Synthetic Patient Trajectories from a Generative Machine-Learning Model: A Framework for Generating Population-Adjusted Survival Outcomes in Chronic Lymphocytic Leukemia

Talha Munir;[1] Paolo Sportoletti;[2] Leyla Mohseninejad;[3] Stephen Opat;[4] Igor Chebuniaev;[5] Samuel Aballéa;[5] Rami Zouari;[5] Noemi Kreif;[6] Mondher Toumi;[5,7] Marcus Lefebure;[8] Rhys Williams;[9] Sheng Xu;[10] Anna Maria Frustaci;[11] Loïc Ysebaert;[12] Mazyar Shadman[13]

[1]Leeds Teaching Hospitals NHS Trust, Leeds, UK; [2]University of Perugia and Santa Maria della Misericordia Hospital, Perugia, Italy; [3]BeOne Medicines, Schiphol, Netherlands; [4]Monash Health, Melbourne, Australia; [5]InovIntell, Krakow, Poland; [6]University of Washington School of Pharmacy, Seattle, Washington, USA; [7]Aix-Marseille University, Marseille, France; [8]BeOne Medicines, London, UK; [9]BeOne Medicines, San Carlos, California, USA; [10]BeOne Medicines, Shanghai, China; [11]ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy; [12]IUCT Oncopole, Toulouse, France; [13]Fred Hutchinson Cancer Center, University of Washington, Seattle, Washington, USA

CONCLUSIONS

- This study demonstrates the ability of a generative artificial intelligence (AI)-based latent ODE model to reproduce survival outcomes in previously unseen CLL patient profiles, supporting the generation of covariate-conditional longitudinal patient trajectories
- To our knowledge, this represents the first generative AI framework in CLL designed to create longitudinal patient trajectories for comparative evidence generation, providing a new approach to exploring clinical questions that cannot readily be addressed through head-to-head randomized trials
- This framework could support population-adjusted indirect treatment comparisons or external control arms, and could be extended to prospective outcome prediction, virtual cohort generation, treatment-sequence modeling and evaluation of hypothetical clinical scenarios
- Future work will include external validation against clinical trial data not used in model training

INTRODUCTION

- Treatment options in chronic lymphocytic leukemia (CLL) are expanding, including covalent and noncovalent Bruton tyrosine kinase (BTK) inhibitors and B-cell lymphoma 2 (BCL2) inhibitors, used as monotherapies or in combination, and as continuous or fixed-duration treatment strategies
- The number of clinically relevant questions exceeds what randomized trials can answer: many pairwise comparisons have never been made head-to-head, and questions about treatment sequencing are unlikely to be answered by clinical trials.
- Generative machine-learning models can simulate a complete patient trajectory, including longitudinal measurements and the timing of clinical events, given a baseline profile. Proposed uses of the resulting synthetic data include synthetic control arms,[1] indirect treatment comparisons,[2] data augmentation for under-accruing trials,[3] and in silico trial simulation,[4–6] as well as data sharing
- The credibility of such a framework rests on showing that generated outcomes reproduce observed outcomes — both in the patient groups the model was tuned on, and in groups not used to select the model configuration
- We developed and validated a generative modeling framework to produce CLL patient trajectories and survival curves conditional on specified baseline characteristics

METHODS

- **Data.** Individual patient data (IPD) from two phase 3 trials of zanubrutinib in CLL, pooled for training (N = 1,399): SEQUOIA (756 treatment-naïve patients), including patients without del(17p) randomized 1:1 to zanubrutinib or bendamustine plus rituximab, and patients with del(17p) treated with zanubrutinib alone or with venetoclax;[7] and ALPINE (643 patients with relapsed or refractory disease, randomized to zanubrutinib or ibrutinib).[8] The zanubrutinib-plus-venetoclax arm contributed to training only and was not evaluated
- **Model.** A latent ordinary differential equation (ODE) model was used.[9-13] An encoder maps baseline covariates to the distribution of an initial latent state; the state is augmented with the baseline covariates and a hazard accumulator and evolved continuously in time by a neural ODE; a decoder reads the evolved state at any requested time and emits laboratory values (absolute lymphocyte and neutrophil counts, platelets, hemoglobin, lactate dehydrogenase), response status, and the hazards of progression, next treatment, death and COVID-19. A complete trajectory can be generated for any baseline patient profile, observed or constructed
- **Baseline profiles.** Baseline profiles for validation were generated using a separate sequential classification and regression trees (CART) model (synthpop),[14] trained on pooled baseline characteristics from SEQUOIA and ALPINE. For the confirmatory analysis, 10,000 profiles were generated, and every synthetic patient was simulated from one of them rather than from a real patient's record, allowing trajectory generation to be evaluated across covariate combinations not necessarily observed in the training data
- **Fine-tuning.** The model configuration was fine-tuned using a two-stage process
  - Stage 1: a wide automated hyperparameter search aiming to minimize the 1-Wasserstein distance between synthetic and observed Kaplan–Meier (KM) curves across 16 "**fine-tuning subgroups**"
  - Stage 2: shortlisted configurations from stage 1 were re-evaluated with more iterations and ranked by overlap of 95% confidence bands around KM curves from observed vs synthetic data
- **Subgroups**
  - 16 **fine-tuning** subgroups defined on del(17p)/*TP53* status, IGHV mutational status, complex karyotype, number of prior therapies and β2-microglobulin
  - 14 **test** subgroups defined on age, Binet stage, bulky disease and sex, excluded from every stage of fine-tuning and evaluated only after the configuration was fixed
- **Uncertainty.** Quantified by bootstrap: 100 resamples with three training runs per resample, alongside 100 training runs on the original data (400 model training runs per analysis). 95% confidence bands around KM curves were produced
- **Metrics**
  - The absolute difference in survival probability between synthetic and observed curves, computed at every 6-month grid point in every subgroup × arm combination
  - Band containment: The mean fraction of the synthetic 95% confidence band lying inside the observed 95% confidence band computed with Greenwood's method
  - Observed inside band: proportion of time points at which the observed estimate falls inside the synthetic 95% band
  - Prognostic contrasts: the hazard ratio between the two levels of each subgroup-defining variable separately in the synthetic and the observed data
- Survival curves were truncated when the observed number at risk fell below 25

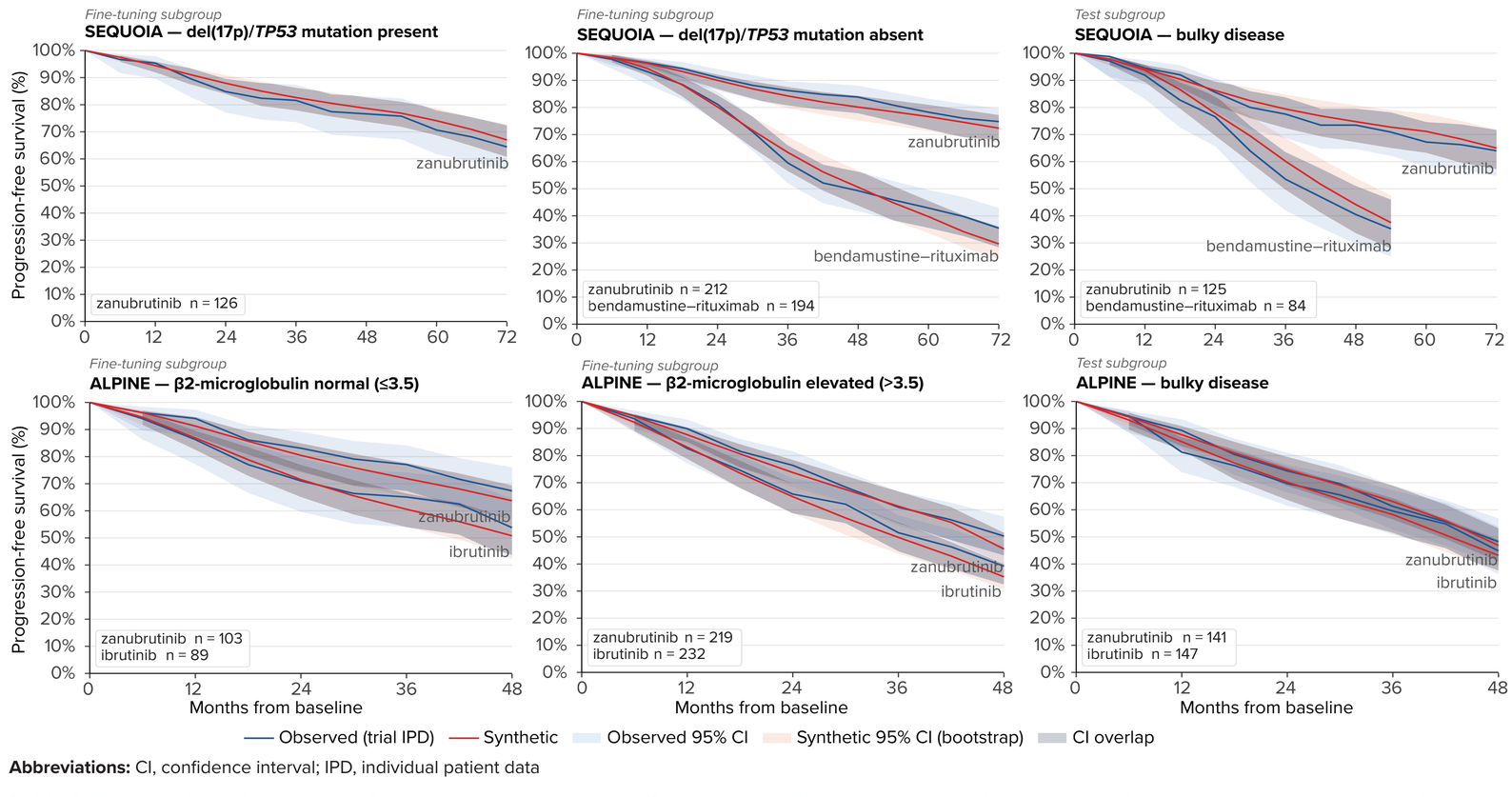
RESULTS

Fidelity of synthetic survival curves

- Across 30 subgroups, both trial arms and up to six years of follow-up, the median absolute difference in survival probability was **0.016 for progression-free survival (PFS) PFS and 0.012 for overall survival (OS) OS**, with band containment of **92.7%** and **92.4%** (**Table 1**; **Figure 1**)

Figure 1. Synthetic versus observed progression-free survival in six example subgroups

n = observed trial patients. The SEQUOIA del(17p) cohort was non-randomized (zanubrutinib only). Curves truncated when <25 patients remained at risk



Abbreviations: CI, confidence interval; IPD, individual patient data

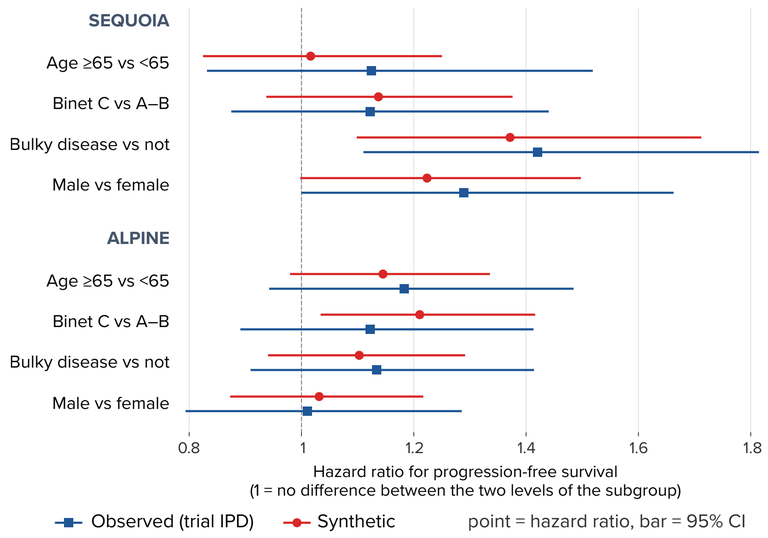
Table 1. Fidelity of synthetic survival curves: summary across all subgroups, pooled over evaluation timepoints

Subgroup set	Endpoint	Time-points	Median |Δ|	Max |Δ|	Band contain-ment (%)	Observed inside band (%)
All (30)	PFS	541	0.016	0.147	92.7	90.8
Fine-tuning (16)	PFS	277	0.020	0.147	91.7	87.0
Test (14)	PFS	264	0.014	0.107	93.7	94.7
All (30)	OS	549	0.012	0.115	92.4	84.2
Fine-tuning (16)	OS	282	0.011	0.115	92.9	79.8
Test (14)	OS	267	0.012	0.070	91.8	88.8

Abbreviations: PFS, progression-free survival; OS, overall survival

- Performance in test subgroups was comparable to the subgroups used for fine-tuning: median absolute difference 0.014 versus 0.020 for PFS, band containment 93.7% versus 91.7%, and the observed estimate inside the synthetic band at 94.7% versus 87.0% of timepoints. Residual error was concentrated in comparator arms, in smaller subgroups and at later follow-up times
- Across the 16 prognostic contrasts, synthetic and observed log hazard ratios agreed closely, to a median of 0.052 log-HR units for PFS and 0.071 for OS, equivalent to hazard ratios matching within about 5% and 7%. All 16 PFS contrasts agreed in direction, and all 32 pairs of confidence intervals (CI) largely overlapped
- The eight contrasts defined on characteristics withheld from fine-tuning are shown in **Figure 2**: the synthetic and observed estimates agreed in direction throughout, and synthetic CIs were largely contained within the observed CIs

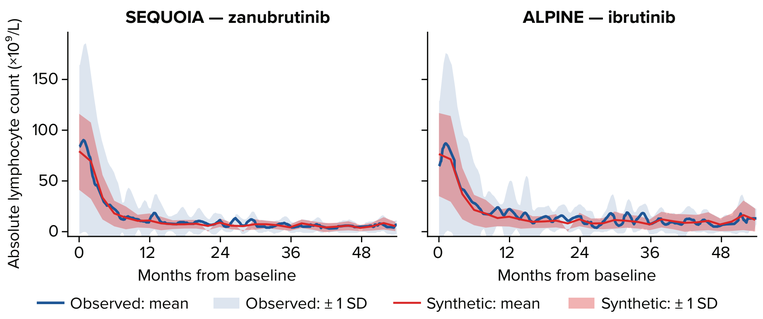
Figure 2. Hazard ratios for PFS between test subgroups



Fidelity of laboratory trajectories

- Mean laboratory trajectories tracked the observed data closely, although between-patient variability was lower in generated data (e.g., absolute lymphocyte count [ALC]; **Figure 3**)

Figure 3. ALC, synthetic versus observed (mean ± 1 standard deviation [SD]): two examples



Limitations

- **Computation.** 400 model training runs per analysis. Analysis (training nested in the bootstrap) mean long run times, which constrain how frequently the framework can be re-tuned
- **Laboratory trajectory variability.** Between-patient variability was underestimated; however, part of this variability is measurement error, which the model does not aim to reproduce
- **Time horizon.** Trajectory generation was validated over a follow-up period of up to 6 years; longer-term trajectories will require additional validation as data become available

References

1. Elvatun S et al. *PLOS Digit Health*. 2025;4(1):e0000581. **2.** Aballéa S et al. *Value Health*. 2025;28:S541. 3. El Kababji S et al. *J Med Internet Res.* 2025;27:e66821. **4.** Eckardt JN et al. *NPJ Digit Med.* 2024;7(1):76. **5.** D'Amico S et al. *JCO Clin Cancer Inform.* 2023;7:e2300021. **6.** Piciocchi A et al. *eJHaem.* 2024;5(2):353–9. **7.** Tam CS et al. *Lancet Oncol.* 2022;23(8):1031–43. **8.** Brown JR et al. *N Engl J Med.* 2023;388(4):319–32. **9.** Kingma D, Welling M. *arXiv*:1312.6114; ICLR 2014. **10.** Chen RTQ et al. *Adv Neural Inf Process Syst.* 2018;31:6571–83. **11.** Rubanova Y et al. *Adv Neural Inf Process Syst.* 2019;32. **12.** De Brouwer E et al. *Adv Neural Inf Process Syst.* 2019;32. **13.** Moon I et al. *Proc Mach Learn Res (MLHC).* 2022;182. **14.** Nowok B et al. *J Stat Softw.* 2016;74(11):1–26.

Acknowledgments

The authors thank the patients and their families, investigators, co-investigators, and the study teams at each of the participating centers. This study was sponsored by BeOne Medicines, Ltd. Editorial assistance was provided by SNELL Medical Communication Inc.

Disclosures

TM: Consulting: AbbVie, Alexion Pharmaceuticals, AstraZeneca, BeOne Medicines, Janssen-Cilag, Lilly, Roche, SOBI; Speakers' bureau: AbbVie, Alexion Pharmaceuticals, AstraZeneca, Gilead Sciences, Janssen-Cilag, SOBI; Travel or meeting support: AbbVie, AstraZeneca, Alexion Pharmaceuticals, Janssen-Cilag; Research support: AbbVie, Janssen-Cilag; **PS:** Honoraria: AbbVie, AstraZeneca, BeOne Medicines, J&J, Lilly; Data safety or advisory board: AbbVie, AstraZeneca, BeOne Medicines; **LM:** Employee and stockholder: BeOne Medicines; **SO:** Consulting: AbbVie, AstraZeneca, BeOne Medicines, J&J, Lilly; Honoraria: AbbVie, AstraZeneca, BeOne Medicines, J&J; Travel or meeting support: AstraZeneca, BeOne Medicines; Data safety or advisory board: Merck; Research grants: AbbVie, AstraZeneca, BeOne Medicines, J&J, Roche; **IC, SA, RZ:** Research support: BeOne Medicines; **NK:** Nothing to disclose; **MT:** Research support: BeOne Medicines; **ML:** Employee and stockholder: BeOne Medicines, Roche; **RW, SX:** Employee and stockholder: BeOne Medicines; **AF:** Consulting: AstraZeneca, BeOne, Janssen, Lilly; Honoraria: AbbVie, AstraZeneca, BeOne Medicines, Janssen, Lilly; Travel or meeting support: AbbVie, BeOne Medicines, Janssen, Lilly; **LY:** Consulting: AbbVie, AstraZeneca, BeOne Medicines, BMS, Gilead/Kite, J&J, Lilly, Roche; Honoraria: AbbVie, AstraZeneca, BeOne Medicines, BMS, Gilead/Kite, J&J, Lilly, Roche; Research grants: AbbVie, AstraZeneca, BeOne Medicines, BMS, Gilead/Kite, J&J, Lilly, Roche; **MS:** Consulting or advisory role: AbbVie, ADC Therapeutics, Ascentage, AstraZeneca, BeOne Medicines, Bristol Myers Squibb, Eli Lilly, Fate Therapeutics, Genentech, Genmab, Janssen, Kite Pharma, Merck, MorphoSys/Incyte, Nurix, Pfizer, Pierre Fabre; Research support: AstraZeneca, BeOne Medicines, Genentech, Genmab, Janssen, Merck, MorphoSys/Incyte, Nurix, Sana Biotechnology; Stock: Koi Biotherapeutics; Employment: BMS.

CORRESPONDENCE: Samuel Aballéa, sab@inovintell.com

Presented at ISPOR EU-26; November 8-11, 2026, Vienna, Austria