Neuro-Oncology Advances

7(1), vdaf145, 2025 | https://doi.org/10.1093/noajnl/vdaf145 | Advance Access date 1 July 2025

Deep learning volumetrics reveal distinct clinical trajectories for pediatric low-grade gliomas under surveillance: A multicenter study

Juan Carlos Climent Pardo, Anna Zapaishchykova, Aidan Boyd, Divyanshu Tak, John Zielke, Maryam Mahootiha, Zezhong Ye, Sridhar Vajapeyam, Jacquelyn Jones, Ceilidh Smith, Ariana M. Familiar, Ali Nabavizadeh, Pratiti Bandopadhayay, Sabine Mueller, Hugo J.W.L. Aerts, Daphne A. Haas-Kogan, Franziska Michor, Keith L. Ligon, Tina Y. Poussaint, Shahrooz Faghihroohi, Benjamin H. Kann

All author affiliations are listed at the end of the article

Corresponding Author: Benjamin H. Kann, MD, Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, 221 Longwood Avenue, Suite 442, Boston, MA, 02115, MA, USA (bejamin_kann@dfci.harvard.edu).

Abstract

Background: Pediatric low-grade gliomas (**pLGGs**) have heterogeneous clinical presentations, and given the morbidity of treatment, some patients receive observation with magnetic resonance (**MR**). The natural histories of untreated pLGGs remain understudied. We leveraged deep learning-based volumetrics to analyze longitudinal growth trajectories and progression risk factors for untreated pLGGs.

Methods: We conducted a pooled, retrospective study of radiographically diagnosed pLGG patients from two institutions diagnosed between 1992 and 2020 who were surveilled for at least 1 year post-diagnosis. Tumor segmentation was applied to longitudinal T2-weighted MR to calculate 3D tumor volumes. We assessed volume trajectories, disease progression, and associated risk factors using Cox-Hazards regression, survival analysis, and time-series forecasting with autoregressive integrated moving average (ARIMA). Patients were categorized based on volumetric changes into progression (≥25%), regression (≤–25%), or stability.

Results: Of 99 patients (970 scans; median follow-up: 7.0 years; median diagnosis age: 12.0 years), 55 (55.5%) had tumors that volumetrically progressed, 28 (28.3%) remained stable, and 16 (16.2%) regressed. 42 (42.4%) patients initiated treatment. Risk factors associated with progression included infancy/preschool age, cortical location, and female sex ($p \leq 0.05$ for each). Most progressions occurred within five years of diagnosis (80.0%), most commonly in school-aged children (7-13 years old). Time-series forecasting predicted future tumor volume with a mean absolute error of 2.04 cm^3.

Conclusion: Deep learning enables systematic, longitudinal, pLGG growth tracking and characterization of patients on surveillance, yielding insights into untreated tumor trajectories and progression risk. This pipeline is useful at population-level to study growth trends and at patient-level to guide personalized management.

Key Points

- Deep learning enables longitudinal volumetric analysis of untreated, surveilled pLGG patients
- 55.5% of tumors progressed, 28.3% were stable, 16.2% regressed over 7-years median follow-up
- Pipeline developed for longitudinal volumetric tracking applicable to various cancers

© The Author(s) 2025. Published by Oxford University Press, the Society for Neuro-Oncology and the European Association of Neuro-Oncology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Importance of the Study

This is the first study to integrate deep learning volumetrics, longitudinal imaging analysis, and time-series forecasting analysis to systematically investigate the natural histories of pediatric low-grade gliomas (pLGGs). By leveraging a validated deep learning segmentation algorithm, the research enables comprehensive volumetric analysis of pLGGs under surveillance, uncovering distinct tumor phenotypes and their associated clinical risk factors. The study's approach, combining various statistical and machine learning techniques, enhances our understanding of pLGG progression patterns and 3D volume-based predictive capabilities. The study provides new insights into pLGG growth dynamics and yields a modular software pipeline applicable to various longitudinal segmentation models, to enable generation of patient-level volumetric tracking to guide tailored surveillance and management strategies.

Pediatric low-grade gliomas (pLGGs) are the most common pediatric brain tumors, constituting up to 40% of tumors in children.[1–5] They are less aggressive than their high-grade counterparts[6,7] and heterogeneous in their clinical presentation, location, histology, genomics, and natural history.[1,2,5,8,9] Given the long survivorship horizon of patients with pLGGs and the morbidity associated with aggressive therapies, clinical management is complex, balancing preservation of quality of life with the pursuit of disease control.[10–12] While some pLGGs will require surgery and, sometimes, combined modality therapy,[13] others are found incidentally, present with minimal symptoms, or located in areas that preclude safe resection. For these patients, upfront surveillance is often pursued with periodic magnetic resonance (**MR**). Some of these patients may never require treatment, while others, sometimes soon after diagnosis, will develop growing tumors and/or symptoms requiring intervention,[14] which may include surgery, radiation, chemotherapy, or targeted agents.[15–20]

While it is felt that initial surveillance in many of these patients is safe, the natural histories of these tumors have never been systematically studied, leaving several open questions regarding how pLGGs change longitudinally, and who will ultimately progress and require therapy. A major barrier to longitudinal analysis is that pediatric brain tumors are not quantitatively annotated and analyzed in routine clinical practice, largely due to resource, time, and expertise constraints.[21–23] Accordingly, there have been only a few studies of longitudinal volumetric trajectories for pLGGs to date.[24]

In recent years, imaging-based deep learning (**DL**)[25,26] has demonstrated the ability to accurately segment and calculate 3D brain tumor volumetrics,[27,28] including in pediatric low-grade gliomas.[29,30] Building upon the work in,[29] we sought to evaluate the validated DL segmentation tool in a large, multicenter cohort of untreated, clinically suspected pLGGs under MR surveillance.

We hypothesized that the application of DL segmentation algorithms to a multicenter, longitudinal pLGG surveillance cohort would enable a systematic study of natural histories of surveilled tumors, reveal distinct phenotypes, and discover risk factors for progression. We demonstrate the potential for DL to enable practical, longitudinal, tumor volumetric tracking and analysis for individual patients under surveillance and at scale. Finally, we investigate the potential for statistical modeling of volumetric trajectories to predict future growth dynamics of tumors to better inform clinical management.

MATERIALS AND METHODS

Study Design and Datasets

This study was conducted in accordance with the Declaration of Helsinki guidelines[31] and after the Institutional Review Board (**IRB**)'s approval. Waiver of consent was obtained from IRB prior to research initiation due to use of public datasets and minimal risk of this retrospective study. We report our results in accordance with the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) guidelines.[32] The study includes patients with radiographically suspected and clinically diagnosed pLGG diagnosed between 1992 and 2020 who underwent initial MR surveillance for at least one year prior to biopsy or treatment pooled from two institutional datasets: Dana-Faber Cancer Institute/Boston Children's Hospital (DF/BCH) and the Children's Brain Tumor Network (CBTN). All patient imaging was collected from diagnosis up until progression, treatment receipt, or last clinical follow-up, whichever came first. Patients with optic gliomas were excluded due to their exclusion from training data of the auto-segmentation algorithm, as were those with neurofibromatosis syndrome due to expected differences in natural histories compared to traditional pLGGs. The inclusion rationale can be found in Supplemental A.1. Radiographic diagnosis of pLGG was based on established T2-weighted MRI characteristics including well-circumscribed hyperintense lesions, absence of significant mass effect or invasive features, anatomic location consistent with pLGG, and slow or absent growth on serial imaging prior to inclusion.

Imaging Quality Control

All T2-weighted MR imaging data available was converted from native DICOM to NifTI format and underwent manual review jointly by a trained research fellow and board-certified radiation oncologist (B.H.K.) to assess the validity and quality (i.e. artifact, incorrect body part, wrong sequence) of T2W sequences prior to algorithm implementation. Following review, patients were excluded if they had

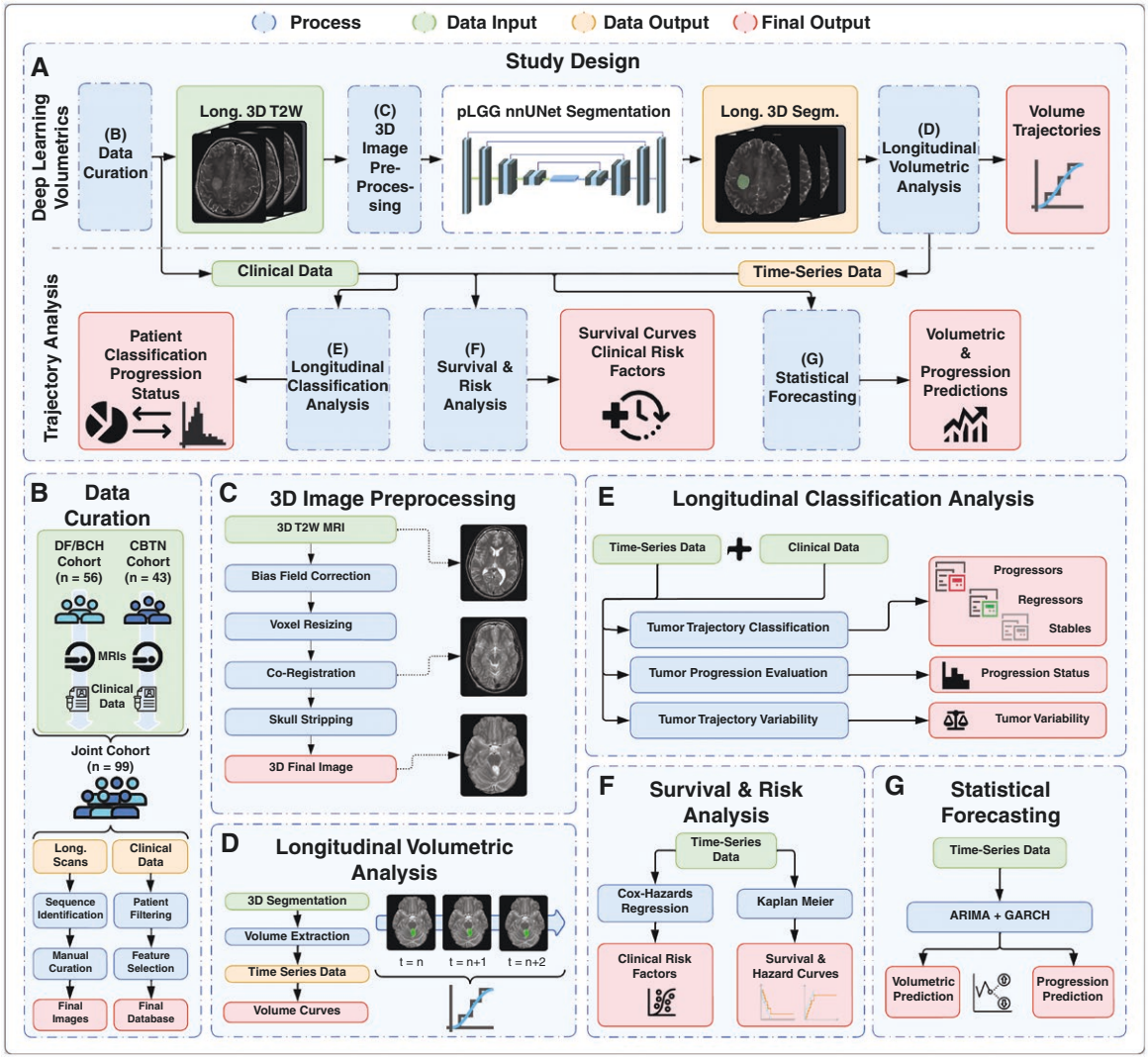

Figure 1. Study design overview (A) separated into its two main steps, deep learning (DL) Volumetrics and Trajectory Analysis. Moreover, (B–G) show the individual, detailed pipeline modules. Across the entire figure the same color coding is utilized: developed processes are marked in blue, input data in green, intermediate results in orange, and final outputs in red. Data Curation (B) is conducted as the first step in deep learning (DL) Volumetrics, including a manual scan review to assess the correct sequences and the quality of clinical features. The selected T2W images are preprocessed in (C) and then segmented through the validated pLGG auto-segmentation algorithm. The output of the segmentation algorithm consists of 3D binary masks used in the Longitudinal Volumetric Analysis module (D) to extract the volumes, plot the volume trajectories over time for each patient, and create a time-series database. The second step, Trajectory Analysis, is comprised of the Longitudinal Classification Analysis (E), which classifies patients based on their trajectories and investigates the progression status over time, the Survival & Risk Analysis (F), which assesses clinical risk factors and survival probabilities, and the Statistical Forecasting (F), which performs volumetric and progression predictions through ARIMA and GARCH models.

less than three MR scans to facilitate longitudinal trend analyses (See Supplemental A.1 for further cohort selection details).

Clinical risk factors such as tumor location, tumor histology, symptoms at diagnosis, and treatment details were abstracted by trained clinical research coordinators. To facilitate mutational subtype analyses, tumor BRAF mutational status was inferred radiographically from a BRAF mutational status DL prediction algorithm, developed, and validated in prior work.[33] For the inference, each longitudinal image was processed through a 3D image pipeline and the average prediction was assigned as the BRAF mutational status. Statistical tests were performed to verify significant differences between CBTN and DF/BCH datasets and to determine whether they could be pooled reasonably. MRI acquisition details can be found in Supplemental A.2.

DL Pipeline

A multistep pipeline was developed for the study (Figure 1A). The pipeline, in full, is published open source for the

scientific and clinical communities under (https://github.com/jc-cp/mri-longitudinal-analysis).

DL Volumetrics Following data curation and quality review (Figure 1B), T2-weighted images underwent preprocessing (Figure 1C), including bias field correction with the Simple-ITK package,[34] voxel resizing to isotropic resolution (1 × 1 × 1mm), image co-registration (to an unbiased standard MR imaging template of brain volume for pediatric data in the 4.5 to 18.5y age range from NIH[35]), and skull stripping via the HD-BET package.[36]

An nnUNet-based 3D tumor auto-segmentation model, previously developed, externally validated, and clinically benchmarked for pLGGs[29] was applied to all images and outputs corresponding 3D tumor segmentations longitudinally. All segmentations were reviewed by a board-certified radiation oncologist (B.H.K.) and cases with failed segmentations or without visible tumor were removed from the analysis (Supplemental A.1, second Q&A).

The obtained segmentations were used in the *Longitudinal Volumetric Analysis* (Figure 1D) to extract the tumor volumes in mm^3, which were subsequently plotted over time (patient age at scan date) for each patient to reveal volumetric tumor trajectories, in absolute terms and normalized to initial tumor volume. The output also included a time-series dataset for each patient, containing tumor volume *V*, volume change between scans ($\Delta V = V_{1-}V_2$, where V_2 is the volume at the later timepoint than V_1), and volume change rate ($rate = \frac{\Delta V}{\Delta t}$, where Δt is the time difference). Additionally, a moving average smoothing technique was applied to each trajectory.

Volumetric Trajectory Analysis The extracted patient-based time-series data were merged with individual clinical data, which contained information about patient age and sex, tumor location, symptomatic presentation, and later received treatment if that was the case (Figure 1E). This enabled tumor trajectory categorization as progressors, stables, or regressors, according to two endpoints, volumetric progression, and composite progression. Volumetric progression was defined as a 25% or more volumetric increase at any point, and volumetric regression as a volume decrease of 25% or more, consistent with 2D RAPNO criteria[37] and 3D RAPNO recommendations.[24] Patients with volume changes within, but never exceeding, these thresholds were considered stable. Composite progression was assigned when a patient had either volumetric progression and/or received tumor-directed treatment. The longitudinal volumetric and composite progression statuses were assessed from the time since diagnosis. Intra-patient tumor variability from scan-to-scan was also evaluated using the coefficient of variation (CoV), as follows: $CV = \frac{\sigma}{\mu}$, where σ is the standard deviation and μ the mean of the volumes over the course of a tumor's trajectory.

Progression Risk Factors and Forecasting Clinical risk factors for progression were investigated with univariable and multivariable proportional Cox hazards models. Baseline variables like age at diagnosis, tumor location, tumor baseline volume, and patient sex were investigated for association with volumetric and composite progression. Additionally, BRAF mutational status (wildtype, fusion, or V600E) was inferred for all patients based on a previously validated, imaging-based mutational status prediction tool.[33] The time elapsed between diagnosis and composite progression, and between a smaller volume change (increase of ≥ 10%) and volumetric progression, were also analyzed. Composite progression-free survival (PFS; from initial scan date) was evaluated with the Kaplan–Meier survival method and compared across clinical variables with log-rank tests (Figure 1F) to account for censoring. The survival curves were collated with the survival and hazard estimates from the multivariate Cox-regression for completeness. Finally, we analyzed the mutational status of tumors that were ultimately resected, as available.

We applied predictive forecasting modeling of volumetric trajectories to determine if past tumor volumetrics could accurately predict future states. A hybrid statistical time-series algorithm combining autoregressive integrated moving average (ARIMA)[38] from the *statsmodel* package[39] and generalized autoregressive conditional heteroskedasticity (GARCH)[40] form the *arch* package[41] were evaluated to predict future volumetric changes. Each volumetric trajectory extracted from the MRI scans was interpolated and separated into two segments. The first 80% of each trajectory was used to fit ARIMA and ARIMA + GARCH models. The last 20% of the trajectory was used for validation of those fittings. Both models were evaluated in the validation segment through different error and model selection metrics such as MSE, RMSE, MAE, AIC, BIC, and HQIC. Finally, both methods were used to perform out-of-the-curve forecasts, which indicated the mean direction and confidence of future volume change per trajectory. Performance of both models was compared in validation fitting and out of the curve predictions.

Statistical Analysis Cohort compatibility for a pooled analysis was assessed through several statistical tests. For categorical variables, Fisher's exact test was used to compare the proportions between the cohorts if both cohorts had binary categories. If either cohort had more than two categories, the Chi-squared test was used to compare the distributions between the cohorts. For continuous variables, if both cohorts passed the normality assumption (assessed using the Shapiro–Wilk test), the independent *t*-test was used to compare the means between the cohorts. If either cohort failed the normality assumption, the Mann–Whitney *U* test was used to compare the distributions between the cohorts. Statistical analyses were conducted using the SciPy software package.[42] A p-value < 0.05 was considered significant (Table 1).

For risk factor association, continuous variables were scaled by removing the mean and scaling to unit variance (later marked with (+) in Figure 4A). After evaluating the variables in the univariate model, the variance inflation factor was investigated to assess multicollinearity among the independent variables. Statistically significant differences between the age groups were calculated through a Kruskal–Wallis H-test, while the post hoc

Table 1. The cohort is comprised of 99 patients surveilled post-diagnosis and before treatment initiation. We considered the longitudinal variables of age and volume, and baseline clinical risk factors such as sex, tumor location, symptomatic presentation, BRAF mutational status (which was radiographically inferred) and the received treatment (which is the eventual treatment after at least 1-year surveillance). For all statistical analyses, a two-sided p-value < 0.05 was considered statistically significant. *More information on the concrete mutational status for the subset of patients who ultimately underwent surgery can be found in Supplemental A.6.

Category	Boston Children's Hospital (BCH)	Children's Brain Tumor Network (CBTN)	Pooled dataset	*P*-values
Cohort (n)				
Patients	56	43	99	
Scans	504	466	970	
Median / avg. per patient	7.0/ 9.0	7.0/ 10.8	7.0/ 9.8	
Clinical follow-up				
Median, years (range)	7.0 (1.0-19.0)	6.0 (1.0-17.0)	7.2 (1.0-19.5)	.745
Median time between scans, months (range)	4.14 (0.03-110.29)	3.35 (0.03-98.88)	3.68 (0.03-110.29)	.225
Tumor volume				.096
Volume median, cm^3 (range)	3.27 (0.03-59.28)	5.11 (0.33-39.36)	4.18 (0.03-59.28)	.101
Normalized volume median, cm^3 (range)	1.00 (0.03-7.48)	1.09 (0.10-21.16)	1.01 (0.03-21.16)	.321
Baseline volume median, cm^3 (range)	3.44 (0.04-59.28)	4.42 (0.55-25.99)	3.91 (0.04-59.28)	.221
Age				.318
Median, years (range)	11.9 (1.08-28.92)	12.1 (0.50-28.01)	12.0 (0.50-28.92)	.659
Median at diagnosis, years (range)	7.6 (1.08-19.05)	7.0 (0.50-16.91)	7.5 (0.50-19.05)	.040
Age group at diagnosis, n (%)				
Infant [0-2y]	4 (7.14%)	9 (20.93%)	13 (13.13%)	.048
Preschool [3-6y]	7 (12.50%)	11 (25.58%)	18 (18.18%)	
School Age [7-13y]	29 (51.78%)	17 (39.53%)	46(46.46%)	
Adolescent [14-18y]	14 (25.00%)	6 (13.95%)	20 (20.20%)	
Young Adult [18 + y]	2 (3.57%)	0 (00.00%)	2 (2.02%)	
Sex, n (%)				
Female	26 (46.42%)	19 (44.18%)	54 (54.5%)	.842
Male	30 (53.57%)	24 (55.81%)	45 (45.5%)	
Location, n (%)				
Basal ganglia /Thalamus	10 (17.85%)	5 (11.63%)	15 (15.15%)	.465
Cerebellum	8 (14.28%)	11 (25.58%)	19 (19.19%)	
Cortical	18 (32.14%)	10 (23.25%)	28 (28.28%)	
Midline	19 (33.92%)	15 (34.88%)	34 (34.34%)	
Other	1 (1.78%)	2 (4.65%)	3 (3.03%)	
Presentation, n (%)				
Asymptomatic (Incidentally found)	7 (12.50%)	17 (39.53%)	24 (24.24%)	.004
Symptomatic	49 (87.50%)	26 (60.47%)	75 (75.75%)	
*BRAF status, n (%)**				
Wildtype	38 (67.85%)	24 (55.81%)	62 (62.62%)	.046
V600E	7 (12.5%)	14 (32.55%)	21 (21.21%)	
Fusion	11 (19.64%)	5 (11.62%)	16 (16.16%)	
Received treatment, n (%)				
Yes	27 (48.21%)	15 (34.88%)	42 (42.42%)	.221
No	29 (51.87%)	28 (65.11%)	57 (57.57%)	
Treatment type, n (%)				
Surgery only	22 (39.28%)	8 (18.60%)	30 (30.30%)	.071
Chemotherapy only	–	2 (4.65%)	2 (2.02%)	
Radiation only	–	–	-	
Surgery + chemo.	3 (5.35%)	4 (9.30%)	7 (7.07%)	
Surgery + radiation	–	1 (2.32%)	1 (1.01%)	
Chemo. + radiation	–	–	–	
No treatment at all	29 (51.78%)	28 (65.11%)	57 (57.57%)	
Surgery + radiation + chemo.	2 (3.57%)	–	2 (2.02%)	

Dunn's test specified that all groups differ from each other. A p-value < 0.05 was considered significant. *BRAF mutational status was inferred via a previously-validated deep learning algorithm for the purposes of this study.

RESULTS

Patient and Tumor Characteristics

We identified 99 patients (n = 56 DF/BCH; n = 43 CBTN) with 970 unique MR scans (median 7 scans per patient, Supplemental A.1) who met eligibility criteria and passed the quality review for the study (Supplemental A.1). Median imaging and clinical follow-up were 7.2 years (range: 1.0-19.5) (Table 1, Supplemental A.1) with a median time interval between scans of 3.68 months (range: 0.03-110.29 months) (Supplemental A.1). The median tumor volume at diagnosis (baseline) was 3.91cm^3 (range: 0.04-59.28.). Patient and tumor characteristics were similar between the cohorts, except for median and grouping age at diagnosis, symptomatic presentation, and inferred BRAF mutational status ($p < 0.05$ for each; Table 1). Midline location comprises brainstem-related, ventricular, meningeal, and suprasellar tumors.

Volumetric Trajectories and Patterns for surveilled pLGGs

Of all patients, 55 (55.5%) had volumetric tumor progression, 28 (28.3%) had stable tumors, and 16 (16.2%) had tumor regression (Figure 2A.1 and C.1). For these three classes, the median normalized volume change per scan was 0.145cm^3 (5.65%), 0.001cm^3 (< 0.01%), and –0.077cm^3 (-2.69%), respectively. In total 42 patients (42.4%) initiated treatment (Figure 2A.2 and C.1), out of which 95.2% underwent surgery. No patients died under surveillance. Out of the patients who had volumetric progression (n = 55), 43.6% (n = 24) underwent treatment, and out of the patients who did not volumetrically progress (n = 44), 40.9% (n = 18) underwent treatment (Figure 2C.2). A total of 73 patients (73.7%) progressed considering the composite endpoint (i.e. via volumetrics or receipt of treatment) (Figure 2A.3 and C.1). The DL pipeline enabled generation of patient-level volumetric trajectories, highlighting distinct natural histories and tumor phenotypes like the described progressors (Figure 2B.1) and regressors (Figure 2B.2), but also including 14 patients (14.14%%) who had tumors that initially progressed and then regressed, i.e. a "waxing-waning" phenotype (Figure 2B.3).

Of 55 patients with volumetric progression, 5 (9.09%) progressed during infancy (ages 0-2), 10 (18.18%) during preschool ages (ages 3-6), 28 (50.90%) progressed during school ages (ages 7-13 years), 10 (18.18%%) during adolescence (ages 14-18), and 2 (3.63%) during young adulthood (age >=18) (Figure 3A.1). The oldest age at progression was 21 years. Most patients progressed within five years of diagnosis (44 patients, 80.0%), and 20 patients (36.36%) progressed within the first year of diagnosis (Figure 3A.2). Patterns were similar for composite progression. (Figure 3A.3-4). The progression status stratified by tumor location showed midline and cortical tumors having the highest amount of progressors (Supplemental A.3). The median time to volumetric progression was 1.1 years (mean and std. dev.: 2.16 ± 2.1y), whereas for composite progression, it was 1.5 years (mean and std. dev.: 2.5y ± 2.8y) (Figure 4B). For tumors that ultimately progressed volumetrically, median time from 10% volume increase to 25% volume increase was 0.38 years (mean and std. dev.: 0.86y ± 0.91y).

Tumor volume change from scan to scan was generally low across the cohort, with only a few cases showing high variability (CoV > 0.7) with a mean coefficient of variation of 0.353 (std. dev. ± 0.250) and a rightward skew in the distribution (Figure 3B). Tumor volume variability was highest in Infants and Preschool patients (ages 0-2 and 2-6, respectively), with the median CoV generally decreasing over the temporal ordered age groups (Infants: 0510, Preschool: 0.525, School Age: 0.350, Adolescents: 0.264, and Young Adults: 0.486, $p < 0.001$ and H-Statistic = 151.88) (Figure 3B). Differences between groups were all $p < 0.05$ except for Young Adults who present $p > 0.05$ compared to all groups except vs Adolescents.

Risk Factors for Progression & Time to Event

On univariable Cox regression, all variables besides the adolescent age group and inferred BRAF Fusion mutational status had significant associations with composite progression ($p < 0.05$ for each, Figure 4A). No variable exceeded a variance inflation factor of 1.5, suggesting no multicollinearity between variables. On multivariable analysis, factors associated with composite progression were younger age at diagnosis (infancy HR: 38.22, $p < 0.01$ and preschool ages HR: 8.71, $p < 0.01$, respectively), cortical tumor location (HR: 2.37, $p < 0.01$, respectively), and female sex (HR: 1.74, $p < 0.01$). Factors associated with lower risk of progression were older age at diagnosis (adolescence HR: 0.22, $p < 0.01$, and young adulthood HR: 0.26, $p < 0.01$), inferred BRAF V600E and inferred BRAF Fusion mutational status (HR: 0.49, $p < 0.01$, and HR: 0.39, $p < 0.01$, respectively), and baseline tumor volume modeled as a continuous variable (HR: 0.67, $p < 0.01$). The multivariable model showed an internal discriminatory performance of C-index 0.735 (95% CI: 0.716 to 0.754) in predicting composite progression. Similar trends were found for the volumetric progression endpoint (Supplemental A.5). Additional hazard and derived survival curves from Cox-regression and a Kaplan–Meier estimator are in Supplemental A.4 and A.5.

The median time to volumetric progression was 1.09 years (mean and std. dev.: 2.06y ± 2.09y), whereas for composite progression it was 1.49 years (mean and std. dev.: 2.53y ± 2.84y) (Figure 4 B). For tumors that ultimately progressed volumetrically, median time from 10% volume increase to 25% volume increase was 0.38 years (mean and std. dev.: 0.86y ± 0.91y).

For the BCH patients who ultimately underwent surgery (n = 27), 21 (77.7%) had genomic data available via next generation sequencing or targeted FISH assays. Of these, NF1 alterations were most prevalent (n = 6, 28.57%), followed by BRAF aberrations including V600E mutations (n = 3, 14.28%), fusions (n = 3, 14.28%), duplications (n = 2, 9.52%), and one case with concurrent BRAF V600E and NF1 mutations (4.76%), while FGFR1 and IDH1 mutations each occurred in two cases (9.52% each). The remaining cases (n = 2, 9.52%) were BRAF wildtype, without other

identified driver mutation (Supplemental A.6). For CBTN patients, the mutational status was gathered for 23 out of the 43 total patients. NF1 alterations were most prevalent (n = 6, 28.08%) although the patients hadn't been clinically diagnosed with NF1 before, followed by BRAF Fusion and BRAF V600E mutations (each n = 5, 11.62%). There were 4 (17.39%) BRAF wildtype, 2 (4.65%) BRAF V600E with CDKN2A/B alterations and 1 (2.32%) case of FGFR1. Notably, NF1 alterations were the most prevalent molecular finding in surgically treated patients from both cohorts (28.57% BCH, 28.08% CBTN), despite clinical exclusion of NF syndrome. This suggests potential germline or somatic NF1 alterations in patients without sufficient clinical manifestations for syndrome diagnosis. Limited sample size precludes definitive conclusions.

Volumetric Forecasting Analysis

For volumetric forecasting, all individual trajectories were interpolated before applying the time-series models (Figure 5A). ARIMA marginally demonstrated better performance in forecasting volume changes during rolling validation over all trajectories in the cohort, considering traditional error metrics (winning 64.6% of cases, Wilcoxon $p < 0.01$) and information criteria assessments (Wilcoxon $p < 0.01$). The ARIMA + GARCH model showed superior practical utility with a mean absolute error of 2.04cm^3 (std. dev. ± 59.6cm^3), compared to ARIMA's 2.80cm^3 (std. dev. ± 92.7cm^3), representing a 27.1% improvement, although both models do not fully capture some outliers (Figure 5B). Systematic bias analysis revealed that both models tend to underpredict volumes during validation, with ARIMA + GARCH showing less systematic bias (mean difference: 1.06cm^3, 95% agreement limits: -115.8 to 118.0cm^3) compared to ARIMA (mean difference: 2.69cm^3, 95% agreement limits: -179.0 to 184.4cm^3), suggesting more consistent predictions. McNemar's test revealed no significant pattern in model superiority ($p > 0.05$). Both models demonstrated high reliability, with 96.0% and 94.6% of predictions within 0.5 cm^3 for ARIMA and ARIMA + GARCH, respectively. In terms of forecast uncertainty, ARIMA + GARCH produced more controlled CIs, with final widths of 0.85 ± 2.34 cm^3 compared to ARIMA's 2.48 ± 8.74 cm^3, indicating higher prediction precision in out-of-curve forecasting (Figure 5C). For forecasting absolute volume changes, both models predicted similar mean magnitudes (ARIMA: 0.39 ± 2.40cm^3; ARIMA + GARCH: 0.41 ± 2.42cm^3), with percentage changes of 8.2% (± 65.1%) and 5.6% (± 39.7%) respectively. Using again a ± 25% threshold on volume change forecasting, both models showed strong agreement, predominantly predicting volume progression in 62.6% of cases, with stability and regression in 29.3% and 8.1% of cases, respectively. Additional information on the model comparison, validation fitting and forecasting evaluation can be found in Supplemental A.7.

Discussion

Pediatric low-grade gliomas are heterogeneous in natural history and clinical presentation, leading to complex management. Some patients elect for upfront surveillance due to the morbidity associated with treatment, especially in the setting if patients are not symptomatic, but the clinical trajectories of these tumors have been unclear. Here, we present the first systematic, longitudinal volumetric analysis of pLGGs that were managed with upfront surveillance, enabled by the use of an AI tumor segmentation tool specialized for pLGG. Furthermore, we develop a scan-to-analysis pipeline that includes statistical modeling to track and predict tumor progression and make this publicly available for the oncologic community. Levering this pipeline, we revealed heretofore unknown characteristics and patterns of pLGGs managed with surveillance. We found that approximately 55% of pLGGs volumetrically progress, and that around 70% require treatment and/or progress at some point, usually within the first five years of diagnosis. Once patients reached 18 years old or had been surveilled for > 5 years, progression was exceedingly rare. The study yields several hypothesis-generating findings regarding risk factors for progression, including imaging-inferred BRAF wild type status, female sex, and younger age at diagnosis. Finally, using statistical forecasting, we demonstrate the feasibility of longitudinal volumetric patterns in predicting future trends. The findings of this study and resulting pipeline will help promote real-time, 3D volumetric assessments for pediatric brain tumors, providing clinicians and patients with data to inform prognosis and treatment decision-making.

This study marks the first longitudinal volumetric assessment of untreated pLGGs. While bidirectional 2D measurements remain the standard-of-care for response assessment in pLGG,[43] this is based on practicality of measurement, rather than biological basis, given the challenges and resource-intensiveness of 3D manual tumor segmentation. Recent studies have suggested that 3D volumetrics would provide a more standardized, accurate, and sensitivity measure of response compared to 2D measurements.[24,44–46] Recent investigations analyzed volumetric trends for manually segmented pLGGs in the setting of targeted therapy, yielding informative data regarding tumor burden during and post-therapy.[47] Prior work used a large database to yield insights about the natural histories of pLGG and probabilities of progression or treatment receipt.[48] This study, while representing a different population compared to our subset of surveilled tumors, found similarly that early growth behavior portended future progression and that diagnosis in infancy predicted for the worst outcomes.

Notably, our study found that larger tumors initially surveilled were less likely to show true volumetric progression (+ 25% increase). This may be due to the fact that larger tumors, in general, are less likely to be surveilled, and that a 25% increase in a large tumor represents substantially more of an absolute cellular increase, when compared to a smaller tumor. Quantitative 3D volumetrics will allow for the study of more granular, biologically based measurement of tumors, which may help risk-stratify and guide management in the future. The relatively high coefficients of variation (CoV) of the trajectories in early ages (infancy and preschool) are indicative of progression in later ages (school age, adolescence).

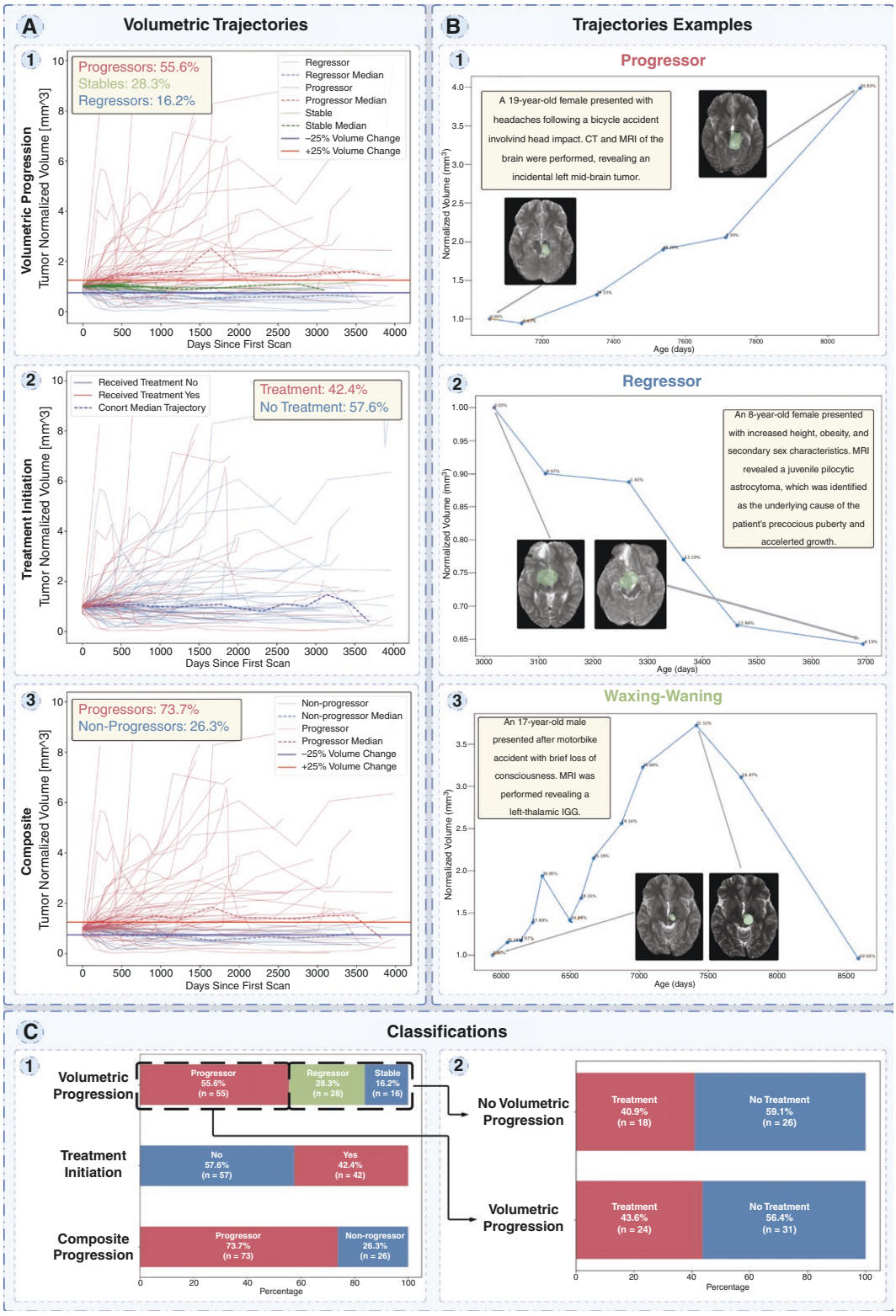

Figure 2: (A) depicts the joint cohort curves based on normalized volume, stratified by volumetric classification, treatment initiation, and composite classification. Volumetric classification defines progression and regression based solely on an increase/decrease of ± 25% from baseline volume at any time point. Treatment initiation indicates whether patients underwent treatment after surveillance. Composite classification combines volumetric classification and treatment initiation; patients with volumetric progression or treatment initiation were considered progressors. Volumetric classification yielded 55.5% progressors, 28.3% stable cases, and 16.2% regressors, while composite classification yielded 73.7% progressors vs. 26.3% non-progressors. Overall, 42.4% of patients initiated treatment. (B) shows exemplary curves for a volumetric progressor, a regressor, and a waxing-waning case (first volumetric progression and then tumor volume decrease), with corresponding axial views of T2W scans at diagnosis and selected follow-up scans. Brief case descriptions are included. (C) shows the percentages of the classifications and the treatment initiation, and a more granular view on the patients who did or did not volumetrically progress and undergo treatment.

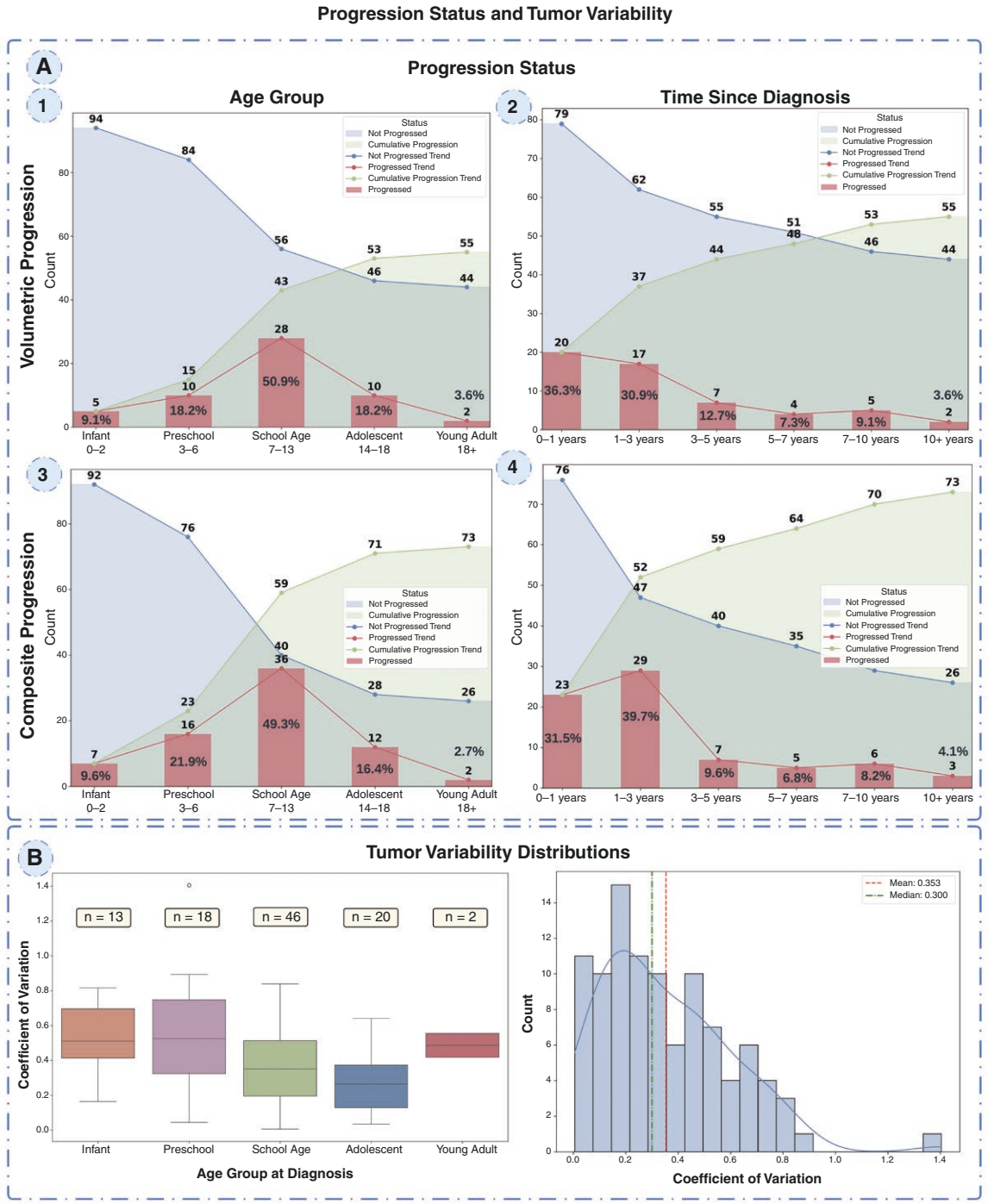

Figure 3. The graphs in (A) show the volumetric and composite progression status over time considering different age groups and the time transcurred since diagnosis. The non-progressed patients are shown in shaded blue, decreasing over the different time components, while the green shades show the cumulative progression over time. The red bars show the current progression magnitude at the specific time point period. (B) focuses on the longitudinal variability in the patient's curves, which is represented by the coefficient of variation (CoV). The right-skewed distribution hints at slow variability in general, with a few exceptions. The stratified coefficient of variation by the age groups at diagnosis reveals statistically significant differences assessed through a Kruskal–Wallis H-Test and a post hoc Dunn's test. All groups are different from each other (*P*-values < .05) except for the young Adults.

Beyond volumetric characterization, our study incorporates predictive forecasting conducted by statistical volumetric forecasting. While external validation of these techniques is critical, the study introduces a pipeline for further investigation. In our internally validated models, we find that statistical modeling on the initial 80% of

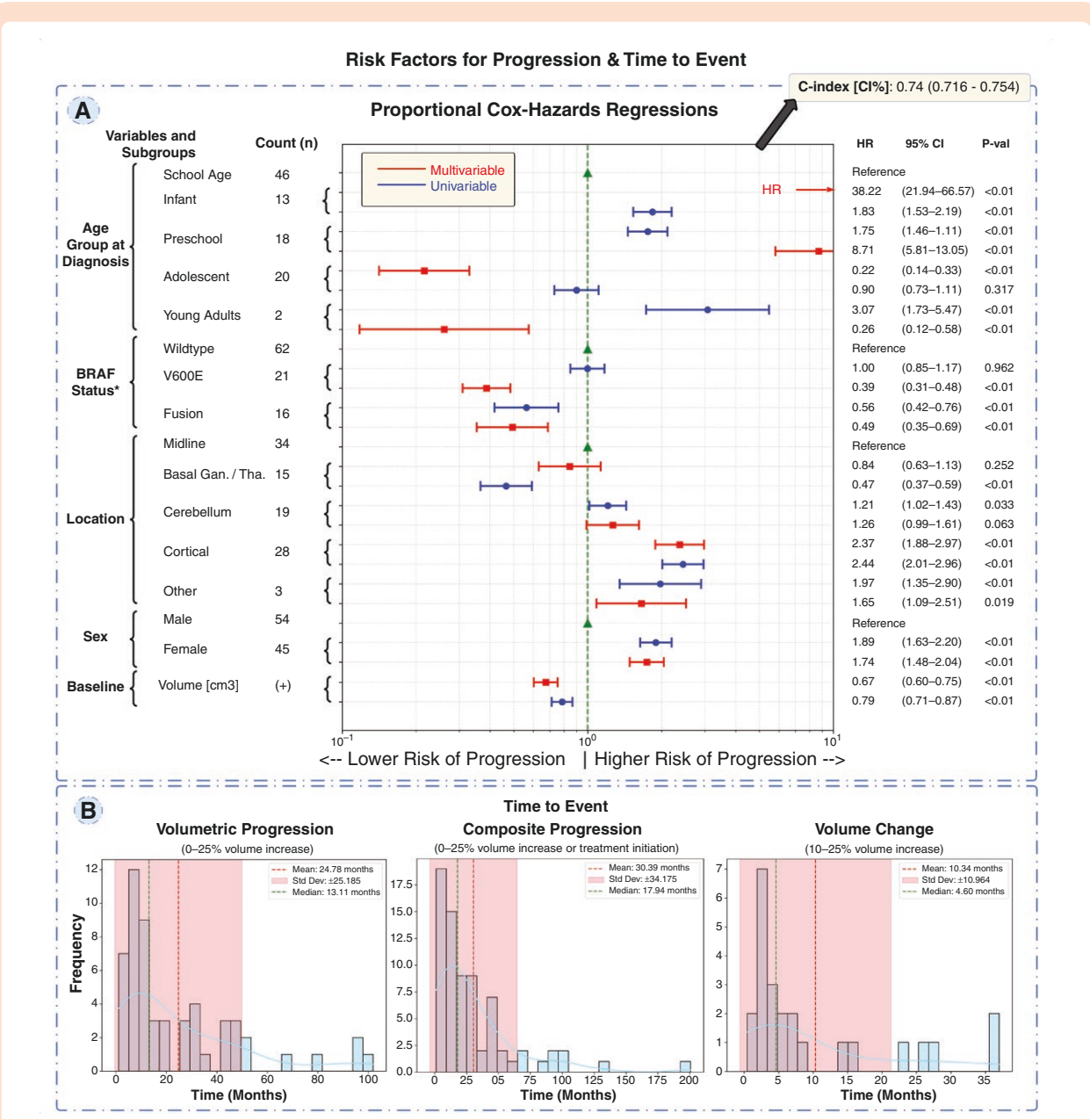

Figure 4. Progression association of baseline clinical risk factor was assessed through a univariable and multivariable proportional Cox-hazards regression (C-Index: 0.735) (A), with age and age grouping being most associated with composite progression (HR: 38.22 for infants, 8.71 for preschool-aged patients, 0.22 for adolescents and 0.26 for young adults). Variables with (+) are continuous and one unit increase needs to be considered when evaluating the risk factor. (B) shows the time distributions for volumetric and composite progression and the difference between a clinically relevant volume increase of 10% to volumetric progression (25%).

volumetric timepoints can predict future state within an average margin of 2.04cm^3 error, which demonstrates fair predictive abilities, helping anticipate clinical changes and derive future volume changes in out-of-curve predictions. It is likely that more complex models incorporating clinical, genomic, and other multimodal data would further improve these longitudinal predictions.

This study has limitations. While the 3D auto-segmentation tool performs well in general, manual quality assurance of generated volumes is still necessary, particularly on lower-quality and older scans. The long observation period introduces inherent variability in imaging parameters and quality. While our preprocessing pipeline mitigated some of these differences, older scans generally demonstrated lower signal-to-noise ratios and spatial resolution, potentially affecting segmentation precision. Future work could explore quantitative image quality metrics and their correlation with segmentation performance across scanner generations.

Since the focus was on the volumetric information extraction, a comparison to auto-2D-RAPNO measurements was out of scope for the project. Additionally, while the

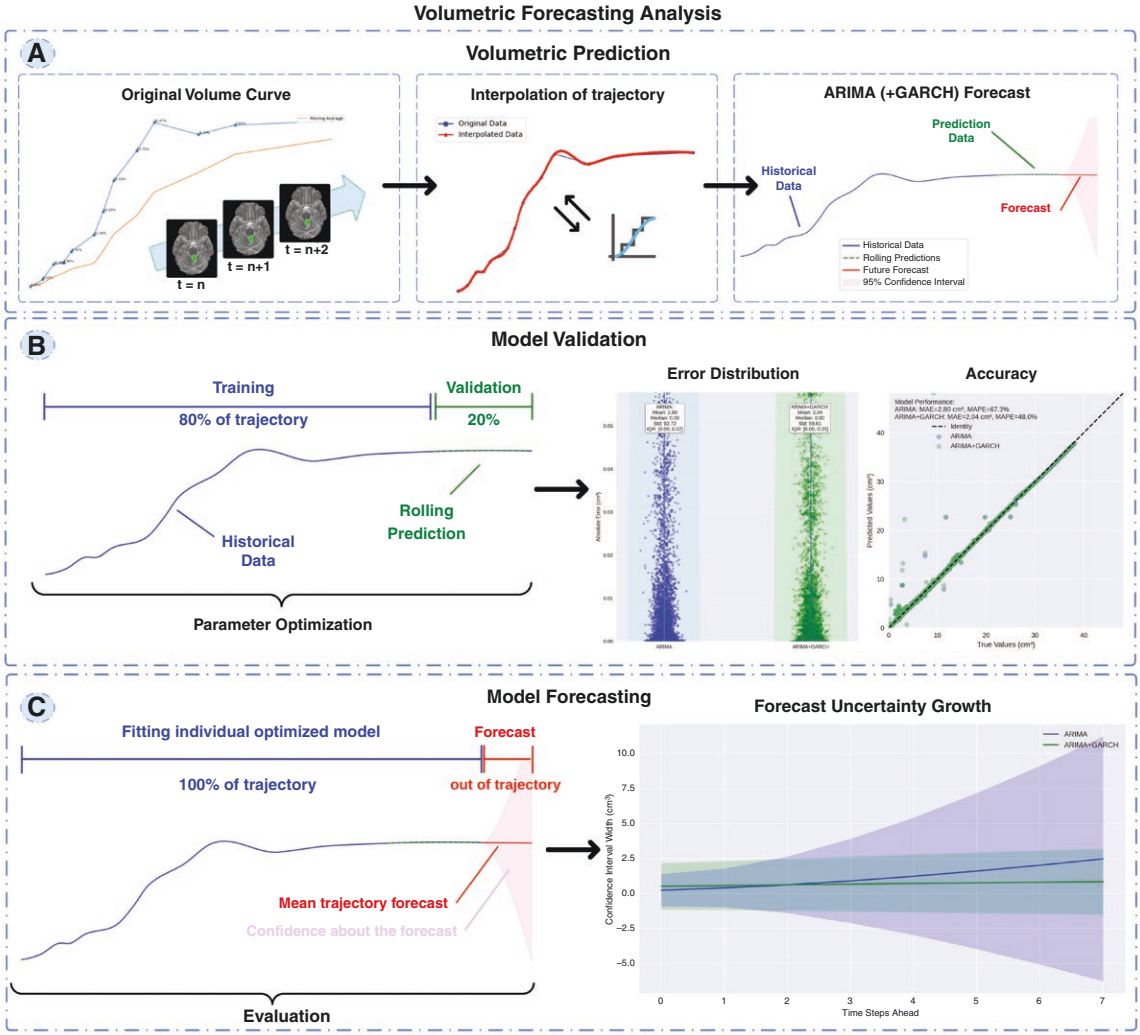

Figure 5. (A) Original volumetric curves get interpolated for the forecasting models to perform out-of-curve predictions. Two models: ARIMA and ARIMA + GARCH are evaluated. (B) Model validation by using the first 80% of the curve for training and initial fitting and the last 20% for validation. Parameters are optimized. Comparison of true and predicted Mean absolute error over cohort is lower in ARIMA + GARCH model: 2.04cm^3. (C) Individual optimized models (one per trajectory) get refitted on the whole data to get out-of-curve forecasts, which are identified by a mean trajectory and its CI. ARIMA + GARCH shows narrower CI width throughout the predictions in the cohort with a final CI width of 0.85cm^3.

RAPNO guidelines recommend multi-sequence assessment including post-contrast imaging, our study utilized only T2-weighted sequences due to the specificity of our validated segmentation algorithm. This approach may underestimate tumor components best visualized on other sequences, particularly contrast-enhancing regions, though T2-weighted imaging remains the primary sequence for delineating the full extent of pLGGs.

The study's retrospective nature subjects it to selection biases in cohort identification, which may influence our identification of particular risk factors for progression and distribution of progression subgroups. While larger samples will be needed to validate these findings, to-date, this represents the largest analysis of untreated pLGGs, indicating the relative rarity of this clinical scenario. Our study focused exclusively on pLGGs managed with upfront surveillance, which represents a specific subset of the broader pLGG population. This creates an inherent selection bias toward patients with less aggressive clinical presentations deemed appropriate for observation, limiting generalizability to all pLGGs, particularly those requiring immediate intervention. The findings should therefore be interpreted as representing the natural history of surveilled pLGGs rather than the entire pLGG population.

The volumetric statistical forecasting techniques require independent validation given the sample size and the inconsistent imaging follow-up intervals between patients and institutions may have affected model performance. While increasing MRI availability would likely be beneficial and obviate interpolation techniques, the optimal number and steady-state point is affected by the follow-up intervals, which are not uniform in clinical, real-world practice.

Scarce data and class imbalance is also an important consideration in trajectory classification and progression association as it may compromise the statistical power of the analysis by inflating standard errors. Additionally, the scope of the prediction algorithms is limited to ML techniques that only provide a general sense of the future mean trajectory. The further and more long-term the prediction, the broader the CIs and prediction instability. Although ARIMA and ARIMA + GARCH combined models adequately fit the curves, further work should investigate predictions that include imaging data directly, not solely the derived volume measurement. Multimodal data and neural network-based algorithms trained on larger datasets, as well as models trained to separate tumor-subregions (e.g. solid vs. cystic components), may improve performance and deliver insights into the distinct biological and growth dynamics. Our current approach does not differentiate between solid and cystic tumor components, which may influence volumetric assessment when cystic changes occur without true tumor growth or response. Future iterations of our pipeline should incorporate multi-compartment segmentation to allow separate tracking of such components, potentially providing more nuanced assessment aligned with emerging response criteria that distinguish these components.

The lack of universal pathological confirmation represents another significant limitation. While T2-weighted radiographic features can suggest pLGG with reasonable confidence, definitive molecular and histopathological diagnosis remains the gold standard. As noted in our Results section, a subset of patients (42.42%) ultimately underwent treatment with surgical intervention, providing pathological confirmation in some of these cases. Building upon our group's previous work in radiographic prediction of BRAF status,[33] future work should integrate comprehensive molecular data with volumetric trajectories. Expanded cohorts with uniform molecular characterization would enable more robust analysis of how specific alterations (BRAF, NF1, FGFR1, etc.) correlate with growth patterns. Additional radiomic features beyond volumetrics could further enhance molecular prediction capabilities.

Finally, future work should integrate complementary MRI characteristics including texture features, heterogeneity metrics, peritumoral environment assessment, and when available, advanced imaging parameters such as apparent diffusion coefficient values, perfusion metrics, or spectroscopic data. A multidimensional approach combining these features with volumetrics would likely yield more robust predictive models to guide clinical decision-making regarding surveillance intervals and treatment timing.

In summary, this study presents a systematic, longitudinal, volumetric analysis of untreated pLGGs on surveillance, generating novel insights regarding tumor growth patterns and associated risk factors for progression. The study yields an automated platform for volumetric trajectory generation and analysis for individual patients and at-scale. The platform can be adapted for use with other medical segmentation models across imaging modalities and diseases. With further validation, volumetric trajectory analysis, as demonstrated in this study, may lead to improved characterization of malignancies and clinical decision-making.

Supplementary Material

Supplementary material is available online at *Neuro-Oncology Advances* (https://academic.oup.com/noa).

Keywords

artificial intelligence | longitudinal analysis | natural history | pediatric low-grade glioma | progression prediction

List of Abbreviations:

pLGG, pediatric Low-Grade Glioma;
MRI, Magnetic Resonance Imaging;
ML, Machine Learning;
DL, Deep Learning;
BCH, Boston Children's Hospital;
CBTN, Children's Brain Tumor Network;
ARIMA, Autoregressive Integrated Moving Average;
GARCH, Generalized Autoregressive Conditional Heteroskedasticity;
AIC, Akaike Information Criterion;
BIC, Bayesian Information Criterion;
HQIC, Hannan–Quinn Information Criterion;
MAE, Mean Absolute Error;
(R)MSE, (Root) Mean Squared Error

Funding

The study is supported by National Institutes of Health [K08DE030216]; National Cancer Institute [P50CA165962, U54 CA274516]; St. Baldrick's Research Foundation; William M. Wood Foundation; Botha-Chan Low-Grade Glioma Consortium.

Conflict of interest statement

All authors declare no competing interests.

Author contributions

Study design: J.C.C.P., B.H.K.; code design, implementation and execution: J.C.C.P.; acquisition, analysis or interpretation of data: J.C.C.P., B.H.K.; image review: J.C.C.P., B.H.K.; writing of manuscript: J.C.C.P., B.H.K.; critical revision of manuscript for important intellectual content: all authors; statistical analysis: J.C.C.P., study supervision: B.H.K., H.J.W.L.A, D.H.K.

Data availability

CBTN data may be requested from the network itself (https://cbtn.org/). DF/BCH data cannot be shared.

Code availability

The code for the pipeline is found under https://github.com/jc-cp/mri-longitudinal-analysis.

Affiliations

Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, Boston, MA, United States (J.C.C.P., A.Z., A.B., D.T., J.Z., M.M., Z.Y., H.J.W.L.A., B.H.K.); Department of Radiation Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States (J.C.C.P., A.Z., A.B., D.T., J.Z., M.M., Z.Y., H.J.W.L.A., D.H.-K., B.H.K.); Radiology and Nuclear Medicine, CARIM & GROW, Maastricht University, Maastricht, Netherlands (H.J.W.L.A.); Boston Children's Hospital, Boston, MA, United States (S.V., P.B., D.H.-K., T.P.); Dana-Farber Cancer Institute, Boston, MA, United States (J.J., P.B., D.H.-K., F.M., K.L., T.P., B.H.K.); Brigham and Women's Hospital, Boston, MA, United States (C.S.); Children's Hospital of Philadelphia, Philadelphia, PA, United States (A.F., A.N.); University of Pennsylvania, PA, 19104, United States (A.N.); Department of Neurology, Neurosurgery and Pediatrics, University of California, San Francisco, CA, United States (S.M.); Chair for Computer Aided Medical Procedures (CAMP), Technical University of Munich, Munich, Germany (J.C.C.P., S.F.)

References

1. Chalil A, Ramaswamy V. Low grade gliomas in children. *J Child Neurol.* 2016;31(4):517–522.
2. Fangusaro J, Jones DT, Packer RJ, et al. Pediatric low-grade glioma: State-of-the-art and ongoing challenges. *Neuro Oncol.* 2024;26(1):25–37.
3. Ostrom QT, Price M, Neff C, et al. CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2015-2019. *Neuro Oncol.* 2022;24(Suppl 5):v1–v95.
4. Goodenberger ML, Jenkins RB. Genetics of adult glioma. *Cancer Genet.* 2012;205(12):613–621.
5. Louis DN, Perry A, Reifenberger G, et al. The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* 2016;131(6):803–820.
6. Packer RJ, Pfister S, Bouffet E, et al. Pediatric low-grade gliomas: implications of the biologic era. *Neuro Oncol.* 2017;19(6):750–761.
7. Schwartzentruber J, Korshunov A, Liu XY, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature.* 2012;482(7384):226–231.
8. Zhang J, Wu G, Miller CP, et al; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project. Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nat Genet.* 2013;45(6):602–612.
9. Ryall S, Zapotocky M, Fukuoka K, et al. Integrated molecular and clinical analysis of 1,000 pediatric low-grade gliomas. *Cancer Cell.* 2020;37(4):569–583.e5.
10. Ullrich NJ, Embry L. Neurocognitive dysfunction in survivors of childhood brain tumors. *Semin Pediatr Neurol.* 2012;19(1):35–42.
11. Gajjar A, Pfister SM, Taylor MD, Gilbertson RJ. Molecular insights into pediatric brain tumors have the potential to transform therapy. *Clin Cancer Res.* 2014;20(22):5630–5640.
12. Pollack IF, Jakacki RI. Childhood brain tumors: epidemiology, current management and future directions. *Nat Rev Neurol.* 2011;7(9):495–506.
13. Merchant TE, Pollack IF, Loeffler JS. Brain tumors across the age spectrum: biology, therapy, and late effects. *Semin Radiat Oncol.* 2010;20(1):58–66.
14. Angelini ED, Delon J, Bah AB, Capelle L, Mandonnet E. Differential MRI analysis for quantification of low grade glioma growth. *Med Image Anal.* 2012;16(1):114–126.
15. Hawkins C, Walker E, Mohamed N, et al. BRAF-KIAA1549 fusion predicts better clinical outcome in pediatric low-grade astrocytoma. *Clin Cancer Res.* 2011;17(14):4790–4798.
16. Srikanthan D, Taccone MS, Van Ommeren R, et al. Diffuse intrinsic pontine glioma: current insights and future directions. *Chin Neurosurg J.* 2021;7(1):6.
17. Lassaletta A, Scheinemann K, Zelcer SM, et al. Phase II weekly vinblastine for chemotherapy-naïve children with progressive low-grade glioma: A Canadian pediatric brain tumor consortium study. *J Clin Oncol.* 2016;34(29):3537–3543.
18. Ater JL, Zhou T, Holmes E, et al. Randomized study of two chemotherapy regimens for treatment of low-grade glioma in young children: a report from the Children's Oncology Group. *J Clin Oncol.* 2012;30(21):2641–2647.
19. Jones DTW, Kocialkowski S, Liu L, et al. Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas. *Cancer Res.* 2008;68(21):8673–8677.
20. Talloa D, Triarico S, Agresti P, et al. BRAF and MEK targeted therapies in pediatric central nervous system tumors. *Cancers (Basel).* 2022;14(17):4264.
21. Grill J, Renaux VK, Bulteau C, et al. Long-term intellectual outcome in children with posterior fossa tumors according to radiation doses and volumes. *Int J Radiat Oncol Biol Phys.* 1999;45(1):137–145.
22. Huang RY, Young RJ, Ellingson BM, et al. Volumetric analysis of IDH-mutant lower-grade glioma: a natural history study of tumor growth rates before and after treatment. *Neuro Oncol.* 2020;22(12):1822–1830.
23. Zhao J, Feng Z, Gallie BL. Natural history of untreated retinoblastoma. *Cancers (Basel).* 2021;13(15):3646.
24. Von Reppert M, Ramakrishnan D, Brüningk SC, et al. Comparison of volumetric and 2D-based response methods in the PNOC-001 pediatric low-grade glioma clinical trial. *Neurooncol. Adv..* 2024;6(1).
25. Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: Overview, challenges and the future BT – classification in bioapps: automation of decision making. *Springer.* 2018;26.
26. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60–88.
27. Hosny A, Bitterman DS, Guthier CV, et al. Clinical validation of deep learning algorithms for radiotherapy targeting of non-small-cell lung cancer: an observational study. *Lancet Digit Health.* 2022;4(9):e657–e666.
28. Jain A, Huang J, Ravipati Y. *Head and Neck Primary Tumor and Lymph Node Autosegmentation for PET/CT Scans.* Cham: Springer; 2023:61–69.
29. Boyd A, Ye Z, Prabhu S, Tjong MC, Zha Y, Zapaishchykova A, Vajapeyam S, Hayat H, Chopra R, Liu KX, Nabavidazeh A, Resnick Adam, Mueller Sabine, Haas-Kogan Daphne, Aerts Hugo J.W.L., Poussaint Tina, Kann Benjamin H. Expert-level pediatric brain tumor segmentation in a limited data scenario with stepwise transfer learning. 2023.
30. Vafaeikia P, Wagner MW, Hawkins C, et al. Improving the segmentation of pediatric low-grade gliomas through multitask learning. In: *2022 44th*

Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Glasgow, Scotland, United Kingdom: IEEE; 2022.

31. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. World Medical Association website. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/. Accessed June 18, 2024.
32. Mongan J, Moy L, Kahn CE, Jr. Checklist for artificial intelligence in medical imaging (CLAIM): A guide for authors and reviewers. *Radiol Artif Intell*. 2020;2(2):e200029.
33. Tak D, Ye Z, Zapaischykova A, et al. Noninvasive molecular subtyping of pediatric low-grade glioma with self-supervised Transfer Learning. *Radiol Artif Intell*. 2024;6(3):e230333.
34. SimpleITK - Home. SimpleITK website. https://simpleitk.org/. Accessed August 16, 2024.
35. NIHPD. McConnell Brain Imaging Centre website. https://www.mcgill.ca/bic/software/tools-data-analysis/anatomical-mri/atlases/nihpd. Accessed August 16, 2024.
36. MIC-DKFZ/HD-BET: MRI brain extraction tool. GitHub website. https://github.com/MIC-DKFZ/HD-BET. Accessed August 16, 2024.
37. Erker C, Tamrazi B, Poussaint TY, et al. Response assessment in paediatric high-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group. *Lancet Oncol*. 2020;21(6):e317–e329.
38. Hayes A. Autoregressive integrated moving average (ARIMA) prediction model. Investopedia website. https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp. Accessed June 12, 2024.
39. statsmodels 0.14.1. statsmodels website. https://www.statsmodels.org/stable/index.html. Accessed August 16, 2024.
40. Skiena HCS. ARIMA and GARCH models. Stony Brook University website. https://www3.cs.stonybrook.edu/~skiena/691/lectures/lecture16.pdf. Accessed June 12, 2024.
41. Introduction to arch. arch 7.2.0 website. https://arch.readthedocs.io/en/stable/index.html. Accessed June 12, 2024.
42. SciPy 1.14.1. SciPy website. https://scipy.org/. Accessed June 12, 2024.
43. Gerstner ER. Volumetric measurements in low-grade glioma: Are we there yet? *Neuro-Oncology*. 2022;24(5):779–780.
44. Reuter M, Gerstner ER, Rapalino O, et al. Impact of MRI head placement on glioma response assessment. *J Neurooncol*. 2014;118(1):123–129.
45. Vos MJ, Uitdehaag BMJ, Barkhof F, et al. Interobserver variability in the radiological assessment of response to chemotherapy in glioma. *Neurology*. 2003;60(5):826–830.
46. Haas-Kogan DA, Aboian MS, Minturn JE, et al. Everolimus for children with recurrent or progressive Low-Grade glioma: Results from the phase II PNOC001 trial. *J Clin Oncol*. 2024;42(4):441–451.
47. Tsai JW, Choi JJ, Ouaalam H, et al. Integrated response analysis of pediatric low-grade gliomas during and after targeted therapy treatment. *Neurooncol. Adv.*. 2022;5(1):vdac182.
48. Goebel AM, Gnekow AK, Kandels D, et al. Natural history of pediatric low-grade glioma disease – first multi-state model analysis. *J Cancer*. 2019;10(25):6314–6326.