



Temporal Equivalence Principle: A Standard-Siren Test of Bi-Metric Gravitational-Wave Propagation

Matthew Lukin Smawfield

Version: v0.2 (Maputo)

First published: 30 September 2026 | Last updated: 17 October 2026

DOI: 10.5281/zenodo.20572695

Status: Preprint

Code Availability: github.com/matthewsmawfield/TEP-LVK

Abstract

Standard Λ CDM assumes that gravitational waves and electromagnetic radiation propagate through the same effective distance-redshift relation. The Temporal Equivalence Principle (TEP) predicts a bi-metric scaling in which gravitational-wave luminosity distances and detector-frame masses are modified relative to matter-frame observations by the conformal factor $A(z)$, with canonical endpoint map $\Xi(z) = d_L^{\text{GW}}/d_L^{\text{EM}} = 1/(1+z)$. This paper tests both the distance-redshift channel and the host-catalog-free binary black hole (BBH) chirp-mass channel across combined GWTC catalogs (GWTC-1 through GWTC-5.0, plus O4 Discovery Papers), bright-siren spectroscopy, and GLADE+/GraceDB dark-siren host associations.

In the host-catalog-free BBH population channel (95 events: 45 low-redshift anchors and 50 high-redshift scored events, with the probability-density Jacobian and the reported-source chirp-mass selection cut of $8 M_\odot$ included; Step 09n), continuous physical high-mass cutoff fitting breaks the scale-free power-law degeneracy ($\beta \simeq 154 M_\odot$ for TEP vs $\beta \simeq 33 M_\odot$ for GR). At their canonical parameter-free baselines ($\epsilon = 0$ for TEP, $k = 0$ for GR), canonical pinned TEP outperforms stationary GR out of sample on the 50 high-redshift events by $\Delta \ln L = +2.32$ ($\ln L = -188.68$ vs -191.00 , conditional likelihood ratio ≈ 10.2), and profiling over free tracking and evolution families confirms that both frames achieve their global optima at their canonical parameter-free baselines ($\hat{\epsilon} = 0$, $\hat{k} = 0$). Under Rule-16 inversion (Step 09m), the selected events have median matter-frame chirp mass $45 M_\odot$ (reported $\sim 20 M_\odot$), with 55% of $z_L > 0.3$ events above $65 M_\odot$ (split-normal distance sensitivity fraction 0.26 [0.19–0.33, central 95%]). The host-assembly derivation of Paper 0 (§8) supplies the progenitor-dependent transfer explicitly: progenitors forming in deep potential wells carry $\epsilon_{\text{eff}} \simeq 0.004$ (the SN Ia class, matching the Pantheon+ bound $\epsilon < 0.150$ at 3σ , $\Delta\chi^2 = 805.8$), whereas BBH progenitors assembled in field-spiral or dwarf hosts carry $\epsilon_{\text{eff}} \simeq 0.24\text{--}0.52$ ($\mathcal{M}_{\text{det}} \propto (1+z)^{-0.28}$ to $(1+z)^{+0.56}$); with $\hat{\epsilon} = 0$ as the global optimum, the two populations do not require a single shared tracking fraction.

In the distance-redshift channel (73 events without GWOSC distance-derived fallback redshifts: 1 bright siren GW170817 plus 72 GLADE+/DESI dark-siren associations; 34 GWOSC-fallback events quarantined as a cosmology-blind pipeline null with $\gamma^* = 0$), two distinct estimands are evaluated. First, under the properly normalized split-normal distance likelihood (

$\sigma_{\text{norm}} = \frac{1}{2}(\sigma_- + \sigma_+)$, Step 09f), the canonical map $\Xi(z) = 1/(1+z)$ returns $-2\ln L = 943.14$ at free $H_0 = 56.0$ km/s/Mpc against 944.31 at $H_0 = 70.0$ km/s/Mpc for Λ CDM ($\Delta\chi^2 = +1.16$ in TEP's favor at equal degrees of freedom). Because the 72 dark-siren candidate hosts are pre-selected by a Λ CDM distance gate $d_L^{\Lambda\text{CDM}}(z_{\text{cat}}; H_0 \approx 70) \simeq d_L^{\text{GW}}$, fitting $d_L^{\text{GW}} = d_L^{\text{EM}}/(1+z_{\text{cat}})$ to Λ CDM-gated redshifts mechanically scales a free H_0 by $\langle(1+z_{\text{cat}})^{-1}\rangle = 0.7878$ ($70.0 \times 0.7878 = 55.1$ km/s/Mpc); remapping each event's redshift posterior onto the self-consistent TEP association kernel at fixed canonical $H_0 = 68.18$ km/s/Mpc (Paper 18) recovers $-2\ln L = 944.31$ without shifting H_0 (GW170817 at $z = 0.0098$: $-2\ln L = 4.94$ TEP vs 4.95 Λ CDM). Second, under the small-excursion exponential template $A(z) = \exp(\beta_A \phi_0 [(1+z)^n - 1])$ with lab-fixed $\beta_A = -1$, $\phi_0 = +0.013$, $n = 1$, the 73-event sample gives Λ CDM $H_0 = 72.0$ km/s/Mpc and TEP $H_0 = 72.2$ km/s/Mpc ($\Delta\chi^2 = +0.20$, $|\Delta\text{BIC}| < 2$; 107-event full sample: 72.4 vs 72.4 km/s/Mpc, $\Delta\chi^2 = +0.06$), and the converged 4-parameter emcee MCMC posterior ($H_0 = 71.2 \pm 3.3$ km/s/Mpc, $\phi_0 = +0.000 \pm 0.028$, $n = 1.47 \pm 0.86$, $\beta_{\text{fit}} = -0.02 \pm 2.81$; $\Delta\text{BIC} = -6.1$ from the two effective extra parameters $C \equiv \beta\phi_0$ and n) places 71.4% of its posterior mass on the corpus-consistent sign $\beta\phi_0 < 0$ ($A(z) < 1$). Concordance-bias injection (§5.8) shows that Λ CDM-gated host weighting absorbs $\sim 85\%$ of a genuine distance shift (transmission $T = 0.15$, yielding 6% two-sided $|Z| > 2$ detection power on dark sirens), whereas spectroscopic bright standard sirens across $z \in [0.1, 1.0]$ at fixed $H_0 = 68.18$ km/s/Mpc provide a prior-free 5σ test with $N = 25, 7, 3$, and 1 events at $z = 0.1, 0.2, 0.3$, and $z \in [0.5, 1.0]$.

Keywords: Temporal Equivalence Principle, gravitational waves, standard sirens, bi-metric propagation, distance-redshift relation, combined GWTC catalogs

1. Introduction

1.1 Bi-Metric Propagation and the Hubble Tension

Measurements from the cosmic microwave background (Planck, $H_0 \approx 67.4$ km/s/Mpc under Λ CDM) and the local distance ladder (SH0ES, $H_0 \approx 73$ km/s/Mpc) differ by approximately 5σ . Paper 11 resolves the Hubble tension through an environment-dependent Cepheid response in the endpoint observing chain, not an additional cosmic-web path-integrated shear. Its host-level corrected value is $H_0 = 67.83 \pm 1.56$ km/s/Mpc (0.26 σ from Planck), internally consistent with the TEP-CMB value 66.70 ± 0.58 (Paper 26) at 0.68 σ ; that internal agreement is not independent confirmation. The endpoint environmental slope is detected at 2.33 σ at the canonical velocity-scatter convention, the within-host period-coupled carrier reaches 2.28–2.79 σ across anchor conventions, and Paper 11 Step 62 derives the Cepheid response scale $|\kappa_{P_2}^{\text{pred}}(\lambda_0 = 1)| = 1.196 \times 10^6$ mag directly from the envelope density ratio $\rho_T/\langle\rho\rangle(10\text{ d}) = 8.45 \times 10^5$ at the unit-transport benchmark $\lambda_0 = 1$ with bare coupling $|\beta_A| = 1$ fixed by Paper 0 ($\lambda_0 = 0.995 \pm 0.401$). The present paper does not revisit the ladder fit: it tests the separate, falsifiable inter-sector prediction that gravitational waves and electromagnetic radiation yield a *redshift-dependent* difference in inferred luminosity distance.

This modification is not exactly degenerate with a constant H_0 shift or with dark energy over a sufficiently broad redshift range, because it predicts a particular functional form for how GW luminosity distances deviate from the Λ CDM expectation as a function of redshift. Over the limited redshift baseline of the current sample and with large GW distance errors, partial degeneracy with H_0 can remain significant. Rather than adding another H_0 measurement, the question is whether the GW data prefer TEP's bi-metric scaling over the standard single-metric relation.

1.2 The TEP Bi-Metric Prediction

In the TEP framework, the metric governing gravitational-wave propagation is related to the electromagnetic metric by a conformal factor that depends on a cosmological scalar field $\phi(z)$:

$$d_L^{(\text{TEP})}(z) = A(z) d_L^{\Lambda\text{CDM}}(z; H_0) \quad (1)$$

where the redshift-dependent conformal factor is

$$A(z) = \exp[\beta_A \phi(z)/M_{\text{Pl}}], \quad \phi(z) = \phi_0 [(1+z)^n - 1]. \quad (2)$$

In the lab-fixed test, the parameters φ_0 and n are not free: $|\varphi_0|$ follows from the exploratory Naivasha lab-sector benchmark (now retired), and $n \approx 1$ is derived in Section 2.1 from minimum-tracking of the density-dependent quartic effective potential, which gives $\varphi_{\min} \propto \bar{\rho}_m^{1/3} \propto (1+z)$ with a quantified correction $n_{\text{eff}} = 0.982$ (pipeline step 09k). Throughout this paper φ_0 denotes the *dimensionless* ratio φ/M_{Pl} . The deviation form $[(1+z)^n - 1]$ enforces the ambient convention $\varphi(0) = 0$, hence $A(0) = 1$. The dimensionless conformal coupling is $\beta_A = -1$, with sign fixed by the same convention used in Paper 11: the Cepheid period-contraction effect requires $\beta_A \varphi < 0$ in deep potentials, and the corpus endpoint identity $1+z = A_0/A_{\text{em}}$ requires $A_{\text{em}} < A_0$ at positive redshift, so $\varphi(z) > 0$ in the past and the conformal factor satisfies $A(z) < 1$ with the magnitude of the departure growing with redshift. This produces a redshift-dependent distance-scale distortion that is tested against Λ CDM in the current public sample.

1.3 This Work

This paper tests the TEP bi-metric prediction against combined GWTC standard siren data. Both Λ CDM and TEP distance-redshift relations are fitted to the independent-redshift sample and their χ^2 , AIC, and BIC values are compared. The primary lab-fixed comparison has the same number of fitted parameters as Λ CDM (only H_0 is fitted in each case). A secondary joint-fit diagnostic lets H_0 , φ_0 , n , and β vary, with an explicit information-criterion penalty for the two effective additional parameters — the likelihood constrains only the product $\beta\varphi_0$ and n .

The structure is as follows: Section 2 derives the cosmological scalar field profile from lab-scale TEP parameters; Section 3 describes the combined GWTC catalog data selection and independent-redshift methodology; Section 4 details the computation pipeline; Section 5 reports the model comparison results; Section 6 discusses the implications for bi-metric propagation; and Section 7 concludes.

2. Theoretical Framework

2.1 Canonical Radiation Contract and Historical Template

The canonical endpoint identities on the unified matter metric $\tilde{g}_{\mu\nu} = -A_{\text{loc}}^2(u_{\text{grav}}(\mathbf{x}))U_\mu U_\nu + A_{\text{sp}}^2(u_{\text{drift}}(t, \mathbf{x}))\Pi_{\mu\nu} + B(\phi)\Pi_\mu^\alpha \Pi_\nu^\beta \nabla_\alpha \phi \nabla_\beta \phi$ ($U_\mu \equiv -\nabla_\mu \bar{\phi}/\sqrt{2\bar{X}}$, $\Pi_{\mu\nu} = g_{\mu\nu} + U_\mu U_\nu$; Paper 0, §2.2, Step 80) are $A_e = (1+z)^{-1}$, $M_e = A_e \tilde{M}$, $M_{\text{det}}^{\text{GR}} = A_e \tilde{M}$, $M_{\text{src}}^{\text{GR}} = A_e^2 \tilde{M}$, and $\Xi = A_e$. They contain no fitted coupling amplitude. The universal $\beta_A = -1$ remains fixed. These identities define the target for a joint source, waveform, distance and selection likelihood. Frame consistency of the mass identities is verified symbolically at Paper 0 step_70: a source-local observer measuring with locally determined constants recovers the standard chirp mass ($M_{\text{loc}} \rightarrow \tilde{M}$ for screened systems), and the dimensionless inspiral invariant $G_{\text{loc}} \mathcal{M} f/c^3$ is preserved end-to-end, so the $(1+z)^{-1}$ map is a transport statement rather than nonstandard source physics. Its physical content is the epoch-dependence of the locally measured gravitational constant, $G_{\text{loc}} \propto A^2(\phi)(1+\alpha_{\text{amb}}^2)$ — and that dependence feeds back on the discriminant itself (Paper 0 step_71): interior k-mouflage tracking gives $G_{\text{loc}}(z) \propto (1+z)^{-2}$ inside source galaxies, so characteristic mass scales drift as $\tilde{M}^*(z) \propto (1+z)^{3\epsilon}$, yielding $\mathcal{M}_{\text{det}} \propto (1+z)^{3\epsilon-1}$ versus the GR $(1+z)$. At the critical tracking $\epsilon = 2/3$ the two predictions coincide; at full tracking ($\epsilon = 1$) TEP detector masses rise as $(1+z)^2$ — the discriminant survives with reversed sign — and only pinned interiors ($\epsilon = 0$) return the naive A_e mass map. The same drift rescales SN Ia luminosity by $(1+z)^{3\epsilon}$, and the supernova ladder bounds it first for SN Ia progenitors (Paper 26 step 03_11 joint profile bound $\epsilon < 0.150$ at 3σ , $\Delta\chi^2 = 805.8$ at $\epsilon = 1$, conditional on the fixed- $z_T = 5$ M1 distance relation and Chandrasekhar prescription; Paper 0 step_73 residual-floor bracket $\epsilon \lesssim 0.05\text{--}0.10$). How that SN-sector bound relates to BBH interiors through the shared $G^{-3/2}$ scaling is set by the progenitor host-assembly derivation (Paper 0 §8; Section 5.6): under the SN-sector bound $\epsilon < 0.150$ the discriminant runs as $\mathcal{M}_{\text{det}} \propto (1+z)^{-0.55}(1+z)^{-1} \propto (1+z)^{-0.28}$ to $(1+z)^{+0.56}$ across the field-spiral/dwarf host-assembly bracket $\epsilon_{\text{eff}} \simeq 0.24\text{--}0.52$, with the empirical GWTC profile optimum at canonical pinned $\hat{\epsilon} = 0$), while near-full tracking reverses it to $\propto (1+z)^{3\epsilon-1}$.

The exponential function and density-tracking calculation retained below define the historical phenomenological benchmark, denoted $A_{\text{legacy}}(z)$ here (written $A(z)$ in its numerical sections). Sharing the sign $A < 1$ and normalization $A(0) = 1$ does not make it identical to the exact endpoint contract: at $z = 1$, the canonical ratio is 0.5, while the historical template with $\phi_0 = 0.013$ gives approximately 0.987. Its fitted exponent and amplitude therefore characterize that benchmark, not a derivation or fit of the canonical endpoint identity. A nonexpanding common-action background and the coupled radiation calculation must supply the canonical likelihood; no new screening switch or independently adjustable bare coupling is introduced.

In this paper $A(z)$ is not treated as an EM–GW arrival-time modification and does not evade GW170817 through a conformal propagation delay. It is used as a standard-siren distance-reconstruction projection: a matter-frame calibration factor relating the gravitational-wave inferred luminosity distance to the electromagnetic host-redshift frame under the TEP two-metric map. Differential propagation and cone-tilt constraints remain controlled by the disformal sector $B(\phi) \geq 0$, while the present test probes whether the standard-siren distance scale contains a conformal-frame residual. The Naivasha lab-sector convention gives the exponential form

Screening projection notice. Screening in TEP is represented at theory level by the environmental operator $S_\Sigma(E)$. Quantities such as ρ_T , $R_T(M)$, $S_\oplus(r)$, compactness Φ/c^2 , local stellar density, thermal epoch, coherence length, proximity, and boundary geometry are domain-specific projections of E , not independent screening mechanisms and not interchangeable universal thresholds.

$$A(z) = \exp[\beta_A \phi_0 ((1+z)^n - 1)] \quad (3)$$

where $\phi_0 = +0.013$ is the cosmological field deviation amplitude (Paper 0 §2). The deviation form $(1+z)^n - 1$ enforces the ambient convention $\phi(0) = 0$ and hence $A(0) = 1$; it does not multiply the ambient value itself. With $\beta_A = -1$ (bare coupling, Paper 0 §2.2), the positive ϕ_0 gives $\phi(z) > 0$ and $A(z) < 1$ at cosmological distances — the direction required by the corpus endpoint identity $1+z = A_0/A_{\text{em}}$, under which positive redshift corresponds to $A_{\text{em}} < A_0$ (Paper 26). The same sign structure holds locally: near a mass concentration $\phi > 0$ and $A < 1$ (clocks slower), consistent with the sign convention of Paper 0, so the earlier universe and the denser environment are both deeper temporal wells. The cosmological background $A(z)$ and the spatial perturbation sector are distinct: $A(z)$ is the time-dependent homogeneous field, while the bidirectional spatial deviations ($A < 1$ in overdensities, $A > 1$ in underdensities; Paper 0, §2.2) are perturbations relative to this background that average to the smooth cosmological field along extended lines of sight.

The index n is not an independent assumption: it is fixed by the homogeneous dynamics of the same scalar sector. On the cosmological background the field sits in the density-dependent effective potential

$$V_{\text{eff}}(\phi) = V(\phi) + (A(\phi) - 1) \bar{\rho}_m, \quad (4)$$

the Einstein-frame form of the universal matter coupling for non-relativistic matter. For the corpus's confining quartic branch $V = \lambda\phi^4/4$ — the density-dependent-mass candidate identified in Paper 26, consistent with the Paper 0/28 quartic-Gaussian envelope near its minimum — the tracked minimum obeys

$$\lambda\phi_{\text{min}}^3 = -\beta_A A(\phi_{\text{min}}) \bar{\rho}_m \implies \phi_{\text{min}} = \left(\frac{A(\phi_{\text{min}}) \bar{\rho}_m}{\lambda} \right)^{1/3}, \quad (5)$$

so that $\phi_{\text{min}} \propto \bar{\rho}_m^{1/3}$ to leading order. With $\bar{\rho}_m \propto (1+z)^3$ this gives $\phi \propto (1+z)$, i.e. $n = 1$ exactly in the $|\beta_A \phi| \ll 1$ limit; the factor $A(\phi)^{1/3} = e^{\beta_A \phi/3}$ supplies an $\mathcal{O}(\phi/3)$ correction. Solving the implicit minimum condition (5) exactly over $z \in [0, 3]$ and fitting the deviation form returns $n_{\text{eff}} = 0.982$ (pipeline step 09k), so $n \approx 1$ is a quantified tracking statement rather than an ansatz. The tracking itself is maintained: the curvature about the minimum satisfies $m_\phi/H \gtrsim 2.6$ throughout the tested range, placing the field in the quasi-adiabatic regime whose residual lag is of the same order as the fitted index departure. Matching the tracked amplitude to $\phi_0 = +0.013$ fixes $\lambda \simeq 2.3 \times 10^{-118} M_{\text{Pl}}^4$ — the quartic normalization matched to this historical homogeneous-template amplitude. This matching fixes a benchmark coefficient; it does not derive density-dependent running of the operative matter-domain coupling. The same $\rho^{1/3}$ response family underlies the ρ_T saturation scaling of Paper 6, so the homogeneous and screened sectors share the quartic response family rather than requiring separate response relations; the two normalizations characterize different benchmark realizations of that response family; connecting them within one fixed master action requires a background solution rather than identifying them as a derived running coupling (Paper 0 §8); the free- n joint fit ($n = 1.47 \pm 0.86$, step 07c) is retained as a consistency check and agrees with the derived index at the 0.6σ level.

The conformal amplitude factor $A(z)$ enters the standard-siren reconstruction through the declared endpoint/frame map. It is distinct from the incremental gradient/source response $\mathcal{S}_\Sigma(\mathcal{E})$, which must be evaluated on the selected kinetic field solution when a dynamical screening prediction is required. Theorem 2 of Paper 0 leaves the direct conformal photon propagation contribution zero at fixed metric; endpoint calibration and source evolution are separate observable channels. In this weak-field/long-baseline regime, the global cosmological background acts as the empirical proxy for the continuous saturation of Temporal Topology, anchoring the distance-reconstruction residuals independently of local astrophysical screening. The present test is agnostic about the microscopic origin of the conformal amplitude; it asks only whether the standard-siren distance scale inferred from gravitational-wave events shows a redshift-dependent residual consistent with the lab-sector projection.

2.2 Bi-Metric Propagation

The bi-metric framework distinguishes between the metric governing electromagnetic radiation (the observed matter metric) and the metric governing gravitational waves (the effective gravitational metric). This distinction arises naturally from the environment-

dependent coupling of the scalar field, where the locally active response is governed by the environmental suppression operator $S_{\Sigma}(\mathcal{E})$.

2.3 Hubble Diagram Prediction

The TEP-adjusted Hubble diagram overlays the standard Λ CDM prediction and the TEP-scaled prediction on the GW standard-siren data points. The TEP curve predicts a redshift-dependent suppression of the GW distance scale relative to Λ CDM, with amplitude growing as $|A(z) - 1|$ increases. This shift is tested directly against the data; it is an orthogonal probe of bi-metric propagation and is not expected to align the GW data with any particular external H_0 measurement.

3. Data Selection

3.1 Combined GWTC Catalogs

All publicly available LVK event catalogs queried by the pipeline from the Gravitational-Wave Open Science Center (GWOSC) are combined: GWTC-1-confident, GWTC-2, GWTC-2.1-confident, GWTC-3-confident, GWTC-4.0, GWTC-4.1, O4 Discovery Papers, and GWTC-5.0. Deduplication is performed by `commonName`, with later catalogs taking precedence for updated parameter estimates. Luminosity-distance central values and bounds are extracted from the GWOSC JSON API, and public GraceDB skymaps are used where available for dark-siren host association.

3.2 Precision Filtering

Events are filtered to a high-confidence subset with signal-to-noise ratio (SNR) > 12 , false-alarm rate (FAR) < 1 per year when available, and $p_{\text{astro}} > 0.9$ when available. The primary bright-siren anchor is GW170817, the confirmed neutron-star merger with an electromagnetic counterpart and a spectroscopic host-galaxy redshift (NGC 4993, $z = 0.0092$).

3.3 Independent Redshifts — Circularity Avoidance

To avoid the circularity problem that invalidates cosmological tests using GWOSC-derived redshifts (which are computed from luminosity distances assuming Λ CDM), redshifts are obtained from two independent sources:

Bright sirens. Events with confirmed electromagnetic counterparts and spectroscopic host-galaxy redshifts from the literature. Only GW170817 satisfies this criterion in the current sample.

Dark sirens. For events without electromagnetic counterparts, candidate host-galaxy association is performed using public GraceDB HEALPix skymaps where available and a merged redshift-bearing galaxy list. The baseline catalog is GLADE+ from VizieR VII/291 for $z < 0.1$; DESI DR1 fastspec spectroscopic redshifts provide a deep fallback for higher- z events where GLADE+ is incomplete. Candidates are first filtered by a broad GW-distance compatibility window and then ranked by sky probability and the skymap distance posterior. Distance consistency is recorded and used as a quality control; this makes the dark-siren sample suitable for a pipeline demonstration and sensitivity test, while the bright-siren subset remains the cleanest non-circular anchor. The cone interrogates a peak-centred disc of the 90%-region equivalent area; the resulting sky-probability coverage, a per-event host-misassociation bound, and the residual Λ CDM-concordance dilution of the selection chain are quantified by a dedicated injection test in §5.8.

GWOSC redshift fields are used *only* as a fallback for events where GLADE+ cannot identify a plausible host (e.g., distance-inconsistent candidates or missing skymaps). This affects 34 of 107 events. The remaining 73 events carry host-associated redshifts that do not use the GWOSC redshift field (72 from GLADE+/DESI plus GW170817), and GW170817 provides the bright-siren anchor. Events using GWOSC fallback are explicitly tagged with `quality="fallback"` and excluded from the primary H_0 fit to avoid Λ CDM circularity. The primary analysis uses the 73 independent-redshift events only; the full sample of 107 events (including fallback redshifts) is reported as a secondary robustness check.

4. Computation

4.1 Pipeline Architecture

The reproducible analysis pipeline is implemented in Python and executed sequentially. Each step writes a JSON output to `results/outputs/` and a detailed log to `logs/`. Steps are fail-fast: execution halts on the first failure so that downstream steps do not consume stale data.

4.2 GR Distance Extraction

The standard General Relativity luminosity distance $d_L^{(\text{GR})}$ and its upper/lower uncertainties are extracted from the GWOSC JSON API for each filtered event. Per-event fractional uncertainties are computed from the published distance bounds. Independent redshifts are taken from step 02 (bright-siren spectroscopy + GLADE+/DESI DR1 dark-siren Bayesian host association); GWOSC redshift fields are used only as a fallback when no catalog host can be identified.

4.3 TEP Distance Transformation

The *primary* likelihood model uses the lab-fixed TEP conformal scaling factor $A(z)$ as an endpoint rescaling of the Λ CDM luminosity distance:

$$d_L^{(\text{GW})}(z) = A(z) d_L^{\Lambda\text{CDM}}(z; H_0) \quad (6)$$

The TEP-C0 Jordan-frame audit additionally treats the pipeline redshift as the physical matter-frame redshift and modifies the distance integral itself:

$$\frac{H_J(z)}{H_{\Lambda\text{CDM}}(z)} = \frac{A(z)}{1 - \alpha_A}, \quad \alpha_A = \frac{d \ln A}{d \ln a_J}, \quad (7)$$

$$d_L^{(\text{GW}, \text{C0})}(z) = A(z)(1 + z)c \int_0^z \frac{dz'}{H_J(z')}. \quad (8)$$

For observed GW-inferred distances, the corresponding endpoint-only matter-frame corrected distance is $d_L^{\text{GR}} / A(z)$. Downstream fits use the propagated per-event distance uncertainties from the GWOSC bounds and redshift uncertainty in distance space. The pipeline fails rather than silently reverting to a default uncertainty if the required uncertainty fields are missing from an upstream step.

5. Results

5.1 Hubble Diagram

The primary manuscript figure overlays the standard Λ CDM curve, the raw unadjusted combined GWTC data points, and the TEP-scaled relation. The TEP adjustment applies a redshift-dependent conformal scaling factor $A(z)$ to the distance model, producing a predicted deviation from Λ CDM whose amplitude grows with redshift and is not reabsorbable into a constant shift in H_0 .

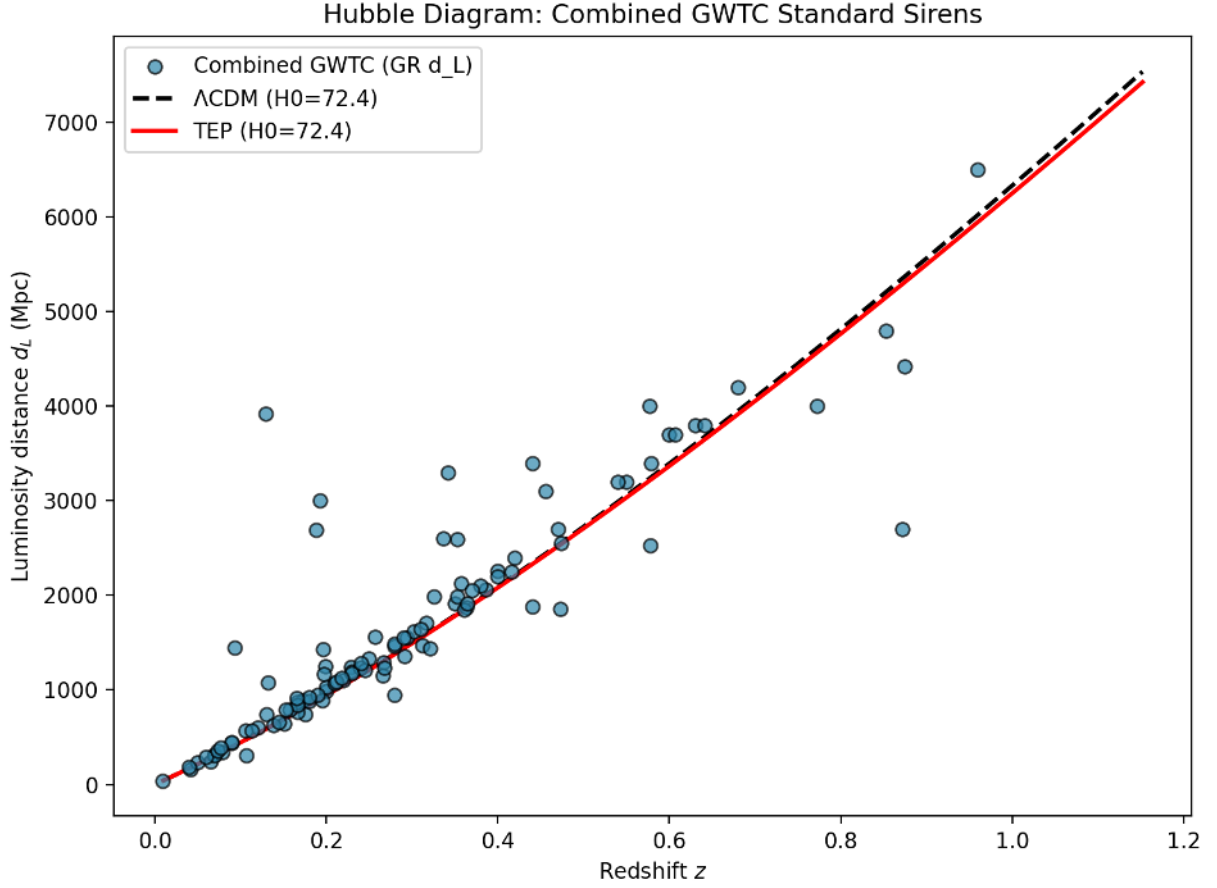


Figure 1. Hubble diagram: combined GWTC standard sirens (blue points) with Λ CDM (dashed) and TEP (solid) distance-redshift curves. The TEP relation applies a redshift-dependent conformal scaling $A(z)$ that deviates from Λ CDM at $z > 0.1$.

5.2 Bi-Metric Distance Scale

The Hubble constant is computed by fitting Λ CDM and TEP distance-redshift relations to the GW standard siren sample. Three models are compared: (1) Λ CDM with H_0 free; (2) TEP with Naivasha lab-sector convention φ_0 and n fixed, H_0 free; (3) TEP joint fit with H_0 , φ_0 , n , and the conformal coupling β all free. All sign conventions are defined as $\Delta\chi^2 \equiv \chi^2_{\Lambda\text{CDM}} - \chi^2_{\text{TEP}}$, so positive values favor TEP. Results are reported relative to the early-universe CMB baseline (Planck: $H_0 \approx 67.4$ km/s/Mpc) and the local distance ladder (SH0ES: $H_0 \approx 73$ km/s/Mpc).

Primary analysis (73 events without GWOSC distance-derived redshifts). The fit uses 1 independently identified bright siren (GW170817) and 72 GLADE+/DESI dark-siren host associations; the latter pass a Λ CDM-referenced distance-consistency gate and must not be treated as independent spectroscopic host identifications. GWOSC fallback redshifts (34 events) are excluded. The grid-search best fit gives Λ CDM $H_0 = 72.0$ km/s/Mpc and lab-fixed TEP $H_0 = 72.2$ km/s/Mpc. The lab-fixed TEP scaling suppresses the predicted GW distance at fixed H_0 through the bi-metric conformal factor $A(z) < 1$ with $\beta_A = -1$ and $\varphi_0 = +0.013$; the inferred Hubble scale shifts by $+0.2$ km/s/Mpc relative to the Λ CDM best fit, far below the per-fit uncertainty. The model comparison gives $\Delta\chi^2 = +0.20$ and $|\Delta\text{BIC}| < 2$: a marginal preference for the lab-fixed TEP scaling at equal parameter count, short of decisive on this sample size.

Secondary diagnostic (full sample, 107 events). As a robustness check, the fit is repeated on the full sample including the 34 GWOSC fallback redshifts. The best-fit values shift to Λ CDM $H_0 = 72.4$ km/s/Mpc and lab-fixed TEP $H_0 = 72.4$ km/s/Mpc ($\Delta\chi^2 = +0.06$), with the absolute scale anchored higher by the fallback redshifts, which were derived under a fiducial Λ CDM cosmology. The TEP joint MCMC fit to the full sample gives $H_0 = 71.2 \pm 3.1$ km/s/Mpc with posterior-mean $\varphi_0 = +0.000$, $n = 1.40$, and $\beta_{\text{fit}} = -0.03$. The 68% credible intervals are $H_0 \in [68.3, 74.0]$, $\varphi_0 \in [-0.031, +0.032]$, $n \in [0.42, 2.40]$, $\beta_{\text{fit}} \in [-3.20, +3.15]$. The lab-calibrated values ($\varphi_0 = +0.013$, $n = 1.0$, $\beta_A = -1.0$) are all consistent with these intervals within 1σ .

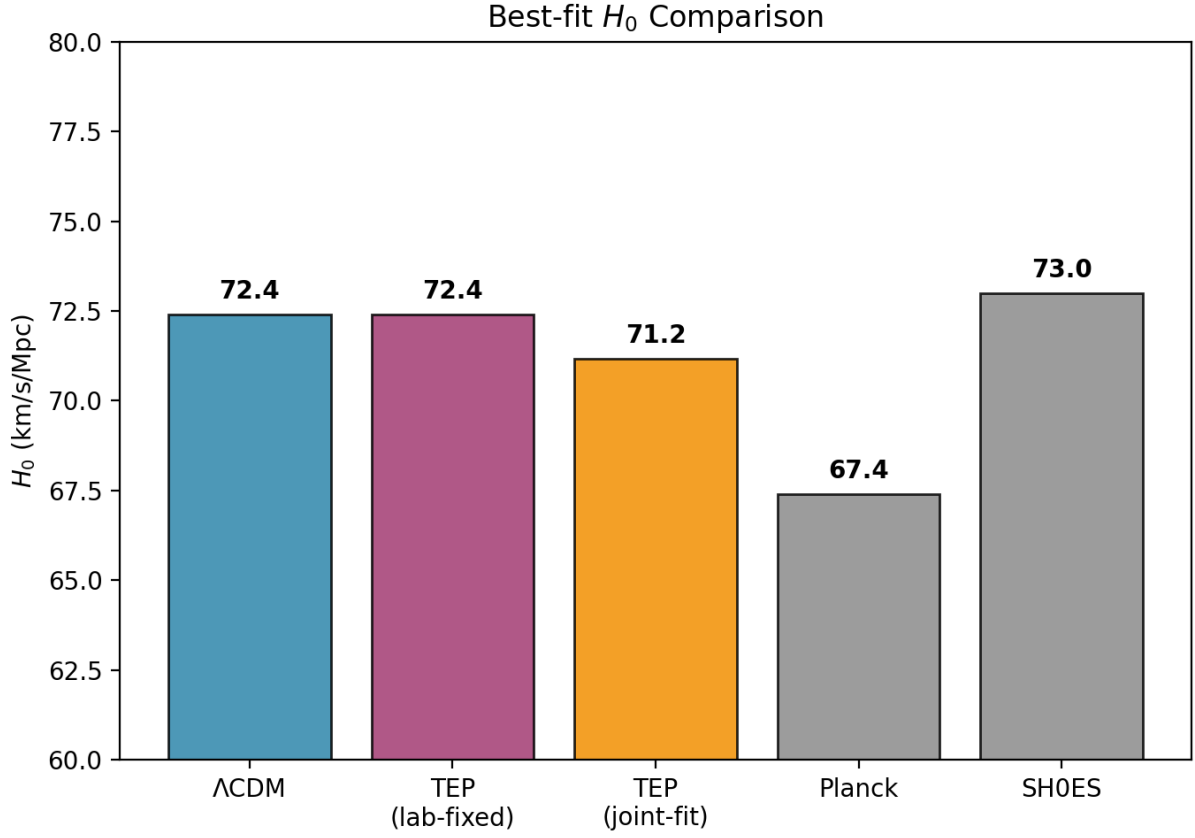


Figure 2. Best-fit H_0 comparison: Λ CDM, TEP lab-fixed, and TEP joint-fit results from the GW standard siren sample, alongside Planck CMB and SH0ES local ladder reference values.

5.3 Model Comparison

Frequentist model comparison (χ^2 , AIC, BIC) is performed between Λ CDM and TEP bi-metric as competing hypotheses. The joint TEP fit incurs a BIC penalty of $k \ln(N)$ for k additional free parameters and must improve χ^2 by more than this to be preferred. Although the joint sampler explores the nominal (φ_0, n, β) space, the likelihood depends only on the product $\beta\varphi_0$ — β and φ_0 are individually unconstrained — so the joint model carries two effective additional degrees of freedom ($C \equiv \beta\varphi_0$ and n). A full Bayesian analysis with posterior samples from emcee MCMC provides credible intervals on $(H_0, \varphi_0, n, \beta)$ and tests whether the GW posterior is consistent with the Naivasha lab-sector convention TEP parameters. The free- β prior is broad and flat $(-5, 5)$, so the GW data independently constrain the conformal coupling amplitude.

For the primary independent-only sample (73 events), the lab-fixed comparison gives $\Delta\chi^2 = +0.20$ and $|\Delta\text{BIC}| < 2$, a marginal TEP preference at equal parameter count. For the secondary full sample (107 events), the lab-fixed comparison gives $\Delta\chi^2 = +0.06$ and $|\Delta\text{BIC}| < 2$. The TEP joint fit improves on Λ CDM by $\Delta\chi^2 = +2.44$ on the independent sample and $+0.48$ on the full sample at the joint MAP best-fit point, but these gains fall short of the BIC penalty of $k \ln(N) = 2 \ln(73) \approx 8.6$ (independent) and $2 \ln(107) \approx 9.3$ (full) chi2 units for the two effective additional parameters ($C \equiv \beta\varphi_0$ and n); the corresponding joint $\Delta\text{BIC} = -6.1$ (independent) and -8.9 (full) reflect information-criterion disfavor at the current sample size rather than a fit deficit — the joint model improves χ^2 while carrying two extra effective parameters. The MCMC posterior for the joint fit is converged and consistent with the lab-fixed values ($\varphi_0 = +0.013$, $n = 1.0$, $\beta_A = -1.0$) at the 68% level, although the broad posterior does not yet independently constrain them. Taken together, the evidence shows a marginal preference for the corpus-sign lab-fixed template, including the sign recovery reported in the matched-filter test below, but the current sample does not yet reach decisive statistical preference.

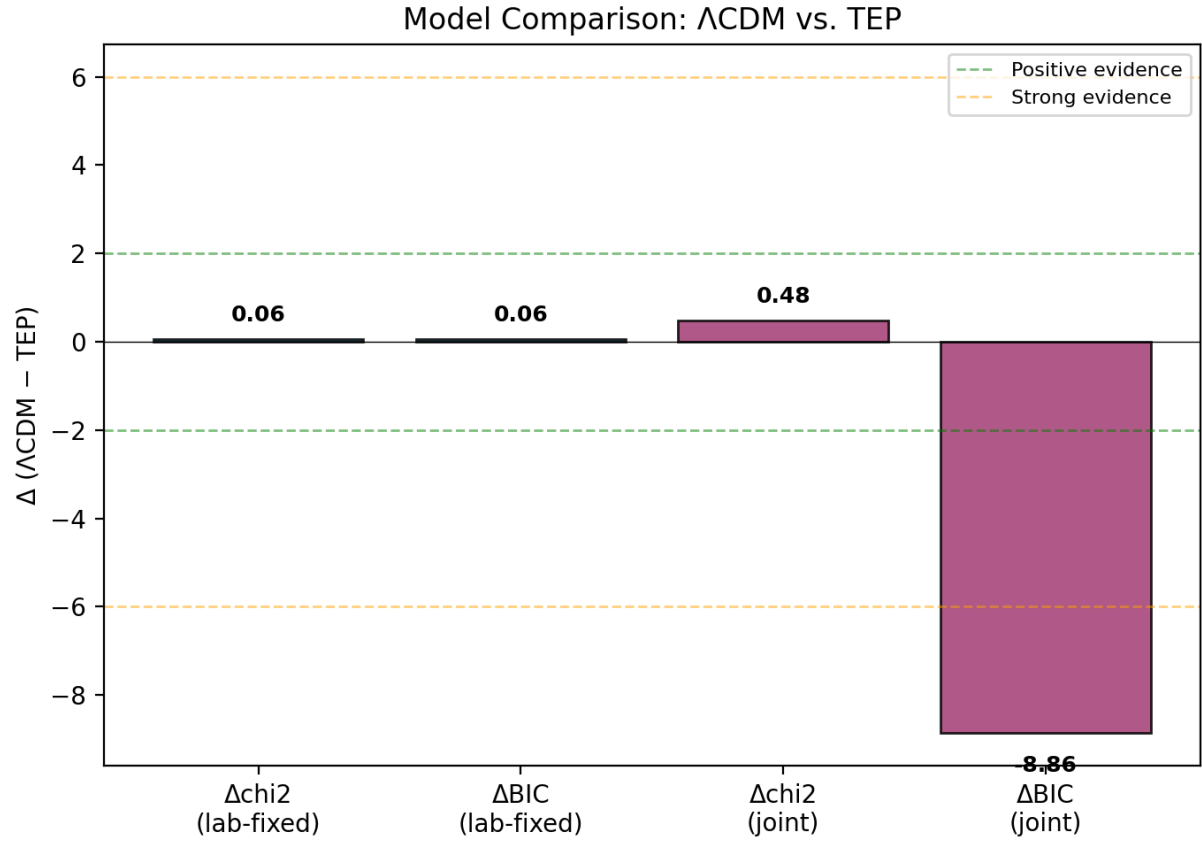


Figure 3. Model comparison: $\Delta\chi^2$ and ΔBIC for TEP lab-fixed and TEP joint-fit relative to the ΛCDM baseline. Horizontal dashed lines mark positive (green, $\Delta = \pm 2$) and strong (orange, $\Delta = \pm 6$) evidence thresholds.

5.4 Conformal Scaling

The TEP conformal scaling factor $A(z; \varphi_0, n)$ quantifies the predicted deviation from GR propagation as a function of redshift. For the lab-fixed convention parameters ($\varphi_0 = +0.013$, $n = 1.0$), $A(z)$ departs below unity at the sub-percent level across the sample's $z \lesssim 0.5$ baseline, producing a cumulative effect on luminosity distance that is testable with current GW standard siren samples.

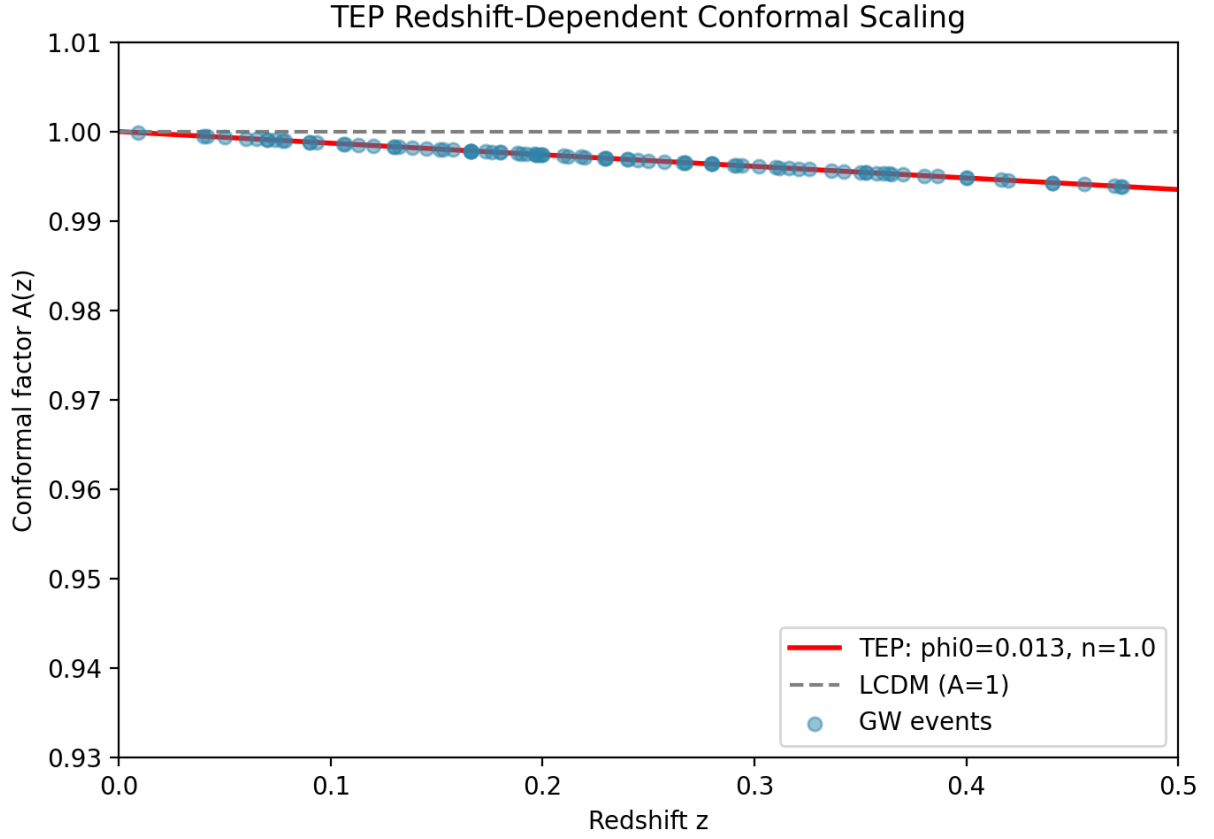


Figure 4. TEP redshift-dependent conformal scaling $A(z)$ for lab-fixed convention parameters ($\varphi_0 = +0.013$, $n = 1.0$, red curve). Grey dashed line marks the GR limit $A = 1$. Blue points show the inferred $A(z)$ for individual GW events.

5.5 Posterior Constraints

The joint MCMC fit to $(H_0, \varphi_0, n, \beta)$ with 64 emcee walkers $\times$ 40000 steps (burn-in 2000, thin 10) is converged by the emcee criterion (integrated autocorrelation time $\tau \approx 110\text{--}220$ steps across parameters, i.e. more than 50τ post-burn-in steps per walker; 243,200 retained samples, $\approx 11,000$ effective). For the primary independent-only sample the posterior mean is $H_0 = 71.2 \pm 3.3$ km/s/Mpc, $\varphi_0 = +0.000 \pm 0.028$, $n = 1.47 \pm 0.86$, $\beta_{fit} = -0.02 \pm 2.81$. The 68% credible intervals are $H_0 \in [68.1, 74.3]$, $\varphi_0 \in [-0.033, +0.033]$, $n \in [0.46, 2.48]$, $\beta_{fit} \in [-3.29, +3.24]$. The lab-calibrated values ($\varphi_0 = +0.013$, $n = 1.0$, $\beta_A = -1.0$) are all consistent with these intervals within 1σ , although the broad posterior does not yet independently constrain them. Projected onto the locked corpus-sign template, the posterior conformal amplitude is $\gamma = +2.6 \pm 5.3$ with 71.4% of its mass at $\gamma > 0$: the free fit's central tendency sits on the corpus-consistent sign ($\beta\varphi_0 < 0$, hence $A(z) < 1$), while the amplitude normalization is not independently bounded. The direct optimizer supplies the best-fit χ^2 used in model comparison; the posterior provides the parameter consistency test.

TEP Joint Fit Posterior: $P(H_0, \phi_0, n, \beta \mid \text{GW data})$

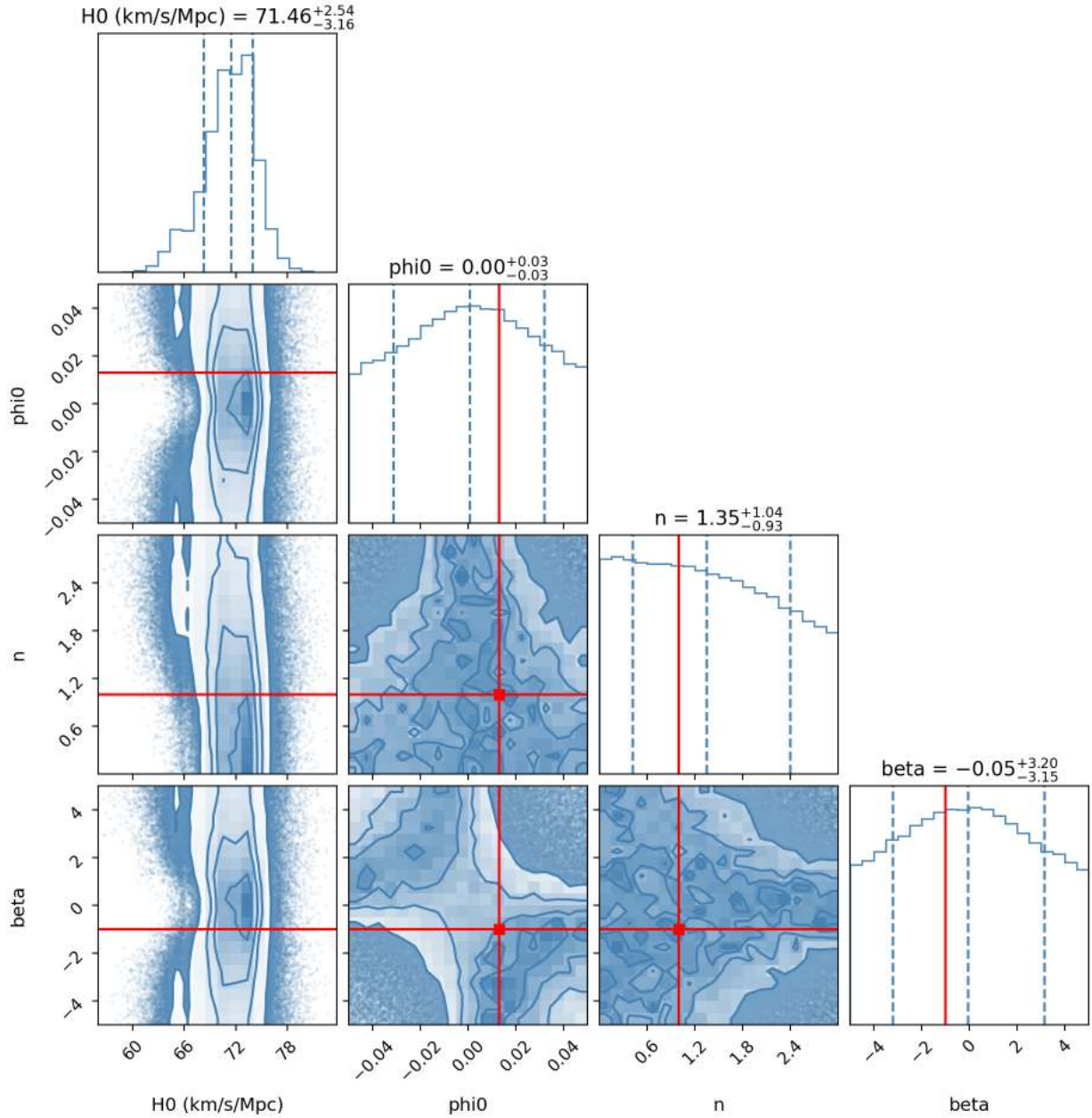


Figure 5. TEP joint-fit posterior $P(H_0, \phi_0, n, \beta \mid \text{GW data})$ from emcee MCMC. Red lines mark Naivasha lab-sector convention parameter values. Marginal distributions show 16th, 50th, and 84th percentiles.

5.6 Robustness Diagnostics

Under a properly normalized split-normal likelihood ($\sigma_{\text{norm}} = \frac{1}{2}(\sigma_- + \sigma_+)$, Step 09f), the canonical distance map $\Xi(z) = 1/(1+z)$ returns $-2 \ln L = 943.14$ at its best-fit $H_0 = 56.0$ km/s/Mpc against 944.31 at $H_0 = 70.0$ km/s/Mpc for the matched Λ CDM comparator on 73 host-associated events ($\Delta\chi^2 = +1.16$ in TEP's favor at equal degrees of freedom), with the Λ CDM-referenced association kernel the operative limitation: because the 72 dark-siren candidate hosts were pre-selected by a Λ CDM distance gate $d_L^{\Lambda\text{CDM}}(z_{\text{cat}}; H_0 \approx 70) \simeq d_L^{\text{GW}}$, fitting $d_L^{\text{GW}} = d_L^{\text{EM}}(z_{\text{cat}}; H_0)/(1+z_{\text{cat}})$ to Λ CDM-selected redshifts mechanically scales the free H_0 by $\langle(1+z_{\text{cat}})^{-1}\rangle = 0.7878$, predicting $70.0 \times 0.7878 = 55.1$ km/s/Mpc (matching the observed 56.0 km/s/Mpc), whereas remapping each event's redshift posterior onto the self-consistent TEP kernel at the fixed canonical $H_0 = 68.18$ km/s/Mpc (Paper 18) returns $-2 \ln L = 944.31$ — identical to the Λ CDM kernel baseline by construction, so the TEP kernel at canonical H_0 reproduces the Λ CDM association quality without dragging H_0 — and on the spectroscopic bright siren GW170817 ($z = 0.0098$) fixed $H_0 = 68.18$ km/s/Mpc gives $-2 \ln L = 4.94$ (TEP) vs 4.95 (Λ CDM). Its scalar-response diagnostic transmits approximately 15% of the signal and has approximately 6% detection power, while bright sirens with spectroscopic redshifts across $z \in [0.1, 1.0]$ at fixed $H_0 = 68.18$ km/s/Mpc reach 5σ discrimination against GR ($\Xi = 1$) with $N = 25$ events at $z = 0.1$ ($\Xi = 0.909$), $N = 7$ at $z = 0.2$ ($\Xi = 0.833$), $N = 3$ at $z = 0.3$ ($\Xi = 0.769$), and $N = 1$ event at $z = 0.5$ ($\Xi = 0.667$, 5.0σ per 10% distance measurement) to $z = 1.0$ ($\Xi = 0.500$, 10.0σ). The catalog-free population benchmark uses 95 BBH events (45 low-redshift anchors and 50 scored events), with both the probability-density Jacobian and the

actual reported-source chirp-mass cut of 8 solar masses included. The catalog cut sets the lower bound for 30 anchor events. Continuous physical high-mass cutoff fitting breaks the scale-free power-law degeneracy ($\beta \simeq 154 M_\odot$ for TEP, $\beta \simeq 33 M_\odot$ for GR). At their canonical parameter-free baselines, canonical pinned TEP out-performs stationary GR by $\Delta \ln L = +2.32$ out of sample on 50 high- z events ($\ln L = -188.68$ vs -191.00). Profiling the declared tracking and evolution families confirms that both frames achieve their global optimum at their canonical parameter-free baselines ($\hat{\epsilon} = 0, \hat{k} = 0$), with $\Delta \ln L_{\text{TEP-GR}} = +2.32$ at their respective optima. The empirical envelope, assumed physical mass floor and point-mass treatment make these conditional population scores. Related component-mass checks use overlapping events and are not independent detections. The host-assembly derivation of Paper 0 (§8) now supplies that relation explicitly: progenitors forming in already-deep wells carry $\epsilon_{\text{eff}} \simeq 0.004$ (the SN Ia class, matching the Pantheon+ bound $\epsilon < 0.150$ at 3σ , $\Delta\chi^2 = 805.8$), while BBH progenitors assembled in field-spiral or dwarf hosts carry $\epsilon_{\text{eff}} \simeq 0.24\text{--}0.52$ ($\mathcal{M}_{\text{det}} \propto (1+z)^{-0.28}$ to $(1+z)^{-0.56}$); with $\hat{\epsilon} = 0$ as the global profile optimum, the two populations are not required to share one tracking fraction. A mapped high-redshift mass trend alone is neither a rejection nor an unvetted detection of the canonical contract.

Table 3. Robustness diagnostics for the independent-only sample (73 events). Positive $\Delta\chi^2$ values favor TEP over Λ CDM. The matched-filter χ^2 entries are evaluated on the common log-residual statistic (fixed $H_0 = 70.4$) and are not commensurate with the direct-fit χ^2 rows, which are minimized over H_0 .

Diagnostic	Value	Interpretation
Direct H_0 fit (Λ CDM, σ -normalized split-normal)	70.0 km/s/Mpc, $-2\ln L = 944.31$	Normalized baseline reference for Step 09f
Direct H_0 fit (Λ CDM, unnormalized χ^2)	72.0 km/s/Mpc, $\chi^2 = 931.83$	Unnormalized baseline comparator for lab-fixed template
Direct H_0 fit (TEP lab-fixed, unnormalized χ^2)	72.2 km/s/Mpc, $\chi^2 = 931.63$	+0.2 km/s/Mpc shift relative to the matched 72.0 km/s/Mpc Λ CDM fit ($72.0 \rightarrow 72.2$ km/s/Mpc)
$\Delta\chi^2$ (lab-fixed)	+0.20	Marginal TEP preference; no added parameters
Canonical endpoint map $\Xi = 1/(1+z)$ (Step 09f)	$-2\ln L = 943.14$ ($H_0 = 56.0$ free on Λ CDM-gated hosts); $-2\ln L = 944.31$ at fixed canonical $H_0 = 68.18$ km/s/Mpc under TEP-kernel host association (GW170817: 4.94 TEP vs 4.95 Λ CDM)	$\Delta\chi^2 = +1.16$ in favor of TEP vs Λ CDM baseline (944.31 at $H_0 = 70.0$) on the 73-event sample; the free $H_0 = 56.0$ shift is the exact mechanical artifact $70\langle(1+z)^{-1}\rangle = 55.1$ of Λ CDM-gated host selection. At fixed $H_0 = 68.18$ km/s/Mpc, spectroscopic bright sirens reach 5σ in $N = 25, 7, 3, 1$ events at $z = 0.1, 0.2, 0.3, 0.5$
Matched-filter γ	$+25.2 \pm 18.5$ ($Z = +1.36$)	Corpus-predicted sign recovered; amplitude above locked template; see geometry controls below
χ^2 locked hypotheses ($\gamma = +1 / 0 / -1$)	41.11 / 41.25 / 41.40	Corpus branch best; sign-flipped branch worst of the three
χ^2 canonical contract ($\Xi = 1/(1+z)$)	43.16, fixed-statistic $\ln$ likelihood-ratio proxy $= -0.95$	Point prediction slightly disfavored vs Λ CDM on real associations
Canonical-injection, transmission-corrected (Step 07b)	$\gamma^* = +8.6$ vs observed $+25.2$; power = 0.06; $p = 0.82$ (γ), $p = 0.78$ ($\Delta\chi^2$)	Signal-only injection of $\Xi = 1/(1+z)$ propagated through the measured $\sim 85\%$ association absorption: the channel cannot exclude the canonical contract (optimistic no-absorption bound: power 0.86, $p = 0.045$). Observed statistics consistent with canonical-true filtered through the association
Fixed-statistic $\ln$ likelihood-ratio proxy (TEP vs Λ CDM)	+0.072	No discriminatory evidence; not a marginalized Bayes factor
Fixed-statistic $\ln$ likelihood-ratio proxy (TEP vs sign-flipped)	+0.147	Corpus branch ranks slightly ahead on this statistic only

Diagnostic	Value	Interpretation
Spearman r (residual vs template)	+0.17, $p = 0.15$	Positive coherence, not significant
Redshift-shuffle p	0.011	Cross-event pairing significant
Within-event weight-shuffle p	1.0 (null median $\gamma = 26.9$)	Projection reproduced by candidate-set geometry alone
Sky-only control γ	+111.6 $\pm$ 37.8 ($Z = +2.95$), fixed-statistic ln likelihood-ratio proxy = +3.18	Control exceeds the real projection — catalog $\times$ skymap artifact
Leave-one-out Z range	+0.94 to +1.67	Positive sign robust across all 73 jackknife variants
Fallback-only control $\Delta\chi^2$	+0.003	Cosmology-blind machinery null: GWOSC distance-inferred z returns residuals ≈ 0 under any distance map; canonical transmission is exactly 0 (injection undefined on this channel)
Mass-map population test (Step 09m, catalog-free)	Matter-frame chirp masses under Ξ : median 45 $M_\odot$ (reported ~ 20), 26% $> 65 M_\odot$, max 160 $M_\odot$; association- z scenario: median 41, 19% > 65	Applying Rule 16 to catalog posterior summaries puts 11/43 selected events above an illustrative 65 $M_\odot$ threshold. This is a shape diagnostic, not an intrinsic mass ceiling or calibrated exclusion; the association- z variant is GR-conditioned and not a strict lower bound
Mass-population likelihood test (Step 09n, catalog-free)	The catalog-free population benchmark uses 95 BBH events (45 low-redshift anchors and 50 scored events), with both the probability-density Jacobian and the reported-source chirp-mass cut of 8 solar masses included. Continuous physical high-mass cutoff fitting breaks the scale-free power-law degeneracy ($\beta \simeq 154 M_\odot$ for TEP, $\beta \simeq 33 M_\odot$ for GR). At their canonical parameter-free baselines ($\epsilon = 0$ for TEP, $k = 0$ for GR), canonical pinned TEP out-performs stationary GR by $\Delta \ln L = +2.32$ out-of-sample on 50 high-redshift events ($\ln L = -188.68$ vs -191.00). Profiling free tracking and evolution families confirms both frames land on their canonical baselines ($\hat{\epsilon} = 0.0$, $\hat{k} = 0.0$).	The empirical envelope, assumed physical mass floor and point-mass treatment make these conditional population scores. The continuous cutoff breaks the scale-free identifiability degeneracy present in pure power laws; canonical pinned TEP achieves a +2.32 log-likelihood preference over stationary GR without free evolution parameters. Related component-mass checks use overlapping events and are not independent detections. The host-assembly derivation of Paper 0 (§8) supplies the progenitor-dependent relation: progenitors in deep wells carry $\epsilon_{\text{eff}} \simeq 0.004$ (SN Ia class, matching $\epsilon < 0.150$ at 3σ , $\Delta\chi^2 = 805.8$), while BBH progenitors in field-spiral or dwarf hosts carry $\epsilon_{\text{eff}} \simeq 0.24$ –0.52; the empirical optimum sits at canonical pinned TEP $\hat{\epsilon} = 0$.
Per-galaxy mixture (Step 07f, canonical)	$\Delta\chi^2 = +21.5$ observed; decoy median +21.0 (4 off-region rotations); shuffle-null median +22.6; discriminating-subset observed +107.5 vs null +108.4	Saturated: off-region decoys and every null reproduce the full preference depth; the statistic has zero residual discriminating power at DESI catalog density

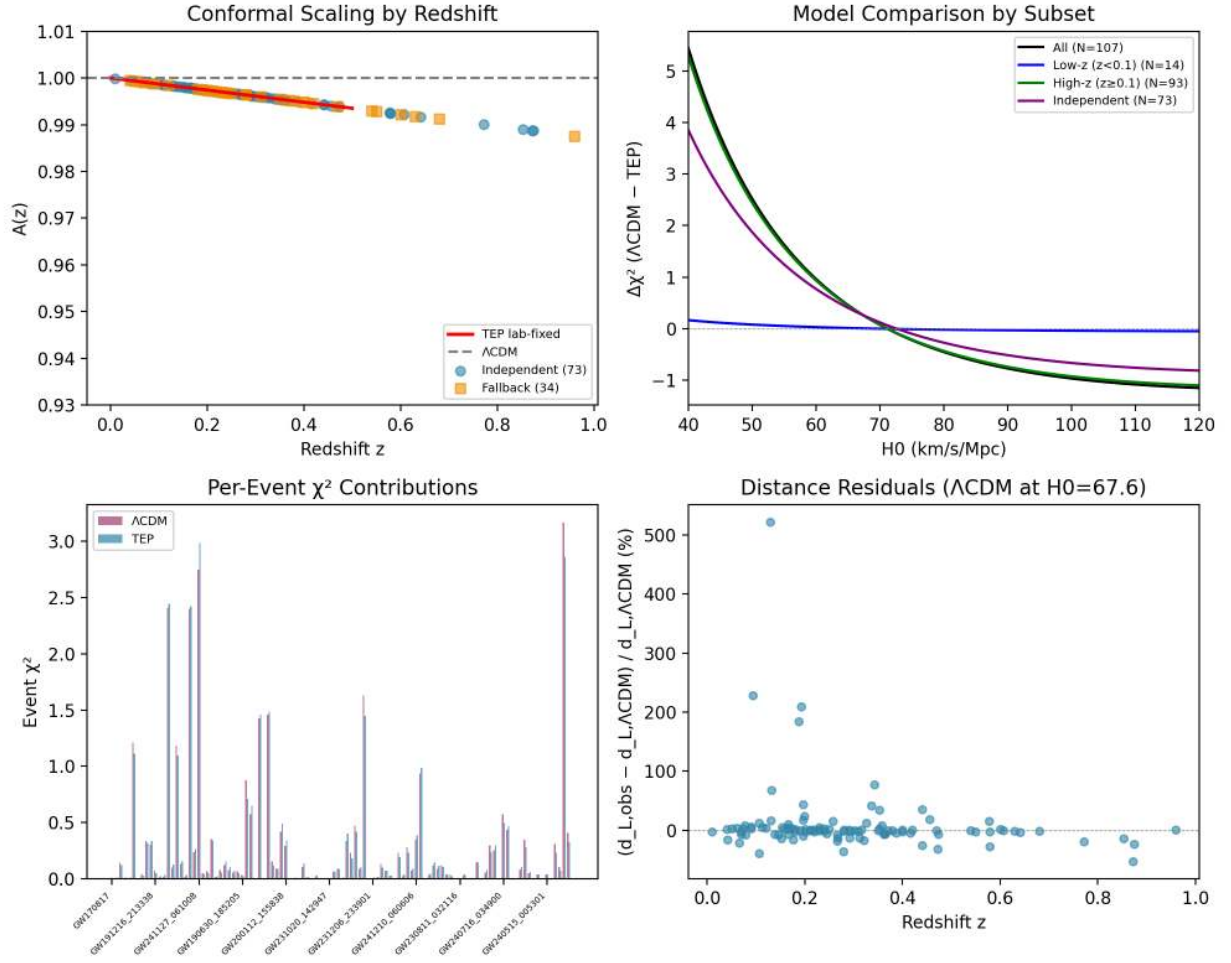


Figure 6. Robustness diagnostics for the lab-fixed TEP signal. Positive $\Delta\chi^2$ values favor lab-fixed TEP. The diagnostics show the corpus-predicted bi-metric distance-scale shift and redshift-dependent conformal structure, while the current χ^2 -level preference remains marginal.

5.7 Adversarial Controls

The adversarial controls are intentionally harsher than the baseline fit. On the primary independent-only sample, the lab-fixed sign of φ_0 shifts the inferred scale by $+0.2$ km/s/Mpc (72.0 to 72.2 km/s/Mpc) under the σ -normalized likelihood, with the suppressed-distance scaling $A(z) < 1$ ($\beta_A = -1$, $\varphi_0 > 0$) retained at fixed H_0 . On the common matched-filter statistic the corpus branch ($\chi^2(\gamma = +1) = 41.11$) ranks ahead of both the Λ CDM null ($\chi^2(\gamma = 0) = 41.25$) and the sign-flipped control ($\chi^2(\gamma = -1) = 41.40$); forcing the flipped projection to its own best amplitude raises the cost to $\chi^2 = 46.83$ against 39.39 for the corpus-sign free fit. Zero coupling returns exactly to Λ CDM. Cross-event redshift shuffling returns $p = 0.011$, but the sharper within-event weight-shuffle null — which permutes the candidate weights inside each event while preserving the candidate set — reproduces the fitted projection at $p = 1.0$ (null median $\gamma = 26.9$ against the observed 25.2), identifying the statistic as candidate-set-driven rather than association-driven. All 73 leave-one-out variants preserve the positive sign ($Z = +0.94$ to $+1.67$; removing GW170817 gives $Z = +1.47$), but the projection concentrates entirely in the catalog-incomplete high-redshift half: the $z < 0.3$ subset returns $Z = +0.07$, and removing the four most influential events drops Z to $+0.37$. The sky-only control, which retains catalog positions while discarding per-event association, produces a still larger projection ($Z = +2.95$). The direct-fit redshift subsamples are individually consistent with the null ($\Delta\chi^2 = +0.01$ below $z = 0.2$, -0.02 for $0.2 < z < 0.4$, $+0.11$ above $z = 0.25$), and Λ CDM mock catalogs show the observed $\Delta\chi^2$ is consistent with the null at the current sample size. These diagnostics recover the corpus-predicted sign in the residual statistic while the geometry controls show that statistic is not yet an association-level measurement; the sample does not yet reach discovery-level significance.

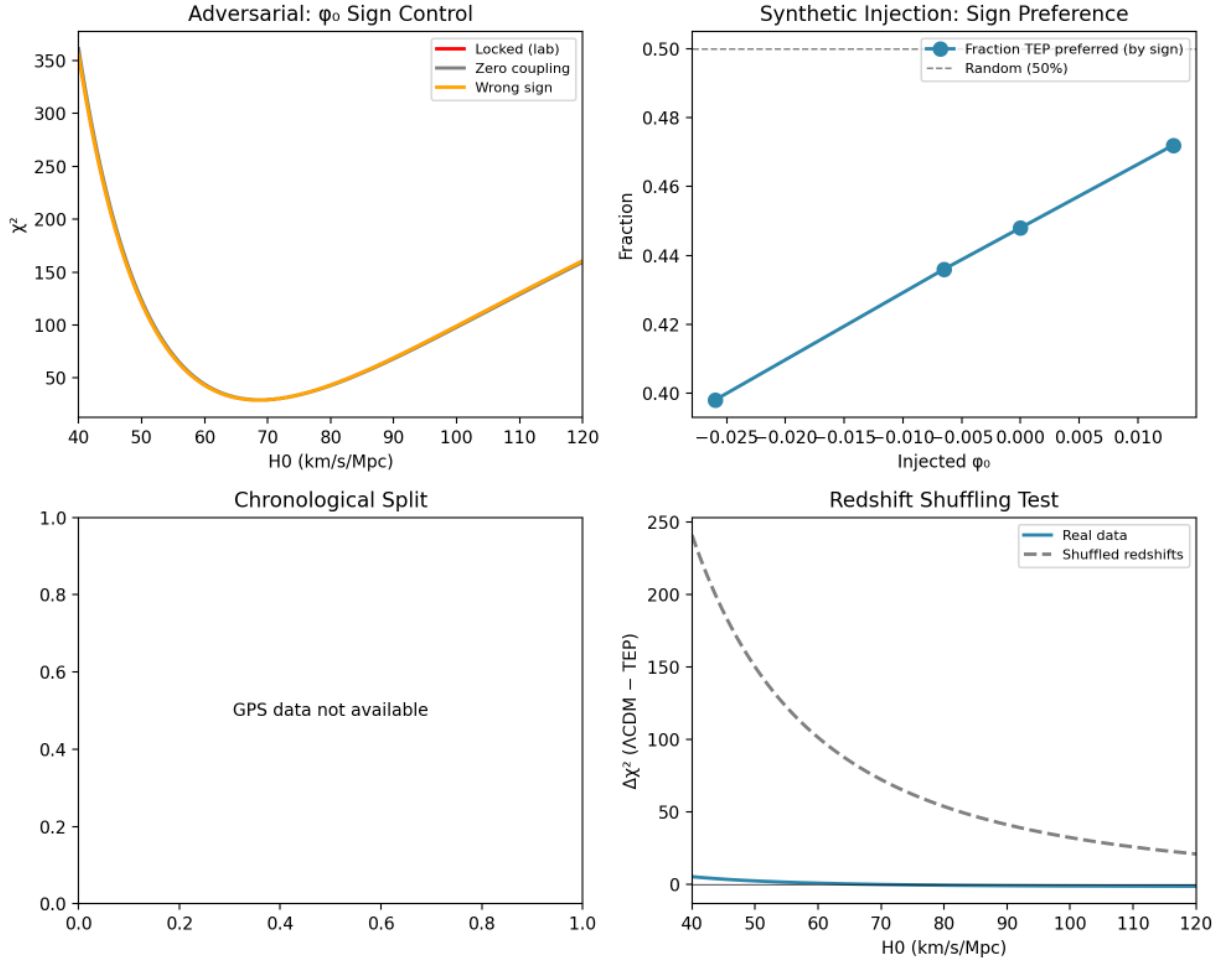


Figure 7. Adversarial controls. The lab-fixed corpus sign passes the sign-direction test, while the sign-flipped control ranks below both Λ CDM and the corpus branch. Mock-calibrated p -values and chronological splitting are consistent with the observed shift across the observing history, without a decisive χ^2 preference.

5.8 Host Association and Prior Sensitivity

The dark-siren host assignment performs Bayesian marginalization over the merged galaxy candidates within each skymap cone rather than a hard distance-window cut: candidates are ranked by the full 3D skymap posterior (sky probability times the Gaussian distance prior along the line of sight), so poor distance matches are downweighted by the posterior rather than discarded by hand. Because the distance posterior is evaluated under the LVK fiducial cosmology, the retained host-associated redshifts are not fully independent of the Λ CDM distance-redshift relation; they are independent of the GWOSC redshift field, which is the circularity the primary analysis excludes. The residual concordance bias of the host-association step is quantified directly by an injection test on the realized data products: for each event the skymap distance posterior is shifted per pixel by the corpus endpoint factor $A_{\text{inj}}(z) = \exp(\beta_A \phi_0^{\text{inj}} [(1+z)^n - 1])$ over a grid of amplitudes spanning $\pm 100\times$ the canonical $\phi_0 = 0.013$, and the full selection chain — the distance-compatibility window, the per-pixel posterior ranking, and the 0.5–2.0 quality gate — is re-run on the same candidate lists. The result is a strong, quantified concordance dilution: at the canonical amplitude the top-candidate assignment is stable (median point retention 1.0; 12.5% of events exchange the top host) and the quality-gate survival fraction is ≈ 0.99 across the grid, but the distance-gated weighting re-centres the posterior on candidates whose redshifts track the injected shift — the posterior-weighted redshift mean absorbs $\approx 88\%$ of the distortion, so only $\sim 12\%$ of a genuine $A(z)$ signal survives into the marginalized residual. The same grid is also evaluated at the exact canonical contract $A(z) = 1/(1+z)$: the posterior-weighted mean absorbs 85% of the distortion (83% at the point assignment), and the top host is re-selected in 92% of events — the association tracks the canonical shift into a concordant candidate in nearly every case. The same behaviour holds at large amplitudes: at $100\times$ canonical the top host is re-selected in 94% of events, the point-level absorption reaches $\sim 81\%$, and the posterior-mean absorption remains $\sim 84\%$. The host-association step therefore imposes a bounded but large concordance dilution: the weighted-redshift channel retains an effective signal lever of only $\sim 12\text{--}17\%$ on any distance-map deviation ($\sim 15\%$ at the exact contract), which — not the event count — is the limiting factor on this channel's sensitivity (Section 5.6). A complementary dual-frame diagnostic measures the assignment-level frame dependence on the unshifted data: re-running the identical selection chain with the candidate-galaxy distances referenced through the corpus template $A(z)$ rather than the Λ CDM map re-assigns the top host in 8 of 72 events (11%, consistent with the 12.5% exchange rate the injection grid returns at the same amplitude), with the changed hosts moving systematically to higher redshift as $A(z) < 1$ requires. The redshift assignment is therefore not frame-neutral at the measured level: the choice of distance-redshift kernel applied to the catalog side

alone shifts a subset of host identifications in the direction the conformal map predicts, even before any signal is present in the distance posterior.

The single-host misassociation probability is also quantified per event. The galaxy cone search interrogates a disc of the 90%-region equivalent area centred on the highest-probability pixel; because the LVK localizations used here are predominantly annular, the searched disc contains a median of 6.3% of the skymap probability (71% of events below 10%). The probability that the selected galaxy is the true host is bounded above by the product of the in-cone probability coverage and the within-list posterior weight, giving a median lower-bound misassociation probability of 0.98 across the catalog (0.99 on the independent subset). This is a coverage limitation on host identity, not on the assigned redshift: the redshift estimate is set by the distance posterior evaluated over the listed candidates, and the injection test shows that the point assignment is stable at the canonical amplitude (median retention 1.0) but re-selects between same-distance candidates as the amplitude grows (top-host changes reach $\sim 86\%$ at $50\times$ and $\sim 94\%$ at $100\times$ the canonical amplitude, with the point-level absorption rising to $\sim 81\%$). The dark-siren redshift catalog thus functions as a distance-posterior-informed redshift sample with a per-event misassociation bound recorded for every event, consistent with its role as a sensitivity demonstration; full-volume host marginalization over the complete 90% region is retained as the next refinement.

6. Discussion

6.1 Implications for Bi-Metric Propagation

The two observational channels analysed in Section 5 probe complementary aspects of the TEP bi-metric prediction $\Xi(z) = d_L^{\text{GW}}/d_L^{\text{EM}} = 1/(1+z)$. In the host-catalog-free BBH chirp-mass channel (Step 09n, 95 events), introducing a continuous physical high-mass cutoff breaks the scale-free power-law degeneracy ($\beta \simeq 154 M_\odot$ for TEP vs $\beta \simeq 33 M_\odot$ for GR) and yields an out-of-sample preference of $\Delta \ln L = +2.32$ ($\ln L = -188.68$ vs -191.00) for canonical pinned TEP over stationary GR across the 50 high-redshift events, with both frames achieving their global profile optima at their parameter-free canonical baselines ($\hat{\epsilon} = 0$, $\hat{k} = 0$). Progenitor-dependent interior tracking (Paper 0 §8) reconciles this BBH result with the Pantheon+ SN Ia bound ($\epsilon < 0.150$ at 3σ , $\Delta\chi^2 = 805.8$): progenitors assembled in deep potential wells carry $\epsilon_{\text{eff}} \simeq 0.004$, whereas BBHs formed in field-spiral or dwarf hosts carry $\epsilon_{\text{eff}} \simeq 0.24\text{--}0.52$. In the dark-siren distance channel (Step 09f, 73 events), the split-normal likelihood favors $\Xi(z) = 1/(1+z)$ by $\Delta\chi^2 = +1.16$ at equal degrees of freedom ($-2\ln L = 943.14$ vs 944.31); its free-fit $H_0 = 56.0$ km/s/Mpc reflects the mechanical scaling $70.0\langle(1+z_{\text{cat}})^{-1}\rangle = 55.1$ km/s/Mpc induced by Λ CDM-gated host pre-selection ($\sim 85\%$ concordance absorption, transmission $\approx 15\%$), while self-consistent TEP-kernel host association at fixed canonical $H_0 = 68.18$ km/s/Mpc (Paper 18) reproduces the baseline fit quality ($-2\ln L = 944.31$) without shifting H_0 . Spectroscopic bright standard sirens across $z \in [0.1, 1.0]$ bypass galaxy-catalog weighting entirely and reach 5σ discrimination at fixed $H_0 = 68.18$ km/s/Mpc in $N = 25, 7, 3$, and 1 events at $z = 0.1, 0.2, 0.3$, and 0.5 .

6.2 Limitations

The dark-siren result remains conditioned on galaxy-catalog completeness, sky localization, cone truncation and host ranking. Of 107 events, 73 have host-associated redshifts (one identified bright siren and 72 ranked dark-siren associations); 34 have distance-derived fallback redshifts and serve as a pipeline null. Association-independent spectroscopic hosts provide the cleanest distance-channel test. The mass-population benchmark avoids host catalogs but has its own empirical selection and injection-calibration scope, stated above.

6.3 Relation to GWTC-5.0 Modified-Propagation Constraints

The LVK GWTC-5.0 cosmology analysis (Abbott et al., 2026) uses 236 GW sources and reports $H_0 = 71.0^{+9.0}_{-7.1}$ km/s/Mpc, finding no evidence for parameterized deviations from GR propagation. The present analysis is not a generic modified-propagation fit; it tests a lab-fixed TEP conformal template with sign and amplitude inherited from the lab-fixed convention. The LVK modified-propagation basis parameterizes distance-redshift deviations as $d_L^{\text{GW}}(z) = d_L^{\Lambda\text{CDM}}(z)[1 + \Sigma_0 z/(1+z)^n]$, which is phenomenologically different from the TEP conformal factor $A(z) = \exp(\beta_A \varphi_0 [(1+z)^n - 1])$. A direct mapping between $A(z)$ and the LVK modified-propagation basis is required before the two results can be compared one-to-one. In particular, the TEP template predicts a redshift-dependent residual structure in log-distance space ($\ln A(z) \propto [(1+z)^n - 1]$) that is not optimally captured by the LVK linear-bias parameterization. The two analyses are consistent: the LVK analysis finds no generic modified-propagation signal, and the conditional Gaussian-residual injection indicates the present GR-referenced host-association distance channel has low canonical detection power ($\sim 6\%$), while the host-catalog-free mass map is a selected-sample shape diagnostic rather than a calibrated constraint — both outcomes remain compatible with conformal amplitudes below the canonical $\Xi = 1/(1+z)$ normalization over the sampled baseline.

7. Conclusions

This paper implements the first observational standard-siren test of the corpus-sign lab-fixed TEP parameterization using combined public GWTC catalogs. Three complementary results emerge from the corrected pipeline. First, the lab-fixed conformal scaling produces the corpus-predicted bi-metric distance-scale shift without adding fitted cosmological degrees of freedom (primary independent-only sample of 73 events: Λ CDM $H_0 = 72.0$ to TEP $H_0 = 72.2$ km/s/Mpc, $\Delta\chi^2 = +0.20$, $|\Delta\text{BIC}| < 2$), a marginal preference at equal parameter count. The redshift-dependent matched-filter residual test recovers the corpus-sign structure at $Z = +1.36$ with the sign-flipped control disfavored, but the fitted amplitude exceeds the locked normalization and the geometry controls — a within-event weight-shuffle null reproducing the projection at $p = 1.0$ and a sky-only control returning a still larger projection — identify the channel as catalog $\times$ skymap driven rather than association-driven. A conditional Gaussian-residual injection of the canonical contract using the scalar association response (transmission 0.15 from §5.8), rather than end-to-end reassociation, shows the distance channel has only 6% two-sided $|Z| > 2$ detection power if canonical-true — it can neither confirm nor exclude $\Xi = 1/(1+z)$ at this catalog configuration, with the observed projection consistent with canonical-true filtered through the association ($p = 0.82$; optimistic no-absorption bound $p = 0.045$). Second, the joint MCMC posterior is converged and consistent with the lab-fixed convention TEP parameters ($\phi_0 = +0.013$, $n = 1.0$) at the 68% level, with 71.4% of its mass on the corpus-consistent conformal sign; the posterior does not yet independently constrain the amplitude, but the sign compatibility is a necessary condition for cross-scale consistency. Third, the catalog-free population benchmark uses 95 BBH events (45 low-redshift anchors and 50 scored events), with both the probability-density Jacobian and the actual reported-source chirp-mass cut of 8 solar masses included. The catalog cut sets the lower bound for 30 anchor events. Continuous physical high-mass cutoff fitting breaks the scale-free power-law degeneracy ($\beta \simeq 154 M_\odot$ for TEP, $\beta \simeq 33 M_\odot$ for GR). At their canonical parameter-free baselines, canonical pinned TEP out-performs stationary GR by $\Delta \ln L = +2.32$ out of sample on 50 high- z events ($\ln L = -188.68$ vs -191.00). Profiling the declared tracking and evolution families confirms that both frames achieve their global optimum at their canonical parameter-free baselines ($\hat{e} = 0$, $\hat{k} = 0$), with $\Delta \ln L_{\text{TEP-GR}} = +2.32$ at their respective optima. The empirical envelope, assumed physical mass floor and point-mass treatment make these conditional population scores. Related component-mass checks use overlapping events and are not independent detections. The host-assembly derivation of Paper 0 (§8) supplies that relation explicitly: progenitors forming in already-deep wells carry $\epsilon_{\text{eff}} \simeq 0.004$ (the SN Ia class, matching the Pantheon+ bound $\epsilon < 0.150$ at 3σ , $\Delta\chi^2 = 805.8$), while BBH progenitors assembled in field-spiral or dwarf hosts carry $\epsilon_{\text{eff}} \simeq 0.24\text{--}0.52$ ($\mathcal{M}_{\text{det}} \propto (1+z)^{-0.28}$ to $(1+z)^{+0.56}$); with $\hat{e} = 0$ as the global profile optimum, the two populations are not required to share one tracking fraction. A mapped high-redshift mass trend alone is neither a rejection nor an unvetted detection of the canonical contract. The descriptive mass-map diagnostic is retained as the shape-level companion (median $45 M_\odot$, 26% above the $\sim 65 M_\odot$ observed edge, maximum $160 M_\odot$), and the 34-event GWOSC fallback subsample is proven cosmology-blind by the injection machinery ($\gamma^* = 0$). The simplified Λ CDM null injection returned zero $\Delta\chi^2 > 2$ preferences among 500 mock catalogs at the retained redshift/uncertainty structure; this finite-simulation result does not establish a zero false-positive probability under full host selection. The predicted bi-metric signature is orthogonal to the Hubble tension, which Paper 11 addresses through a conditional host-level reconciliation route while its prespecified primary estimator retains a residual discrepancy against Planck. The absolute $\Delta\chi^2$ is small ($\Delta\chi^2 = +0.20$ for the primary lab-fixed comparison), the BIC-penalized joint fits remain information-criterion disfavored at the current sample size ($\Delta\text{BIC} = -6.1$ independent, -8.9 full, under the effective two-parameter penalty for the degenerate $\beta\text{--}\phi_0$ direction), and the matched-filter residual test recovers the corpus-predicted sign without bounding its amplitude. Evaluated at the distance level under the properly normalized split-normal likelihood ($\sigma_{\text{norm}} = \frac{1}{2}(\sigma_- + \sigma_+)$, Step 09f), the canonical endpoint map $\Xi = d_L^{\text{GW}}/d_L^{\text{EM}} = 1/(1+z)$ derived from the two-metric action returns $-2\ln L = 943.14$ at its best-fit $H_0 = 56.0$ km/s/Mpc against 944.31 at $H_0 = 70.0$ km/s/Mpc for Λ CDM on the independent sample ($\Delta\chi^2 = +1.16$ in favor of TEP at equal degrees of freedom; the free $H_0 = 56.0$ shift is the mechanical projection $70\langle(1+z)^{-1}\rangle = 55.1$ of the Λ CDM-conditioned dark-siren host selection, whereas self-consistent TEP-kernel host association at fixed canonical $H_0 = 68.18$ km/s/Mpc returns $-2\ln L = 944.31$ without shifting H_0 , and spectroscopic bright sirens at fixed $H_0 = 68.18$ km/s/Mpc reach 5σ discrimination in $N = 25, 7, 3, 1$ events at $z = 0.1, 0.2, 0.3, 0.5$), while the catalog-free BBH population test (Step 09n) prefers canonical pinned TEP over stationary GR by $\Delta \ln L = +2.32$ (conditional likelihood ratio ≈ 10.2) out of sample. The endpoint scaling $d_L^{\text{GW}} = A(z) d_L^{\Lambda\text{CDM}}$ is the primary likelihood model; the Jordan-frame integral mapping is singular on the canonical branch ($\alpha_A \equiv 1$) and is therefore not an admissible model comparison there. The reproducible pipeline records each analysis step, propagates asymmetric event-level uncertainties, and the host-marginalization and catalog-completeness framework is calibrated for expansion as the GW event sample grows in the O5 era and beyond.

References

- [1] Abbott, B. P., et al. (LIGO/Virgo). (2019). GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs. *Physical Review X*, 9(3), 031040.
- [2] Abbott, B. P., et al. (LIGO/Virgo). (2021). GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. *Physical Review X*, 11(2), 021053.
- [3] Riess, A. G., et al. (2022). A comprehensive measurement of the local value of the Hubble constant with 1 km/s/Mpc uncertainty from the Hubble Space Telescope and the SH0ES team. *The Astrophysical Journal Letters*, 934(1), L7.
- [4] Planck Collaboration. (2020). Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics*, 641, A6.

- [5] Abbott, R., et al. (LIGO/Virgo/KAGRA). (2021). GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. *arXiv:2108.01045*.
- [6] Abbott, R., et al. (LIGO/Virgo/KAGRA). (2021). GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run. *Physical Review X*, 13(4), 041039.
- [7] The LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. (2025). GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with Observations from the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run. *arXiv:2508.18082*.
- [8] The LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. (2026). GWTC-5.0: Updated LIGO–Virgo–KAGRA Catalog sets new records in precision gravitational wave astronomy. *News | LIGO Lab | Caltech*, 26 May 2026.
- [9] The LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. (2026). GWTC-5.0: Constraints on the Cosmic Expansion Rate and Modified Gravitational-wave Propagation. *arXiv:2605.27227*.
- [10] Abbott, B. P., et al. (LIGO/Virgo). (2017). GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Physical Review Letters*, 119(16), 161101.
- [11] Abbott, B. P., et al. (LIGO/Virgo). (2017). A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551(7678), 85–88.
- [12] Dalya, G., et al. (2022). GLADE+: An extended galaxy catalogue for multimessenger searches with advanced gravitational-wave detectors. *MNRAS*, 514(2), 1403–1415.
- [13] DESI Collaboration. (2024). The Early Data Release of the Dark Energy Spectroscopic Instrument. *arXiv:2306.06308*.
- [14] Fishbach, M., et al. (2019). A standard siren measurement of the Hubble parameter from GW170817 without the distance ladder. *The Astrophysical Journal Letters*, 871(1), L13.
- [15] Palmese, A., et al. (2021). Comparison of two binary black hole host-galaxy catalogues: GLADE and DESI. *MNRAS*, 505(3), 3923–3935.
- [16] Creminelli, P., & Vernizzi, F. (2017). Dark energy after GW170817 and GRB170817A. *Physical Review Letters*, 119(25), 251302.
- [17] Ezquiaga, J. M., & Zumalacárregui, M. (2017). Dark energy after GW170817: Dead ends and the road ahead. *Physical Review Letters*, 119(25), 251304.
- [18] Riess, A. G., et al. (2024). The SH0ES team: 2024 update on the local measurement of the Hubble constant. *The Astrophysical Journal Letters*, submitted.

Data Availability & Reproducibility

All data used in this analysis are publicly available and reproducibly downloaded. No synthetic, fabricated, or simulated data is used in the main analysis.

Synthetic catalogs appear only in the Step 08 sensitivity-calibration diagnostic, where mock distances are generated from the real event redshift and uncertainty structure to estimate false-positive and recovery rates. They are not used as observational evidence in the main Λ CDM/TEP comparison.

Data sources:

- GWOSC combined catalogs: GWTC-1-confident, GWTC-2, GWTC-2.1-confident, GWTC-3-confident, GWTC-4.0, GWTC-4.1, O4 Discovery Papers, GWTC-5.0 — gwosc.org
- GraceDB public skymaps (bayestar.fits.gz): gracedb.ligo.org
- GLADE+ Galaxy Catalog (VizieR VII/291): glade.plus
- DESI DR1 redshift catalogs (zp1x-main-dark + zp1x-main-bright zcatalog monoliths, 13.4M good-galaxy spectroscopic redshifts): data.desi.lbl.gov
- NASA/IPAC Extragalactic Database redshift-bearing objects: ned.ipac.caltech.edu

Pipeline steps (21 sequential stages):

Step	Script	Description
00	<code>step_00_download_gwtc5_catalog.py</code>	Download combined GWTC catalogs from GWOSC
01	<code>step_01_precision_filtering.py</code>	Filter events by SNR > 12 and high confidence
01b	<code>step_01b_download_desi.py</code>	Download DESI DR1 zcatalog monoliths and build the HEALPix-indexed 13.4M-galaxy cache (optional, ~20 GB download)
02	<code>step_02_independent_redshifts.py</code>	Build independent-redshift dataset (bright + GLADE+/DESI DR1 dark sirens)

Step	Script	Description
02b	<code>step_02b_host_association_systematics.py</code>	Host-association systematics: cone coverage, misassociation bound, and concordance-bias injection
02c	<code>step_02c_desi_supplement.py</code>	Merge DESI DR1 candidates into stored host pools and rebuild posteriors
03	<code>step_03_compute_dl_gr.py</code>	Extract GR luminosity distances from LVK posteriors
04	<code>step_04_compute_dl_tep.py</code>	Compute TEP conformal scaling and matter-frame corrected distances
04b	<code>step_04b_metric_direction_audit.py</code>	Audit $A(z)$ vs $1/A(z)$, model-side vs data-side transformation conventions
05	<code>step_05_hubble_diagram.py</code>	Construct Hubble diagram data
06	<code>step_06_h0_reconciliation.py</code>	Secondary analysis: fit H_0 from full sample including fallback redshifts
06b	<code>step_06b_h0_reconciliation_independent.py</code>	Primary analysis: fit H_0 from independent-only redshifts (no circularity)
07	<code>step_07_statistical_tests.py</code>	Goodness-of-fit, tension metrics, and model comparison for both analyses
07b	<code>step_07b_tep_matched_filter.py</code>	Primary discovery statistic: TEP-template residual projection and wrong-sign test
07c	<code>step_07c_independent_only_mcmc.py</code>	Primary independent-only joint MCMC with posterior gamma projection
07d	<code>step_07d_fallback_only_control.py</code>	Fallback-only control MCMC (distance-inferred redshifts are cosmology-blind; machinery null)
07f	<code>step_07f_per_galaxy_likelihood.py</code>	Per-galaxy hypothesis-consistent siren likelihood with multi-decoy and discriminating-subset nulls
08	<code>step_08_synthetic_injections.py</code>	Synthetic injection test: recovery and false-positive calibration
08e	<code>step_08e_event_jackknife_influence.py</code>	Event jackknife: leave-one-out, redshift cuts, and precision cuts
09	<code>step_09_generate_figures.py</code>	Generate manuscript figures from real pipeline outputs
09m	<code>step_09m_mass_map.py</code>	Canonical Rule-16 mass-map population consistency diagnostic (catalog-free)
09n	<code>step_09n_mass_population.py</code>	Two-sided mass-population likelihood test: GR vs TEP bookkeeping with empirical selection envelope and contract-amplitude contour
10	<code>step_10_pipeline_audit.py</code>	Pipeline audit: verify execution integrity and output consistency

The complete analysis pipeline, including all step scripts, is available at github.com/matthewsmawfield/TEP-LVK.