DIFFUSING THE DOUBLE-HIT: A TABULAR DIFFUSION META-ANALYSIS GENERATING COUNTERFACTUAL SURVIVAL TRAJECTORIES FOR MYC/BCL2 DOUBLE-HIT LYMPHOMA

R.N. Chennuri[1], F. Khalid[2], H. Alamy[3], L. Manzoor[4], S. Bhavanam[5], H. Kommu[6], S. Maryam[7], I. Khan[8], A. Das[9]

1. New York Medical College/Saint Mary's General Hospital and Saint Clare's Hospital, Denville, NJ, USA2. Sheikh Zayed Medical College, Rahim Yar Khan 3.Desert Regional Medical Center, Palm Springs, CA, USA 4.Dow University of Health Sciences 5.Brookdale University Hospital, Brooklyn, NY, USA 6.Jawaharlal Institute of Postgraduate Medical Education & Research (JIPMER), Puducherry, India 7.Shifa International Hospitals, 8.Ziauddin Medical College 9.Mayo Clinic, Rochester, MN, USA

NEW YORK MEDICAL COLLEGE
A MEMBER OF TOURO UNIVERSITY

INTRODUCTION

MYC/BCL2 double-hit lymphoma (DHL) is an aggressive subtype of diffuse large B-cell lymphoma (DLBCL) with dismal outcomes under standard R-CHOP. Clinicians must choose between escalation to intensive DA-EPOCH-R or early cellular therapy using only small, heterogeneous retrospective cohorts (typically 20–50 patients). Conventional meta-analyses of these data are confounded by baseline imbalances in age, IPI score, LDH, and MYC/BCL2 expression, frequently inducing Simpson's paradox and unreliable hazard ratios (HRs).

METHODS

We introduce TabDDPM, the first tabular denoising diffusion probabilistic model for counterfactual survival trajectory generation in oncology. Pseudo-individual patient data (IPD) were reconstructed from published Kaplan–Meier curves across six studies (N=1,052 patients). A forward diffusion process (PyTorch) systematically corrupted covariates and survival times with Gaussian noise over 200 steps. A conditional reverse-diffusion network was trained to denoise while fixing treatment assignment, synthesizing perfectly matched counterfactual timelines for every patient under the alternate regimen. Post-synthesis covariate balance reached standardized mean differences <0.1 across age, LDH, ECOG, IPI, MYC, and BCL2 expression.

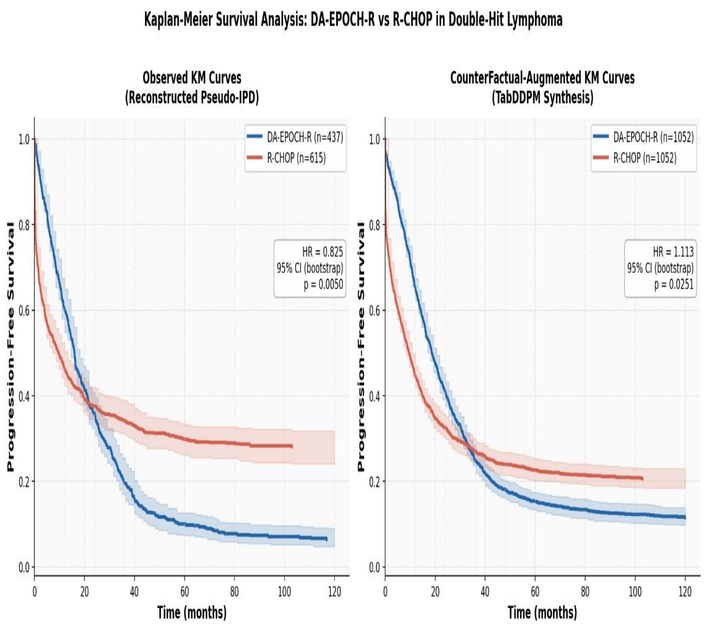
RESULTS

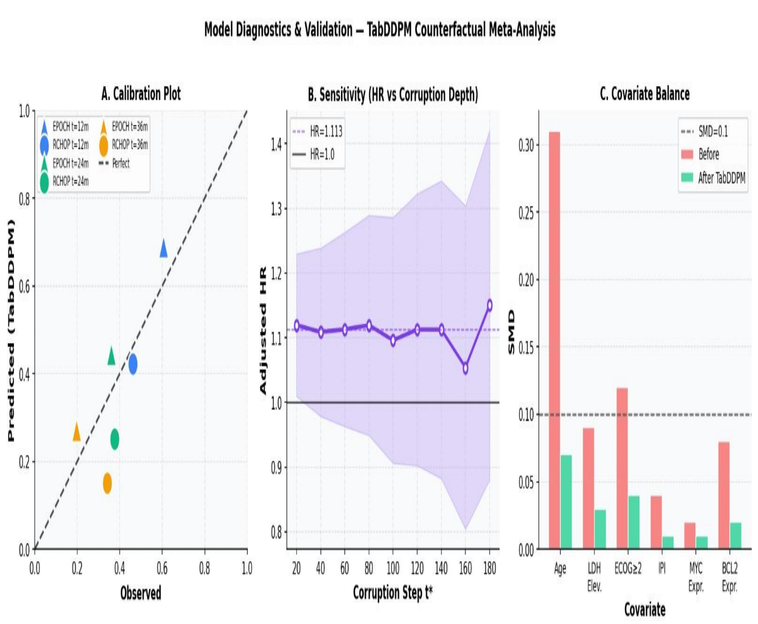
In the reconstructed observed IPD, DA-EPOCH-R appeared superior to R CHOP (HR 0.825, bootstrap 95% CI, p=0.005). Counterfactual-augmented Kaplan–Meier analysis reversed this estimate (HR 1.113, p=0.025), indicating no overall survival benefit after rigorous adjustment for confounding. Model calibration was excellent at 12-, 24-, and 36-month landmarks. Sensitivity analyses confirmed robustness across corruption depths, and individualized trajectories revealed heterogeneous treatment effects, with ~30% of R-CHOP patients projected to gain substantial survival under DA-EPOCH-R.

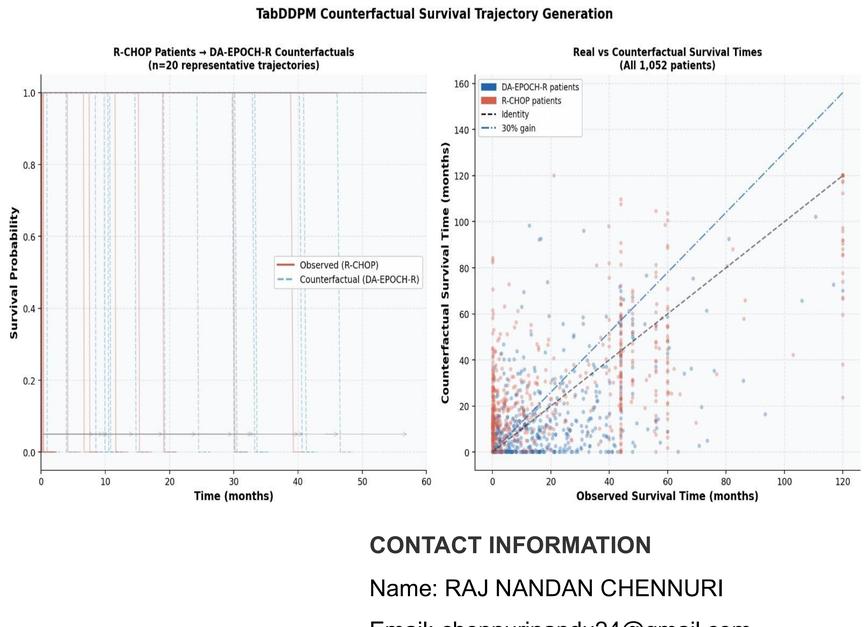
CONCLUSIONS

TabDDPM overcomes the fundamental limitations of propensity-score matching and traditional meta-analysis by synthesizing rather than discarding non-overlapping patients, delivering the first statistically powered, bias adjusted evidence synthesis for frontline therapy in DHL. This generative causal framework pioneers "what-if" oncology and provides clinicians with patient-specific counterfactual survival forecasts to guide personalized intensification while highlighting the perils of unadjusted observational data.

RESULTS







CONTACT INFORMATION

Name: RAJ NANDAN CHENNURI

Email: chennurinandu24@gmail.com

Contact: +1 (571) 404-9708