What the US reporting-rate analysis establishes—and what it does not

Submitted by the SPC

Evidence from 30 exact BET pairs | Interpretation note | 19 August 2026

Key conclusion. The US tests are limited in their approach for understanding this sensitivity and address a narrower question because they do not consider full model fits. The US direct RR0/RR1 comparison (RR=0 inclusion, RR=1 exclusion) fits only a constant fishing mortality parameter and does not estimate other parameters in a full model. It also does not present the re-estimated stock status or management quantities. Their results, while useful, do not fully represent the propagation of a bias into the key stock status quantities. Our 30 exact pairs (i.e., models with common grid settings but different RR flag settings) show why this matters: after the full model is re-estimated, RR1 ('exclusion' of reporting rate adjustment, i.e. RR assumed is 100% in mixing periods) gives more pessimistic recent depletion (SB/SBF=0) in five pairs and the more pessimistic F/FMSY in four. Importantly this demonstrates that the sensitivity of the stock status quantities is not inevitably all a one-way bias under the RR1 scenario as stated by the US.

What RR0 and RR1 actually do

RR0 (the report's 'include' setting) and RR1 (its 'exclude' setting) are flag labels, not reporting-rate values. During the early period before tagged fish are assumed to be fully mixed, RR0 expands reported returns R to approximately R/X, where X is the fitted reporting rate; RR1 uses R directly for that removal calculation. RR1 neither discards the tag observations nor assumes complete reporting throughout the assessment: fitted reporting rates remain active after mixing under both settings.

Why the exact BET pairs matter

The original 34-versus-46 comparison did not compare the same model configurations. To isolate the flag—and answer the report's call for paired refits (p. 14)—we used each retained RR0 model as an anchor and reran the identical BET configuration under RR1. Only the flag changed; BET was fully re-estimated. Thirty of 34 reruns produced usable pairs. This test measures how a complete BET fit responds to the flag on common model settings; it does not determine which result is true. The median RR1-minus-RR0 change in SBrecent/SBF=0 was about 0.018. Figure 1, ordered by natural mortality M, shows that clear separation is concentrated mainly in the lower tail, while most remaining trajectories are very similar. Five biomass-ratio pairs and four F/FMSY pairs move in the opposite direction. Figure 2A also shows variation with the tag-mixing cutoff K. The effect is real, but the wider BET configuration can amplify, dampen or reverse it.

Local tag bias is not stock-status bias

For fixed R and X≤1, R/X cannot be smaller than R. This identifies the larger premixing removal target; it does not determine which complete BET result must have higher F/FMSY or biomass. A full rerun adjusts many parameters and recalculates FMSY and unfished biomass. The US report's own fixed-parameter test found no change in F/FMSY (p. 9), and p. 14 states that the downstream cause was not established. Page 12 nevertheless invokes 'by construction' and presents the 15% lower F/FMSY association as a pseudo-data cost. The **US preliminary mixing-period simulation (HTML)** demonstrates a local under-removal mechanism under simplified assumptions. It does not test whether RR0 gives unbiased BET management quantities or RR1 gives biased stock status. The exact pairs measure the total fitted response to the flag, but contain no known stock truth and do not separate its internal pathways.

The comparison is also numerically selected

The planned ensemble contained 50 configurations under each flag, but retained only 34 RR0 and 46 RR1 models (p. 2). Figure 2B–C shows the resulting selection in M: RR0 retained 13 of 24 lower-M configurations, compared with 24 of 26 under RR1; the corresponding higher-M counts were 21 of 26 and 22 of 24. Under RR0, a very small fitted X can make R/X very large and the calculation unstable. The code provides a credible numerical route, although the available logs do not prove that every rejected fit overflowed. The PDF itself says the failures were not random (p. 12) and describes the original comparison as observational because the retained groups were selected and not paired model-by-model (p. 14). Part of the original contrast could therefore reflect which M configurations survived. Figure 2A shows that a flag response remains after exact pairing, so selection does not explain it away. But the original comparison cannot identify the effect over the planned grid, and an RR0-only ensemble would retain the more selectively filtered result set.

What the evidence supports

The paired results measure sensitivity, not bias. Local bias in reconstructed tag removals does not tell us whether F/FMSY, SB/SBF=0 or management advice is biased. Neither analysis knows the true stock status. That requires a full BET-like

simulation: generate data from known stock and reporting-rate values, refit both settings across low reporting rates and relevant M/K values, compare recovery of the key quantities and decisions, and include numerical failures after the calculation is stabilized. Until that test is done, the evidence does not establish RR0 as unbiased, RR1 as biased, or an RR0-only ensemble as scientifically necessary.

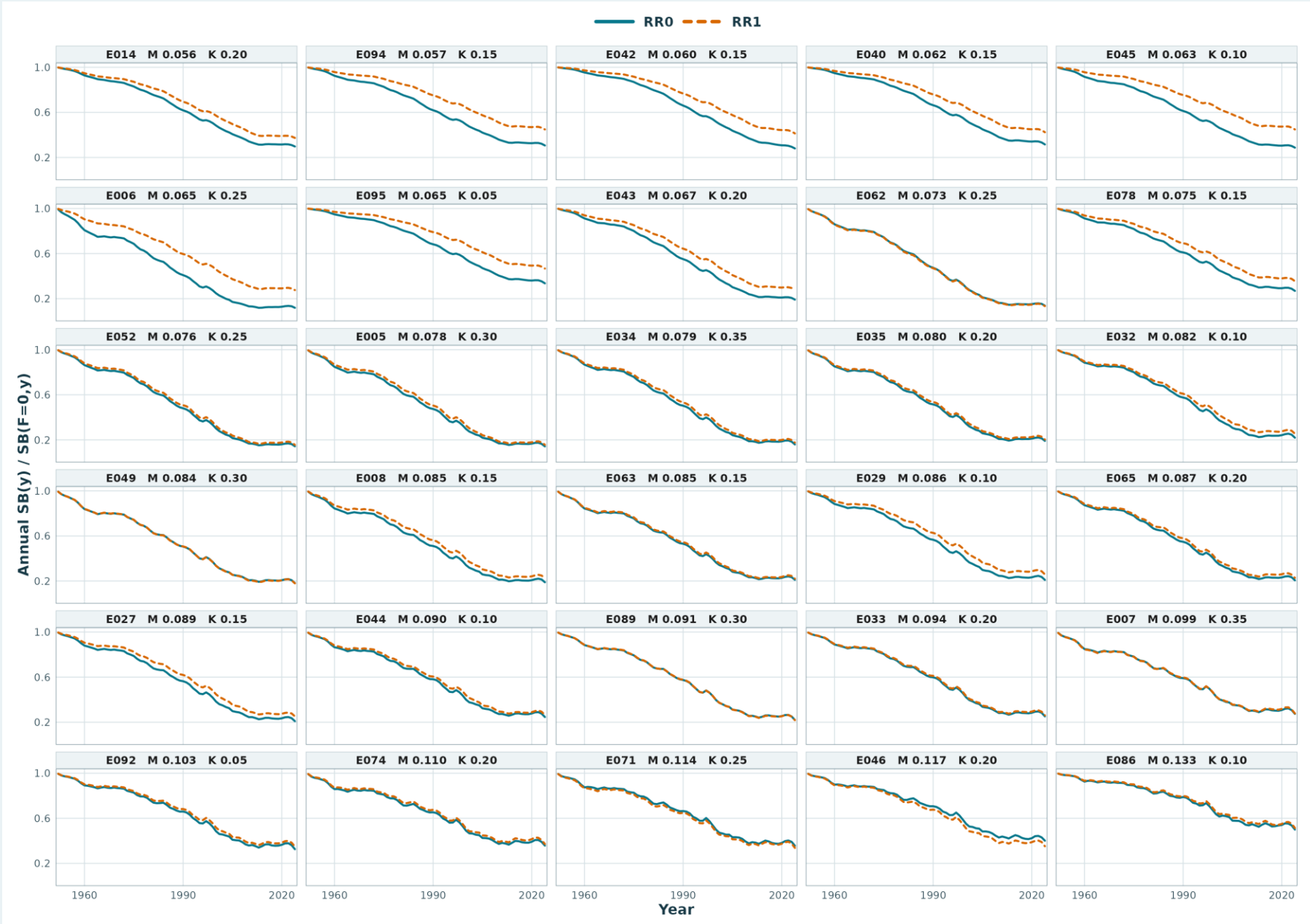

Figure 1. Annual depletion in 30 exact RR0/RR1 BET pairs, ordered by M. Within each panel the data and grid settings are identical and only the premixing RR flag differs; labels give model ID, M and K. RR0 is the solid teal line and RR1 the dashed orange line. All displayed fits passed MGC. These are central estimates for models that succeeded under both settings, not the full planned grid; Hessian uncertainty is not shown.

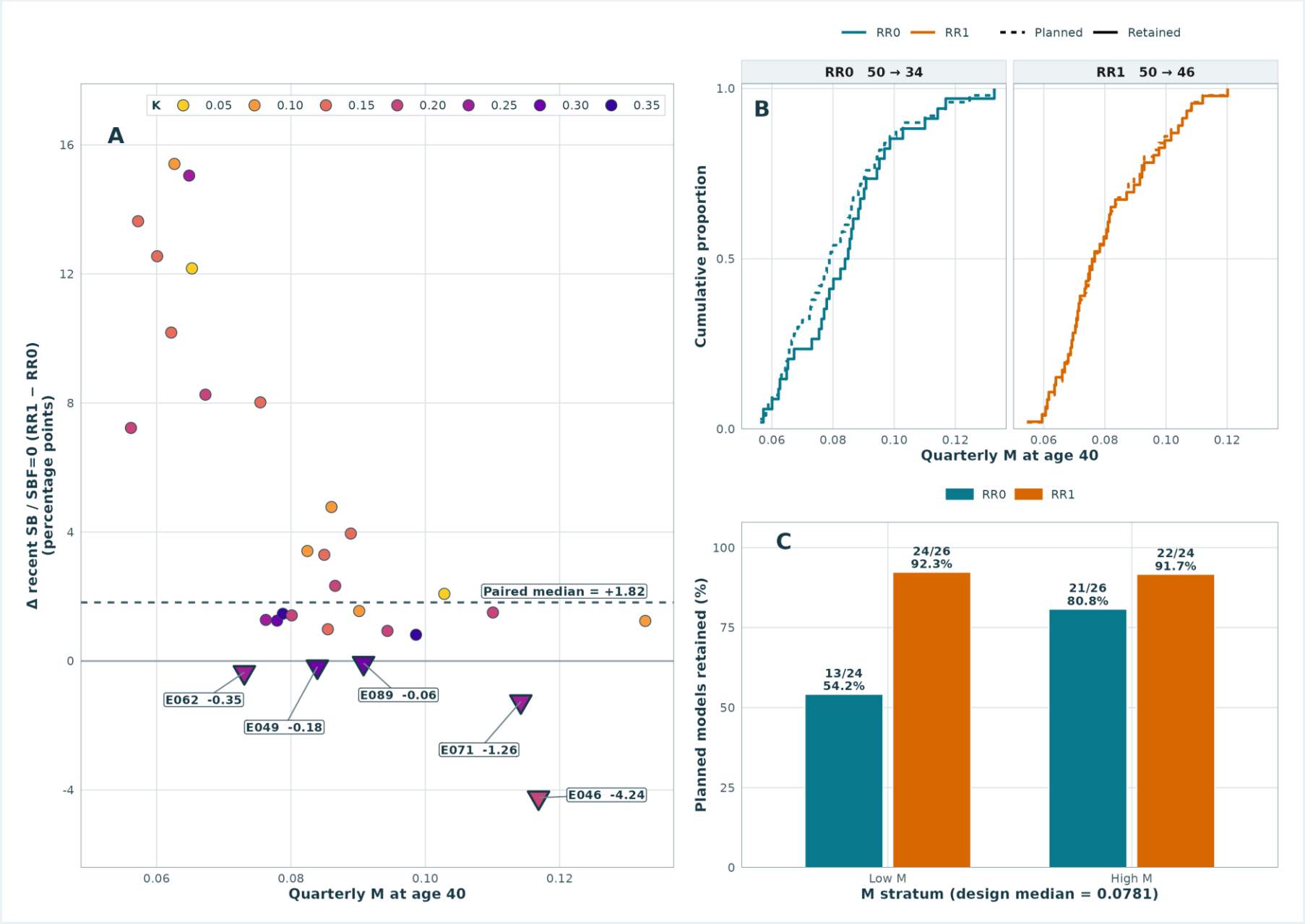

Figure 2. Heterogeneity and numerical selection. (A) RR1-minus-RR0 change in recent SB/SBF=0 for each exact pair; the dashed line is the paired median, colours show K, and labelled triangles identify the five opposite-direction results. (B) Planned and retained M distributions. (C) Retention in the lower and higher halves of planned M. The paired results describe models that succeeded under both settings; central estimates only, without Hessian uncertainty.