HPR
Hospital Practices and Research
http://www.jhpr.ir
Hosp Pract Res. 2026 June;11(2):847-850
doi 10.30491/hpr.2026.248291
Editorial

On Where, and Whether, the Sex Difference Disappears: Denominators and Conditioning in the Timing of ICU Treatment-Limitation Decisions

Mostafa Akbariqomi [1], **Amir Vahedian-Azimi** [2*]

[1] Tissue Engineering and Regenerative Medicine Research Center, New Health Technologies Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran
[2] Nursing Care Research Center, Clinical Sciences Institute, Nursing Faculty, Baqiyatallah University of Medical Sciences, Tehran, Iran

***Corresponding Author:** Amir Vahedian-Azimi, Ph.D., Nursing Care Research Center, Clinical Sciences Institute, Nursing Faculty, Baqiyatallah University of Medical Sciences, Tehran, Iran. Email: Amirvahedian63@gmail.com

Online Published June 30, 2026

Amacher and colleagues have assembled a dataset and posed exactly the right question: not merely *whether* women and men differ in how often life-sustaining therapy is curtailed, but *when* along the care pathway that divergence takes shape.[1] Their headline finding is elegant in its symmetry. A pronounced female excess at admission (adjusted odds ratio 1.26, 95% CI 1.24–1.28) gives way, they report, to parity once treatment is

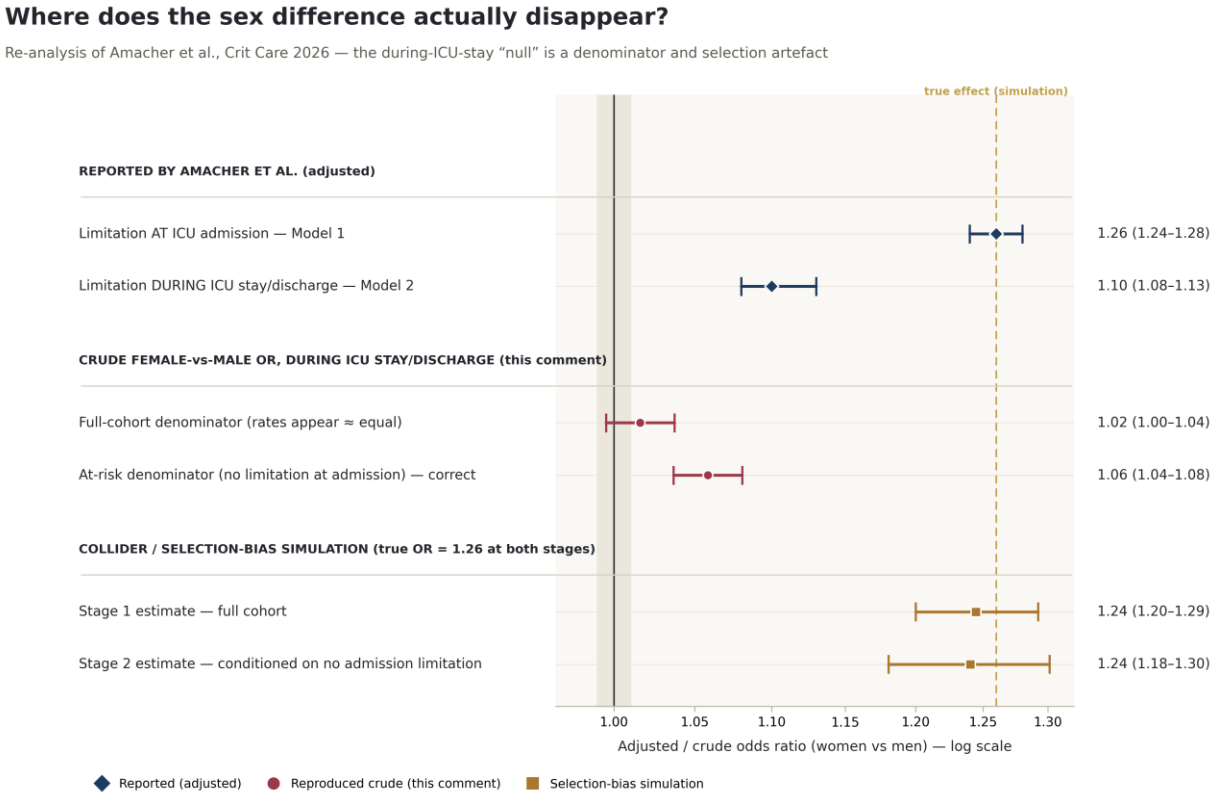

Figure 1. The During-Stay "null" Under Three Lenses. Female-versus-male odds ratios for treatment limitation, plotted on a logarithmic scale. Top band: the adjusted estimates reported by Amacher et al. for limitations at admission and during the ICU stay. Middle band: crude odds ratios for the during-stay outcome recomputed from the published counts (Woolf 95% CIs), first against the full admitted cohort and then against the correct at-risk denominator; the latter excludes unity and matches the authors' adjusted figure. Bottom band: a generative experiment in which the true sex effect is fixed at 1.26 at both stages (dashed line), with the unmeasured shared factor at a plausible strength; the stage-1 (full-cohort) and stage-2 (admission-conditioned) estimates are both close to 1.24, with stage 2 lying no higher than stage 1, while the systematic gap that opens between the stages as that factor strengthens is shown in Figure 3. Across the three bands the apparent "null" tracks the choice of denominator and the conditioning step rather than clinician behavior.

underway, the during-stay rates standing at "5.5% versus 5.5%." From this, they infer that decisions grow more even-handed as objective clinical information accrues. We admire the study but believe two linked problems undermine that inference. The first is arithmetical and present on the face of Table 1; the second is structural and inheres in the staged design itself. Both deserve closer scrutiny than they have so far received.

A Denominator that Shifts Beneath the Argument

The during-stay comparison rests on a denominator that is never fixed. The text reports 5.5% in each sex, a figure recoverable only against the full admitted cohort. Yet the second regression, by the authors' own specification, is restricted to patients admitted *without* a limitation; that is, the population genuinely at risk of acquiring one later.[1] Computed correctly against that at-risk denominator, the crude rates are 6.96% in women and 6.61% in men. The discrepancy is not cosmetic. As Figure 1 shows, the crude female-versus-male odds ratio for a during-stay limitation is 1.02 (0.995–1.04) when the full cohort is taken as the base, but 1.06 (1.04–1.08) once the base is the at-risk group. The latter excludes the null and aligns with the authors' own adjusted estimate of 1.10 (1.08–1.13). The model the authors fitted therefore contradicts the "no difference" they describe, and the reader is left with one outcome and two incompatible messages.

The table compounds the confusion. The "Total number" column in the risk panel reads 243,334 women and 371,563 men, which match neither the primary cohort (259,220 and 395,440) nor the at-risk cohort (228,091 and 361,409). Each equals the cohort total minus that sex's later (during-stay or discharge) events, a quantity with no clear reading as a base. The two percentages set beside it, 12.0%/8.6% at admission and 6.1%/6.0% during the stay, are both taken against the primary cohort rather than against the "Total number" printed in the same row, so the displayed base and the operative one disagree; and the during-stay figure is in any case the full-cohort rate, not the at-risk rate the second model estimates.

Why this matters is shown most economically in Figure 2, which sets the relative and absolute scales side by side. On the absolute scale the admission gap is substantial, a risk difference of 3.40 percentage points (3.25–3.56). The during-stay gap is small but genuinely non-zero: 0.36 points (0.23–0.49) against the at-risk base, collapsing to a non-significant 0.09 points (−0.03–0.21) only when the full-cohort framing dilutes it. The choice of denominator thus does real interpretive work, converting a modest but statistically supported residual into an apparent vanishing. The honest reading is that the sex difference *attenuates but persists*, not that it disappears.

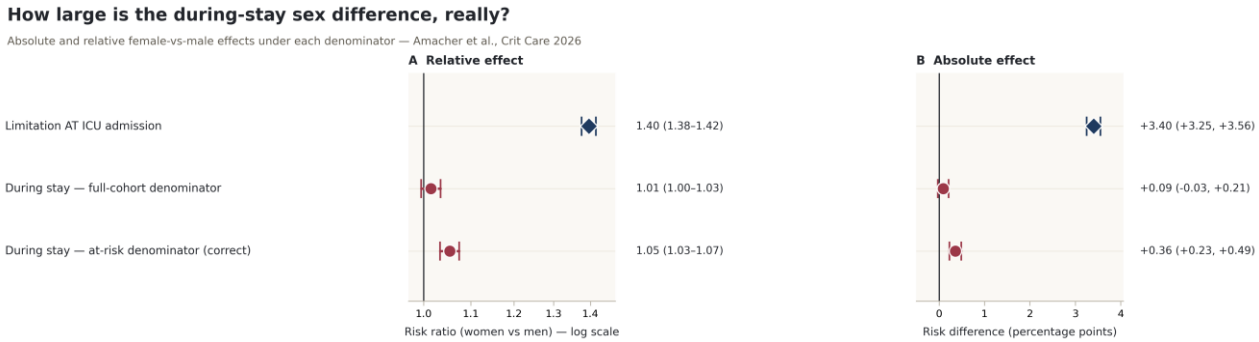

Figure 2. Relative and Absolute Female-versus-Male Effects Under Each Denominator. Panel A: risk ratios (logarithmic scale, log 95% CIs). Panel B: risk differences in percentage points (Wald 95% CIs). Each row gives the limitation present at admission, the during-stay outcome under the full-cohort denominator, and the during-stay outcome under the at-risk denominator. The admission difference is large on both scales; the during-stay difference is modest yet non-null on the correct denominator and is flattened toward zero only when the full cohort is used as the base.

Conditioning on a Downstream Event is not Innocuous

The deeper issue concerns the staged design that gives the paper its structure. Stage 2 is estimated only among patients who reached the ICU without an admission limitation. That restriction conditions the analysis on a variable, the presence of an admission limitation, that is itself a consequence of sex and of unmeasured patient characteristics such as frailty, functional reserve, and previously stated preferences. Documented sex differences in the intensity of organ support delivered in intensive care[2] make such latent imbalance more than a formal possibility. An admission limitation is therefore a common effect of the exposure and of latent prognostic factors. Restricting to its absence opens precisely the kind of selection (collider) path that the causal-inference literature has long cautioned against.[3-5] Once such a path is opened, an association between sex and the latent factors is induced within the surviving subset, and the stage-2 sex coefficient need no longer reflect the decision process at all.[6] A fall from 1.26 to 1.10 cannot, on this design, be read as evidence that bedside decisions have become "more balanced." Part of the descent may simply record who was left in the room.

To show that the concern is not merely theoretical, we ran a transparent generative experiment, the results of which occupy the lower band of Figure 1. The true sex effect was held identical at both stages (odds ratio 1.26), and an unmeasured factor was allowed to influence limitation at admission and during the stay alike. Estimated on the full cohort, stage 1 recovers the truth (1.24, 1.20–1.29). Re-estimated after conditioning on the absence of an admission limitation, exactly as the authors do, the stage-2 estimate is no higher than stage 1 (1.24, 1.18–1.30) at a plausible confounder strength, and across the full range of strengths, it sits at or below it, never above (Figure 3). The displacement, slight at this setting and wider as the shared factor grows, is manufactured by the conditioning rather than by any softening of bedside behavior.

Figure 3 establishes that this is not an artefact of one convenient parameter choice. Sweeping the strength of the unmeasured shared factor across a wide range, the stage-2 curve sits at or below the stage-1 curve at every positive value, and the two coincide only when the factor is switched off entirely. The shaded gap between them is the bias the design introduces. Its magnitude in our setting is moderate rather than dramatic, so we make the careful claim rather than the extravagant one: collider-induced selection is sufficient to bias the stage-2 estimate downward and therefore to forbid a causal reading of the observed attenuation, even if it is unlikely to account for the full 1.26-to-1.10 gap on its own. The point is qualitative and direction-fixed: the staged comparison is not bias-free, and the disappearance of the difference is not self-evidently a property of clinical decision-making.

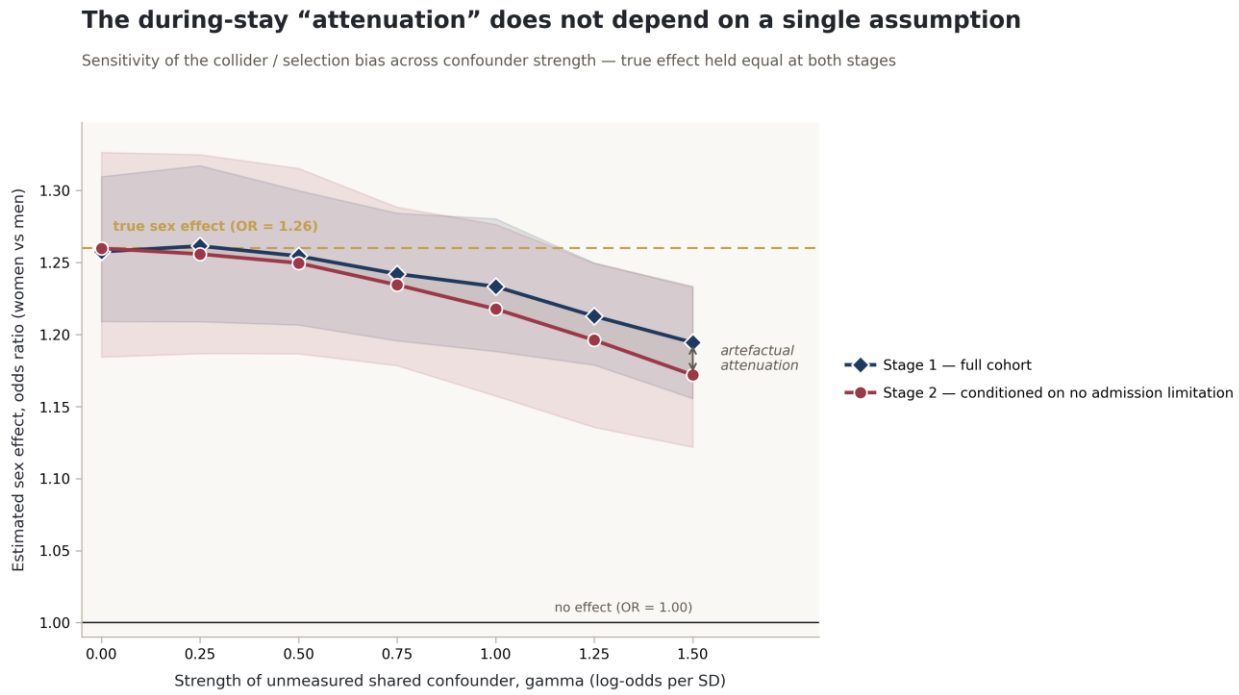

Figure 3. The Attenuation Does Not Hinge on a Single Assumption. Estimated female-versus-male odds ratio as the strength of the unmeasured shared confounder is varied along the horizontal axis. The full-cohort estimate (stage 1) and the admission-conditioned estimate (stage 2) are each plotted with their simulation bands; the true effect (1.26) and the null (1.00) are marked for reference. The stage-2 curve lies at or below the stage-1 curve for every positive confounder strength and meets it only at zero, so the downward bias from conditioning is present across the plausible range rather than at one chosen value.

Implications

Three correctives follow. The during-stay analysis should report a single, explicitly defined denominator, and we would urge the at-risk base, with rates and an absolute risk difference alongside the odds ratio, so that the residual female excess is neither concealed nor overstated. The inferential weight placed on the admission -versus-during-stay contrast should be tempered by recognition that conditioning on the admission decision is a selection step, best handled through joint modelling of both stages or through inverse-probability weighting that restores the full cohort.[5,7,8] And the interpretation should be recast: the data are consistent with a difference that narrows yet endures, and with a comparison whose later “equality” is, in part, an architectural feature of the analysis rather than a finding about clinicians.

None of this diminishes the value of the cohort or the importance of the question. It does suggest that the answer is less tidy and rather more interesting than a clean disappearance.

Availability of Data and Materials

All quantities used here were taken directly from the published article by Amacher et al. and its tables. The code generating the figures and the simulation is available from the authors on reasonable request.

Acknowledgments

The authors would like to thank the Clinical Research Development Unit of Baqiyatallah Hospital, for all their support and guidance during carrying out this study.

Conflict of Interest Disclosures

All authors declared that they have no conflict of interest.

References

1. Amacher SA, Gebert P, Tröster V, Mikail N, Vinzens A, Regitz-Zagrosek V, et al. Timing matters: sex differences in treatment limitation decisions in intensive care. Crit Care. 2026;30(1):331. doi:10.1186/s13054-026-06139-x
2. Modra LJ, Higgins AM, Pilcher DV, Bailey M, Bellomo R. Sex differences in vital organ support provided to ICU patients. Crit Care Med. 20241;52(1):1-10. doi:10.1097/ccm.0000000000006058
3. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10(1):37-48.
4. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615-25. doi:10.1097/01.ede.0000135174.63482.43
5. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. Int J Epidemiol. 2010;39(2):417-20. doi:10.1093/ije/dyp334
6. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. Am J Epidemiol. 2013;177(4):292-8. doi:10.1093/aje/kws412
7. Hernández-Díaz S, Schisterman EF, Hernán MA. The birth weight "paradox" uncovered?. Am J Epidemiol. 2006;164(11):1115-20. doi:10.1093/aje/kwj275
8. Lederer DJ, Bell SC, Branson RD, Chalmers JD, Marshall R, Maslove DM, et al. Control of confounding and reporting of results in causal inference studies. Guidance for authors from editors of respiratory, sleep, and critical care journals. Ann Am Thorac Soc. 2019;16(1):22-8. doi:10.1513/AnnalsATS.201808-564PS