# Online Resource 2. Supplementary methods, tables and figure

**Journal:** Journal of Occupational Rehabilitation

**Article title:** Employment pathways in the five years after workers' compensation medical care ends: a sequence analysis of two Korean national cohorts

**Authors:** Ning Huo, Zhihan Liu

**Corresponding author:** Zhihan Liu, Pathology and Pathophysiology, School of Basic Medicine, Guizhou Medical University, Guiyang, Guizhou Province, China. E-mail: liuzhihan9222@gmail.com

**Contents.** S1 Supplementary methods (S1.1–S1.10; S1.10, post hoc analyses); S2 Supplementary tables (Tables S1–S19; Tables S16–S19, post hoc analyses); S3 Supplementary figure (Fig. S1). References cited here are listed in the main article.

## S1 Supplementary methods

### S1.1 Survey design, samples and weights

Each PSWCI cohort is a sample of workers whose compensated medical care for an approved occupational injury or disease ended in one calendar year. Workers with unknown addresses, foreign workers and residents of Jeju Island (and, in cohort 1, five workers with disability grades 1–3 who had not used rehabilitation services) were excluded from the survey population, which comprised 82,493 workers whose care ended in 2012 (cohort 1) and 75,392 workers whose care ended in 2017 (cohort 2). Cohort 1 (2,000 participants) was stratified by region (nine areas), disability grade (six categories) and use of rehabilitation services, with 1,650 participants with and 350 without a disability grade. Cohort 2 (3,294 participants) was stratified by disability grade (six categories), sex and age group, with priority allocation to grades 1–3 (88 participants) and to workers without a disability grade (650), square-root allocation across the remaining grades, and region and use of rehabilitation services as implicit stratification variables. Interviews were conducted face to face every August to October (cohort 1, 2013–2017; cohort 2, 2018–2022).

The analytic sample comprised participants who responded at all five interviews (Table S1). We used the weights supplied with the data without modification: the longitudinal weight for continuous respondents (positive only for these participants and summing to the survey population of each cohort) for prevalence, descriptive estimates and SA5, and the first-interview weight for SA6, which included all participants. Variable definitions are given in Table S2.

### S1.2 Sequence distances

Each participant's sequence comprised the states at the five annual interviews. Optimal matching (OM) used a constant substitution cost of 2 and an insertion–deletion cost of 1. For sequences of equal length these costs make the OM distance equal to $2 \times (5 - \mathrm{L})$, where L is the length of the longest common subsequence, so the distance reflects the order and duration of states rather than their exact timing. SA1 used the Hamming distance (the number of interviews at which two sequences differ). SA2 used OM with substitution costs $2 - p(j|i) - p(i|j)$, where $p(j|i)$ is the proportion of transitions from state i at one interview to state j at the next, pooled over interviews, and an insertion–deletion cost of half the largest substitution cost.

### S1.3 Clustering and choice of the number of pathways

Both cohorts were clustered together, without survey weights. Computations used the 389 distinct sequences, each weighted by the number of participants who had it; unit tests confirmed that this gives the same results as using every participant. For each number of pathways k from 2 to 10, Ward's hierarchical clustering with frequency weights (equivalent to R hclust with method "ward.D2" applied to all participants) was cut into k groups. The medoid of each group—the sequence with the smallest weighted sum of distances to the other members—was the starting point for partitioning around medoids (PAM). In each PAM iteration, the single exchange of a medoid and a non-medoid that most reduced the total weighted distance of participants to their nearest medoid was made, until no exchange reduced it. Ties were broken by the smallest index, so results are deterministic.

Quality indices were computed over all pairs of participants. The silhouette of a participant compares the mean distance to the other members of the same pathway (a) with the mean distance to the members of the nearest other pathway (b) as $(b - a)/\max(a, b)$; the average silhouette width (ASW) is its mean. The point-biserial correlation (PBC) is the correlation between pairwise distances and an indicator that the pair belongs to different pathways, and Hubert's C compares the sum of within-pathway distances with its smallest and largest possible values (0 is best). The pre-specified rule considered $k = 3$–8, kept solutions in which every

pathway included at least 3% of participants, and chose the smallest k whose ASW was within 0.02 of the highest ASW among them (Table S3, Fig. S1). Pathways were named from their medoids and state distributions. The reference pathway was defined as the pathway with the highest share of person-years in state O and the persistent non-employment pathway as the pathway whose medoid had at least four years in states U or N.

### S1.4 Stability and replication

Stability was assessed in 200 bootstrap samples of participants drawn with replacement. In each sample, Ward's method and PAM were rerun with $k = 4$ on the distinct sequences drawn, weighted by the number of draws. For each original pathway, the Jaccard coefficient with each pathway of the bootstrap solution was computed among the participants included in the sample, each counted once, and the largest value was kept (Hennig 2007). Mean values $\geq 0.75$ were pre-specified to indicate a stable pathway, 0.60–0.75 a pattern without stability and $< 0.60$ an unstable pathway (Table S4). For replication, each cohort was clustered separately with the same procedure and k, and agreement with the combined typology within the cohort was measured with the adjusted Rand index (ARI; SA10, Table S13).

### S1.5 Association models

The multinomial logistic regression was fitted by Newton–Raphson iterations with step halving, without weights, and with heteroscedasticity-consistent (HC0) sandwich standard errors. It included all factors in Table S2, cohort and use of rehabilitation services (27 parameters per equation including the intercept). The pre-specified procedure for non-convergence or quasi-complete separation (any coefficient with an absolute value $> 10$)—successively merging transport with other industries, firm sizes $< 5$ and 5–29 employees, and the tenure categories $\geq 1$ year—was not needed: the model converged, and the largest absolute coefficient was 4.04. Average marginal effects (AMEs) were obtained by setting a factor to the comparison level and then to the reference level for every participant, keeping all other variables at their observed values, and averaging the differences in predicted probabilities; standard errors used the delta method with a numerical Jacobian. Interactions of cohort with age group, disability grade and employment status at injury were added one at a time and tested jointly across equations with Wald tests (Table S7). In SA5 the model and the AMEs were weighted with the longitudinal weight; SA7 excluded participants aged $\geq 60$ years and SA8 excluded occupational diseases and commuting injuries (Table S8).

### S1.6 Employment persistence (H2)

Participants employed at the first interview (states O, R or S) were followed for an exit from employment, defined as being unemployed or out of the labour force at any of the second to fifth interviews. Modified Poisson regression (log link, robust standard errors; Zou 2004) included the type of employment at the first interview, all factors at injury, cohort, use of rehabilitation services and the three health measures at the first interview; heterogeneity by cohort was tested with a Wald test of two exposure $\times$ cohort product terms (Table S9). Within each cohort, logistic regression with the same covariates (except cohort) modelled being employed at every later interview. Standardised probabilities were obtained by setting every participant's exposure to O, R and S in turn and averaging the predicted probabilities (g-computation; Snowden et al. 2011), with percentile 95% CIs from 1,000 bootstrap resamples of participants (Fig. 2). E-values (VanderWeele and Ding 2017) were computed for the adjusted risk ratios and for the confidence limit closer to 1.

### S1.7 Additional and sensitivity analyses

Among participants not employed at the first interview, weighted proportions employed at any later interview and at both the fourth and fifth interviews were estimated with bootstrap CIs (Table S10). Weighted transition matrices between consecutive interviews are shown in Table S11, and exits from employment between the second and third interviews are compared between cohorts in Table S12.

Typology sensitivity analyses (Table S13) repeated the clustering with the Hamming distance (SA1) and transition-rate costs (SA2), compared the solutions with one pathway fewer and one more (SA3; Table S15), fitted latent class models with 3–8 classes, 50 random starts each and selection by the Bayesian information criterion, assigning participants to the class with the highest posterior probability (SA4), and repeated the procedure and the selection rule for three-state (employed, unemployed, out of the labour force) and two-state (employed, not employed) sequences (SA9).

SA6 included all 5,294 participants (Table S14). (a) Missed interviews were treated as a sixth state with a substitution cost of 2 to every other state, and the clustering and selection rule were repeated. (b) Missed states were multiply imputed by chained multinomial logistic regression. Missing states were first filled by carrying the previous state forward; then, in each of 10 iterations, the state at each interview from the

second to the fifth was modelled on all factors at injury, cohort, use of rehabilitation services and the states at the previous and (except for the fifth interview) next interviews, and missing values were drawn from the predicted probabilities after drawing the coefficients from a multivariate normal distribution centred on their estimates with the robust covariance matrix. 20 imputed data sets were created. Each imputed sequence was assigned to the nearest medoid of the main typology (OM distance), prevalence was weighted with the first-interview weight, and estimates were combined with Rubin's rules, which add the variance between imputations to the average variance within imputations; the within-imputation variance of a weighted proportion was estimated by linearisation without design strata. The state at the first interview was never missing.

### S1.8 Pre-specification record and amendment

The analysis plan was kept in a local git repository and was not registered with a third party (for example, a public registry). The commits below record when the plan and the pathway names were fixed. The timestamps are those of the authors' repository and have not been certified by a third party; the full commit history will be made available with the code. Analyses added after an internal review of the draft are described separately as post hoc analyses (S1.10).

**Pre-specification record.** Commits that fixed the analysis plan and the pathway names

| Commit | Time | Content |
|---|---|---|
| 7e2ce1254e21 | 2026-11-11 10:55:04 (UTC+9) | Statistical analysis plan v1.0 frozen together with the data-preparation code, the sequence and regression functions and their unit tests, before employment states were examined |
| 8389c0070854 | 2026-11-11 10:56:56 (UTC+9) | Names of the four pathways selected by the pre-specified rule and the clustering script, before any regression model was fitted |
| 3d5cecc421c4 | 2026-11-11 11:27:43 (UTC+9) | Scripts for the descriptive, model, secondary, sensitivity and figure analyses and Amendment A1, committed after these analyses had been run |

Commit identifiers are the first characters of the git SHA-1 hash

Amendment A1 (11 November 2026). In SA6(b) the chained multinomial imputation models could not be fitted because the information matrix was singular at one interview. A small L2 penalty ($\lambda = 0.001$) on the coefficients was added to these imputation models only; all other models are unpenalised. The change was made before any result of SA6(b) had been produced and is recorded in the analysis plan. No other change was made to the pre-specified analyses after the plan was frozen; the post hoc analyses in S1.10 were added later and replace none of them.

### S1.9 Software, random numbers and validation

Analyses used Python 3.10.11 with pandas 2.3.3, NumPy 2.2.6, SciPy 1.15.3, statsmodels 0.15.0 (modified Poisson and logistic models), pyreadstat 1.3.6 (reading SPSS files) and matplotlib 3.10.9 (figures). Sequence distances, Ward's method with frequency weights, PAM, the quality indices, the ARI, the bootstrap Jaccard coefficients, latent class analysis and the multinomial logistic regression were written by the authors. Unit tests checked that (1) OM distances agree with hand-calculated examples, with the longest-common-subsequence distance and with a naive dynamic-programming implementation and satisfy the properties of a metric; (2) transition-rate costs agree with their definition; (3) weighted Ward clustering of distinct sequences agrees with scipy on the expanded data; (4) ASW, PBC and Hubert's C on distinct sequences equal those on the expanded data; (5) PAM reaches a local optimum and finds the optimum of small problems solved by brute force; (6) the ARI reproduces known values; (7) the bootstrap Jaccard coefficient is 1 for well-separated clusters; (8) latent class analysis recovers two simulated classes; and (9) the multinomial regression reproduces the coefficients, HC0 standard errors and AMEs of statsmodels, gives identical point estimates with integer weights and expanded data, codes reference categories correctly, and computes Wald tests and E-values correctly. Post hoc, the main multinomial model was also fitted to the study data with statsmodels MNLogit (HC0 standard errors): coefficients and robust standard errors agreed to within $10^{-13}$.

The global random seed was 20261111; derived seeds were used for stability (+1), bootstrap CIs of weighted prevalence (+2), g-computation (+3), multiple imputation (+4), latent class analysis (+5), other bootstrap CIs (+6) and the post hoc stratified bootstrap (+7). All tables and figures are produced from the raw data files by a single script (run_all.sh).

### S1.10 Post hoc analyses

The analyses in this section were added on 16 November 2026, after an internal review of the draft, and were not part of the analysis plan. A separate script (07b_posthoc.py) produces them and stores its results apart from those of the pre-specified analyses, which did not change.

*Typology and the dominant state (Table S16).* With a substitution cost of 2 and an insertion–deletion cost of 1, the OM distance between a sequence and a sequence of five identical states (such as OOOOO) equals 2 × (5 − the number of years spent in that state). Because all four medoids of the main typology were such sequences, assigning a worker to the nearest medoid amounts to choosing the medoid state in which the worker spent the most years, with ties resolved by the order of the medoids in the algorithm. We compared the typology with this rule (rule A) and with a variant that counted unemployment and being out of the labour force together (rule B). Rule A agreed for 4,076 of 4,087 workers (99.7%; ARI 0.997); all 11 discordant workers had equally many years in self-employment and out of the labour force. Rule B agreed for 97.5% (104 discordant workers); most discordant workers were in the different-employer pathway (83) or the injury-employer pathway (15) although they were unemployed or out of the labour force in three or more years, because unemployment counts towards no medoid and because of how ties were resolved.

*Unadjusted associations (Table S17).* Each factor was entered alone in the multinomial model, so the AMEs equal differences in observed proportions. They had the same sign as the adjusted AMEs for 87% of the 100 estimates and were usually larger, especially for job characteristics, which are correlated with one another (trade union and the injury-employer pathway: 42.2 vs. 11.0 percentage points adjusted; tenure $<1$ year: $-49.6$ vs. $-31.0$; age $\geq 60$ years and persistent non-employment: 34.1 vs. 27.4).

*Events per parameter (Table S18).* The self-employment equation of the main model had 9.3 events per parameter (251 workers, 27 parameters). In a reduced model with 17 parameters per equation (14.8 events per parameter), AMEs for self-employment changed by at most 2.9 and those for persistent non-employment by at most 3.9 percentage points; self-employment RRRs changed by factors of 0.64 to 2.00, mainly for employment status and firm size, because the reduced model omits correlated job characteristics.

*Missed interviews and deaths (Table S19).* The data record a reason for each missed interview, including death. Of the participants who missed at least one interview (486 in cohort 1 and 721 in cohort 2), 16 and 54 were recorded as having died. Because the main analyses included only participants interviewed five times, no deceased participant was in the main sample. In SA6(b) the missing states of deceased participants had been imputed as if they were alive; recomputing the prevalence from the stored imputations without them gave persistent non-employment of 11.6% in cohort 1 and 17.8% in cohort 2, compared with 11.8% and 18.1% with them (Table S14). At the first interview, 31.7% (cohort 1) and 34.8% (cohort 2) of participants who later missed an interview were not employed, compared with 28.7% and 38.2% of continuous respondents (unweighted).

*Bootstrap within approximate strata.* The public data contain no stratum identifiers. We approximated the sampling strata by disability grade (none, grades 10–14, 4–9 and 1–3) × use of rehabilitation services in cohort 1 (7 strata) and by disability grade × sex × age group in cohort 2 (31 strata), resampled participants within these strata 1,000 times and recomputed the weighted pathway prevalence. The percentile 95% CIs were 0.95 to 1.03 times as wide as those in Table 2 of the main article. Neither bootstrap re-estimated the non-response adjustment of the survey weights.

## S2 Supplementary tables

**Table S1** Participation at the five annual interviews, by cohort

| | Cohort 1 (care ended 2012; interviews 2013–2017), n (%) | Cohort 2 (care ended 2017; interviews 2018–2022), n (%) |
|---|---|---|
| Participants at the first interview | 2,000 | 3,294 |
| Responded at interview 2 | 1,803 (90.2) | 2,965 (90.0) |
| Responded at interview 3 | 1,704 (85.2) | 2,823 (85.7) |
| Responded at interview 4 | 1,660 (83.0) | 2,797 (84.9) |
| Responded at interview 5 | 1,616 (80.8) | 2,728 (82.8) |
| Responded at all five interviews (analytic sample) | 1,514 (75.7) | 2,573 (78.1) |
| Responded at exactly four interviews | 155 (7.8) | 215 (6.5) |
| Responded at exactly three interviews | 77 (3.9) | 111 (3.4) |
| Responded at exactly two interviews | 108 (5.4) | 154 (4.7) |
| Responded at exactly one interview | 146 (7.3) | 241 (7.3) |
| Participation pattern: complete or monotone dropout | 1,877 (93.8) | 3,116 (94.6) |
| Participation pattern: intermittent (responded again after a missed interview) | 123 (6.2) | 178 (5.4) |

Percentages are of participants at the first interview. Interview 1 took place about 8 to 21 months after compensated medical care ended.

**Table S2** Definitions, categories and sources of the variables

| Variable | Definition and categories | Item (cohort 1; cohort 2) | Source |
|---|---|---|---|
| Employment state (each interview) | Economic activity in the week before the interview: O, working for the employer at the time of injury; R, working for a different employer; S, self-employed or unpaid family worker; U, unemployed; N, not in the labour force. Respondents employed at the previous interview were first asked whether they still held the same job (dependent interviewing) | emp06; emp6 | Survey |
| Sex | Male (ref.), female | gender; gender | Survey data, basic variables |
| Age group, years | ≤39 (ref.), 40–49, 50–59, ≥60; completed years in August of the first survey year (age groups only in the public data; cohort 1's two youngest groups merged) | age; age4 | Survey data, basic variables |
| Education | Middle school or less, high school, college or higher (ref.) | edu; edu | Survey data, basic variables |
| Occupational disease | Occupational disease vs. injury (ref.); commuting injuries in cohort 2 counted as injury | accident; accident | Insurance records |
| Commuting injury | Used only to exclude commuting injuries in SA8 (category exists in cohort 2 only) | –; accident | Insurance records |
| Duration of compensated medical care | ≤6 months (ref.), >6–12 months, >12 months | con; con16 | Insurance records |
| Disability grade | None (ref.), grades 10–14, grades 1–9 (grades 1–3 merged with 4–9 in models; shown separately in Table 1) | disa15; disa15 | Insurance records |
| Employment status at injury | Regular (ref.), temporary or other (including self-employed at injury), daily | C001007; C005001 | First interview |
| Firm size at injury, employees | <5, 5–29, 30–299, ≥300 (ref.) | C002004; C024001 | First interview |
| Trade union at the workplace of injury | Yes, no (ref.), unknown or not asked (cohort 2 participants who were self-employed at injury were not asked) | C002019; C011001 | First interview |
| Industry at injury | Manufacturing (ref.), construction, transport, other | C001003; C002002 | First interview |
| Job tenure at injury, years | <1, 1–5, ≥5 (ref.) | workperiod; workperiod14 | Insurance records |
| Married at injury | Married vs. not married (ref.) | A003001; A002001 | First interview |

| | | | |
|---|---|---|---|
| Use of rehabilitation services | Any use vs. none; design variable only (cohort 1 over-sampled service users) | service; service | Insurance records |
| High pain interference | Pain from the injury interfered with daily life for a considerable time or always (highest two of four categories); secondary analysis only | Ga001004; G003001 | First interview |
| Physical limitation | Long-lasting limitation in walking, climbing stairs or lifting; secondary analysis only | Gb003002; G009002 | First interview |
| Chronic disease | Current chronic disease; secondary analysis only | Gb001003; G006001 | First interview |
| Longitudinal weight | Weight for continuous respondents; positive only for participants who responded at all five interviews; sums to the survey population (82,493 and 75,392) | wt051; wt051 | Survey weights |
| First-interview weight | Used for the prevalence estimates of SA6, which include all participants | wt01; wt01 | Survey weights |

ref., reference category. Insurance records were extracted by the survey from the Korea Workers' Compensation & Welfare Service (cohort 1: December 2012; cohort 2: April 2018). Categories were merged on the basis of covariate counts in the analytic sample before employment states were examined. No variable had missing values in the analytic sample.

**Table S3** Cluster quality by number of pathways (k)

| k | ASW | PBC | HC | Smallest pathway, % | Eligible under the rule | Medoid sequences |
|---|---|---|---|---|---|---|
| 2 | 0.483 | 0.610 | 0.109 | 32.6 | No | OOOOO NRRRR |
| 3 | 0.621 | 0.818 | 0.046 | 22.9 | Yes | OOOOO RRRRR NNNNN |
| 4[a] | 0.686 | 0.875 | 0.012 | 6.1 | Yes | OOOOO RRRRR SSSSS NNNNN |
| 5 | 0.558 | 0.735 | 0.079 | 6.1 | Yes | OOOOO RRRRR SSSSS NRRRR NNNNN |
| 6 | 0.591 | 0.775 | 0.045 | 6.1 | Yes | OOOOO RRRRR SSSSS URRNN NRRRR NNNNN |
| 7 | 0.583 | 0.718 | 0.056 | 4.1 | Yes | OOOOO OORRR RRRRR RURNN SSSSS NRRRR NNNNN |
| 8 | 0.597 | 0.717 | 0.043 | 3.5 | Yes | OOOOO OORRR RRRRR RURNN SSSSS NRRRR NUNRR NNNNN |
| 9 | 0.610 | 0.708 | 0.038 | 3.5 | No | OOOOO OOORR RRRRR RURNN SSSSS URRRR NRRRR NUNRR NNNNN |
| 10 | 0.614 | 0.707 | 0.029 | 2.7 | No | OOOOO OOONN OORRR RRRRR RURNN SSSSS URRRR NRRRR NUNRR NNNNN |

Optimal matching (substitution cost 2, insertion–deletion cost 1), Ward's method followed by partitioning around medoids, n = 4,087 participants (389 distinct sequences). ASW, average silhouette width; PBC, point-biserial correlation (higher is better for both); HC, Hubert's C (lower is better). The pre-specified rule considered k = 3–8 with every pathway including at least 3% of participants and chose the smallest k whose ASW was within 0.02 of the highest ASW among them. [a] Selected solution. States: O, working for the employer at injury; R, working for a different employer; S, self-employed or unpaid family worker; U, unemployed; N, not in the labour force

**Table S4** Stability of the four pathways in bootstrap resamples

| Pathway | Medoid | Mean Jaccard | Median Jaccard | 5th percentile | Resamples with Jaccard < 0.5, % |
|---|---|---|---|---|---|
| Sustained work for the employer at injury | OOOOO | 1.000 | 1.000 | 1.000 | 0.0 |
| Work for a different employer | RRRRR | 0.990 | 1.000 | 0.950 | 0.0 |
| Self-employment | SSSSS | 0.957 | 0.958 | 0.907 | 0.0 |

| | | | | | |
|---|---|---|---|---|---|
| Persistent non-employment | NNNNN | 0.976 | 0.987 | 0.911 | 0.0 |

B = 200 bootstrap resamples of participants. In each resample, Ward's method and partitioning around medoids were rerun with k = 4; for each original pathway, the largest Jaccard coefficient with any pathway of the resample solution was recorded (Hennig 2007). Mean values ≥0.75 indicate a stable pathway (pre-specified)

**Table S5** Weighted distribution of employment states at each interview, by cohort (%)

| Cohort, interview (year) | O | R | S | U | N | Employed (O, R or S) |
|---|---|---|---|---|---|---|
| Cohort 1, interview 1 (2013) | 40.3 | 34.5 | 4.3 | 5.2 | 15.7 | 79.1 |
| Cohort 1, interview 2 (2014) | 34.8 | 45.7 | 5.4 | 4.2 | 10.0 | 85.8 |
| Cohort 1, interview 3 (2015) | 31.3 | 47.0 | 7.6 | 4.0 | 10.1 | 85.9 |
| Cohort 1, interview 4 (2016) | 29.9 | 48.8 | 7.5 | 2.0 | 11.7 | 86.3 |
| Cohort 1, interview 5 (2017) | 28.7 | 48.7 | 8.4 | 2.5 | 11.7 | 85.8 |
| Cohort 2, interview 1 (2018) | 35.6 | 31.4 | 3.8 | 6.7 | 22.4 | 70.9 |
| Cohort 2, interview 2 (2019) | 31.3 | 38.7 | 6.1 | 6.0 | 18.0 | 76.0 |
| Cohort 2, interview 3 (2020) | 27.2 | 42.8 | 7.0 | 4.7 | 18.4 | 76.9 |
| Cohort 2, interview 4 (2021) | 25.8 | 44.5 | 8.1 | 2.5 | 19.1 | 78.4 |
| Cohort 2, interview 5 (2022) | 23.5 | 44.4 | 8.8 | 2.9 | 20.4 | 76.7 |

Participants who responded at all five interviews (cohort 1, n = 1,514; cohort 2, n = 2,573), weighted with the longitudinal weight. O, working for the employer at injury; R, working for a different employer; S, self-employed or unpaid family worker; U, unemployed; N, not in the labour force

**Table S6** Relative risk ratios (95% CI) from the multinomial logistic regression of pathway membership (reference: sustained work for the employer at injury)

| Factor (reference) | Work for a different employer | Self-employment | Persistent non-employment |
|---|---|---|---|
| **Age group, years (ref. ≤39)** | | | |
| 40–49 | 0.79 (0.61–1.03) | 0.49 (0.31–0.78) | 0.85 (0.56–1.29) |
| 50–59 | 0.72 (0.55–0.94) | 0.35 (0.22–0.55) | 1.51 (1.02–2.24) |
| ≥60 | 0.95 (0.68–1.33) | 0.45 (0.26–0.80) | 6.30 (4.07–9.76) |
| **Education (ref. college or higher)** | | | |
| Middle school or less | 1.19 (0.90–1.58) | 1.55 (0.95–2.53) | 1.74 (1.17–2.58) |
| High school | 0.93 (0.74–1.16) | 1.01 (0.67–1.53) | 1.23 (0.86–1.74) |
| **Duration of medical care (ref. ≤6 months)** | | | |
| >6–12 months | 1.42 (1.16–1.74) | 1.52 (1.08–2.13) | 1.86 (1.43–2.40) |
| >12 months | 1.53 (1.07–2.18) | 1.68 (0.93–3.04) | 7.22 (5.04–10.36) |
| **Disability grade (ref. none)** | | | |
| Grades 10–14 | 1.15 (0.91–1.44) | 1.02 (0.69–1.49) | 1.10 (0.81–1.51) |
| Grades 1–9 | 0.98 (0.69–1.38) | 0.98 (0.55–1.76) | 2.39 (1.59–3.60) |
| **Employment status at injury (ref. regular)** | | | |
| Temporary or other | 2.08 (1.54–2.81) | 2.10 (1.30–3.41) | 2.04 (1.41–2.94) |
| Daily | 1.82 (1.38–2.39) | 2.64 (1.67–4.18) | 1.76 (1.26–2.47) |
| **Firm size at injury, employees (ref. ≥300)** | | | |
| <5 | 1.46 (1.00–2.15) | 3.67 (1.79–7.52) | 2.05 (1.24–3.40) |
| 5–29 | 1.58 (1.13–2.22) | 2.12 (1.07–4.20) | 1.69 (1.07–2.67) |
| 30–299 | 1.59 (1.14–2.22) | 1.56 (0.79–3.08) | 1.68 (1.07–2.65) |
| **Trade union at workplace (ref. no)** | | | |
| Yes | 0.46 (0.31–0.66) | 0.60 (0.27–1.32) | 0.65 (0.41–1.03) |
| Unknown or not asked | 1.53 (1.01–2.33) | 2.82 (1.63–4.88) | 1.70 (1.03–2.79) |
| **Industry (ref. manufacturing)** | | | |
| Construction | 1.74 (1.31–2.30) | 1.04 (0.65–1.66) | 1.74 (1.23–2.45) |
| Transport | 0.77 (0.49–1.21) | 1.43 (0.74–2.76) | 0.89 (0.50–1.59) |
| Other | 0.95 (0.77–1.16) | 1.04 (0.72–1.51) | 1.06 (0.81–1.39) |
| **Job tenure at injury, years (ref. ≥5)** | | | |
| <1 | 6.80 (5.16–8.97) | 4.26 (2.51–7.24) | 3.80 (2.69–5.36) |
| 1–5 | 2.75 (2.09–3.62) | 2.24 (1.31–3.82) | 1.49 (1.05–2.11) |
| Female (ref. male) | 1.42 (1.12–1.80) | 0.51 (0.30–0.87) | 2.29 (1.70–3.08) |

| | | | |
|---|---|---|---|
| Occupational disease (ref. injury) | 1.51 (1.07–2.13) | 1.18 (0.58–2.37) | 2.17 (1.48–3.19) |
| Married at injury (ref. not married) | 0.72 (0.59–0.88) | 1.42 (0.98–2.06) | 0.54 (0.42–0.69) |
| Cohort 2 (ref. cohort 1) | 1.26 (1.04–1.52) | 1.41 (1.02–1.95) | 1.89 (1.47–2.44) |

n = 4,087; all factors were entered simultaneously (27 parameters per equation including the intercept). Intercepts and the coefficients for use of rehabilitation services, included only as a design variable because cohort 1 over-sampled service users, are not shown. Robust (HC0) standard errors. The model converged without the pre-specified category merges

**Table S7** Wald tests of interactions between cohort and pre-specified factors

| Interaction | $\chi^2$ | df | P |
|---|---|---|---|
| Age group × cohort | 14.97 | 9 | 0.09 |
| Disability grade × cohort | 2.86 | 6 | 0.83 |
| Employment status at injury × cohort | 4.66 | 6 | 0.59 |

Each set of product terms was added separately to the main multinomial model and tested jointly across the three equations with robust standard errors

**Table S8** Associations of pre-specified factors with pathway membership in the main model and in sensitivity analyses SA5, SA7 and SA8 (95% CI)

| Factor | Pathway | Main (n = 4,087) | SA5 weighted (n = 4,087) | SA7 age <60 (n = 3,038) | SA8 excl. disease and commuting (n = 3,799) |
|---|---|---|---|---|---|
| **A. Relative risk ratios (reference: sustained work for the employer at injury)** | | | | | |
| Age group, years: ≥60 | Work for a different employer | 0.95 (0.68–1.33) | 0.76 (0.45–1.31) | – | 0.89 (0.63–1.26) |
| | Self-employment | 0.45 (0.26–0.80) | 0.59 (0.25–1.40) | – | 0.41 (0.23–0.72) |
| | Persistent non-employment | 6.30 (4.07–9.76) | 5.01 (2.39–10.50) | – | 5.81 (3.70–9.14) |
| Duration of medical care: >12 months | Work for a different employer | 1.53 (1.07–2.18) | 2.35 (1.22–4.52) | 1.62 (1.09–2.39) | 1.46 (1.01–2.12) |
| | Self-employment | 1.68 (0.93–3.04) | 2.41 (0.97–5.98) | 1.54 (0.82–2.90) | 1.56 (0.84–2.89) |
| | Persistent non-employment | 7.22 (5.04–10.36) | 6.70 (3.57–12.56) | 8.03 (5.24–12.31) | 6.55 (4.47–9.58) |
| Disability grade: Grades 1–9 | Work for a different employer | 0.98 (0.69–1.38) | 1.06 (0.65–1.75) | 1.04 (0.71–1.54) | 0.94 (0.66–1.33) |
| | Self-employment | 0.98 (0.55–1.76) | 0.92 (0.39–2.16) | 1.15 (0.60–2.20) | 0.93 (0.51–1.67) |
| | Persistent non-employment | 2.39 (1.59–3.60) | 2.53 (1.44–4.45) | 2.37 (1.42–3.96) | 2.26 (1.48–3.46) |
| Employment status at injury: Temporary or other | Work for a different employer | 2.08 (1.54–2.81) | 2.59 (1.62–4.14) | 2.35 (1.64–3.36) | 2.00 (1.48–2.71) |
| | Self-employment | 2.10 (1.30–3.41) | 1.45 (0.63–3.36) | 2.00 (1.13–3.54) | 2.12 (1.30–3.46) |
| | Persistent non-employment | 2.04 (1.41–2.94) | 3.68 (1.93–7.02) | 2.48 (1.54–3.98) | 2.01 (1.39–2.91) |
| Employment status at injury: Daily | Work for a different employer | 1.82 (1.38–2.39) | 2.11 (1.35–3.29) | 2.14 (1.55–2.96) | 1.86 (1.41–2.47) |
| | Self-employment | 2.64 (1.67–4.18) | 2.81 (1.34–5.88) | 2.70 (1.60–4.55) | 2.70 (1.70–4.30) |
| | Persistent non-employment | 1.76 (1.26–2.47) | 2.44 (1.39–4.28) | 2.04 (1.34–3.12) | 1.87 (1.33–2.64) |
| Firm size at injury, employees: <5 | Work for a different employer | 1.46 (1.00–2.15) | 2.22 (1.16–4.27) | 1.60 (1.04–2.46) | 1.56 (1.04–2.35) |
| | Self-employment | 3.67 (1.79–7.52) | 6.41 (2.28–17.97) | 4.13 (1.90–8.97) | 4.25 (1.93–9.36) |
| | Persistent non-employment | 2.05 (1.24–3.40) | 3.26 (1.41–7.57) | 2.10 (1.07–4.13) | 1.89 (1.12–3.19) |
| Female | Work for a different employer | 1.42 (1.12–1.80) | 1.58 (1.07–2.34) | 1.45 (1.10–1.91) | 1.44 (1.13–1.85) |
| | Self-employment | 0.51 (0.30–0.87) | 0.45 (0.20–1.02) | 0.55 (0.30–1.01) | 0.48 (0.28–0.85) |
| | Persistent non-employment | 2.29 (1.70–3.08) | 2.49 (1.52–4.09) | 2.90 (1.98–4.25) | 2.37 (1.74–3.23) |
| Occupational disease | Work for a different employer | 1.51 (1.07–2.13) | 1.80 (1.03–3.15) | 1.49 (1.02–2.17) | – |
| | Self-employment | 1.18 (0.58–2.37) | 0.41 (0.16–1.03) | 1.26 (0.60–2.66) | – |
| | Persistent non-employment | 2.17 (1.48–3.19) | 2.24 (1.14–4.40) | 1.98 (1.24–3.18) | – |
| Cohort 2 | Work for a different employer | 1.26 (1.04–1.52) | 1.11 (0.82–1.50) | 1.24 (1.01–1.53) | 1.27 (1.04–1.54) |
| | Self-employment | 1.41 (1.02–1.95) | 1.00 (0.56–1.77) | 1.62 (1.13–2.32) | 1.40 (1.00–1.96) |
| | Persistent non-employment | 1.89 (1.47–2.44) | 1.81 (1.18–2.80) | 1.71 (1.25–2.33) | 1.87 (1.44–2.42) |
| **B. Average marginal effects, percentage points** | | | | | |
| Age group, years: ≥60 | Sustained work for the employer at injury | −5.1 (−9.4 to −0.7) | −2.0 (−9.6 to 5.6) | – | −3.8 (−8.2 to 0.7) |
| | Work for a different employer | −14.7 (−20.5 to −9.0) | −16.5 (−25.7 to −7.3) | – | −15.0 (−21.0 to −9.0) |
| | Self-employment | −7.6 (−11.4 to −3.8) | −4.3 (−9.6 to 1.0) | – | −8.3 (−12.3 to −4.3) |
| | Persistent non-employment | 27.4 (23.0–31.8) | 22.8 (15.2–30.3) | – | 27.1 (22.5–31.6) |
| Duration of medical care: >12 months | Sustained work for the employer at injury | −12.7 (−16.6 to −8.8) | −15.4 (−22.2 to −8.6) | −13.0 (−17.8 to −8.1) | −11.8 (−16.0 to −7.6) |
| | Work for a different employer | −12.5 (−17.6 to −7.4) | −1.1 (−11.0 to 8.9) | −8.9 (−14.9 to −2.8) | −12.3 (−17.7 to −7.0) |
| | Self-employment | −1.0 (−3.6 to 1.6) | 0.4 (−4.9 to 5.7) | −1.2 (−4.1 to 1.8) | −1.1 (−3.9 to 1.6) |
| | Persistent non-employment | 26.2 (21.6–30.7) | 16.1 (9.0–23.1) | 23.0 (17.9–28.1) | 25.3 (20.5–30.1) |
| Disability grade: Grades 1–9 | Sustained work for the employer at injury | −3.4 (−8.1 to 1.4) | −3.7 (−10.8 to 3.5) | −3.7 (−9.5 to 2.1) | −2.5 (−7.5 to 2.4) |
| | Work for a different employer | −8.4 (−13.9 to −2.9) | −5.5 (−13.5 to 2.6) | −6.0 (−12.4 to 0.5) | −8.6 (−14.3 to −2.9) |
| | Self-employment | −1.1 (−4.0 to 1.8) | −1.6 (−6.0 to 2.8) | −0.1 (−3.6 to 3.3) | −1.3 (−4.5 to 1.8) |
| | Persistent non-employment | 12.9 (8.3–17.4) | 10.7 (5.0–16.5) | 9.8 (4.9–14.7) | 12.5 (7.7–17.2) |

| | | | | | |
|---|---|---|---|---|---|
| Employment status at injury: Temporary or other | Sustained work for the employer at injury | −10.7 (−14.7 to −6.8) | −15.4 (−22.0 to −8.9) | −13.2 (−18.1 to −8.3) | −10.4 (−14.4 to −6.4) |
| | Work for a different employer | 7.4 (2.6–12.1) | 10.2 (2.1–18.4) | 9.7 (4.0–15.4) | 6.7 (1.8–11.6) |
| | Self-employment | 1.1 (−1.3 to 3.5) | −1.8 (−5.4 to 1.9) | 0.4 (−2.4 to 3.3) | 1.3 (−1.2 to 3.8) |
| | Persistent non-employment | 2.2 (−1.5 to 6.0) | 7.0 (0.6–13.4) | 3.1 (−1.0 to 7.2) | 2.4 (−1.4 to 6.2) |
| Employment status at injury: Daily | Sustained work for the employer at injury | −9.5 (−13.4 to −5.6) | −13.4 (−20.2 to −6.5) | −12.3 (−17.1 to −7.6) | −9.9 (−13.9 to −6.0) |
| | Work for a different employer | 4.9 (0.4–9.4) | 6.5 (−1.2 to 14.2) | 7.9 (2.5–13.3) | 4.9 (0.2–9.5) |
| | Self-employment | 3.3 (0.6–6.1) | 3.6 (−1.8 to 9.0) | 3.0 (−0.3 to 6.3) | 3.3 (0.6–6.1) |
| | Persistent non-employment | 1.2 (−2.2 to 4.7) | 3.2 (−1.2 to 7.7) | 1.4 (−2.1 to 5.0) | 1.8 (−1.8 to 5.3) |
| Firm size at injury, employees: <5 | Sustained work for the employer at injury | −8.9 (−14.7 to −3.1) | −17.3 (−28.3 to −6.4) | −10.9 (−17.9 to −3.9) | −9.6 (−15.7 to −3.4) |
| | Work for a different employer | −1.5 (−8.4 to 5.3) | 5.4 (−6.2 to 16.9) | 0.7 (−7.1 to 8.5) | 0.0 (−7.3 to 7.3) |
| | Self-employment | 5.8 (2.4–9.2) | 6.4 (2.1–10.6) | 7.1 (3.1–11.2) | 6.5 (2.9–10.0) |
| | Persistent non-employment | 4.7 (−0.7 to 10.1) | 5.6 (−0.8 to 12.0) | 3.1 (−2.2 to 8.3) | 3.1 (−2.7 to 9.0) |
| Female | Sustained work for the employer at injury | −5.6 (−8.7 to −2.6) | −6.7 (−12.1 to −1.2) | −6.5 (−10.2 to −2.7) | −5.8 (−9.0 to −2.7) |
| | Work for a different employer | 1.2 (−2.9 to 5.3) | 5.1 (−1.7 to 12.0) | 1.1 (−3.8 to 6.0) | 1.2 (−3.1 to 5.5) |
| | Self-employment | −4.3 (−5.8 to −2.8) | −5.5 (−8.3 to −2.8) | −4.3 (−6.2 to −2.4) | −4.6 (−6.1 to −3.0) |
| | Persistent non-employment | 8.7 (5.1–12.3) | 7.1 (2.0–12.1) | 9.7 (5.5–13.9) | 9.2 (5.4–13.0) |
| Occupational disease | Sustained work for the employer at injury | −6.6 (−10.4 to −2.7) | −7.6 (−14.6 to −0.6) | −6.3 (−11.1 to −1.6) | – |
| | Work for a different employer | 1.0 (−5.2 to 7.1) | 8.0 (−2.1 to 18.2) | 2.6 (−4.4 to 9.5) | – |
| | Self-employment | −1.1 (−4.3 to 2.0) | −5.2 (−7.6 to −2.8) | −0.6 (−4.5 to 3.4) | – |
| | Persistent non-employment | 6.7 (1.8–11.6) | 4.8 (−2.6 to 12.2) | 4.3 (−0.6 to 9.2) | – |
| Cohort 2 | Sustained work for the employer at injury | −5.0 (−7.7 to −2.3) | −2.8 (−7.5 to 1.9) | −4.7 (−7.8 to −1.5) | −5.0 (−7.8 to −2.1) |
| | Work for a different employer | −1.4 (−4.6 to 1.8) | −1.5 (−6.7 to 3.8) | −0.5 (−4.1 to 3.1) | −1.1 (−4.4 to 2.2) |
| | Self-employment | 0.6 (−1.0 to 2.2) | −0.9 (−4.3 to 2.5) | 1.7 (−0.2 to 3.5) | 0.5 (−1.2 to 2.2) |
| | Persistent non-employment | 5.8 (3.2–8.3) | 5.1 (1.7–8.5) | 3.5 (0.9–6.0) | 5.5 (2.9–8.2) |

SA5, model and marginal effects weighted with the longitudinal weight; SA7, participants aged <60 years at the first interview; SA8, participants with injuries other than commuting injuries (occupational diseases excluded). All models include all factors at injury, cohort and use of rehabilitation services. –, not estimable because the factor was excluded

**Table S9** Type of employment at the first interview and exit from employment at the second to fifth interviews

| Employment at the first interview | n | Exit from employment, n (%) | Adjusted RR (95% CI) | P | E-value (CI limit) |
|---|---|---|---|---|---|
| Employer at injury (reference) | 1,314 | 280 (21.3) | 1 | | |
| Different employer | 1,215 | 350 (28.8) | 1.13 (0.97–1.32) | 0.11 | 1.52 (1.00) |
| Self-employed or unpaid family worker | 141 | 18 (12.8) | 0.52 (0.33–0.82) | 0.005 | 3.28 (1.74) |

Exit from employment: unemployed or out of the labour force at any of the second to fifth interviews (n = 2,670, 648 exits). Modified Poisson regression with robust standard errors, adjusted for all factors at injury, cohort, use of rehabilitation services and pain interference, physical limitation and chronic disease at the first interview. Exposure × cohort interaction: $\chi^2 = 1.76$, df = 2, P = 0.41. E-values are given for the point estimate and, in parentheses, for the confidence limit closer to 1 (1 when the CI includes 1). Standardised probabilities by cohort are shown in Fig. 2

**Table S10** Later employment among participants not employed at the first interview

| Cohort | n not employed at interview 1 | Outcome | n with outcome | Weighted % (95% CI) |
|---|---|---|---|---|
| Cohort 1 | 434 | Employed at any of interviews 2–5 | 291 | 75.1 (67.8–81.6) |
| Cohort 1 | 434 | Employed at both interviews 4 and 5 | 206 | 52.7 (44.0–61.1) |
| Cohort 2 | 983 | Employed at any of interviews 2–5 | 573 | 69.8 (65.9–74.1) |
| Cohort 2 | 983 | Employed at both interviews 4 and 5 | 377 | 46.6 (41.8–51.6) |

Weighted with the longitudinal weight; percentile 95% CIs from 1,000 bootstrap resamples within cohort

**Table S11** Weighted transitions between employment states at consecutive interviews, by cohort (row %)

| | To O | To R | To S | To U | To N |
|---|---|---|---|---|---|
| **Cohort 1, interview 1 → 2 (2013–2014)** | | | | | |
| From O | 83.8 | 8.7 | 0.6 | 3.5 | 3.4 |
| From R | 0.5† | 91.9 | 1.2 | 3.5 | 2.9 |

| | | | | | |
|---|---|---|---|---|---|
| From S | 0.0 | 19.2 | 80.8 | 0.0 | 0.0 |
| From U | 1.2† | 69.0 | 8.9† | 5.1 | 15.8 |
| From N | 4.6 | 38.7 | 5.0 | 8.1 | 43.6 |
| **Cohort 1, interview 2 → 3 (2014–2015)** | | | | | |
| From O | 87.7 | 8.6 | 0.9† | 1.6 | 1.1 |
| From R | 1.1 | 87.2 | 2.7 | 5.6 | 3.4 |
| From S | 1.0† | 8.6 | 89.1 | 1.3† | 0.0 |
| From U | 1.0† | 46.5 | 28.4† | 4.2 | 19.9 |
| From N | 2.2† | 17.6 | 0.8† | 6.4 | 73.0 |
| **Cohort 1, interview 3 → 4 (2015–2016)** | | | | | |
| From O | 92.8 | 3.5 | 1.5† | 1.1 | 1.2 |
| From R | 0.3† | 93.4 | 0.5† | 1.4 | 4.4 |
| From S | 2.8† | 6.2† | 86.8 | 0.4† | 3.8† |
| From U | 13.5† | 38.6 | 4.0† | 9.6 | 34.3 |
| From N | 0.0 | 18.0 | 0.6† | 6.4 | 75.0 |
| **Cohort 1, interview 4 → 5 (2016–2017)** | | | | | |
| From O | 95.6 | 1.3 | 0.7 | 0.3† | 2.1 |
| From R | 0.0 | 93.3 | 1.1 | 2.1 | 3.6 |
| From S | 0.0 | 0.4† | 93.0 | 0.0 | 6.6† |
| From U | 1.0† | 48.8 | 0.0 | 29.7 | 20.6 |
| From N | 0.6† | 14.9 | 5.6 | 6.9 | 71.9 |
| **Cohort 2, interview 1 → 2 (2018–2019)** | | | | | |
| From O | 81.7 | 7.2 | 1.5 | 2.6 | 7.0 |
| From R | 0.0† | 84.3 | 2.3 | 5.6 | 7.8 |
| From S | 0.0 | 5.8 | 92.5 | 0.6† | 1.1† |
| From U | 7.0 | 46.4 | 4.2 | 21.9 | 20.4 |
| From N | 7.4 | 28.2 | 4.4 | 8.1 | 51.8 |
| **Cohort 2, interview 2 → 3 (2019–2020)** | | | | | |
| From O | 84.8 | 5.7 | 1.0 | 2.4 | 6.1 |
| From R | 0.4† | 87.3 | 0.7 | 3.6 | 8.0 |
| From S | 0.0 | 4.5 | 91.4 | 0.0 | 4.0 |
| From U | 1.1† | 45.2 | 7.6 | 22.7 | 23.5 |
| From N | 2.5 | 23.4 | 2.3 | 6.4 | 65.4 |
| **Cohort 2, interview 3 → 4 (2020–2021)** | | | | | |
| From O | 91.3 | 1.9 | 0.9 | 1.4 | 4.4 |
| From R | 0.5† | 88.2 | 2.5 | 1.5 | 7.3 |
| From S | 0.2† | 1.5 | 89.6 | 2.6† | 6.1 |
| From U | 5.6 | 44.6 | 3.4† | 12.6 | 33.8 |
| From N | 2.7 | 22.2 | 1.9 | 3.7 | 69.4 |
| **Cohort 2, interview 4 → 5 (2021–2022)** | | | | | |
| From O | 89.1 | 4.5 | 0.2† | 0.1† | 6.1 |
| From R | 0.3† | 87.1 | 1.2 | 3.3 | 8.0 |
| From S | 0.0 | 5.9 | 90.2 | 0.0 | 3.9 |
| From U | 1.7† | 31.0 | 7.3† | 32.4 | 27.7 |
| From N | 2.0 | 16.8 | 3.7 | 2.9 | 74.6 |

Participants who responded at all five interviews, weighted with the longitudinal weight; each row sums to 100%. O, working for the employer at injury; R, working for a different employer; S, self-employed or unpaid family worker; U, unemployed; N, not in the labour force. † Based on fewer than five participants

**Table S12** Exit from employment between the second and third interviews

| | Years | n employed at interview 2 | Weighted % unemployed or out of the labour force at interview 3 (95% CI) |
|---|---|---|---|
| Cohort 1 | 2014–2015 | 1,223 | 6.0 (4.0–8.2) |
| Cohort 2 | 2019–2020 | 1,777 | 9.7 (7.7–11.6) |
| Difference, cohort 2 − cohort 1 (percentage points) | | | 3.7 (0.8–6.5) |

Weighted with the longitudinal weight; percentile 95% CIs from 1,000 bootstrap resamples. For cohort 2 the interval covers the first year of the COVID-19 pandemic. Descriptive comparison only

**Table S13** Sensitivity of the typology to distance measures, number of pathways, method, state definitions and cohort

| Analysis | k (by rule) | ASW | ARI with main typology | Medoid sequences with % of participants |
|---|---|---|---|---|
| SA1 Hamming distance | 4 (4) | 0.665 | 0.990 | OOOOO 27.7; RRRRR 44.4; SSSSS 6.7; NNNNN 21.2 |
| SA2 OM, transition-rate costs[a] | 4 (4) | 0.687 | 0.959 | OOOOO 26.5; RRRRR 45.3; SSSSS 6.1; NNNNN 22.1 |
| SA3 k = 3 | 3 | – | 0.915 | OOOOO 31.2; RRRRR 45.9; NNNNN 22.9 |

| | | | | |
|---|---|---|---|---|
| SA3 k = 5 | 5 | – | 0.702 | OOOOO 27.4; RRRRR 19.7; SSSSS 6.1; NRRRR 27.5; NNNNN 19.3 |
| SA4 Latent class analysis[b] | 8 | – | 0.726 | NNRRR 6.6; RRRRR 35.4; SSSSS 5.0; RRSSS 2.1; OORRR 4.7; NNNNN 16.2; OOOOO 23.9; RRRNN 6.0 |
| SA9 Three states[c] | 6 (6) | 0.666 | 0.259 | EEEEE 59.2; EEENN 5.5; UEEEE 4.2; NEEEE 10.7; NUNEE 4.4; NNNNN 16.0 |
| SA9 Two states[c] | 7 (7) | 0.760 | 0.239 | EEEEE 53.1; EEEEX 6.1; EEXXX 5.6; EXEEE 3.4; XEEEE 11.0; XXEEE 4.9; XXXXX 15.9 |
| SA10 Cohort 1 clustered separately[d] | 4 | 0.723 | 0.992 | OOOOO 31.4; RRRRR 47.2; SSSSS 6.4; NNNNN 14.9 |
| SA10 Cohort 2 clustered separately[d] | 4 | 0.666 | 0.997 | OOOOO 25.5; RRRRR 43.1; SSSSS 6.5; NNNNN 24.9 |

Main typology: k = 4, ASW 0.69. k (by rule): number of pathways chosen by the pre-specified rule when the analysis repeated the selection; SA1 and SA2 report the solution with the main k. ARI, adjusted Rand index between the analysis and the main typology for the same participants. Percentages are unweighted. [a] Substitution costs: O–R 1.95; O–S 1.99; O–U 1.94; O–N 1.93; R–S 1.94; R–U 1.49; R–N 1.74; S–U 1.95; S–N 1.94; U–N 1.67; insertion–deletion cost half the largest substitution cost. [b] 3–8 classes, 50 random starts each, chosen by the Bayesian information criterion (BIC: 3 classes 36,167; 4 classes 33,291; 5 classes 32,733; 6 classes 32,459; 7 classes 32,371; 8 classes 32,339); sequences shown are the most typical sequence of each class and percentages are estimated class sizes. [c] E, employed (O, R or S); U, unemployed; N, not in the labour force; X, not employed (U or N). [d] ARI computed within the cohort. O, working for the employer at injury; R, working for a different employer; S, self-employed or unpaid family worker; U, unemployed; N, not in the labour force

**Table S14** Pathway prevalence when all participants are included (SA6)

| | Cohort 1 (n = 2,000) | Cohort 2 (n = 3,294) |
|---|---|---|
| **A. Missed interviews as a sixth state (M); pathways re-selected by the pre-specified rule** | | |
| Medoid OOOOO | 529 (28.1) | 731 (25.0) |
| Medoid RRRRR | 763 (38.2) | 1,118 (35.1) |
| Medoid RMMMM | 309 (15.2) | 469 (15.7) |
| Medoid SSSSS | 108 (6.1) | 185 (6.0) |
| Medoid NNNNN | 291 (12.3) | 791 (18.1) |
| **B. Multiple imputation of missed interviews (m = 20); main pathways** | | |
| Sustained work for the employer at injury | 31.8 (28.4–35.2) | 28.6 (26.2–31.0) |
| Work for a different employer | 49.5 (45.9–53.2) | 46.1 (43.5–48.8) |
| Self-employment | 6.8 (5.0–8.7) | 7.2 (5.8–8.6) |
| Persistent non-employment | 11.8 (9.6–14.0) | 18.1 (16.1–20.0) |

Panel A: unweighted n (percentage weighted with the first-interview weight). Panel B: weighted percentage (95% CI) combined over imputations with Rubin's rules; each imputed sequence was assigned to the nearest medoid of the main typology. M, missed interview; O, working for the employer at injury; R, working for a different employer; S, self-employed or unpaid family worker; U, unemployed; N, not in the labour force

**Table S15** Cross-classification of the solutions with three and five pathways (SA3) with the main typology (number of participants)

| | Sustained work for the employer at injury | Work for a different employer | Self-employment | Persistent non-employment |
|---|---|---|---|---|
| **Solution with k = 3** | | | | |
| Pathway with medoid OOOOO | 1,131 | – | 146 | – |
| Pathway with medoid RRRRR | – | 1,827 | 48 | – |
| Pathway with medoid NNNNN | – | – | 57 | 878 |
| **Solution with k = 5** | | | | |
| Pathway with medoid OOOOO | 1,119 | – | – | – |
| Pathway with medoid RRRRR | – | 807 | – | – |
| Pathway with medoid SSSSS | – | – | 250 | – |
| Pathway with medoid NRRRR | 12 | 1,020 | – | 89 |
| Pathway with medoid NNNNN | – | – | – | 789 |

–, fewer than five participants (including none). O, working for the employer at injury; R, working for a different employer; S, self-employed or unpaid family worker; U, unemployed; N, not in the labour force

**Table S16** Pathway membership and the state in which workers spent the most years (post hoc; number of participants)

| Pathway (main typology) | O | R | S | N (rule A); U or N (rule B) |
|---|---|---|---|---|
| **Rule A: state among O, R, S and N with the most years** | | | | |
| Sustained work for the employer at injury | 1,131 | – | – | – |

| | | | | |
|---|---|---|---|---|
| Work for a different employer | – | 1,827 | – | – |
| Self-employment | – | – | 251 | – |
| Persistent non-employment | – | – | 11 | 867 |
| **Rule B: as rule A, with U and N counted together** | | | | |
| Sustained work for the employer at injury | 1,116 | – | – | 15 |
| Work for a different employer | – | 1,744 | – | 83 |
| Self-employment | – | – | 250 | – |
| Persistent non-employment | – | – | 5 | 873 |

Columns give the state with the most years at the five annual interviews; ties were assigned in the order O, R, S, N. Agreement with the main typology: rule A, 4,076 of 4,087 workers (99.7%; all 11 discordant workers had equally many years in two states); rule B, 97.5%. –, fewer than five participants (including none). O, working for the employer at injury; R, working for a different employer; S, self-employed or unpaid family worker; U, unemployed; N, not in the labour force

**Table S17** Unadjusted average marginal effects of factors at injury on the probability of each pathway (post hoc; percentage points, 95% CI)

| Factor (reference) | Sustained work for the employer at injury | Work for a different employer | Self-employment | Persistent non-employment |
|---|---|---|---|---|
| Female (ref. male) | −4.4 (−7.8 to −0.9) | −0.8 (−4.8 to 3.1) | −4.5 (−5.9 to −3.0) | 9.7 (6.1–13.3) |
| **Age group, years (ref. ≤39)** | | | | |
| 40–49 | 4.8 (0.1–9.5) | −3.3 (−8.3 to 1.7) | −2.7 (−5.4 to 0.0) | 1.2 (−1.7 to 4.1) |
| 50–59 | −3.0 (−7.1 to 1.2) | −3.4 (−7.9 to 1.1) | −3.7 (−6.1 to −1.3) | 10.1 (7.1–13.0) |
| ≥60 | −17.0 (−21.1 to −12.9) | −12.8 (−17.6 to −8.1) | −4.3 (−6.8 to −1.8) | 34.1 (30.5–37.8) |
| **Education (ref. college or higher)** | | | | |
| Middle school or less | −18.1 (−22.1 to −14.0) | −1.6 (−6.0 to 2.9) | −0.9 (−3.1 to 1.3) | 20.5 (17.2–23.7) |
| High school | −4.5 (−8.7 to −0.3) | −1.5 (−5.9 to 2.8) | −0.5 (−2.7 to 1.6) | 6.6 (3.7–9.5) |
| Occupational disease (ref. injury) | 11.2 (5.3–17.1) | −14.8 (−20.4 to −9.1) | −2.7 (−5.1 to −0.4) | 6.3 (0.9–11.7) |
| **Duration of medical care (ref. ≤6 months)** | | | | |
| >6–12 months | −7.6 (−10.7 to −4.5) | 0.2 (−3.3 to 3.6) | 0.2 (−1.5 to 1.9) | 7.2 (4.5–9.9) |
| >12 months | −16.0 (−19.6 to −12.4) | −19.9 (−24.2 to −15.7) | −2.2 (−4.2 to −0.3) | 38.1 (33.7–42.5) |
| **Disability grade (ref. none)** | | | | |
| Grades 10–14 | −3.4 (−7.1 to 0.4) | 1.4 (−2.7 to 5.4) | −0.8 (−2.9 to 1.2) | 2.8 (0.0 to 5.7) |
| Grades 1–9 | −13.0 (−17.4 to −8.7) | −14.9 (−19.7 to −10.0) | −2.2 (−4.6 to 0.2) | 30.1 (25.8–34.4) |
| **Employment status at injury (ref. regular)** | | | | |
| Temporary or other | −25.9 (−29.6 to −22.3) | 14.9 (10.1–19.6) | 0.8 (−1.4 to 3.1) | 10.2 (6.1–14.4) |
| Daily | −30.2 (−32.8 to −27.6) | 17.2 (13.8–20.6) | 2.5 (0.7–4.2) | 10.5 (7.6–13.4) |
| **Firm size at injury, employees (ref. ≥300)** | | | | |
| <5 | −34.4 (−40.1 to −28.7) | 18.7 (13.0–24.3) | 6.9 (4.1–9.8) | 8.8 (4.0–13.5) |
| 5–29 | −32.0 (−37.2 to −26.8) | 22.8 (18.0–27.6) | 2.6 (0.5–4.7) | 6.6 (2.6–10.6) |
| 30–299 | −25.2 (−30.8 to −19.6) | 18.4 (13.1–23.6) | 0.9 (−1.2 to 3.1) | 5.9 (1.6–10.2) |
| **Trade union at workplace (ref. no)** | | | | |
| Yes | 42.2 (37.0–47.3) | −30.6 (−34.9 to −26.4) | −3.3 (−5.2 to −1.3) | −8.3 (−12.1 to −4.4) |
| Unknown or not asked | −13.3 (−17.2 to −9.4) | 2.0 (−3.9 to 7.9) | 3.2 (−0.2 to 6.7) | 8.0 (2.7–13.4) |
| **Industry (ref. manufacturing)** | | | | |
| Construction | −27.9 (−31.0 to −24.8) | 17.1 (13.4–20.9) | 0.7 (−1.1 to 2.6) | 10.1 (6.9–13.2) |
| Transport | 3.8 (−4.3 to 12.0) | −5.5 (−13.3 to 2.4) | 3.9 (−0.9 to 8.6) | −2.2 (−8.0 to 3.6) |
| Other | −8.9 (−12.4 to −5.4) | 1.5 (−2.2 to 5.1) | 0.1 (−1.7 to 1.9) | 7.4 (4.4–10.4) |
| **Job tenure at injury, years (ref. ≥5)** | | | | |
| <1 | −49.6 (−53.5 to −45.7) | 37.3 (33.9–40.7) | 3.7 (2.0–5.3) | 8.6 (5.4–11.9) |
| 1–5 | −25.4 (−30.2 to −20.5) | 21.7 (17.4–26.0) | 2.6 (0.6–4.7) | 1.0 (−2.7 to 4.7) |
| Married at injury (ref. not married) | 10.6 (7.8–13.4) | −9.8 (−13.2 to −6.5) | 1.5 (0.0 to 3.0) | −2.3 (−5.1 to 0.5) |
| Cohort 2 (ref. cohort 1) | −6.0 (−8.9 to −3.1) | −4.3 (−7.5 to −1.2) | 0.4 (−1.1 to 1.9) | 9.9 (7.4–12.4) |

n = 4,087. Each factor was entered alone in a multinomial logistic regression, so each effect equals the difference in observed proportions between the factor level and the reference level; the effects in a row sum to zero. CIs use robust standard errors and the delta method. Adjusted effects are shown in Table 3 of the main article

**Table S18** Full and reduced multinomial models: self-employment and persistent non-employment (post hoc; 95% CI)

| Factor | Pathway | Full model (27 parameters per equation) | Reduced model (17 parameters per equation) |
|---|---|---|---|
| **A. Average marginal effects, percentage points** | | | |
| Age group, years: 40–49 | Self-employment | −4.4 (−8.0 to −0.8) | −2.9 (−5.8 to 0.0) |
| | Persistent non-employment | 0.6 (−2.7 to 3.8) | −0.3 (−3.4 to 2.9) |
| Age group, years: 50–59 | Self-employment | −6.5 (−10.1 to −3.0) | −4.2 (−6.9 to −1.6) |
| | Persistent non-employment | 7.5 (4.3–10.8) | 6.8 (3.8–9.9) |
| Age group, years: ≥60 | Self-employment | −7.6 (−11.4 to −3.8) | −4.8 (−7.5 to −2.0) |
| | Persistent non-employment | 27.4 (23.0–31.8) | 27.8 (24.0–31.7) |
| Duration of medical care: >6–12 months | Self-employment | 0.7 (−1.1 to 2.5) | 0.6 (−1.2 to 2.4) |

| | | | |
|---|---|---|---|
| | Persistent non-employment | 4.4 (1.8–7.1) | 4.1 (1.4–6.7) |
| Duration of medical care: >12 months | Self-employment | −1.0 (−3.6 to 1.6) | −0.9 (−3.6 to 1.8) |
| | Persistent non-employment | 26.2 (21.6–30.7) | 26.4 (21.8–31.0) |
| Disability grade: Grades 10–14 | Self-employment | −0.4 (−2.6 to 1.7) | −0.5 (−2.7 to 1.6) |
| | Persistent non-employment | 0.1 (−3.1 to 3.4) | 0.1 (−3.1 to 3.3) |
| Disability grade: Grades 1–9 | Self-employment | −1.1 (−4.0 to 1.8) | −1.2 (−4.2 to 1.7) |
| | Persistent non-employment | 12.9 (8.3–17.4) | 13.1 (8.5–17.7) |
| Employment status at injury: Temporary or other | Self-employment | 1.1 (−1.3 to 3.5) | 1.5 (−1.0 to 4.0) |
| | Persistent non-employment | 2.2 (−1.5 to 6.0) | 4.7 (1.2–8.3) |
| Employment status at injury: Daily | Self-employment | 3.3 (0.6–6.1) | 2.2 (0.4–4.0) |
| | Persistent non-employment | 1.2 (−2.2 to 4.7) | 5.2 (2.5–7.8) |
| Firm size at injury, employees: <5 | Self-employment | 5.8 (2.4–9.2) | 5.9 (2.9–8.9) |
| | Persistent non-employment | 4.7 (−0.7 to 10.1) | 6.8 (2.1–11.4) |
| Firm size at injury, employees: 5–29 | Self-employment | 2.0 (−0.7 to 4.6) | 2.1 (−0.2 to 4.4) |
| | Persistent non-employment | 2.2 (−2.6 to 7.1) | 4.0 (0.0 to 8.0) |
| Firm size at injury, employees: 30–299 | Self-employment | 0.5 (−2.1 to 3.0) | 0.7 (−1.7 to 3.2) |
| | Persistent non-employment | 2.4 (−2.5 to 7.3) | 3.5 (−0.7 to 7.8) |
| Female | Self-employment | −4.3 (−5.8 to −2.8) | −4.1 (−5.6 to −2.6) |
| | Persistent non-employment | 8.7 (5.1–12.3) | 10.4 (7.0–13.7) |
| Occupational disease | Self-employment | −1.1 (−4.3 to 2.0) | −1.3 (−4.4 to 1.7) |
| | Persistent non-employment | 6.7 (1.8–11.6) | 6.2 (1.5–10.8) |
| Cohort 2 | Self-employment | 0.6 (−1.0 to 2.2) | 0.3 (−1.3 to 1.9) |
| | Persistent non-employment | 5.8 (3.2–8.3) | 5.8 (3.2–8.3) |
| **B. Relative risk ratios (reference: sustained work for the employer at injury)** | | | |
| Age group, years: 40–49 | Self-employment | 0.49 (0.31–0.78) | 0.47 (0.31–0.72) |
| | Persistent non-employment | 0.85 (0.56–1.29) | 0.67 (0.45–0.99) |
| Age group, years: 50–59 | Self-employment | 0.35 (0.22–0.55) | 0.40 (0.27–0.59) |
| | Persistent non-employment | 1.51 (1.02–2.24) | 1.26 (0.88–1.79) |
| Age group, years: ≥60 | Self-employment | 0.45 (0.26–0.80) | 0.62 (0.39–0.98) |
| | Persistent non-employment | 6.30 (4.07–9.76) | 5.67 (3.90–8.23) |
| Duration of medical care: >6–12 months | Self-employment | 1.52 (1.08–2.13) | 1.42 (1.02–1.99) |
| | Persistent non-employment | 1.86 (1.43–2.40) | 1.70 (1.33–2.18) |
| Duration of medical care: >12 months | Self-employment | 1.68 (0.93–3.04) | 1.69 (0.94–3.05) |
| | Persistent non-employment | 7.22 (5.04–10.36) | 7.01 (4.97–9.89) |
| Disability grade: Grades 10–14 | Self-employment | 1.02 (0.69–1.49) | 0.98 (0.67–1.44) |
| | Persistent non-employment | 1.10 (0.81–1.51) | 1.08 (0.80–1.45) |
| Disability grade: Grades 1–9 | Self-employment | 0.98 (0.55–1.76) | 0.98 (0.55–1.73) |
| | Persistent non-employment | 2.39 (1.59–3.60) | 2.42 (1.65–3.55) |
| Employment status at injury: Temporary or other | Self-employment | 2.10 (1.30–3.41) | 3.30 (2.07–5.25) |
| | Persistent non-employment | 2.04 (1.41–2.94) | 3.54 (2.50–5.02) |
| Employment status at injury: Daily | Self-employment | 2.64 (1.67–4.18) | 5.30 (3.75–7.48) |
| | Persistent non-employment | 1.76 (1.26–2.47) | 5.29 (4.03–6.95) |
| Firm size at injury, employees: <5 | Self-employment | 3.67 (1.79–7.52) | 6.07 (3.24–11.36) |
| | Persistent non-employment | 2.05 (1.24–3.40) | 3.90 (2.54–5.98) |
| Firm size at injury, employees: 5–29 | Self-employment | 2.12 (1.07–4.20) | 3.57 (1.98–6.44) |
| | Persistent non-employment | 1.69 (1.07–2.67) | 3.18 (2.18–4.64) |
| Firm size at injury, employees: 30–299 | Self-employment | 1.56 (0.79–3.08) | 2.32 (1.24–4.36) |
| | Persistent non-employment | 1.68 (1.07–2.65) | 2.57 (1.74–3.81) |
| Female | Self-employment | 0.51 (0.30–0.87) | 0.56 (0.33–0.94) |
| | Persistent non-employment | 2.29 (1.70–3.08) | 2.54 (1.94–3.34) |
| Occupational disease | Self-employment | 1.18 (0.58–2.37) | 0.76 (0.38–1.51) |
| | Persistent non-employment | 2.17 (1.48–3.19) | 1.38 (0.96–1.97) |
| Cohort 2 | Self-employment | 1.41 (1.02–1.95) | 1.27 (0.93–1.74) |
| | Persistent non-employment | 1.89 (1.47–2.44) | 1.77 (1.39–2.26) |

n = 4,087. The self-employment equation had 9.3 events per parameter in the full model (251 workers in the self-employment pathway) and 14.8 in the reduced model, which contained age group, duration of care, disability grade, occupational disease, employment status at injury, firm size, sex, cohort and use of rehabilitation services (coefficients not shown). RRRs for employment status and firm size changed more than the average marginal effects because the reduced model omits correlated job characteristics (tenure, trade union, industry, education and marital status)

**Table S19** Reasons recorded for missed interviews (post hoc)

| | Cohort 1 (n = 486) | Cohort 2 (n = 721) |
|---|---|---|
| **Reason recorded at the first missed interview** | | |
| Died | 15 (3.1) | 50 (6.9) |
| Ill, disabled or in an institution | 10 (2.1) | 20 (2.8) |
| Refused or too busy | 236 (48.6) | 296 (41.1) |
| Not contacted or moved | 161 (33.1) | 321 (44.5) |
| Other or not recorded | 64 (13.2) | 34 (4.7) |
| Recorded as having died at any missed interview | 16 (3.3) | 54 (7.5) |

Participants who missed at least one of the five interviews; n (%). Interviewers recorded a reason for each missed interview. Refused or too busy: refusal, distrust of the survey or busy (including requests for another visit); ill, disabled or in an institution; not contacted or moved: absent, long absence, moved (including to Jeju Island or another island), address or telephone number unknown, no answer or no household member able to respond. Continuous respondents, who form the main sample, were alive at the fifth interview. $<5$, fewer than five participants

## S3 Supplementary figure

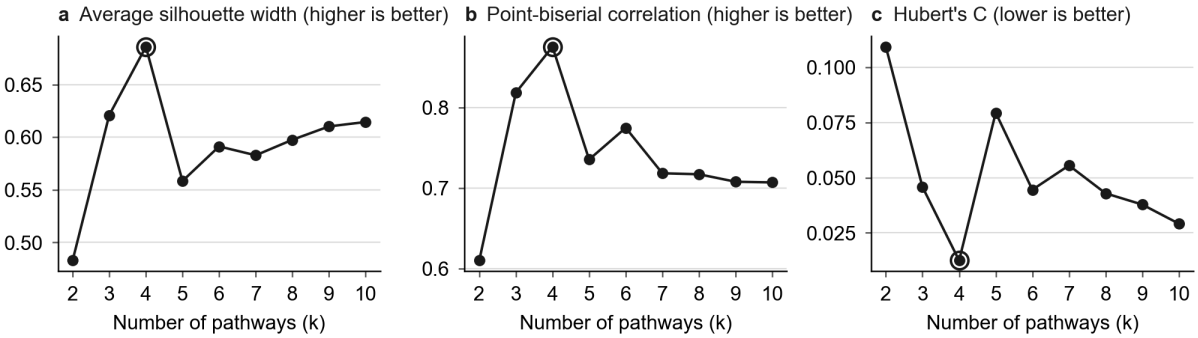


**Fig. S1** Cluster quality indices for solutions with 2 to 10 pathways: average silhouette width (a) and point-biserial correlation (b), for which higher is better, and Hubert's C (c), for which lower is better. The circled point marks the solution selected by the pre-specified rule (k = 4). Source data: Table S3