels, archive URL, retrieval time, content hash, raw label, normalized label, and inclusion flag.
2. Calculate complete supported years, qualifying counts, emergency evidence, washout, linkage, risk-set counts, pre-index leads, date heaping, and post-year row presence in that order.
3. Apply every numeric threshold in `decision_regions` without viewing a post-index LOENMV_13 value.

`first_injury` (treatment, One row per qualifying person.).

Sources: PNR, RECNUM, C_ADIAG, C_DIAG, C_INDM, C_ULYK, D_INDDTO, V_ALDER. Time: Candidate must have full LPR coverage for each day in the 1,826-day lookback. Missingness: Missing diagnosis, accident, emergency, date, age, or PNR evidence cannot qualify.

1. Normalize diagnosis strings to uppercase without punctuation and flag S00-T79 on the same contact using C_ADIAG or summarized C_DIAG.
2. Require summarized valid C_ULYK and official acute/emergency C_INDM on that contact, age 25-54, and valid PNR.
3. Sort by PNR, D_INDDTO, RECNUM and retain the first event preceded by 1,826 clean days.

`riskset_weight` (weight, Person within treated-event stack.).

Sources: PNR, D_INDDTO, V_ALDER, LOENMV_13. Time: All matching attributes precede the treated event date; future injury status is never used. Missingness: Missing matching fields exclude that candidate from that stack.

1. For each treated date enumerate eligible PNRs not previously qualifying-injured and apply identical age, coverage, and pre-record rules.
2. Exact-match on index date, five-year V_ALDER band, positive wage indicators at -2 and -1, and deciles of asinh wage at -1 computed among eligible comparators without post-index values.

3. Retain up to 20 comparator PNRs ordered by SHA256(PNR token || treated PNR token || seed 420904); assign treated weight 1 and each selected comparator weight 1 divided by selected count.

Equation or algorithm

For outcome j in {positive_wage, asinh_wage}, fit $Y_{ist} = \alpha_s + \beta_j I(\text{event_year}=-2)*T_i$

Target statistic: `gate_vector_and_claim_ceiling`. Estimator/API: `statsmodels.formula.api.wls` with `statsmodels.cov_type=cluster` and `groups=PNR`

Weights: Treated weight 1 and comparator weights summing to 1 per treated-event stack. Fixed effects: Risk-set-stack and pre-index calendar-year indicators. References: Event year -1; comparator arm; earliest calendar year lexicographically.

Uncertainty: Two-sided cluster-robust Wald tests and 95% confidence intervals. Clustering: PNR across all reused comparator and treated observations.

Ordered actions **a01 — Freeze authoritative code mappings and a synthetic test fixture.**

Consumes: `gate_lpr_adm`, `gate_lpr_diag`, `gate_lpr_accident`. Produces: `gate_codebook_manifest`, `gate_synthetic`.

1. Retrieve the official DST variable pages linked in the schema snapshot and their referenced
2. Archive <https://www.dst.dk/da/Statistik/dokumentation/Times/sygehusbenyttelse/indm> and requ
3. Generate 50 synthetic contacts covering boundary codes, missingness, duplicate RECNUMs, was

Assertions: Every included code has one official label and version interval; synthetic expected classifications equal computed classifications; no real post-index outcome field was read

Failure route: `mark_infeasible` — Set `codebook_abandon` and prohibit all downstream outcome analysis.

a02 — Construct and reconcile qualifying first injury events.

Consumes: `gate_codebook_manifest`, `gate_synthetic`, `gate_lpr_adm`, `gate_lpr_diag`, `gate_lpr_accident`. Produces: `gate_event_audit`.

1. Reduce diagnoses and accidents separately to RECNUM flags, join to deduplicated LPR_ADM, ap
2. Retain first qualifying events after 1,826 clean covered days and calculate codebook, cover

Assertions: One first event per PNR; joined administrative row count is unchanged; qualifying overall and yearly counts reconcile to source flags

Failure route: `mark_infeasible` — Record `linkage_or_exposure_abandon` and prohibit downstream outcome analysis.

a03 — Construct risk sets and outcome-blind follow-up observability.

Consumes: `gate_event_audit`, `gate_ind`. Produces: `gate_risksets`.

1. Build identical same-date eligibility for each treated event, create frozen exact-match str
2. Read IND values only for years -2 and -1; for +1,+2,+3 read only whether the PNR-year row e

Assertions: No future treatment status enters selection; comparator weights sum to 1 per stack; `post_index_value_read_count` is zero

Failure route: `mark_infeasible` — Set `riskset_or_observability_abandon` and prohibit confirmatory fits.

a04 — Estimate frozen pretrend and anticipation diagnostics.

Consumes: `gate_risksets`. Produces: `gate_pretrends`.

1. Fit the two registered -2 lead models with -1 reference and PNR-clustered covariance; compute
2. Count event dates at first and last calendar days and compare each boundary-day share with t

Assertions: Exactly two lead fits completed; all tests use only pre-index outcome values; heaping calculations retain exact denominators

Failure route: `downgrade_claim` — Set `timing_downgrade` unless exposure or linkage assertions failed, which remains abandonment.

a05 — Apply ordered gates and emit one downstream claim ceiling.

Consumes: `gate_event_audit`, `gate_risksets`, `gate_pretrends`. Produces: `gate_decision`.

1. Apply abandon conditions first, then downgrade conditions, then causal-continuation condition
2. Write the complete gate vector, selected complete index years, sample counts, hashes, value

Assertions: Exactly one decision state; every gate has numerator denominator threshold and result; `post_index_value_read_count` equals zero

Failure route: `stop_experiment` — Stop the lane because no deterministic activation can be established.

Output contract

Output	Path	Grain
<code>gate_codebook_manifest</code>	<code>preregistration/run/gate/codebook_manifest.csv</code>	One code-version interval
<code>gate_synthetic</code>	<code>preregistration/run/gate/synthetic_results.csv</code>	One synthetic assertion
<code>gate_event_audit</code>	<code>preregistration/run/gate/event_audit.parquet</code>	One gate-year plus first-event key
<code>gate_risksets</code>	<code>preregistration/run/gate/risksets.parquet</code>	One person-stack plus arm-horizo
<code>gate_pretrends</code>	<code>preregistration/run/gate/pretrends.csv</code>	One outcome diagnostic
<code>gate_decision</code>	<code>preregistration/run/gate/decision.yaml</code>	One execution decision

Procedure accounting

- `pretrend_models`: 1 calls per unit $\times$ 2 fitted units $\times$ 1 repetitions = 2 fits. One lead regression for each of two pre-index outcome definitions.
- Full-data passes: 7. Largest intermediate: LPR_ADM contact rows left-joined to RECNUM-level diagnosis and accident flags..

Synthetic dry run. The 50-row synthetic fixture exercises every predicate boundary, missing rule, join multiplicity, tie break, event-time mapping, and all three terminal decision states before real keys are processed.

Reconciliation. Source RECNUM counts equal qualifying plus reason-specific exclusions; event-stack records reconcile to treated plus comparator counts and IND presence denominators.

Decision regions.

- **causal_continue** (supports): codebook unambiguous; at least 5 complete years; at least 500 injuries and 50 per retained year; valid C_INDM at least 95 percent and emergency share at least 90 percent; washout-valid linkage at least 95 percent; at least 10 comparators for at least 95 percent; RECNUM reconciled and PNR linkage at least 99 percent; each-arm IND presence at least 95 percent at every horizon with arm difference below 2 percentage points; both individual lead p-values at least 0.05, Holm joint p-value at least 0.10, all standardized differences below 0.10; and no date-heaping flag. — All operational gates pass; activate both confirmatory experiments with conditional causal ceiling.
- **association_only** (inconclusive): No abandon condition holds, but complete years are 3-4; valid C_INDM is 80-94.999 percent; emergency share is 70-89.999 percent; PNR linkage is 95-98.999 percent; comparator support has at least 5 for at least 80 percent; IND presence is 85-94.999 percent or arm gap is 2-4.999 points; or exactly one lead/heaping criterion fails while every absolute standardized difference is below 0.10. — Do not run registered confirmatory causal fits; terminate at outcome-blind feasibility and permit a separately registered future associational plan only.
- **no_study** (feasibility_failure): Codebook ambiguity; fewer than 3 complete years; fewer than 200 injuries; valid mode below 80 percent; emergency share below 70 percent; washout/linkage below 80/95 percent respectively; fewer than 5 comparators for more than 20 percent; any unexplained RECNUM collision; presence below 85 percent or arm gap at least 5 points; multiple lead failures; any standardized difference at least 0.10; or administrative heaping combined with any lead failure. — Stop all outcome analysis and report definitions and gate counts only.

confirm_employment: Confirmatory unconditional positive-wage employment event study

Epistemic role: **confirmatory**. Estimate the three-year employment ATT without conditioning on post-index work or care.

Activation. Run when gate_audit decision equals causal_continue; skip when gate_audit decision is association_only or no_study.

Hypothesis. Injury reduces unconditional positive-wage employment in at least one registered post-index year.

Design and estimand. did_event_study; The treated-person average percentage-point effect on $\Pr(\text{IND.LOENMV}_{13} > 0)$ at event years +1,+2,+3 relative to -1, versus same-date not-yet-treated risk-set comparators.

Identifying variation. Individual variation in each person's own first qualifying injury date within exact-date risk sets; person-event is the assignment unit.

Counterfactual. The treated person's positive-wage employment path had the person remained qualifying-injury-free on that date, represented by contemporaneous eligible not-yet-treated people.

Analysis. Weighted least squares saturated stacked event-study linear probability model.: outcome on stack fixed effects, calendar-year fixed effects, and treated interactions for -2,+1,+2,+3; -1 is

reference.

Pseudocode execution plan

Closed identity. How does first emergency accidental injury change unconditional positive-wage employment in complete post-index years +1 through +3? -> The treated-person average percentage-point effect on $\Pr(\text{IND.LOENMV}_{13} > 0)$ at event years +1,+2,+3 relative to -1, versus same-date not-yet-treated risk-set comparators. -> **positive_wage** -> **employment_att_vector_pp** -> **employment_att**.

Sources

ID	Resource and tables	Fields
confirm_lpr	readable_42_table_schema: LPR_ADM, LPR_DIAG, LPR_ULK	PNR, RECNUM, C_ADL
confirm_ind	readable_42_table_schema: IND	PNR, LOENMV_13

Variable construction **positive_wage** (outcome, Person-stack-year.).

Sources: LOENMV_13. Time: Construct only at -2,-1,+1,+2,+3; year 0 is never constructed. Missingness: Missing IND row remains missing and the row is excluded from the fit.

1. Verify exactly one numeric LOENMV_13 per observed PNR-year.
2. Set positive_wage to 1 when LOENMV_13 is greater than 0 and otherwise set it to 0.
3. Leave positive_wage missing when the IND row is absent.

event_year (event_time, Person-stack-year.).

Sources: D_INDDTO, PNR. Time: Year 0 is excluded because it mixes pre- and post-event time. Missingness: Missing index date or year is an assertion failure.

1. Extract index calendar year from D_INDDTO.
2. Subtract index calendar year from IND dataset year.
3. Retain only values -2,-1,1,2,3.

analysis_weight (weight, Person-stack.).

Sources: PNR, D_INDDTO. Time: Fixed at index and independent of post-index values. Missingness: Missing weight or stack membership stops the experiment.

1. Recover the gate-selected treated and comparator tokens for each stack.
2. Assign treated weight 1 and each comparator weight 1 divided by its stack comparator count.
3. Repeat the person-stack weight unchanged across retained years.

Equation or algorithm

$$Y_{ist} = \alpha_s + \lambda_{\text{calendar_year}} + \sum_{k \in \{-2,1,2,3\}} \beta_k T_{i,k} I(\text{event_year}=k) +$$

Target statistic: **employment_att_vector_pp**. Estimator/API: statsmodels.formula.api.wls followed by `fit(cov_type='cluster', cov_kwds={'groups': PNR_token, 'use_correction': True})`

Weights: Treated 1; each comparator 1 divided by comparator count in its stack; weights repeated by year. Fixed effects: Full categorical stack and calendar-year indicators. References: Comparator arm, event year -1, and lexicographically first calendar year.

Uncertainty: PNR-clustered sandwich covariance, two-sided 95 percent intervals, Holm p-values across three post-event coefficients. Clustering: PNR token across years and every stack in which the person appears.

Ordered actions a01 — Verify activation and reconstruct the immutable analysis roster.

Consumes: `confirm_lpr`, `confirm_ind`. Produces: `employment_roster`.

1. Read `gate_decision` and require decision exactly `causal_continue` with matching upstream and
2. Reconstruct stack membership and compare a sorted SHA-256 digest with the gate roster before

Assertions: Activation state is `causal_continue`; roster hash and retained years match `gate_decision`

Failure route: `stop_experiment` — Emit no employment estimate because activation or roster identity failed.

a02 — Build the employment person-stack-year analysis table.

Consumes: `employment_roster`, `confirm_ind`. Produces: `employment_analysis`.

1. Expand each roster record to -2,-1,+1,+2,+3 and join IND on PNR plus dataset calendar year.
2. Construct `positive_wage`, `event_year`, `treated` interactions, fixed-effect identifiers, and fr

Assertions: Unique PNR-stack-event_year rows; binary outcome; comparator weights sum to one; missing counts equal gate audit

Failure route: `stop_experiment` — Emit no employment estimate and record the failed invariant.

a03 — Fit the primary employment event-study model.

Consumes: `employment_analysis`. Produces: `employment_att`.

1. Fit the exact WLS equation once with PNR-clustered covariance and extract coefficients +1,+2,+3
2. Apply Holm correction to their two-sided p-values and store the complete coefficient covari

Assertions: Exactly one primary fit; finite coefficients and covariance; target rows exactly +1,+2,+3

Failure route: `stop_experiment` — Mark employment confirmation infeasible; no alternate estimator is allowed.

a04 — Execute frozen influence robustness and reconcile outputs.

Consumes: `employment_analysis`, `employment_att`. Produces: `employment_robustness`.

1. Repeat the same model once per retained index year while omitting that cohort year and label
2. Reconcile fitted observations
3. arm-year counts
4. coefficient names
5. and primary estimates against the compact output.

Assertions: One leave-year-out fit per retained year; primary estimates reproduce exactly; all analysis rows are accounted for

Failure route: `downgrade_claim` — Retain primary output but label robustness unavailable and inference inconclusive.

Output contract

Output	Path	Grain
employment_roster	preregistration/run/employment/roster.parquet	One person-stack
employment_analysis	preregistration/run/employment/analysis.parquet	One person-stack-year
employment_att	preregistration/run/employment/att.csv	One post-event coefficient
employment_robustness	preregistration/run/employment/leave_year_out.csv	One omitted cohort-year by ev

Procedure accounting

- **employment_primary**: 1 calls per unit $\times$ 1 fitted units $\times$ 1 repetitions = 1 fits. One outcome, formula, sample, and estimator.
- **employment_leave_year_out**: 1 calls per unit $\times$ 5 fitted units $\times$ 1 repetitions = 5 fits. Exactly five retained complete index years are required for causal activation; omit each once.
- Full-data passes: 4. Largest intermediate: Person-stack-year event-study table with five years per roster membership..

Synthetic dry run. Run all actions on the gate synthetic fixture with planted -3,-2,-1 percentage-point effects and require recovery within 0.1 point under a zero-noise fixture.

Reconciliation. Analysis row totals reconcile by arm and event year; compact estimates reproduce from saved design matrices to floating tolerance 1e-10.

Decision regions.

- **employment_harm** (supports): At least one Holm-adjusted p-value is below 0.05 and its estimate is at or below -2.0 percentage points, with no significantly positive post-year coefficient. — Supports an adverse employment effect in at least one named post-index year under the conditional causal assumptions.
- **employment_null_equivalent** (null_equivalent): Every simultaneous 95 percent confidence interval lies wholly inside [-2.0,2.0] percentage points. — Effects of magnitude two percentage points or larger are incompatible with all three estimates.
- **employment_opposite** (challenges): No employment_harm condition holds and at least one Holm-adjusted p-value is below 0.05 with estimate at or above 2.0 percentage points. — Challenges the directional theory; report without reinterpretation or model change.
- **employment_imprecise** (inconclusive): None of employment_harm, employment_null_equivalent, or employment_opposite holds. — Employment evidence is inconclusive relative to zero and the smallest effect.

confirm_wage_income: Confirmatory unconditional wage-income event study

Epistemic role: **confirmatory**. Estimate the three-year ATT on unconditional asinh wage income without selecting on employment.

Activation. Run when gate_audit decision equals causal_continue; skip when gate_audit decision is association_only or no_study.

Hypothesis. Injury reduces unconditional asinh wage income in at least one registered post-index year.

Design and estimand. `did_event_study`; The treated-person average effect on `asinh(IND.LOENMV_13 / 1000)` at event years +1,+2,+3 relative to -1, versus same-date not-yet-treated risk-set comparators.

Identifying variation. Individual variation in each person's own first qualifying injury date within exact-date risk sets; person-event is the assignment unit.

Counterfactual. The treated person's `asinh` wage-income path had the person remained qualifying-injury-free on that date, represented by contemporaneous eligible not-yet-treated people.

Analysis. Weighted least squares saturated stacked event study on `asinh` wage income.: transformed wage on stack and calendar-year fixed effects plus treated interactions for -2,+1,+2,+3; -1 reference.

Pseudocode execution plan

Closed identity. How does first emergency accidental injury change unconditional annual wage income in complete post-index years +1 through +3? -> The treated-person average effect on `asinh(IND.LOENMV_13 / 1000)` at event years +1,+2,+3 relative to -1, versus same-date not-yet-treated risk-set comparators. -> `asinh_wage` -> `wage_att_vector_asinh_thousand_dkk` -> `wage_att`.

Sources

ID	Resource and tables	Fields
<code>confirm_lpr</code>	<code>readable_42_table_schema: LPR_ADM, LPR_DIAG, LPR_ULK</code>	PNR, RECNUM, C_ADL
<code>confirm_ind</code>	<code>readable_42_table_schema: IND</code>	PNR, LOENMV_13

Variable construction `asinh_wage` (outcome, Person-stack-year.).

Sources: `LOENMV_13`. Time: Construct only at -2,-1,+1,+2,+3; year 0 is excluded. Missingness: Absent IND row remains missing; nonfinite observed value stops the experiment.

1. Verify exactly one finite numeric `LOENMV_13` per observed PNR-year.
2. Divide `LOENMV_13` by 1000.0 without rounding.
3. Apply `numpy.arcsinh` and retain observed zero and negative values; leave absent rows missing.

`event_year` (event_time, Person-stack-year.).

Sources: `D_INDDTO`, PNR. Time: Year 0 is omitted because it mixes pre- and post-event time. Missingness: Missing date or dataset year stops the experiment.

1. Extract index calendar year from `D_INDDTO`.
2. Subtract it from the IND dataset year.
3. Retain only -2,-1,1,2,3.

`analysis_weight` (weight, Person-stack.).

Sources: PNR, `D_INDDTO`. Time: Fixed at index before outcome access. Missingness: Missing weight stops the experiment.

1. Recover frozen gate stack membership.
2. Assign treated weight 1 and comparator weight reciprocal of comparator count.
3. Repeat weights unchanged across years.

winsorized_wage_thousands (outcome, Person-stack-year.).

Sources: LOENMV_13. Time: Bounds use pre-index years only and apply unchanged to all retained years. Missingness: Absent row remains missing.

1. Using only pooled observed -2 and -1 values compute linear-interpolated 1st and 99th percentiles.
2. Divide LOENMV_13 by 1000 and clip every pre and post observed value to those frozen bounds.
3. Retain zeros and negative values and label this robustness-only.

Equation or algorithm

$W_{list} = \alpha_s + \lambda_{calendar_year} + \sum_{k \in \{-2,1,2,3\}} \beta_k * T_i * I(event_year=k) +$

Target statistic: wage_att_vector_asinh_thousand_dkk. Estimator/API: statsmodels.formula.api.wls followed by fit(cov_type='cluster', cov_kwds={'groups': PNR_token, 'use_correction': True})

Weights: Treated 1 and comparator reciprocal-count weights summing to 1 within each stack. Fixed effects: Full categorical stack and calendar-year indicators. References: Comparator arm, event year -1, and lexicographically first calendar year.

Uncertainty: PNR-clustered sandwich covariance, two-sided 95 percent intervals, Holm p-values across three post coefficients. Clustering: PNR token across every reused stack and year.

Ordered actions a01 — Verify activation and reconstruct the immutable wage-analysis roster.

Consumes: confirm_lpr, confirm_ind. Produces: wage_roster.

1. Read gate_decision and require causal_continue with matching input and codebook hashes.
2. Reconstruct sorted stack membership and require its SHA-256 digest and retained five years

Assertions: Activation state is causal_continue; roster and retained-year hashes match gate_decision

Failure route: stop_experiment — Emit no wage estimate because activation or identity failed.

a02 — Build both frozen wage variables on the registered analysis rows.

Consumes: wage_roster, confirm_ind. Produces: wage_analysis.

1. Expand to registered event years, join IND, construct event year and weights, then compute a
2. Compute robustness winsorization bounds from pooled -2 and -1 wages only and archive the tw

Assertions: Unique PNR-stack-event_year; reversible asinh transform; fixed winsor bounds; missing counts equal gate audit

Failure route: stop_experiment — Emit no wage estimate and record the failed construction invariant.

a03 — Fit the primary asinh wage event-study model.

Consumes: wage_analysis. Produces: wage_att.

1. Fit the registered WLS model once and extract +1,+2,+3 coefficients in order.
2. Compute PNR-clustered intervals, Holm-adjusted p-values, and the full covariance matrix.

Assertions: Exactly one primary fit; finite coefficient vector and covariance; exactly three target rows

Failure route: `stop_experiment` — Mark wage confirmation infeasible and do not substitute an estimator.

a04 — Fit registered wage robustness models and reconcile results.

Consumes: wage_analysis, wage_att. Produces: wage_robustness.

1. Fit five leave-one-index-year-out asinh models and one full-sample winsorized-level model
2. Reconcile all counts and reproduce primary estimates from the archived design matrix to tol

Assertions: Exactly six robustness fits; all estimates labeled non-primary; primary output reproduces exactly

Failure route: `downgrade_claim` — Retain the primary result but classify the wage family as inconclusive.

Output contract

Output	Path	Grain	Requ
wage_roster	preregistration/run/wage/roster.parquet	One person-stack	stack
wage_analysis	preregistration/run/wage/analysis.parquet	One person-stack-year	stack
wage_att	preregistration/run/wage/att.csv	One post-event coefficient	event
wage_robustness	preregistration/run/wage/robustness.csv	One procedure and event-year coefficient	proce

Procedure accounting

- `wage_primary`: 1 calls per unit $\times$ 1 fitted units $\times$ 1 repetitions = 1 fits. One primary transformed outcome model.
- `wage_leave_year_out`: 1 calls per unit $\times$ 5 fitted units $\times$ 1 repetitions = 5 fits. One omission for each frozen retained index year.
- `wage_winsorized_level`: 1 calls per unit $\times$ 1 fitted units $\times$ 1 repetitions = 1 fits. One prespecified non-primary scale robustness.
- Full-data passes: 5. Largest intermediate: Person-stack-year wage analysis table containing primary and robustness variables..

Synthetic dry run. Use a zero-noise synthetic fixture with planted asinh effects -0.15,-0.10,-0.05 and require exact recovery to 1e-10 before real outcome access.

Reconciliation. Person-stack-year totals reconcile across outcomes and the employment experiment; compact estimates reproduce from the saved design matrix.

Decision regions.

- **wage_harm** (supports): At least one Holm-adjusted p-value is below 0.05 and its estimate is at or below -0.10, with no significantly positive post-year coefficient. — Supports an adverse wage-income effect in at least one named post year under conditional causal assumptions.
- **wage_null_equivalent** (null_equivalent): Every simultaneous 95 percent confidence interval lies wholly inside $[-0.10, 0.10]$ asinh-thousand-DKK units. — Shifts of magnitude 0.10 or larger are incompatible with all three estimates.
- **wage_opposite** (challenges): No wage_harm condition holds and at least one Holm-adjusted p-value is below 0.05 with estimate at or above 0.10. — Challenges the directional theory and is reported without redesign.
- **wage_imprecise** (inconclusive): None of wage_harm, wage_null_equivalent, or wage_opposite holds. — Wage-income evidence is inconclusive relative to zero and the smallest effect.

Synthesis worlds

infeasible_world

Trigger. gate_audit is association_only or no_study.

Interpretation. The restricted schema or timing diagnostics do not support the registered conditional causal comparison.

Allowed claim. Report only frozen definitions, gate statistics, and the named feasibility ceiling; do not inspect post-index outcomes.

joint_harm_world

Trigger. Both employment_harm and wage_harm occur.

Interpretation. Injury is followed by material adverse changes on both unconditional labor margins under maintained assumptions.

Allowed claim. Conditional causal evidence supports employment and wage-income harm at the specific event years satisfying each family rule.

employment_only_harm_world

Trigger. employment_harm occurs and wage_harm does not occur.

Interpretation. Evidence is consistent with an extensive-margin employment response without registered evidence of material wage harm.

Allowed claim. Claim only the employment decision; do not infer wage protection or mediation.

wage_only_harm_world

Trigger. wage_harm occurs and employment_harm does not occur.

Interpretation. Evidence is consistent with wage-income loss not classified as a material employment-probability effect.

Allowed claim. Claim only the wage decision; do not infer hours, wages, or intensive-margin mechanisms separately.

no_harm_classified_world

Trigger. Neither harm decision occurs and both families are null-equivalent.

Interpretation. Both outcome families exclude effects at least as large as their registered equivalence margins.

Allowed claim. Large effects are incompatible with the estimates; exact zero and effects below the margins are not established.

mixed_nonharm_world

Trigger. Neither harm decision occurs and at least one family is opposite or imprecise.

Interpretation. Results are mixed, surprising, or insufficiently precise for the registered harm claim.

Allowed claim. Report each family decision separately and make no joint protective, null, or mechanism claim.

Frozen machine-readable preregistration

This appendix is authoritative and contains the complete registered plan.

```
schema_version: '1.1'
artifact: preregistration
registration:
  rq_id: restricted42_acute_injury
  title: First emergency accidental injury and subsequent employment and wage income
  mode: schema_first
  version: 1
  status: under_review
  generated_at: '2026-09-04T09:30:00-05:00'
  knowledge_cutoff: '2026-09-04T09:04:07-05:00'
  outcome_blind: true
  input_digests:
    research_brief: 69e67ce2ee2fbd61778275930211e62921b1ae7f6e60dd65df5c0ae1e2084013
    literature_map: 173307d7b9bbf60d9c2c5a10d4205ee7d334c3b65fbec28b24fa8cf69c34b724
    data_feasibility: 4b3a62bc24b7d12732f34e3a8e4816fb8232fc1f4bb71d7a0536aa9166f665f6
    design_space: 53048a8cb29ed09007cc152f99775a782cbf987000b7ed140d20225c65147891
  scope:
    research_question: Among labor-attached people aged 25-54, how does a first hospital-ascerta
    accidental injury, versus remaining free of a qualifying injury on the same index date, cha
    employment and unconditional annual wage income in the first three complete post-index cal
    population: People aged 25-54 at a qualifying LPR contact, directly measured by V_ALDER, with
    person linkage, five complete prior LPR years, and IND wage-income records in both complet
    years; comparators satisfy the same criteria on the treated person's date and are not yet c
    unit: Person within treated-person risk-set stack and complete calendar year.
    setting: Danish register data within the exclusive readable_42_table_schema boundary.
    time_period: Maximal set of complete index calendar years within 2005-2018 that supports five
    LPR years, complete pre-index years -2 and -1, and complete post-index years +1 through +3
```

calendar year is excluded from all outcome contrasts.

primary_estimand: 'Two separately registered ATT objects: the treated-person average change of positive annual wage income and the treated-person average change in asinh annual wage income each at event years +1,+2,+3 relative to -1 and versus same-date not-yet-treated risk-set controls'

inferential_ceiling: Conditional causal ATT only if every exposure, support, linkage, observation, no-anticipation, and pretrend gate passes; otherwise association, description, or feasibility of reporting as explicitly routed. Diagnostics cannot establish exchangeability, latent frailty, job demands remain open, and death and emigration are not identifiable.

exclusions:

- Invalid or non-person PNR under documented CPRTJEK and CPRTYPE values.
- Age outside 25-54 on the candidate contact; PNR is never decoded for age or any attribute.
- No IND record in either full pre-index calendar year.
- Any qualifying injury in the five years before the candidate date.
- Contacts failing the frozen same-record diagnosis, accident-link, and emergency-mode prediction.
- Index years without three complete post-index IND years or two complete pre-index IND years.

theory:

focal_mechanism: Acute injury can reduce functional capacity and impose treatment and recovery costs, causing workforce exit or reduced earning capacity; employment is therefore analyzed unconstrained and never used to select the wage-income sample.

rival_mechanisms:

- Frailty, risk behavior, and deteriorating health can jointly cause injury, care seeking, and earnings decline.
- Physical job demands and occupational sorting can jointly determine accident exposure and earnings.
- Planned contacts or coding intensity can create apparent event timing and ascertainment differences.
- Calendar and seasonal shocks can change injury hazards and labor outcomes together.
- Differential loss of register observability can mimic labor-market effects.

directional_predictions:

- Injury lowers the probability of positive wage income in each of years +1 through +3.
- Injury lowers asinh annual wage income in each of years +1 through +3.
- Effects before injury should be zero under no anticipation and conditional parallel trends assumptions.

scope_conditions:

- Results apply only to hospital-ascertained emergency accidental injuries, not untreated or non-emergency injuries.
- Accident, diagnosis, and admission-mode mappings must be authoritatively documented before outcome value is accessed.
- Annual outcomes cannot resolve within-year timing; year 0 is excluded.
- The treatment-policy estimand retains subsequent injuries rather than censoring at recurrence.

literature_basis:

nearest_work:

- humlum_munch_plato_2025
- fallesen_campos_2020
- graff_et_al_2019
- fouquet_et_al_2024

established_knowledge:

- Danish work accidents and concussion have already been studied in relation to labor outcomes; this study is a narrowed extension, not a claim of an untouched field.
- Acute-admission event studies require conditional timing assumptions; an admission date is

exogenous.

- Conditioning earnings on post-injury employment induces survivor selection, motivating unconfounded outcomes.
- Occupation, prior work absence, physical demands, depression, and health history are plausible and recovery determinants.

open_questions:

- Whether the restricted schema can operationalize an emergency accidental injury consistently enough complete years remains unknown until the pre-outcome gate executes.
- Whether same-date not-yet-treated people approximate untreated trends after observed-history remains unverified.
- Whether unmeasured frailty, physical demands, death, migration, or care ascertainment explain contrast cannot be resolved with these tables.

contribution: An outcome-blind, explicitly gated comparison of a broadly defined first emergency injury with same-date risk sets, separating unconditional employment from unconditional wage and limiting causal language when timing diagnostics fail.

data_basis:

schema_snapshot_ids:

- readable_42_table_schema

population_construction: Build contact-level candidates from LPR_ADM, summarize LPR_DIAG and separately by RECNUM, apply a five-year washout, then form treated-date risk sets from people 25-54 with two prior IND records and no qualifying injury by that date.

linkage_plan: RECNUM links contact components after each one-to-many table is reduced to one per RECNUM; PNR links the resulting person-event file to IND person-years. No identifier is used. Joins are audited before outcome fields are read.

construct_map:

- **construct_id:** acute_accidental_injury
- role:** treatment
- operational_definition:** Earliest LPR_ADM contact whose same RECNUM has an ICD-10 S00-T79 diagnosis in C_ADIAG or LPR_DIAG.C_DIAG, at least one LPR_ULK row whose C_ULK is valid under the official version dictionary, and C_INDM equals string-normalized code "1" (Akut) under the DST TIMES INDM value set, after five clean years. Codes T80-T98 and all non-S00-T79 codes are excluded. Official labels, versions, URLs, retrieval times, and hashes are frozen in gate_codebook before outcome access; ambiguity, missing labels, or inability to distinguish acute/emergency from abandonment.

source_fields:

- LPR_ADM.PNR
- LPR_ADM.RECNUM
- LPR_ADM.C_ADIAG
- LPR_ADM.C_INDM
- LPR_ADM.D_INDDTO
- LPR_ADM.V_ALDER
- LPR_DIAG.RECNUM
- LPR_DIAG.C_DIAG
- LPR_ULK.RECNUM
- LPR_ULK.C_ULK

timing: Exact D_INDDTO date; the first event follows 1,826 days with no qualifying record.

alternatives: []

- **construct_id:** positive_wage_employment

```

role: outcome
operational_definition: Indicator 1 if IND.LOENMV_13 is strictly greater than 0 DKK and 0
    and less than or equal to 0; absent IND row is missing observability, not nonemployment.
source_fields:
- IND.PNR
- IND.LOENMV_13
timing: Complete calendar years -2,-1,+1,+2,+3 relative to index year; year 0 excluded.
alternatives: []
- construct_id: asinh_wage_income
role: outcome
operational_definition: asinh(IND.LOENMV_13 / 1000), retaining observed zero and negative
    unit is inverse-hyperbolic-sine thousands of nominal DKK.
source_fields:
- IND.PNR
- IND.LOENMV_13
timing: Complete calendar years -2,-1,+1,+2,+3 relative to index year; year 0 excluded.
alternatives: []
- construct_id: baseline_labor_health
role: confounder
operational_definition: Pre-index outcomes at -2 and -1, age in five-year bands from V_ALDI
    of LPR contacts and qualifying injury indicators in days -1826 through -1, and exact index
    date strata. No post-index measure enters adjustment.
source_fields:
- LPR_ADM.V_ALDER
- LPR_ADM.D_INDDTO
- LPR_ADM.PNR
- IND.LOENMV_13
timing: Strictly before each treated or pseudo-index date.
alternatives: []
- construct_id: register_observability
role: censoring
operational_definition: Indicator that exactly one IND person-year record is present; absent
    loss of observability and is never coded as zero income, death, or emigration.
source_fields:
- IND.PNR
- IND.LOENMV_13
timing: Each required pre- and post-index calendar year.
alternatives: []
- construct_id: recurrent_injury
role: competing_event
operational_definition: A later event meeting the identical frozen injury predicate more than
    after index; retained in treatment-policy follow-up and reported only as context.
source_fields:
- LPR_ADM.PNR
- LPR_ADM.D_INDDTO
- LPR_ADM.C_ADIAG
- LPR_DIAG.C_DIAG
- LPR_ULK.C_ULK

```

```

    timing: Days 31 through the LPR coverage end after index.
    alternatives: []
    unresolved_quantities:
      - Actual official value labels, version mappings, eligible event counts, risk-set overlap, 1.
        pretrend statistics, outcome variance, and post-index record presence remain unknown until
      - The gate resolves mappings only from archived official documentation and treatment-defining
        before post-index LOENMV_13 values are read; it never chooses mappings from observed effects
    no_outcome_data_accessed: true
  study_dag:
    source_design_space_sha256: 53048a8cb29ed09007cc152f99775a782cbf987000b7ed140d20225c65147891
    estimand_context: For working-age, pre-index labor-attached people eligible on a given calendar date,
      the average effect among those experiencing a first hospital-ascertained emergency accident
      of that injury at that date, versus remaining free of a qualifying injury at that date, on
      annual employment and real earnings over the retained post-index horizon. The causal estimand
      on timing exchangeability; otherwise the same contrast is an adjusted associational event-
    nodes:
      - node_id: baseline_labor
        label: Pre-index employment, earnings, and occupation history
        role: confounder
        observed: true
      - node_id: baseline_health
        label: Pre-index somatic, psychiatric, and primary-care history
        role: confounder
        observed: true
      - node_id: latent_frailty
        label: Unmeasured frailty and risk behavior
        role: confounder
        observed: false
      - node_id: physical_job
        label: Physical job demands and exposure opportunity
        role: confounder
        observed: false
      - node_id: calendar_shocks
        label: Calendar, seasonal, and macroeconomic shocks
        role: confounder
        observed: true
      - node_id: injury
        label: First emergency accidental injury
        role: treatment
        observed: true
      - node_id: functional_loss
        label: Functional limitation and recovery burden
        role: mediator
        observed: false
      - node_id: post_care
        label: Post-index somatic and mental-health care
        role: mediator
        observed: true

```

- node_id: recurrent_injury
label: Later qualifying injury
role: mediator
observed: true
- node_id: employment
label: Unconditional annual employment attachment
role: outcome
observed: true
- node_id: earnings
label: Unconditional annual real earnings
role: outcome
observed: true
- node_id: occupational_mobility
label: Observed occupational or employer transition
role: outcome
observed: true
- node_id: register_observability
label: Presence in required follow-up records
role: selection
observed: true
- node_id: care_ascertainment
label: Hospital and billed-care ascertainment
role: measurement
observed: true

edges:

- from: baseline_labor
to: injury
relation: causal
warrant:
 - basis: prior_evidence
 - statement: Occupation and prior work history can predict injury and return to work.
 - source_ids:
 - occupation_confounding
- from: baseline_labor
to: earnings
relation: causal
warrant:
 - basis: theory
 - statement: Prior earnings and employment predict counterfactual earnings.
- from: baseline_labor
to: employment
relation: causal
warrant:
 - basis: theory
 - statement: Prior labor attachment predicts continued employment.
- from: baseline_health
to: injury
relation: causal

```

    warrant:
      basis: maintained_assumption
      statement: Prior health can alter injury risk and hospital ascertainment.
- from: baseline_health
  to: employment
  relation: causal
  warrant:
    basis: prior_evidence
    statement: Health and recovery determinants predict return to work.
    source_ids:
      - occupation_confounding
- from: baseline_health
  to: earnings
  relation: causal
  warrant:
    basis: theory
    statement: Prior health can affect counterfactual earning capacity.
- from: latent_frailty
  to: injury
  relation: causal
  warrant:
    basis: maintained_assumption
    statement: Unmeasured frailty and behavior may increase injury incidence.
- from: latent_frailty
  to: employment
  relation: causal
  warrant:
    basis: maintained_assumption
    statement: Frailty can reduce employment without the index injury.
- from: physical_job
  to: injury
  relation: causal
  warrant:
    basis: prior_evidence
    statement: Physical job demands are a plausible determinant of injury exposure.
    source_ids:
      - occupation_confounding
- from: physical_job
  to: earnings
  relation: causal
  warrant:
    basis: theory
    statement: Job demands and occupation shape counterfactual earnings paths.
- from: calendar_shocks
  to: injury
  relation: causal
  warrant:
    basis: theory

```

```

    statement: Season and aggregate conditions can change accident hazards.
- from: calendar_shocks
  to: employment
  relation: causal
  warrant:
    basis: theory
    statement: Aggregate conditions change employment opportunities.
- from: calendar_shocks
  to: earnings
  relation: causal
  warrant:
    basis: theory
    statement: Aggregate conditions change wage and hours paths.
- from: injury
  to: functional_loss
  relation: causal
  warrant:
    basis: theory
    statement: Injury can reduce functional capacity.
- from: injury
  to: post_care
  relation: causal
  warrant:
    basis: theory
    statement: Injury creates treatment and recovery needs.
- from: injury
  to: recurrent_injury
  relation: causal
  warrant:
    basis: maintained_assumption
    statement: The index injury can change exposure and recurrence risk.
- from: functional_loss
  to: employment
  relation: causal
  warrant:
    basis: prior_evidence
    statement: Loss of ability can reduce employment.
    source_ids:
      - nearest_danish_overlap
- from: functional_loss
  to: earnings
  relation: causal
  warrant:
    basis: prior_evidence
    statement: Loss of ability can reduce earnings.
    source_ids:
      - nearest_danish_overlap
- from: functional_loss

```

```

to: occupational_mobility
relation: causal
warrant:
  basis: prior_evidence
  statement: Loss of ability can prompt occupational transitions.
  source_ids:
    - nearest_danish_overlap
- from: post_care
to: employment
relation: causal
warrant:
  basis: theory
  statement: Treatment and recovery demands can alter work attachment.
- from: employment
to: earnings
relation: causal
warrant:
  basis: theory
  statement: Employment is a component cause of annual labor earnings.
- from: employment
to: occupational_mobility
relation: selection
warrant:
  basis: prior_evidence
  statement: Occupation is observed only through post-index job states; conditioning can s
    employment.
  source_ids:
    - post_employment_selection
- from: latent_frailty
to: register_observability
relation: selection
warrant:
  basis: maintained_assumption
  statement: Unobserved death or migration may affect record presence.
- from: injury
to: register_observability
relation: selection
warrant:
  basis: maintained_assumption
  statement: Injury may affect unobserved mortality and subsequent record presence.
- from: care_ascertainment
to: injury
relation: measurement
warrant:
  basis: institution
  statement: The exposure exists in the data only when hospital and accident fields record
- from: care_ascertainment
to: post_care

```

```

relation: measurement
warrant:
  basis: institution
  statement: Care outcomes depend on contact, billing, and table coverage.
open_paths:
- injury <- latent_frailty -> employment/earnings remains open because frailty and risk behavior not observed.
- injury <- physical_job -> earnings remains partly open because occupation codes do not fully capture physical demands.
- injury -> register_observability <- latent_frailty creates selection bias if analysis conditional on follow-up presence.
- injury <- care_ascertainment -> measured post_care can mix true utilization with contact and care intensity.
forbidden_adjustment:
- functional_loss
- post_care
- recurrent_injury
- employment
- register_observability
adjustment_sets:
- experiment_id: confirm_employment
  covariates:
  - baseline_labor
  - baseline_health
  - calendar_shocks
  rationale: Uses only observed pre-treatment histories and exact calendar risk sets.
remaining_open_paths:
- injury <- latent_frailty -> employment
- injury <- physical_job -> employment
- experiment_id: confirm_wage_income
  covariates:
  - baseline_labor
  - baseline_health
  - calendar_shocks
  rationale: Uses only observed pre-treatment histories and exact calendar risk sets.
remaining_open_paths:
- injury <- latent_frailty -> earnings
- injury <- physical_job -> earnings
rationale: Risk-set construction addresses observed labor, health, and common-time causes but excludes frailty or physical demands. Functional loss, post-index care, recurrence, employment, and register_observability are downstream or selection variables and are forbidden from total-effect adjustment. This is a complete study-specific DAG copied from the design space without unioning prior-paper graphs. The sole added content is the explicit mapping of its permitted baseline nodes to the two split experiments.
question_map:
- question_id: q_gate
  question_clause: Can the restricted schema support the frozen first-emergency-injury risk set if any post-index outcome value is inspected?

```

```

experiment_id: gate_audit
estimand: The vector of pre-outcome feasibility indicators for codebook validity, complete y
counts, emergency completeness, washout, linkage, risk-set support, pretrends, and follow-u
presence.
output_id: gate_decision
claim_ceiling: feasibility_only
- question_id: q_employment
question_clause: How does first emergency accidental injury change unconditional positive-wa
in complete post-index years +1 through +3?
experiment_id: confirm_employment
estimand: The treated-person average percentage-point effect on  $\Pr(\text{IND.LOENMV}_{13} > 0)$  at even
+1,+2,+3 relative to -1, versus same-date not-yet-treated risk-set comparators.
output_id: employment_att
claim_ceiling: causal
- question_id: q_wage
question_clause: How does first emergency accidental injury change unconditional annual wage
in complete post-index years +1 through +3?
experiment_id: confirm_wage_income
estimand: The treated-person average effect on  $\text{asinh}(\text{IND.LOENMV}_{13} / 1000)$  at event years +1
to -1, versus same-date not-yet-treated risk-set comparators.
output_id: wage_att
claim_ceiling: causal
execution_governance:
authority: This YAML and its bound upstream digests govern execution; no observed result may
estimators, thresholds, or branches.
execution_order:
- gate_audit
- confirm_employment
- confirm_wage_income
environment:
language: Python 3.12
software:
- pandas==2.2.3
- numpy==2.1.3
- scipy==1.14.1
- statsmodels==0.14.4
- pyarrow==18.1.0
environment_capture: Write python --version and pip freeze with SHA-256 hashes to preregis
seed_policy: Set NumPy PCG64 seed 420904 for comparator tie-breaking and 421904 for the syn
dry run; no other randomness is permitted.
time_zone: Europe/Copenhagen
locale: C.UTF-8
data_access_sequence: Resolve and hash official dictionaries; read only keys, treatment fiel
age, and IND row-presence plus pre-index LOENMV_13; run all gates; only after gate_decision
causal_continue may read post-index LOENMV_13 values for confirmatory fits.
global_invariants:
- Never decode PNR or infer sex, birth date, residence, family links, death, or migration.
- Reduce each one-to-many RECNUM table before joining and reconcile source and aggregate coun

```

- Use identical injury predicates for index, washout, future-treatment status, and recurrence
- Exclude event year 0 and never code a missing IND row as zero earnings or nonemployment.
- Do not condition either primary outcome on post-index employment, care, recurrence, or observation

failure_policy: Failed assertions stop the named experiment and follow only its registered default or failure route; no substitute field, horizon, code list, estimator, or sample is allowed

deviation_log: preregistration/run/deviations.yaml

experiments:

- **experiment_id:** gate_audit
 - title:** Outcome-blind exposure, support, timing, and observability audit
 - epistemic_role:** diagnostic
 - specification_status:** conditional
 - unresolved_choices:**
 - Official C_ULYK labels and version mappings are unknown; C_INDM is frozen to code 1 (Akut) resolution algorithm either archives unambiguous accident mappings or abandons.
 - Actual complete years, counts, overlap, linkage, pretrends, and IND row presence are unknown for this outcome-blind gate runs.

purpose: Determine whether the confirmatory design is executable and its maximum claim ceiling by reading post-index outcome values.

dependencies: []

activation:

- when:** Always first
- skip_when:** Never

hypothesis:

- claim:** All registered feasibility and conditional-identification gates pass.
- 'null':** One or more abandon or downgrade conditions occur.
- alternative:** Every continuation condition occurs.
- directional:** false

design:

- family:** descriptive
- estimand:** The vector of pre-outcome feasibility indicators for codebook validity, complete event counts, emergency completeness, washout, linkage, risk-set support, pretrends, and record presence.
- identifying_variation:** No effect is estimated; this audit uses treatment-defining fields and outcomes only.
- counterfactual:** Same-date eligible not-yet-treated people define support and pretrend benchmarks, not an outcome effect in this experiment.
- assignment_mechanism:** A person's own first qualifying injury date defines individual assignment to the contemporaneous risk set.
- assumptions:**
 - Official dictionaries are authoritative for their stated versions
 - pre-index values may be inspected without violating outcome blindness
 - row presence can be read without reading post-index outcome values.
- diagnostics:**
 - Codebook hash audit
 - RECNUM reconciliation
 - exact-date heaping
 - two-lead balance
 - arm-specific IND record presence.

`downgrade_path`: Any downgrade caps downstream claims at association; any abandon condition confirmatory experiments.
`domain_fit`: Combines target-trial time-zero alignment with econometric risk-set and pretrends.
`method_specific`: {}
`operationalization`:
 `sample`: All LPR_ADM contacts in 2000-2019 plus linked diagnosis and accident rows, restricted to registered predicate and support rules.
 `treatment_or_exposure`: First qualifying emergency accidental injury after 1,826 clean days.
 `outcome`: Gate vector only; post-index LOENMV_13 values remain unread.
 `covariates`:
 - V_ALDER five-year band
 - pre-index wage employment at -2 and -1
 - pre-index asinh wage at -2 and -1
 - prior LPR contact count
 `time`: Exact contact date for risk sets; complete calendar years for pretrends and record periods.
 `exclusions`:
 - Invalid PNR
 - age outside 25-54
 - incomplete washout
 - missing pre-index IND records
 - unsupported years
`analysis`:
 `estimator`: Deterministic counts, proportions, standardized differences, and OLS lead regressions.
 `specification`: Two separate weighted lead regressions, one for positive wage and one for asinh wage, at event years -2 with -1 reference after risk-set weighting.
 `uncertainty`: HC1 covariance for gate-only lead coefficients; two-sided Wald tests.
 `clustering`: PNR for lead regressions.
 `missingness`: Missing required keys or pre-index IND rows exclude eligibility; post-index IND values are not read and only row-presence is counted.
 `multiplicity_family`: gate_leads
 `robustness`:
 - Exact date heaping table by day-of-month and month boundary
 - comparator support distribution
 - duplicate-key audit
`prospective_assessment`:
 `expected_direction`: 'null'
 `smallest_effect_of_interest`:
 `status`: fixed
 `value`: 0.1
 `unit`: absolute standardized difference
 `rationale`: Frozen substantive-smallness boundary for pretrend downgrade.
 `precision_plan`: Report exact denominators and Wilson 95% intervals for proportions; feasible to report not p-values, govern operational gates except the frozen lead tests.
 `weak_data_scenario`: Fewer than 200 injuries, fewer than three supported years, weak risk set, poor observability terminates confirmation.
 `surprising_result`: Large pre-index differences or administrative-date heaping challenges undermine timing and prevents causal continuation.
`execution`:

```

identity_chain:
  question_clause: Can the restricted schema support the frozen first-emergency-injury risk
    any post-index outcome value is inspected?
  estimand: The vector of pre-outcome feasibility indicators for codebook validity, completeness
    event counts, emergency completeness, washout, linkage, risk-set support, pretrends, and
    record presence.
  analysis_variable: gate_vector
  target_statistic: gate_vector_and_claim_ceiling
  primary_output_id: gate_decision
sources:
- source_id: gate_lpr_adm
  resource_id: readable_42_table_schema
  tables:
  - LPR_ADM
  fields:
  - PNR
  - RECNUM
  - CPRTJEK
  - CPRTYPE
  - C_ADIAG
  - C_INDM
  - C_HENM
  - D_HENDTO
  - D_INDDTO
  - V_ALDER
  grain: One administrative contact row per RECNUM version.
  keys:
  - RECNUM
  - PNR
  temporal_semantics: D_INDDTO is index candidate date; D_HENDTO is referral date; all 2000
    support washout checks subject to coverage.
  filters:
  - D_INDDTO nonmissing and valid
  - age evaluated only from V_ALDER
- source_id: gate_lpr_diag
  resource_id: readable_42_table_schema
  tables:
  - LPR_DIAG
  fields:
  - RECNUM
  - C_DIAG
  - C_DIAGTYPE
  grain: Zero or more diagnosis rows per RECNUM.
  keys:
  - RECNUM
  temporal_semantics: Diagnosis belongs to the administrative contact sharing RECNUM.
  filters:
  - No frequency-based code selection

```

- source_id: gate_lpr_accident
resource_id: readable_42_table_schema
tables:
 - LPR_ULYK
fields:
 - RECNUM
 - C_ULYK
 - C_ART
 - C_TILULYK
grain: Zero or more accident-code rows per RECNUM.
keys:
 - RECNUM
temporal_semantics: Accident code belongs to the contact sharing RECNUM and is available through March 2019.
filters:
 - No observed-frequency-based value selection
- source_id: gate_ind
resource_id: readable_42_table_schema
tables:
 - IND
fields:
 - PNR
 - LOENMV_13
 - CPRTJEK
 - CPRTYPE
grain: One person-year record after duplicate verification.
keys:
 - PNR
temporal_semantics: Dataset year is the calendar income year; values are read only for -1 while later years supply row presence only.
filters:
 - Years required by candidate index support

joins:

- join_id: gate_diag_join
left_source: gate_lpr_adm
right_source: gate_lpr_diag
keys:
 - left_field: RECNUM
 - right_field: RECNUM
cardinality: one_to_many
retained_side: left
unmatched_rule: Unmatched contact retains C_ADIAG only and has no secondary diagnosis.
validation: LPR_DIAG is reduced to one qualifying-diagnosis flag per RECNUM before joining
row count is unchanged.
- join_id: gate_accident_join
left_source: gate_lpr_adm
right_source: gate_lpr_accident
keys:

- left_field: RECNUM
 - right_field: RECNUM
- cardinality: one_to_many
- retained_side: left
- unmatched_rule: Unmatched contact has accident_evidence=0 and cannot qualify.
- validation: LPR_ULK is reduced to one valid-accident flag per RECNUM before joining and row count is unchanged.
- join_id: gate_ind_join
 - left_source: gate_lpr_adm
 - right_source: gate_ind
- keys:
 - left_field: PNR
 - right_field: PNR
- cardinality: many_to_many
- retained_side: left
- unmatched_rule: Summarize IND to PNR-year first; missing required pre-year excludes elig and missing post-year counts as absent observability.
- validation: Joined row count equals the explicit event-by-required-year grid and every P appears at most once.

variables:

- variable_id: gate_vector
 - role: statistic
 - grain: One row per registered gate.
 - source_ids:
 - gate_lpr_adm
 - gate_lpr_diag
 - gate_lpr_accident
 - gate_ind
 - source_fields:
 - PNR
 - RECNUM
 - C_ADIAG
 - C_DIAG
 - C_INDM
 - C_ULK
 - D_INDDTO
 - V_ALDER
 - LOENMV_13
 - ordered_steps:
 - Resolve official version-specific labels, archive URL, retrieval time, content hash, r normalized label, and inclusion flag.
 - Calculate complete supported years, qualifying counts, emergency evidence, washout, lin counts, pre-index leads, date heaping, and post-year row presence in that order.
 - Apply every numeric threshold in decision_regions without viewing a post-index LOENMV_
- temporal_rule: Pre-index leads use -2 versus -1; follow-up observability reads only row p for +1 through +3.
- missing_rule: Missing code labels fail codebook resolution; missing keys enter linkage f missing post-year rows enter observability denominators.

output_type: Gate table with statistic, numerator, denominator, threshold, result, and c
 validation: Exactly one terminal status is assigned and post_index_value_read_count equal

- variable_id: first_injury
 - role: treatment
 - grain: One row per qualifying person.
 - source_ids:
 - gate_lpr_adm
 - gate_lpr_diag
 - gate_lpr_accident
 - source_fields:
 - PNR
 - RECNUM
 - C_ADIAG
 - C_DIAG
 - C_INDM
 - C_ULK
 - D_INDDTO
 - V_ALDER
 - ordered_steps:
 - Normalize diagnosis strings to uppercase without punctuation and flag S00-T79 on the s
using C_ADIAG or summarized C_DIAG.
 - Require summarized valid C_ULK and official acute/emergency C_INDM on that contact, a
and valid PNR.
 - Sort by PNR, D_INDDTO, RECNUM and retain the first event preceded by 1,826 clean days.
- temporal_rule: Candidate must have full LPR coverage for each day in the 1,826-day lookba
- missing_rule: Missing diagnosis, accident, emergency, date, age, or PNR evidence cannot c
- output_type: Person-event table.
- validation: One row per PNR, unique deterministic tie break, and every event satisfies tl
predicate.
- variable_id: riskset_weight
 - role: weight
 - grain: Person within treated-event stack.
 - source_ids:
 - gate_lpr_adm
 - gate_ind
 - source_fields:
 - PNR
 - D_INDDTO
 - V_ALDER
 - LOENMV_13
 - ordered_steps:
 - For each treated date enumerate eligible PNRs not previously qualifying-injured and ap
age, coverage, and pre-record rules.
 - Exact-match on index date, five-year V_ALDER band, positive wage indicators at -2 and
deciles of asinh wage at -1 computed among eligible comparators without post-index val
 - Retain up to 20 comparator PNRs ordered by SHA256(PNR token || treated PNR token || se
assign treated weight 1 and each selected comparator weight 1 divided by selected count
- temporal_rule: All matching attributes precede the treated event date; future injury stat

```

    used.
missing_rule: Missing matching fields exclude that candidate from that stack.
output_type: Nonnegative numeric analysis weight.
validation: Weights sum to 1 among comparators within every retained stack and no comparators
    already treated.
sample_flow:
- step_id: s01
  input_population: All LPR_ADM rows in 2000-2019
  operation: Deduplicate exact RECNUM versions using highest numeric VERSION and stop on ties.
  output_population: Unique administrative contacts
  exclusion_record: Duplicate-resolution ledger by year
- step_id: s02
  input_population: Unique administrative contacts
  operation: Apply same-record diagnosis, linked accident, emergency, valid-PNR, and age predicate
  output_population: Qualifying candidate contacts
  exclusion_record: Counts by predicate and year
- step_id: s03
  input_population: Qualifying candidate contacts
  operation: Retain first contact after 1,826 fully covered clean days and require two pre-treatment
    records.
  output_population: Eligible treated events
  exclusion_record: Washout and labor-observability exclusions
- step_id: s04
  input_population: Eligible treated events
  operation: Create exact-date risk sets and exact-match strata, then retain stacks meeting selection
    rules.
  output_population: Weighted diagnostic stacks
  exclusion_record: Treated exclusions and comparator counts by stack
model:
  model_id: gate_model
  equation_or_algorithm: For outcome  $j$  in {positive_wage, asinh_wage}, fit  $Y_{ist} = \alpha_s + \beta_j +$ 
     $I(\text{event\_year}=-2) \cdot T_i + \lambda_t + e_{ist}$  on years -2 and -1 with -1 reference and risks.
    separately compute all deterministic gate proportions.
  target_statistic: gate_vector_and_claim_ceiling
  symbol_map:
  - symbol:  $Y_{ist}$ 
    meaning: Pre-index outcome  $j$  for person  $i$  in stack  $s$  and year  $t$ 
    source_fields:
    - LOENMV_13
    construction_variables:
    - gate_vector
  - symbol:  $T_i$ 
    meaning: Indicator for the treated person in a risk-set stack
    source_fields:
    - D_INDDTO
    construction_variables:
    - first_injury

```

- symbol: alpha_s
 - meaning: Risk-set-stack fixed effect
 - source_fields:
 - PNR
 - D_INDDTO
 - construction_variables:
 - riskset_weight
- estimator_api: statsmodels.formula.api.wls with statsmodels cov_type=cluster and groups=
- weights: Treated weight 1 and comparator weights summing to 1 per treated-event stack.
- fixed_effects: Risk-set-stack and pre-index calendar-year indicators.
- reference_levels: Event year -1; comparator arm; earliest calendar year lexicographically.
- uncertainty: Two-sided cluster-robust Wald tests and 95% confidence intervals.
- clustering: PNR across all reused comparator and treated observations.
- diagnostics:
 - Individual lead p-values
 - joint two-outcome Holm test
 - standardized lead differences
 - date heaping
 - leverage and weight sums
- actions:
 - action_id: a01
 - goal: Freeze authoritative code mappings and a synthetic test fixture.
 - consumes:
 - gate_lpr_adm
 - gate_lpr_diag
 - gate_lpr_accident
 - procedure:
 - Retrieve the official DST variable pages linked in the schema snapshot and their refer value dictionaries; save URL, UTC retrieval time, response hash, version interval, raw and label.
 - Archive <https://www.dst.dk/da/Statistik/dokumentation/Times/sygehusbenyttelse/indm> and string-normalized C_INDM code "1" labeled Akut; include diagnosis prefixes S00-T79 and values only when the official version dictionary labels them valid accident codes; reject undocumented, or conflicting mappings.
 - Generate 50 synthetic contacts covering boundary codes, missingness, duplicate RECNUMS boundaries, and deterministic ties using seed 421904.
 - produces:
 - gate_codebook_manifest
 - gate_synthetic
 - assertions:
 - Every included code has one official label and version interval
 - synthetic expected classifications equal computed classifications
 - no real post-index outcome field was read
 - on_failure:
 - route: mark_infeasible
 - detail: Set codebook_abandon and prohibit all downstream outcome analysis.
 - action_id: a02
 - goal: Construct and reconcile qualifying first injury events.

```

consumes:
- gate_codebook_manifest
- gate_synthetic
- gate_lpr_adm
- gate_lpr_diag
- gate_lpr_accident
procedure:
- Reduce diagnoses and accidents separately to RECNUM flags, join to deduplicated LPR_AD,
  the frozen predicate, and sort PNR, D_INDDTO, RECNUM.
- Retain first qualifying events after 1,826 clean covered days and calculate codebook,
  count, emergency, washout, and linkage statistics with denominators.
produces:
- gate_event_audit
assertions:
- One first event per PNR
- joined administrative row count is unchanged
- qualifying overall and yearly counts reconcile to source flags
on_failure:
  route: mark_infeasible
  detail: Record linkage_or_exposure_abandon and prohibit downstream outcome analysis.
- action_id: a03
goal: Construct risk sets and outcome-blind follow-up observability.
consumes:
- gate_event_audit
- gate_ind
procedure:
- Build identical same-date eligibility for each treated event, create frozen exact-match,
  select up to 20 comparators deterministically, and calculate comparators per treated p
- Read IND values only for years -2 and -1; for +1,+2,+3 read only whether the PNR-year
  then tabulate presence by arm and horizon.
produces:
- gate_risksets
assertions:
- No future treatment status enters selection
- comparator weights sum to 1 per stack
- post_index_value_read_count is zero
on_failure:
  route: mark_infeasible
  detail: Set riskset_or_observability_abandon and prohibit confirmatory fits.
- action_id: a04
goal: Estimate frozen pretrend and anticipation diagnostics.
consumes:
- gate_risksets
procedure:
- Fit the two registered -2 lead models with -1 reference and PNR-clustered covariance;
  standardized -2 arm differences using pooled pre-index standard deviations.
- Count event dates at first and last calendar days and compare each boundary-day share v
  mean share of days 2-27 in the same month; flag heaping when any ratio exceeds 2.0 and

```

```

    excess exceeds 1 percentage point.
produces:
- gate_pretrends
assertions:
- Exactly two lead fits completed
- all tests use only pre-index outcome values
- heaping calculations retain exact denominators
on_failure:
  route: downgrade_claim
  detail: Set timing_downgrade unless exposure or linkage assertions failed, which remain
- action_id: a05
goal: Apply ordered gates and emit one downstream claim ceiling.
consumes:
- gate_event_audit
- gate_risksets
- gate_pretrends
procedure:
- Apply abandon conditions first, then downgrade conditions, then causal-continuation conditions
  using the constants in decision_regions; never allow a later pass to override an earlier one
- Write the complete gate vector, selected complete index years, sample counts, hashes, version,
  audit, and exactly one of causal_continue, association_only, or no_study.
produces:
- gate_decision
assertions:
- Exactly one decision state
- every gate has numerator denominator threshold and result
- post_index_value_read_count equals zero
on_failure:
  route: stop_experiment
  detail: Stop the lane because no deterministic activation can be established.
outputs:
- output_id: gate_codebook_manifest
  path: preregistration/run/gate/codebook_manifest.csv
  format: csv
  grain: One code-version interval
  fields:
    - field
    - raw_value
    - label
    - version_start
    - version_end
    - include
    - url
    - retrieved_at
    - sha256
  produced_by: a01
  consumers:
    - a02

```

- output_id: gate_synthetic
 - path: preregistration/run/gate/synthetic_results.csv
 - format: csv
 - grain: One synthetic assertion
 - fields:
 - case_id
 - expected
 - observed
 - pass
 - produced_by: a01
 - consumers:
 - a02
- output_id: gate_event_audit
 - path: preregistration/run/gate/event_audit.parquet
 - format: parquet
 - grain: One gate-year plus first-event key audit
 - fields:
 - gate
 - year
 - numerator
 - denominator
 - value
 - threshold
 - status
 - pnr_token
 - recnum_token
 - produced_by: a02
 - consumers:
 - a03
 - a05
- output_id: gate_risksets
 - path: preregistration/run/gate/risksets.parquet
 - format: parquet
 - grain: One person-stack plus arm-horizon presence
 - fields:
 - stack_id
 - pnr_token
 - treated
 - index_date
 - event_year
 - weight
 - ind_row_present
 - comparator_count
 - produced_by: a03
 - consumers:
 - a04
 - a05
- output_id: gate_pretrends

```

path: preregistration/run/gate/pretrends.csv
format: csv
grain: One outcome diagnostic
fields:
- outcome
- lead
- estimate
- standard_error
- p_value
- standardized_difference
- heaping_ratio
- heaping_excess_pp
produced_by: a04
consumers:
- a05
- output_id: gate_decision
path: preregistration/run/gate/decision.yaml
format: yaml
grain: One execution decision
fields:
- decision
- claim_ceiling
- retained_years
- gate_results
- post_index_value_read_count
- input_hashes
produced_by: a05
consumers:
- causal_continue
- association_only
- no_study
quality_control:
synthetic_dry_run: The 50-row synthetic fixture exercises every predicate boundary, miss-
  join multiplicity, tie break, event-time mapping, and all three terminal decision state
  real keys are processed.
invariant_checks:
- One first event per PNR
- no row multiplication after summarized joins
- no future information in matching
- comparator weights sum to one per stack
- zero post-index outcome values read
reconciliation: Source RECNUM counts equal qualifying plus reason-specific exclusions; e
  records reconcile to treated plus comparator counts and IND presence denominators.
run_log: preregistration/run/gate/run.jsonl
stop_conditions:
- Ambiguous official codes
- fewer than three complete index years
- fewer than 200 injuries

```

- linkage below 95 percent
- irreconcilable keys

compute_plan:

- fit_operations:
 - operation_id: pretrend_models
 - calls_per_unit: 1
 - fitted_units: 2
 - repetitions: 1
 - total_fits: 2
 - justification: One lead regression for each of two pre-index outcome definitions.
- full_data_passes: 7
- largest_intermediate: LPR_ADM contact rows left-joined to RECNUM-level diagnosis and acco
- feasibility_note: Joins are reduced before combination; risk-set enumeration may be large
- partitionable by index date and year.

decision_regions:

- decision_id: causal_continue
- condition: codebook unambiguous; at least 5 complete years; at least 500 injuries and 50 p
- year; valid C_INDM at least 95 percent and emergency share at least 90 percent; washout-
- at least 95 percent; at least 10 comparators for at least 95 percent; RECNUM reconciled a
- linkage at least 99 percent; each-arm IND presence at least 95 percent at every horizon v
- difference below 2 percentage points; both individual lead p-values at least 0.05, Holm
- at least 0.10, all standardized differences below 0.10; and no date-heaping flag.

parameters:

- name: minimum_complete_years
 - value: 5
 - unit: years
 - rationale: Full-strength source-coverage gate.
- name: minimum_injuries
 - value: 500
 - unit: persons
 - rationale: Stable stacked design gate.
- name: minimum_per_year
 - value: 50
 - unit: persons
 - rationale: Avoid sparse cohorts.
- name: minimum_valid_mode
 - value: 95
 - unit: percent
 - rationale: Emergency evidence completeness.
- name: minimum_emergency_share
 - value: 90
 - unit: percent
 - rationale: Unexpected-event classification.
- name: minimum_ten_comparator_share
 - value: 95
 - unit: percent
 - rationale: Contemporaneous support.
- name: minimum_pnr_link

```

    value: 99
    unit: percent
    rationale: Person-linkage integrity.
- name: minimum_presence
    value: 95
    unit: percent
    rationale: Comparable follow-up observability.
- name: maximum_presence_gap
    value: 2
    unit: percentage_points
    rationale: Arm comparability.
- name: individual_lead_alpha
    value: 0.05
    unit: probability
    rationale: Frozen individual lead screen.
- name: joint_lead_alpha
    value: 0.1
    unit: probability
    rationale: Conservative joint lead screen.
- name: maximum_standardized_difference
    value: 0.1
    unit: standard_deviation
    rationale: Substantive lead balance.
- name: heaping_ratio
    value: 2.0
    unit: ratio
    rationale: Administrative-boundary anomaly rule.
- name: heaping_excess
    value: 1
    unit: percentage_point
    rationale: Requires absolute as well as relative excess.
interpretation: All operational gates pass; activate both confirmatory experiments with con
    causal ceiling.
claim_status: supports
- decision_id: association_only
condition: No abandon condition holds, but complete years are 3-4; valid C_INDM is 80-94.99
    emergency share is 70-89.999 percent; PNR linkage is 95-98.999 percent; comparator support
    least 5 for at least 80 percent; IND presence is 85-94.999 percent or arm gap is 2-4.999
    or exactly one lead/heaping criterion fails while every absolute standardized difference
    0.10.
parameters:
- name: minimum_limited_years
    value: 3
    unit: years
    rationale: Limited-window lower bound.
- name: minimum_valid_mode_limited
    value: 80
    unit: percent

```

```

    rationale: Lower operational mode bound.
- name: minimum_emergency_limited
  value: 70
  unit: percent
  rationale: Lower emergency share bound.
- name: minimum_pnr_limited
  value: 95
  unit: percent
  rationale: Lower linkage bound.
- name: minimum_five_comparator_share
  value: 80
  unit: percent
  rationale: Modified support bound.
- name: minimum_presence_limited
  value: 85
  unit: percent
  rationale: Limited observability bound.
- name: maximum_presence_gap_limited
  value: 5
  unit: percentage_points
  rationale: Beyond this the horizon is abandoned.
- name: maximum_standardized_difference_limited
  value: 0.1
  unit: standard_deviation
  rationale: Required for associational downgrade.
interpretation: Do not run registered confirmatory causal fits; terminate at outcome-blind
and permit a separately registered future associational plan only.
claim_status: inconclusive
- decision_id: no_study
condition: Codebook ambiguity; fewer than 3 complete years; fewer than 200 injuries; valid
80 percent; emergency share below 70 percent; washout/linkage below 80/95 percent respect
fewer than 5 comparators for more than 20 percent; any unexplained RECNUM collision; pre
85 percent or arm gap at least 5 points; multiple lead failures; any standardized differ
least 0.10; or administrative heaping combined with any lead failure.
parameters:
- name: absolute_minimum_years
  value: 3
  unit: years
  rationale: Below this longitudinal design is unsupported.
- name: absolute_minimum_injuries
  value: 200
  unit: persons
  rationale: Below this confirmation is abandoned.
- name: absolute_minimum_valid_mode
  value: 80
  unit: percent
  rationale: Exposure evidence floor.
- name: absolute_minimum_emergency

```

```

    value: 70
    unit: percent
    rationale: Unexpected-event floor.
- name: absolute_minimum_washout
  value: 80
  unit: percent
  rationale: First-event observability floor.
- name: absolute_minimum_linkage
  value: 95
  unit: percent
  rationale: Longitudinal linkage floor.
- name: absolute_minimum_presence
  value: 85
  unit: percent
  rationale: Follow-up observability floor.
- name: absolute_maximum_presence_gap
  value: 5
  unit: percentage_points
  rationale: Differential selection ceiling.
interpretation: Stop all outcome analysis and report definitions and gate counts only.
claim_status: feasibility_failure
- experiment_id: confirm_employment
  title: Confirmatory unconditional positive-wage employment event study
  epistemic_role: confirmatory
  specification_status: frozen
  unresolved_choices: []
  purpose: Estimate the three-year employment ATT without conditioning on post-index work or c
  dependencies:
  - gate_audit
activation:
  when: gate_audit decision equals causal_continue
  skip_when: gate_audit decision is association_only or no_study
hypothesis:
  claim: Injury reduces unconditional positive-wage employment in at least one registered pos
    year.
  'null': All three event-year ATT values equal zero.
  alternative: At least one event-year ATT is negative.
  directional: true
design:
  family: did_event_study
  estimand: The treated-person average percentage-point effect on  $\Pr(\text{IND.LOENMV}_{13} > 0)$  at e
    +1,+2,+3 relative to -1, versus same-date not-yet-treated risk-set comparators.
  identifying_variation: Individual variation in each person's own first qualifying injury da
    exact-date risk sets; person-event is the assignment unit.
  counterfactual: The treated person's positive-wage employment path had the person remained
    on that date, represented by contemporaneous eligible not-yet-treated people.
  assignment_mechanism: A person's first hospital-ascertained emergency accidental injury da
    is no lottery or collective assignment and exchangeability is maintained, not guaranteed

```

assumptions:

- Conditional timing exchangeability after frozen histories
- parallel untreated trends
- no anticipation
- no relevant interference
- comparable IND observability
- correct exposure classification
- no already-treated controls

diagnostics:

- Gate audit
- weight and support reconciliation
- event-year coefficient covariance
- influence by stack
- recurrence-context tabulation

downgrade_path: The experiment is skipped unless every causal-continuation gate passes; estimation failure yields no confirmatory employment claim.

domain_fit: A stacked risk-set event study targets a labor-economics ATT while target-trial fixes eligibility and time zero.

method_specific:

- treatment_timing:** Exact first qualifying D_INDDT0; annual event time is IND year minus initial year and year 0 is omitted.
- comparison_group:** Same-date eligible people without a qualifying prior injury, sampled before of future treatment.
- parallel_trends:** Conditional parallel trends after exact-date risk sets and frozen pre-injury
- no_anticipation:** Tested at -2 relative to -1 before activation; year 0 is excluded.
- heterogeneity_estimand:** Stack-specific contrasts are aggregated with equal weight per treatment targeting the retained treated-person ATT.
- pretrend_response:** Any registered failure prevents activation; no post-result model repair

operationalization:

- sample:** Gate-retained treated people and their up to 20 exact-stratum comparators, observed over years -2,-1,+1,+2,+3.
- treatment_or_exposure:** Frozen first emergency accidental injury from gate_codebook_manifestation event algorithm.
- outcome:** Indicator that observed IND.LOENMV_13 is strictly positive; missing IND person-year is zero.

covariates:

- Stack fixed effects
- calendar-year fixed effects
- treated-by-event-year indicators

time: Event years -2,-1,+1,+2,+3 with -1 reference; year 0 and horizons beyond +3 excluded

exclusions:

- Any gate-excluded stack
- any person-year without an IND row
- any malformed duplicate PNR-year

analysis:

- estimator:** Weighted least squares saturated stacked event-study linear probability model.
- specification:** outcome on stack fixed effects, calendar-year fixed effects, and treated indicator for -2,+1,+2,+3; -1 is reference.

uncertainty: PNR-clustered sandwich covariance; simultaneous 95 percent confidence intervals and two-sided p-values across +1,+2,+3.
clustering: PNR, accounting for comparator reuse across stacks and repeated years.
missingness: A missing IND row excludes that person-year; activation already requires presence of IND. No imputation and no zero coding.
multiplicity_family: employment_post_years
robustness:
 - Treatment-policy follow-up retains recurrent injury; report leave-one-index-year-out estimates of non-primary robustness with Holm adjustment
prospective_assessment:
expected_direction: negative
smallest_effect_of_interest:
 status: fixed
 value: -2.0
 unit: percentage_points
 rationale: A two-point employment decline is the fixed smallest policy-relevant adverse effect of interest
precision_plan: Report the covariance matrix and 95 percent intervals; if every interval spans both 0 and -2 points, classify the family as imprecise regardless of p-values.
weak_data_scenario: Wide intervals spanning both zero and the smallest effect yield inconclusive evidence of no effect.
surprising_result: A statistically positive effect challenges the functional-loss prediction of no effect measurement and selection interpretation without changing the model.

execution:
identity_chain:
 question_clause: How does first emergency accidental injury change unconditional positive wage in complete post-index years +1 through +3?
 estimand: The treated-person average percentage-point effect on $\Pr(\text{IND.LOENMV}_{13} > 0)$ at +1,+2,+3 relative to -1, versus same-date not-yet-treated risk-set comparators.
 analysis_variable: positive_wage
 target_statistic: employment_att_vector_pp
 primary_output_id: employment_att
sources: &id001
 - source_id: confirm_lpr
 resource_id: readable_42_table_schema
tables:
 - LPR_ADM
 - LPR_DIAG
 - LPR_ULYK
fields:
 - PNR
 - RECNUM
 - C_ADIAG
 - C_DIAG
 - C_INDM
 - C_ULYK
 - D_INDDTO
 - V_ALDER
grain: Contact components reduced to one frozen first event per person.

```

keys:
- PNR
- RECNUM
temporal_semantics: Exact D_INDDTO anchors index date; the gate-approved predicate and r
years are immutable inputs.
filters:
- Only gate-approved first events and risk-set eligibility
- source_id: confirm_ind
resource_id: readable_42_table_schema
tables:
- IND
fields:
- PNR
- LOENMV_13
grain: One person-year after duplicate assertion.
keys:
- PNR
temporal_semantics: Dataset year is calendar year; only -2,-1,+1,+2,+3 are used and index
is excluded.
filters:
- Gate-retained PNRs and years
joins: &id002
- join_id: confirm_person_year_join
left_source: confirm_lpr
right_source: confirm_ind
keys:
- left_field: PNR
right_field: PNR
cardinality: one_to_many
retained_side: left
unmatched_rule: Create required person-year grid first; unmatched IND years are missing a
never zero.
validation: Each joined PNR-year-stack is unique and the grid reconciles to five times th
count before missing rows.
variables:
- variable_id: positive_wage
role: outcome
grain: Person-stack-year.
source_ids:
- confirm_ind
source_fields:
- LOENMV_13
ordered_steps:
- Verify exactly one numeric LOENMV_13 per observed PNR-year.
- Set positive_wage to 1 when LOENMV_13 is greater than 0 and otherwise set it to 0.
- Leave positive_wage missing when the IND row is absent.
temporal_rule: Construct only at -2,-1,+1,+2,+3; year 0 is never constructed.
missing_rule: Missing IND row remains missing and the row is excluded from the fit.

```

```

    output_type: Binary integer 0 or 1.
    validation: Every nonmissing value is exactly 0 or 1 and source value mapping reconciles
- variable_id: event_year
  role: event_time
  grain: Person-stack-year.
  source_ids:
    - confirm_lpr
    - confirm_ind
  source_fields:
    - D_INDDTO
    - PNR
  ordered_steps:
    - Extract index calendar year from D_INDDTO.
    - Subtract index calendar year from IND dataset year.
    - Retain only values -2,-1,1,2,3.
  temporal_rule: Year 0 is excluded because it mixes pre- and post-event time.
  missing_rule: Missing index date or year is an assertion failure.
  output_type: Categorical integer.
  validation: Support equals a subset of {-2,-1,1,2,3} and every retained stack contains -1
- variable_id: analysis_weight
  role: weight
  grain: Person-stack.
  source_ids:
    - confirm_lpr
  source_fields:
    - PNR
    - D_INDDTO
  ordered_steps:
    - Recover the gate-selected treated and comparator tokens for each stack.
    - Assign treated weight 1 and each comparator weight 1 divided by its stack comparator count
    - Repeat the person-stack weight unchanged across retained years.
  temporal_rule: Fixed at index and independent of post-index values.
  missing_rule: Missing weight or stack membership stops the experiment.
  output_type: Positive floating-point value.
  validation: Comparator weights sum to 1 and treated weight equals 1 within each stack.
sample_flow: &id003
- step_id: s01
  input_population: gate-approved risk-set stacks
  operation: Reconstruct the frozen event-by-person membership and require exact key/hash collision
    with gate_decision.
  output_population: Verified person-stack roster
  exclusion_record: No discretionary exclusions; mismatch stops experiment
- step_id: s02
  input_population: Verified person-stack roster
  operation: Expand to event years -2,-1,+1,+2,+3 and left join unique IND PNR-year records
  output_population: Required person-stack-year grid
  exclusion_record: Missing IND rows by arm stack and horizon
- step_id: s03

```

```

input_population: Required person-stack-year grid
operation: Drop only rows with absent IND records and retain stacks satisfying the already
    observability gate.
output_population: Confirmatory analysis rows
exclusion_record: Exact missing person-years without inferred causes
model:
  model_id: employment_event_model
  equation_or_algorithm:  $Y_{ist} = \alpha_s + \lambda_{calendar\_year} + \sum_{k \in \{-2,1,2,3\}} \beta_k + error_{ist}$ , estimated by WLS on event years -2,-1,+1,+2,+3.
  target_statistic: employment_att_vector_pp
  symbol_map:
    - symbol:  $Y_{ist}$ 
      meaning: Positive-wage indicator for person  $i$  stack  $s$  year  $t$ 
      source_fields:
        - LOENMV_13
      construction_variables:
        - positive_wage
    - symbol:  $T_i$ 
      meaning: Treated-person indicator within stack
      source_fields:
        - PNR
        - D_INDDTO
      construction_variables:
        - analysis_weight
    - symbol:  $k$ 
      meaning: Event year relative to index calendar year
      source_fields:
        - D_INDDTO
        - PNR
      construction_variables:
        - event_year
  estimator_api: 'statsmodels.formula.api.wls followed by fit(cov_type='cluster', cov_kwds={
    PNR_token, 'use_correction': True})'
  weights: Treated 1; each comparator 1 divided by comparator count in its stack; weights 1
    by year.
  fixed_effects: Full categorical stack and calendar-year indicators.
  reference_levels: Comparator arm, event year -1, and lexicographically first calendar year.
  uncertainty: PNR-clustered sandwich covariance, two-sided 95 percent intervals, Holm p-values
    three post-event coefficients.
  clustering: PNR token across years and every stack in which the person appears.
  diagnostics:
    - Design matrix full rank
    - finite covariance
    - coefficient order +1,+2,+3
    - weight reconciliation
    - leave-one-index-year-out influence
  actions:
    - action_id: a01

```

```

goal: Verify activation and reconstruct the immutable analysis roster.
consumes:
- confirm_lpr
- confirm_ind
procedure:
- Read gate_decision and require decision exactly causal_continue with matching upstream
  hashes.
- Reconstruct stack membership and compare a sorted SHA-256 digest with the gate roster l
  post-index LOENMV_13 values.
produces:
- employment_roster
assertions:
- Activation state is causal_continue
- roster hash and retained years match gate_decision
on_failure:
  route: stop_experiment
  detail: Emit no employment estimate because activation or roster identity failed.
- action_id: a02
goal: Build the employment person-stack-year analysis table.
consumes:
- employment_roster
- confirm_ind
procedure:
- Expand each roster record to -2,-1,+1,+2,+3 and join IND on PNR plus dataset calendar y
- Construct positive_wage, event_year, treated interactions, fixed-effect identifiers, and
  weights in the registered order.
produces:
- employment_analysis
assertions:
- Unique PNR-stack-event_year rows
- binary outcome
- comparator weights sum to one
- missing counts equal gate audit
on_failure:
  route: stop_experiment
  detail: Emit no employment estimate and record the failed invariant.
- action_id: a03
goal: Fit the primary employment event-study model.
consumes:
- employment_analysis
procedure:
- Fit the exact WLS equation once with PNR-clustered covariance and extract coefficients
  in that order multiplied by 100.
- Apply Holm correction to their two-sided p-values and store the complete coefficient c
  matrix.
produces:
- employment_att
assertions:

```

- Exactly one primary fit
- finite coefficients and covariance
- target rows exactly +1,+2,+3

on_failure:

- route: stop_experiment
- detail: Mark employment confirmation infeasible; no alternate estimator is allowed.

- action_id: a04

goal: Execute frozen influence robustness and reconcile outputs.

consumes:

- employment_analysis
- employment_att

procedure:

- Repeat the same model once per retained index year while omitting that cohort year and estimate non-primary.
- Reconcile fitted observations
- arm-year counts
- coefficient names
- and primary estimates against the compact output.

produces:

- employment_robustness

assertions:

- One leave-year-out fit per retained year
- primary estimates reproduce exactly
- all analysis rows are accounted for

on_failure:

- route: downgrade_claim
- detail: Retain primary output but label robustness unavailable and inference inconclusive

outputs:

- output_id: employment_roster

path: preregistration/run/employment/roster.parquet

format: parquet

grain: One person-stack

fields:

- stack_id
- pnr_token
- treated
- index_date
- weight
- roster_hash

produced_by: a01

consumers:

- a02

- output_id: employment_analysis

path: preregistration/run/employment/analysis.parquet

format: parquet

grain: One person-stack-year

fields:

- stack_id

- pnr_token
- treated
- calendar_year
- event_year
- positive_wage
- weight
- produced_by: a02
- consumers:
 - a03
 - a04
- output_id: employment_att
 - path: preregistration/run/employment/att.csv
 - format: csv
 - grain: One post-event coefficient
 - fields:
 - event_year
 - estimate_pp
 - standard_error_pp
 - ci_low_pp
 - ci_high_pp
 - p_value
 - holm_p_value
 - n_rows
 - n_persons
 - n_stacks
 - produced_by: a03
 - consumers:
 - a04
 - employment_harm
 - employment_null_equivalent
 - employment_imprecise
 - employment_opposite
- output_id: employment_robustness
 - path: preregistration/run/employment/leave_year_out.csv
 - format: csv
 - grain: One omitted cohort-year by event-year coefficient
 - fields:
 - omitted_index_year
 - event_year
 - estimate_pp
 - standard_error_pp
 - status
 - produced_by: a04
 - consumers:
 - employment_harm
 - employment_null_equivalent
 - employment_imprecise
 - employment_opposite

```

quality_control:
  synthetic_dry_run: Run all actions on the gate synthetic fixture with planted -3,-2,-1 p
    effects and require recovery within 0.1 point under a zero-noise fixture.
  invariant_checks:
    - Roster hash equals gate hash
    - year 0 absent
    - binary outcome
    - weights reconcile
    - no employed-only filtering
  reconciliation: Analysis row totals reconcile by arm and event year; compact estimates r
    from saved design matrices to floating tolerance 1e-10.
  run_log: preregistration/run/employment/run.jsonl
  stop_conditions:
    - Activation mismatch
    - roster mismatch
    - duplicate PNR-year
    - rank deficiency
    - nonfinite covariance
compute_plan:
  fit_operations:
    - operation_id: employment_primary
      calls_per_unit: 1
      fitted_units: 1
      repetitions: 1
      total_fits: 1
      justification: One outcome, formula, sample, and estimator.
    - operation_id: employment_leave_year_out
      calls_per_unit: 1
      fitted_units: 5
      repetitions: 1
      total_fits: 5
      justification: Exactly five retained complete index years are required for causal acti
        omit each once.
  full_data_passes: 4
  largest_intermediate: Person-stack-year event-study table with five years per roster mem
  feasibility_note: The retained index window is exactly five years; if more than five qua
    the latest five complete years. This rule is fixed before outcomes and makes robustness
    exact.
decision_regions:
- decision_id: employment_harm
  condition: At least one Holm-adjusted p-value is below 0.05 and its estimate is at or below
    points, with no significantly positive post-year coefficient.
parameters:
- name: family_alpha
  value: 0.05
  unit: probability
  rationale: Familywise error threshold.
- name: employment_seoi

```

```

    value: -2.0
    unit: percentage_points
    rationale: Fixed smallest adverse effect.
  interpretation: Supports an adverse employment effect in at least one named post-index year
    the conditional causal assumptions.
  claim_status: supports
- decision_id: employment_null_equivalent
  condition: Every simultaneous 95 percent confidence interval lies wholly inside [-2.0,2.0]
    points.
  parameters:
    - name: employment_equivalence_bound
      value: 2.0
      unit: percentage_points
      rationale: Symmetric smallest-effect equivalence margin.
  interpretation: Effects of magnitude two percentage points or larger are incompatible with
    estimates.
  claim_status: null_equivalent
- decision_id: employment_opposite
  condition: No employment_harm condition holds and at least one Holm-adjusted p-value is below
    with estimate at or above 2.0 percentage points.
  parameters:
    - name: opposite_alpha
      value: 0.05
      unit: probability
      rationale: Familywise surprising-direction threshold.
    - name: opposite_bound
      value: 2.0
      unit: percentage_points
      rationale: Material positive effect boundary.
  interpretation: Challenges the directional theory; report without reinterpretation or model
  claim_status: challenges
- decision_id: employment_imprecise
  condition: None of employment_harm, employment_null_equivalent, or employment_opposite hold
  parameters:
    - name: confidence_level
      value: 95
      unit: percent
      rationale: Registered precision level.
  interpretation: Employment evidence is inconclusive relative to zero and the smallest effect
  claim_status: inconclusive
- experiment_id: confirm_wage_income
  title: Confirmatory unconditional wage-income event study
  epistemic_role: confirmatory
  specification_status: frozen
  unresolved_choices: []
  purpose: Estimate the three-year ATT on unconditional asinh wage income without selecting on
  dependencies:
    - gate_audit

```

activation:

- when: gate_audit decision equals causal_continue
- skip_when: gate_audit decision is association_only or no_study

hypothesis:

- claim: Injury reduces unconditional asinh wage income in at least one registered post-indexed event year.
- 'null': All three event-year ATT values equal zero.
- alternative: At least one event-year ATT is negative.
- directional: true

design:

- family: did_event_study
- estimand: The treated-person average effect on $\text{asinh}(\text{IND.LOENMV}_{13} / 1000)$ at event years t relative to -1 , versus same-date not-yet-treated risk-set comparators.
- identifying_variation: Individual variation in each person's own first qualifying injury date defines exact-date risk sets; person-event is the assignment unit.
- counterfactual: The treated person's asinh wage-income path had the person remained qualified on that date, represented by contemporaneous eligible not-yet-treated people.
- assignment_mechanism: A person's first hospital-ascertained emergency accidental injury date is a maintained conditional assumption, not an institutional lottery.

assumptions:

- Conditional timing exchangeability after frozen histories
- parallel untreated trends
- no anticipation
- no relevant interference
- comparable IND observability
- correct exposure classification
- no already-treated controls

diagnostics:

- Gate audit
- weight and support reconciliation
- event-year coefficient covariance
- influence by stack
- outcome transformation domain checks

downgrade_path: The experiment is skipped unless every causal-continuation gate passes; estimation assertion failure yields no confirmatory wage-income claim.

domain_fit: The estimator follows staggered-treatment labor-economics logic while preserving wage-income states including zeros and negatives.

method_specific:

- treatment_timing: Exact first qualifying D_INDDTO; annual event time is IND year minus index year and year 0 is omitted.
- comparison_group: Same-date eligible people without a qualifying prior injury, sampled with replacement information.
- parallel_trends: Conditional parallel trends after exact-date risk sets and frozen pre-indexed event year.
- no_anticipation: Tested at -2 relative to -1 before activation; year 0 is excluded.
- heterogeneity_estimand: Equal treated-person stack aggregation yields the retained treatment effect ATT at each event year.
- pretrend_response: Any registered failure prevents activation and no observed post-result is permitted.

operationalization:

sample: Gate-retained treated people and their up to 20 exact-stratum comparators, observed years -2,-1,+1,+2,+3.
treatment_or_exposure: Frozen first emergency accidental injury from the gate manifest and
outcome: asinh(IND.LOENMV_13 divided by 1000), retaining observed zero and negative wages and only absent IND rows.
covariates:

- Stack fixed effects
- calendar-year fixed effects
- treated-by-event-year indicators

time: Event years -2,-1,+1,+2,+3 with -1 reference; year 0 and horizons beyond +3 excluded
exclusions:

- Any gate-excluded stack
- any person-year without an IND row
- any malformed duplicate PNR-year

analysis:
estimator: Weighted least squares saturated stacked event study on asinh wage income.
specification: transformed wage on stack and calendar-year fixed effects plus treated interaction for -2,+1,+2,+3; -1 reference.
uncertainty: PNR-clustered sandwich covariance; Holm-adjusted two-sided p-values across +1
clustering: PNR across reused stacks and years.
missingness: Missing IND row excludes that person-year; observed zero and negative LOENMV_13 in the outcome.
multiplicity_family: wage_post_years
robustness:

- Leave-one-index-year-out estimates
- non-primary levels model in nominal thousand DKK winsorized at fixed pre-index pooled 1st and 99th percentiles

prospective_assessment:
expected_direction: negative
smallest_effect_of_interest:
status: fixed
value: -0.1
unit: asinh_thousand_dkk
rationale: A 0.10-unit adverse shift is the frozen smallest effect for decision classification
precision_plan: Report covariance and 95 percent intervals; intervals spanning both 0 and -0.10
imprecision:
weak_data_scenario: If all intervals span zero and the adverse bound, report inconclusive result
no effect.
surprising_result: A material positive effect challenges theory and is reported without change in
or the sample.
execution:
identity_chain:
question_clause: How does first emergency accidental injury change unconditional annual wage
in complete post-index years +1 through +3?
estimand: The treated-person average effect on asinh(IND.LOENMV_13 / 1000) at event years +1
relative to -1, versus same-date not-yet-treated risk-set comparators.
analysis_variable: asinh_wage
target_statistic: wage_att_vector_asinh_thousand_dkk

```

    primary_output_id: wage_att
sources: *id001
joins: *id002
variables:
- variable_id: asinh_wage
  role: outcome
  grain: Person-stack-year.
  source_ids:
    - confirm_ind
  source_fields:
    - LOENMV_13
  ordered_steps:
    - Verify exactly one finite numeric LOENMV_13 per observed PNR-year.
    - Divide LOENMV_13 by 1000.0 without rounding.
    - Apply numpy.arcsinh and retain observed zero and negative values; leave absent rows missing.
  temporal_rule: Construct only at -2,-1,+1,+2,+3; year 0 is excluded.
  missing_rule: Absent IND row remains missing; nonfinite observed value stops the experiment.
  output_type: Double-precision numeric.
  validation: Reversing with 1000*sinh(value) recovers LOENMV_13 to relative tolerance 1e-7.
- variable_id: event_year
  role: event_time
  grain: Person-stack-year.
  source_ids:
    - confirm_lpr
    - confirm_ind
  source_fields:
    - D_INDDTO
    - PNR
  ordered_steps:
    - Extract index calendar year from D_INDDTO.
    - Subtract it from the IND dataset year.
    - Retain only -2,-1,1,2,3.
  temporal_rule: Year 0 is omitted because it mixes pre- and post-event time.
  missing_rule: Missing date or dataset year stops the experiment.
  output_type: Categorical integer.
  validation: Support is a subset of {-2,-1,1,2,3} and each stack contains -1.
- variable_id: analysis_weight
  role: weight
  grain: Person-stack.
  source_ids:
    - confirm_lpr
  source_fields:
    - PNR
    - D_INDDTO
  ordered_steps:
    - Recover frozen gate stack membership.
    - Assign treated weight 1 and comparator weight reciprocal of comparator count.
    - Repeat weights unchanged across years.

```

```

temporal_rule: Fixed at index before outcome access.
missing_rule: Missing weight stops the experiment.
output_type: Positive floating-point value.
validation: Treated weight is 1 and comparator weights sum to 1 per stack.
- variable_id: winsorized_wage_thousands
  role: outcome
  grain: Person-stack-year.
  source_ids:
    - confirm_ind
  source_fields:
    - LOENMV_13
  ordered_steps:
    - Using only pooled observed -2 and -1 values compute linear-interpolated 1st and 99th p
    - Divide LOENMV_13 by 1000 and clip every pre and post observed value to those frozen bo
    - Retain zeros and negative values and label this robustness-only.
  temporal_rule: Bounds use pre-index years only and apply unchanged to all retained years
  missing_rule: Absent row remains missing.
  output_type: Double-precision thousand nominal DKK.
  validation: Every nonmissing value lies within archived pre-index bounds.
sample_flow: *id003
model:
  model_id: wage_event_model
  equation_or_algorithm:  $W_{ist} = \alpha_s + \lambda_{calendar\_year} + \sum_{k \in \{-2,1,2,3\}} \text{over } k$ 
    + errorist, fit by WLS; repeat identically with winsorized wage only as robustness.
  target_statistic: wage_att_vector_asinh_thousand_dkk
  symbol_map:
    - symbol: W_ist
      meaning: Asinh wage income for person i stack s year t
      source_fields:
        - LOENMV_13
      construction_variables:
        - asinh_wage
    - symbol: T_i
      meaning: Treated-person indicator within stack
      source_fields:
        - PNR
        - D_INDDTO
      construction_variables:
        - analysis_weight
    - symbol: k
      meaning: Event year relative to index calendar year
      source_fields:
        - D_INDDTO
        - PNR
      construction_variables:
        - event_year
  estimator_api: 'statsmodels.formula.api.wls followed by fit(cov_type='cluster', cov_kw
    PNR_token, 'use_correction': True})'

```

```

weights: Treated 1 and comparator reciprocal-count weights summing to 1 within each stack
fixed_effects: Full categorical stack and calendar-year indicators.
reference_levels: Comparator arm, event year -1, and lexicographically first calendar year
uncertainty: PNR-clustered sandwich covariance, two-sided 95 percent intervals, Holm p-values,
    three post coefficients.
clustering: PNR token across every reused stack and year.
diagnostics:
- Full-rank design
- finite covariance
- reversible asinh transform
- coefficient order +1,+2,+3
- weight reconciliation
actions:
- action_id: a01
  goal: Verify activation and reconstruct the immutable wage-analysis roster.
  consumes:
  - confirm_lpr
  - confirm_ind
  procedure:
  - Read gate_decision and require causal_continue with matching input and codebook hashes
  - Reconstruct sorted stack membership and require its SHA-256 digest and retained five years
    equal the gate record before reading post-index wages.
  produces:
  - wage_roster
  assertions:
  - Activation state is causal_continue
  - roster and retained-year hashes match gate_decision
  on_failure:
    route: stop_experiment
    detail: Emit no wage estimate because activation or identity failed.
- action_id: a02
  goal: Build both frozen wage variables on the registered analysis rows.
  consumes:
  - wage_roster
  - confirm_ind
  procedure:
  - Expand to registered event years, join IND, construct event year and weights, then compute
    exactly.
  - Compute robustness winsorization bounds from pooled -2 and -1 wages only and archive them
    bounds before transforming post-index values.
  produces:
  - wage_analysis
  assertions:
  - Unique PNR-stack-event_year
  - reversible asinh transform
  - fixed winsor bounds
  - missing counts equal gate audit
  on_failure:

```

```

    route: stop_experiment
    detail: Emit no wage estimate and record the failed construction invariant.
- action_id: a03
  goal: Fit the primary asinh wage event-study model.
  consumes:
    - wage_analysis
  procedure:
    - Fit the registered WLS model once and extract +1,+2,+3 coefficients in order.
    - Compute PNR-clustered intervals, Holm-adjusted p-values, and the full covariance matrix.
  produces:
    - wage_att
  assertions:
    - Exactly one primary fit
    - finite coefficient vector and covariance
    - exactly three target rows
  on_failure:
    route: stop_experiment
    detail: Mark wage confirmation infeasible and do not substitute an estimator.
- action_id: a04
  goal: Fit registered wage robustness models and reconcile results.
  consumes:
    - wage_analysis
    - wage_att
  procedure:
    - Fit five leave-one-index-year-out asinh models and one full-sample winsorized-level model using
      the identical design and clustering.
    - Reconcile all counts and reproduce primary estimates from the archived design matrix to within
      1e-10.
  produces:
    - wage_robustness
  assertions:
    - Exactly six robustness fits
    - all estimates labeled non-primary
    - primary output reproduces exactly
  on_failure:
    route: downgrade_claim
    detail: Retain the primary result but classify the wage family as inconclusive.
outputs:
- output_id: wage_roster
  path: preregistration/run/wage/roster.parquet
  format: parquet
  grain: One person-stack
  fields:
    - stack_id
    - pnr_token
    - treated
    - index_date
    - weight

```

```

- roster_hash
produced_by: a01
consumers:
- a02
- output_id: wage_analysis
path: preregistration/run/wage/analysis.parquet
format: parquet
grain: One person-stack-year
fields:
- stack_id
- pnr_token
- treated
- calendar_year
- event_year
- asinh_wage
- winsorized_wage_thousands
- weight
produced_by: a02
consumers:
- a03
- a04
- output_id: wage_att
path: preregistration/run/wage/att.csv
format: csv
grain: One post-event coefficient
fields:
- event_year
- estimate_asinh
- standard_error
- ci_low
- ci_high
- p_value
- holm_p_value
- n_rows
- n_persons
- n_stacks
produced_by: a03
consumers:
- a04
- wage_harm
- wage_null_equivalent
- wage_imprecise
- wage_opposite
- output_id: wage_robustness
path: preregistration/run/wage/robustness.csv
format: csv
grain: One procedure and event-year coefficient
fields:

```

```

- procedure
- omitted_index_year
- event_year
- estimate
- standard_error
- status
- lower_winsor_bound
- upper_winsor_bound
produced_by: a04
consumers:
- wage_harm
- wage_null_equivalent
- wage_imprecise
- wage_opposite
quality_control:
  synthetic_dry_run: Use a zero-noise synthetic fixture with planted asinh effects -0.15,-0.15
    and require exact recovery to 1e-10 before real outcome access.
  invariant_checks:
    - Roster hash matches gate
    - year 0 absent
    - asinh reversible
    - observed zeros retained
    - weights reconcile
reconciliation: Person-stack-year totals reconcile across outcomes and the employment exposure
  compact estimates reproduce from the saved design matrix.
run_log: preregistration/run/wage/run.jsonl
stop_conditions:
- Activation mismatch
- roster mismatch
- duplicate PNR-year
- nonfinite wage
- rank deficiency
- nonfinite covariance
compute_plan:
  fit_operations:
    - operation_id: wage_primary
      calls_per_unit: 1
      fitted_units: 1
      repetitions: 1
      total_fits: 1
      justification: One primary transformed outcome model.
    - operation_id: wage_leave_year_out
      calls_per_unit: 1
      fitted_units: 5
      repetitions: 1
      total_fits: 5
      justification: One omission for each frozen retained index year.
    - operation_id: wage_winsorized_level

```

```

    calls_per_unit: 1
    fitted_units: 1
    repetitions: 1
    total_fits: 1
    justification: One prespecified non-primary scale robustness.
    full_data_passes: 5
    largest_intermediate: Person-stack-year wage analysis table containing primary and robust
    feasibility_note: Exactly the latest five fully supported index years are retained before
        fit counts are therefore fixed.
decision_regions:
- decision_id: wage_harm
  condition: At least one Holm-adjusted p-value is below 0.05 and its estimate is at or below
    with no significantly positive post-year coefficient.
  parameters:
  - name: family_alpha
    value: 0.05
    unit: probability
    rationale: Familywise threshold.
  - name: wage_seoi
    value: -0.1
    unit: asinh_thousand_dkk
    rationale: Fixed smallest adverse shift.
  interpretation: Supports an adverse wage-income effect in at least one named post year under
    causal assumptions.
  claim_status: supports
- decision_id: wage_null_equivalent
  condition: Every simultaneous 95 percent confidence interval lies wholly inside [-0.10,0.10]
    units.
  parameters:
  - name: wage_equivalence_bound
    value: 0.1
    unit: asinh_thousand_dkk
    rationale: Symmetric smallest-effect margin.
  interpretation: Shifts of magnitude 0.10 or larger are incompatible with all three estimates.
  claim_status: null_equivalent
- decision_id: wage_opposite
  condition: No wage_harm condition holds and at least one Holm-adjusted p-value is below 0.05
    estimate at or above 0.10.
  parameters:
  - name: opposite_alpha
    value: 0.05
    unit: probability
    rationale: Familywise surprising-direction threshold.
  - name: opposite_bound
    value: 0.1
    unit: asinh_thousand_dkk
    rationale: Material positive boundary.
  interpretation: Challenges the directional theory and is reported without redesign.

```

```

    claim_status: challenges
- decision_id: wage_imprecise
  condition: None of wage_harm, wage_null_equivalent, or wage_opposite holds.
  parameters:
    - name: confidence_level
      value: 95
      unit: percent
      rationale: Registered precision level.
  interpretation: Wage-income evidence is inconclusive relative to zero and the smallest effect size
  claim_status: inconclusive
experiment_graph:
  entry_experiments:
    - gate_audit
  activations:
    - from: gate_audit
      on_decision: causal_continue
      to: confirm_employment
      meaning: Every causal gate passes and employment confirmation is activated.
    - from: gate_audit
      on_decision: causal_continue
      to: confirm_wage_income
      meaning: Every causal gate passes and wage confirmation is activated.
  terminals:
    - from: gate_audit
      on_decision: association_only
      terminal_state: gated_association_ceiling
      meaning: Outcome-blind gates do not authorize the confirmatory estimators; a new preregistration
        required for association.
      claim_ceiling: feasibility_only
    - from: gate_audit
      on_decision: no_study
      terminal_state: definitions_only
      meaning: Exposure or support is infeasible and no outcome analysis occurs.
      claim_ceiling: feasibility_only
    - from: confirm_employment
      on_decision: employment_harm
      terminal_state: employment_adverse
      meaning: Conditional adverse employment evidence is complete.
      claim_ceiling: causal
    - from: confirm_employment
      on_decision: employment_null_equivalent
      terminal_state: employment_small
      meaning: Employment effects at the registered magnitude are excluded.
      claim_ceiling: causal
    - from: confirm_employment
      on_decision: employment_opposite
      terminal_state: employment_theory_challenged
      meaning: Employment result runs materially opposite prediction.

```

```

    claim_ceiling: causal
- from: confirm_employment
  on_decision: employment_imprecise
  terminal_state: employment_unresolved
  meaning: Employment evidence does not distinguish the registered regions.
  claim_ceiling: causal
- from: confirm_wage_income
  on_decision: wage_harm
  terminal_state: wage_adverse
  meaning: Conditional adverse wage evidence is complete.
  claim_ceiling: causal
- from: confirm_wage_income
  on_decision: wage_null_equivalent
  terminal_state: wage_small
  meaning: Wage effects at the registered magnitude are excluded.
  claim_ceiling: causal
- from: confirm_wage_income
  on_decision: wage_opposite
  terminal_state: wage_theory_challenged
  meaning: Wage result runs materially opposite prediction.
  claim_ceiling: causal
- from: confirm_wage_income
  on_decision: wage_imprecise
  terminal_state: wage_unresolved
  meaning: Wage evidence does not distinguish the registered regions.
  claim_ceiling: causal
completion_rule: The study completes only when the gate reaches one terminal or, after causal
  both confirmatory nodes reach exactly one terminal each.
synthesis_worlds:
- world_id: infeasible_world
  triggering_pattern: gate_audit is association_only or no_study.
  substantive_interpretation: The restricted schema or timing diagnostics do not support the restricted
    conditional causal comparison.
  allowed_claim: Report only frozen definitions, gate statistics, and the named feasibility ceiling.
    not inspect post-index outcomes.
  next_research: Obtain authoritative coding or observability resources and preregister anew.
- world_id: joint_harm_world
  triggering_pattern: Both employment_harm and wage_harm occur.
  substantive_interpretation: Injury is followed by material adverse changes on both unconditional
    margins under maintained assumptions.
  allowed_claim: Conditional causal evidence supports employment and wage-income harm at the study
    event years satisfying each family rule.
  next_research: Study mechanisms with separately registered care, recurrence, hours, and occupation.
- world_id: employment_only_harm_world
  triggering_pattern: employment_harm occurs and wage_harm does not occur.
  substantive_interpretation: Evidence is consistent with an extensive-margin employment response.
    registered evidence of material wage harm.
  allowed_claim: Claim only the employment decision; do not infer wage protection or mediation

```

next_research: Examine transfer income and hours with a new outcome-specific plan.

- world_id: wage_only_harm_world

triggering_pattern: wage_harm occurs and employment_harm does not occur.

substantive_interpretation: Evidence is consistent with wage-income loss not classified as a employment-probability effect.

allowed_claim: Claim only the wage decision; do not infer hours, wages, or intensive-margin separately.

next_research: Distinguish hours, wage rates, and occupation using newly registered measures
- world_id: no_harm_classified_world

triggering_pattern: Neither harm decision occurs and both families are null-equivalent.

substantive_interpretation: Both outcome families exclude effects at least as large as their equivalence margins.

allowed_claim: Large effects are incompatible with the estimates; exact zero and effects below are not established.

next_research: Assess other outcomes or smaller effects only with prospectively justified pre
- world_id: mixed_nonharm_world

triggering_pattern: Neither harm decision occurs and at least one family is opposite or impr

substantive_interpretation: Results are mixed, surprising, or insufficiently precise for the harm claim.

allowed_claim: Report each family decision separately and make no joint protective, null, or claim.

next_research: Audit selection and measurement, then replicate without changing this register

synthesis_resolution:

selection: ordered_precedence

precedence:

- infeasible_world
- joint_harm_world
- employment_only_harm_world
- wage_only_harm_world
- no_harm_classified_world
- mixed_nonharm_world

unmatched_action: If software or assertion failure leaves either activated family without a c label synthesis incomplete and make no cross-family claim.

multiplicity:

families:

- Employment primary family contains event years +1,+2,+3 from one outcome and one primary m
- Wage primary family contains event years +1,+2,+3 from one outcome and one primary model.
- Gate lead family contains the two -2 coefficients for employment and wage.

correction: Holm step-down correction within each three-coefficient primary family at family 0.05; Holm across the two gate leads, with gate joint threshold 0.10. No correction across distinct estimands; joint synthesis enumerates their decisions.

reporting: Report every coefficient, unadjusted and adjusted p-value, confidence interval, c matrix, family membership, fit count, and all robustness results regardless of direction.

deviations:

permitted:

- Pure software syntax repair that leaves sources, rows, variables, formula, estimator, thres and outputs unchanged.
- Security-preserving tokenization of PNR and RECNUM with documented deterministic hashes.

- Performance partitioning by year or date when concatenation reproduces the registered sort order and hash exactly.

prohibited:

- Inspecting post-index LOENMV_13 before causal_continue.
- Changing injury, emergency, accident, washout, comparison, age, time, outcome, scale, estimator, fixed effects, clustering, smallest effects, or decision thresholds after outcome access.
- Decoding identifiers or substituting unavailable death, migration, education, family, presence, monthly labor, or general-population data.
- Coding absent IND records as zero or selecting wage income on post-index employment.
- Adding an unregistered estimator, subgroup, horizon, outcome, or favorable reporting branch.

disclosure: Every permitted implementation deviation is timestamped before rerun with reason, hashes, exact code diff, and proof of scientific equivalence; all other deviations require approval and are labeled exploratory if executed.