Online Resource 6. Sensitivity and robustness analyses for the age–period–cohort model

Article: Age–period–cohort effects on suicide mortality in Andalusia, Spain (2000–2024): demographic masking and sustained pandemic excess

Journal: Social Psychiatry and Psychiatric Epidemiology

Authors: Carlos Romero-Olóriz[1], Esperanza López Hidalgo[2], María Victoria Villalba Soria[2], Miguel Guerrero Díaz[3]*

[1] Andalusian Emergency Health Centre 061 (CES-061), Provincial Service of Malaga, Malaga, Spain. [2] Institute of Legal Medicine and Forensic Sciences of Malaga, Malaga, Spain. [3] Clinical Management Unit of Mental Health, University Hospital Virgen de la Victoria, Malaga, Spain.

*** Corresponding author:** Miguel Guerrero Díaz. Coordinator of Community Mental Health Unit Marbella, Centro de Salud Las Albarizas, Avenida Las Albarizas s/n, 29600-Marbella (Malaga), Spain. Email: miguel.guerrero.diaz.sspa@juntadeandalucia.es. ORCID: 0000-0002-3066-5930

This Online Resource presents the results of sensitivity and robustness analyses conducted to evaluate the stability of the age–period–cohort (APC) model estimates reported in the main text. Five complementary approaches are described: (1) assessment and correction of temporal autocorrelation in trend estimates; (2) sensitivity of APC effects to alternative age-range specifications; (3) residual diagnostics for the intrinsic estimator model; (4) assessment of overdispersion in suicide counts; and (5) Monte Carlo permutation validation of joinpoint locations. Together, these analyses confirm that the principal findings are robust to key methodological choices.

1. Temporal autocorrelation assessment

Log-linear trend models may be affected by autocorrelation in the residuals, which could bias standard errors and confidence intervals. To evaluate this, we computed Durbin–Watson (DW) statistics and Ljung–Box Q tests (5 lags) on OLS residuals, and compared OLS estimates with Prais–Winsten (PW) generalised least-squares estimates that correct for first-order autocorrelation (AR(1)).

The DW statistics indicated mild positive autocorrelation for both sexes combined (DW = 0.984) and for men (DW = 0.972), while women showed acceptable values (DW = 1.543). The Ljung–Box Q test was significant only for men (p = 0.018). Despite this, the Prais–Winsten correction produced negligible changes in point estimates: the maximum absolute difference in EAPC between OLS and PW was 0.053 percentage points (men), and all differences were below 0.06 pp. These results confirm that autocorrelation does not materially affect the trend estimates reported in the main text.

2. Sensitivity of APC period effects to age-range specification

The main APC model used ages 40–84 (I = 9 age groups, J = 5 periods, K = 13 cohorts). To assess the sensitivity of period effects to the choice of age range, we re-estimated the intrinsic estimator under two alternative specifications: a restricted model excluding the oldest group (40–79; I = 8, K = 12) and an extended model including younger adults (30–84; I = 10, K = 14). Table S7a presents the period rate ratios for both sexes combined under each specification, along with model fit indices.

Table S7a. Period effects (rate ratios) of the APC intrinsic estimator model under three age-range specifications, both sexes combined

Period	Main model (40–84)	Restricted (40–79)	Extended (30–84)
R^2	0.983	0.970	0.985
Age groups (I)	9	8	10
Cohorts (K)	13	12	14
2000–2004	1.028	1.023	1.024
2005–2009	1.057	1.056	1.047
2010–2014	1.014	1.009	1.007
2015–2019	0.890	0.893	0.890
2020–2024	**1.020**	**1.028**	**1.042**
Pearson r vs main	—	0.997	0.984

RR: rate ratio from the intrinsic estimator (IE), centred at RR = 1.000 (grand mean). R^2: proportion of variance in log-rates explained by the APC model. Pearson r: correlation between period RR vectors of the main model and each alternative specification. The period nadir at 2015–2019 was replicated across all three specifications (RR = 0.890–0.893).

The three specifications yielded virtually identical period effect patterns. The period nadir at 2015–2019 was consistently identified across all models (RR range: 0.890–0.893), and the Pearson correlation between the main and alternative period RR vectors exceeded

0.98 in both cases. Model fit was high under all specifications (R^2 = 0.970–0.985). These results indicate that the reported period effects are insensitive to the choice of age range in the Lexis table.

Sex-stratified analyses showed the same pattern of stability: the period nadir was located at 2015–2019 for both men (RR = 0.879 in the main model) and women (RR = 0.926), and alternative specifications did not alter these conclusions qualitatively (Figure S16).

3. Assessment of overdispersion

Suicide counts may exhibit extra-Poisson variability (overdispersion), which inflates type I error rates if unaccounted for. We assessed overdispersion by computing the variance-to-mean ratio of annual death counts (ϕ = Var/Mean), testing its significance with a chi-squared goodness-of-fit test, and reporting the variance inflation factor ($\sqrt{\phi}$) by which standard errors should be multiplied.

Table S7b. Assessment of overdispersion in annual suicide counts, Andalusia 2000–2024

	Mean (deaths/yr)	Variance	ϕ (Var/Mean)	χ^2	p-value	$\sqrt{\phi}$
Both sexes	734.0	4,652.0	6.338	152.11	< 0.001	2.518
Men	568.2	3,047.0	5.362	128.69	< 0.001	2.316
Women	165.8	417.1	2.516	60.39	< 0.001	1.586

ϕ: dispersion index (variance-to-mean ratio); values > 1 indicate overdispersion. χ^2: chi-squared goodness-of-fit test for equidispersion (df = 24). $\sqrt{\phi}$: variance inflation factor for correcting standard errors under quasi-Poisson assumptions.

Significant overdispersion was detected in all three strata (p < 0.001), with the dispersion index ranging from 2.52 (women) to 6.34 (both sexes combined). This indicates that year-to-year variability in suicide counts exceeds what would be expected under a Poisson distribution, likely reflecting unmeasured period-specific factors. Standard errors of rate estimates should be inflated by the factor $\sqrt{\phi}$ (1.59–2.52) when computing confidence intervals under Poisson assumptions. The APC model, which accounts for period effects explicitly, substantially absorbs this extra-Poisson variability (model R^2 > 0.90 in all strata).

4. Monte Carlo permutation validation of joinpoint locations

The significance of joinpoints identified by BIC-based model selection (main text) was validated through Monte Carlo permutation tests (4,999 random permutations). Under the null hypothesis of a single linear trend (0 joinpoints), the F-statistic of the observed segmented model was compared against the permutation distribution.

Table S7c. Monte Carlo permutation tests for joinpoint significance in age-standardised suicide rates, Andalusia 2000–2024

Sex	Alternative hypothesis	Joinpoint location(s)	F statistic	Permutation p-value	Result
Both sexes	1 joinpoint	2018	14.185	0.003	**Significant**
Both sexes	2 joinpoints	2019, 2020	10.402	0.007	**Significant**
Men	1 joinpoint	2019	15.041	0.003	**Significant**
Men	2 joinpoints	2019, 2021	10.791	0.005	**Significant**
Women	1 joinpoint	2016	5.122	0.148	Not significant
Women	2 joinpoints	2019, 2020	4.503	0.234	Not significant

H_0: 0 joinpoints (single linear trend); H_1: model with k joinpoints. Permutation p-values based on 4,999 Monte Carlo replicates. Joinpoint locations correspond to the year of maximum likelihood under each alternative hypothesis.

The permutation tests confirmed the presence of at least one significant joinpoint for both sexes combined (p = 0.003) and for men (p = 0.003), consistent with the abrupt rate increase around 2019–2020 identified by BIC-based selection. For women, neither the one-joinpoint (p = 0.148) nor the two-joinpoint (p = 0.234) model reached significance, consistent with the absence of a clear inflection point in age-standardised rates among women. These results reinforce the sex-differential pandemic impact described in the main text.

5. Residual diagnostics for the APC model

The adequacy of the APC intrinsic estimator model was evaluated through inspection of standardised deviance residuals across the 45 Lexis cells (9 age groups × 5 periods) in each sex stratum. Residuals were assessed for normality (Shapiro–Wilk test), the proportion of cells exceeding $|z| > 2$ (expected ≤5% under normality), and the spatial pattern of residuals across the age–period plane.

The Shapiro–Wilk test did not reject normality in any stratum (p = 0.144, 0.921, and 0.357 for both sexes, men, and women, respectively). The proportion of cells with $|z| > 2$ ranged from 4.4% to 6.7%, within the expected range. The largest residual was observed in the cell 45–49 years / 2000–2004 for both sexes combined (z = 3.29), representing a localised underprediction. No systematic spatial clustering of residuals was observed (Figure S14), supporting the assumption that the APC model adequately captures the age–period–cohort structure of the data.

Figures

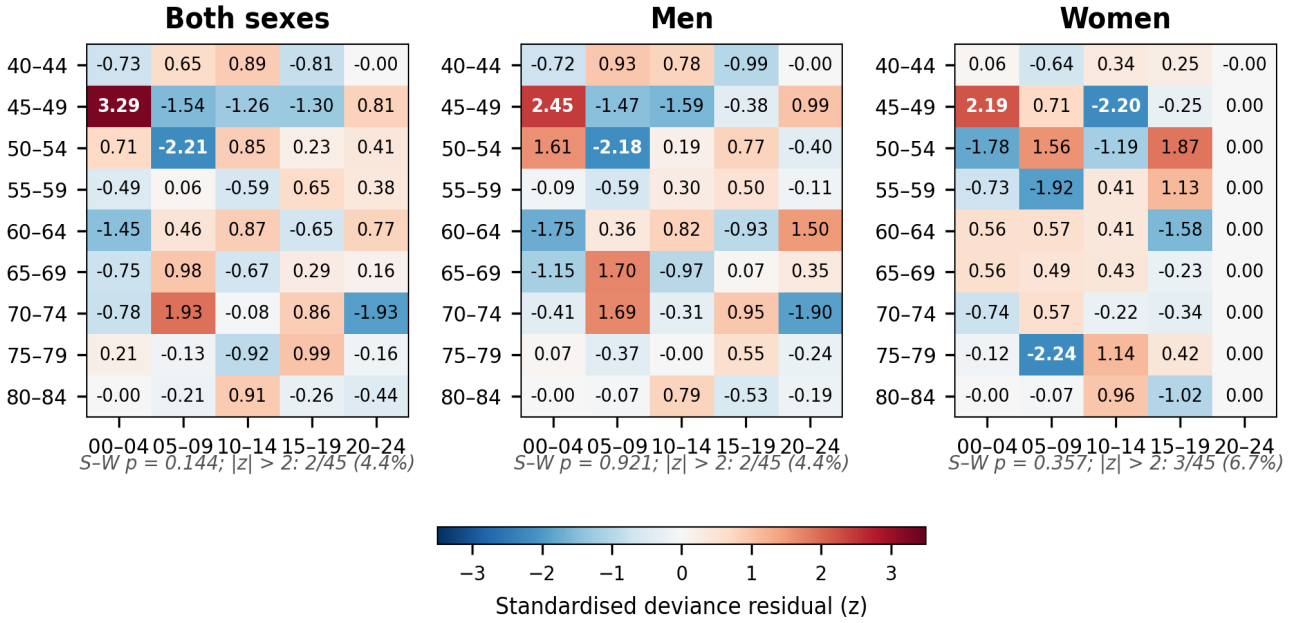

Figure S14. Standardised deviance residuals of the age–period–cohort intrinsic estimator model across 45 Lexis cells (9 age groups × 5 periods), by sex. Colour scale indicates the magnitude and direction of residuals (blue = overprediction; red = underprediction). Values in bold exceed $|z| > 2$. Shapiro–Wilk (S–W) p-values and the proportion of cells exceeding $|z| > 2$ are reported below each panel. All three strata show approximately normal residual distributions with no systematic spatial clustering.

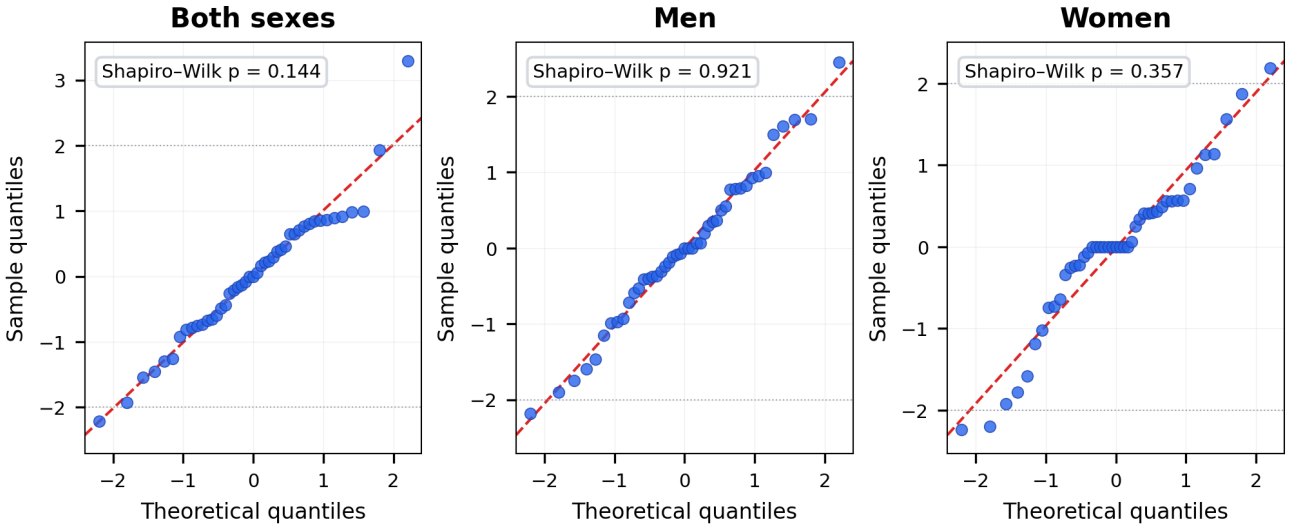

Figure S15. Normal Q–Q plots of standardised deviance residuals from the APC intrinsic estimator model, by sex (n = 45 cells per panel). Dashed red line: theoretical normal reference. Dotted grey lines: ±2 standard deviation thresholds. Shapiro–Wilk p-values are reported in each panel. The close alignment of sample quantiles to the theoretical line confirms approximate normality of residuals in all three strata.

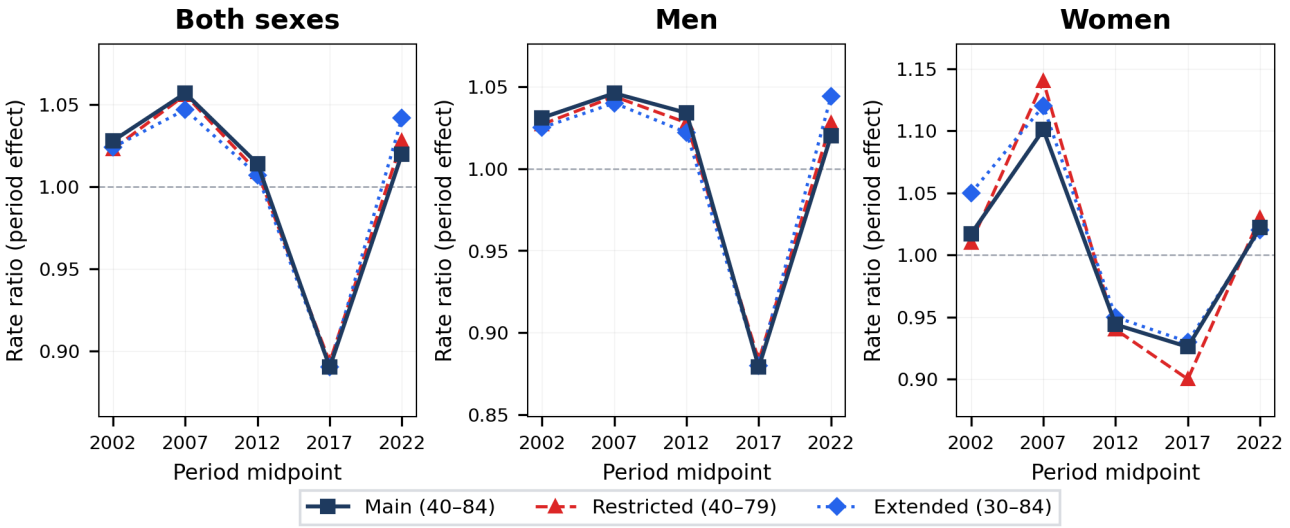

Figure S16. Period effects (rate ratios) from the APC intrinsic estimator model under three age-range specifications, by sex. Main model: 40–84 years (9 age groups); restricted model: 40–79 years (8 age groups, excluding 80–84); extended model: 30–84 years (10 age groups, including 30–39). RR = 1.000 indicates the grand mean. The period nadir at 2015–2019 is consistently identified across all specifications and sex strata, with Pearson correlations between period RR vectors exceeding 0.98.

Summary

All sensitivity and robustness analyses converge on the same conclusions: (1) temporal autocorrelation does not materially affect trend estimates (maximum EAPC difference between OLS and Prais–Winsten: 0.053 pp); (2) APC period effects are insensitive to age-range specification (period nadir RR = 0.890–0.893 across three models, Pearson $r > 0.98$); (3) model residuals are approximately normally distributed with no influential outliers (maximum $|z| = 3.29$, Shapiro–Wilk $p \geq 0.144$); (4) significant overdispersion in annual counts ($\phi = 2.5–6.3$) is effectively captured by the APC period effects; and (5) Monte Carlo permutation tests confirm the significance of joinpoints for both sexes combined and for men ($p \leq 0.007$), while the absence of significant joinpoints for women ($p \geq 0.148$) is consistent with their more stable secular trend. These findings collectively support the robustness of the principal results reported in the main manuscript.