Non-inferiority of a non-contrast MRI follow up protocol for isolated optic pathway gliomas.

Ezekiel Maloney

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2018

Committee:

Amanda Phipps

Jeffrey Jarvik

Ramesh Iyer

Daniel Hippe

Program Authorized to Offer Degree:

Department of Epidemiology

University of Washington

Abstract

Non-inferiority of a non-contrast MRI follow up protocol for isolated optic pathway gliomas.

Ezekiel Maloney

Chair of the Supervisory Committee:

Amanda Phipps

Department of Epidemiology

Background: Pediatric optic pathway gliomas (OPGs) are typically indolent, but have variable clinical course. Treatment is dictated by symptoms and changes on contrast-enhanced magnetic resonance (CEMR) examinations. Gadolinium retention in children has motivated parsimonious use of gadolinium-based contrast agents (GBCAs).

Objectives: 1) Determine surveillance MR factors that motivate changes in tumor-directed therapies (TDTs), establish GBCA exposure burden for patients with OPG, and extrapolate cost-efficacy of a non-contrast follow-up protocol. 2) Perform a pilot series of blinded interpretations of follow-up MRI examination pairs for OPGs with attention to the tumor characteristics identified in Objective 1, and establish variability estimates to power a non-inferiority trial. 3) Build an appropriately powered and balanced case set and perform a non-inferiority, multi-reader-multi-case study of a non-contrast MRI protocol versus standard contrast-enhanced protocol for follow-up of isolated

OPGs with attention to tumor characteristics identified in Objective 1.

Materials and methods: An imaging database search identified patients with isolated OPG and ≥3 follow-up CEMRs. Medical records and imaging were reviewed for: 1) coincident changes on CEMR and TDT; 2) demographics and duration of follow-up; 3) motivations for intervention; 4) assessment of GBCAs utility; 5) healthcare utilization data. Cost impact was assessed in terms of Relative Value Unit (RVU) burden. Subsequently, two blinded readers reviewed 16 exam pairs from 8 individual patients and assessed change in tumor size between exams on both non-contrast and contrast-enhanced sequences. Lastly, a series of MRI brain exam pairs from OPG patients was built and was assessed in blinded fashion by 4 of 7 pediatric radiologists; readers assessed changes in tumor size on both non-contrast and contrast-enhanced versions of each exam pair with at least 1 week of memory washout between interpretations.

Results: The 17 neurofibromatosis type 1 (NF1) and 21 non-NF1 patients included in analyses of objectives 1 and 2 underwent a median 16.9 and 24.3 cumulative CEMR exams over 7.7 and 8.1 years of follow-up, respectively. Eight patients (1 with NF1) had intervention based on CEMR findings alone. For these 8, increased tumor size was the only common feature, and was apparent on non-contrast T2 sequences. For the median patient, a non-contrast follow-up protocol could result in 15.9 (NF1) and 23.3 (non-NF1) fewer GBCA administrations, and a 39% reduction in yearly RVU burden. Based on power calculations from the pilot 2-reader assessments of OPG size changes and available MRI brain exams meeting inclusion criteria, 60 exam pairs were compiled from 42 patients with isolated OPG (19 with NF1), and the non-inferiority threshold was set at 12% for intra-reader agreement. Interim analyses from the multi-reader-multi-case study suggest that a non-contrast protocol might be non-inferior to a contrast-enhanced protocol for assessment of change in

tumor size in routine follow up MRI exams.

Conclusions: Pediatric patients with isolated OPG undergo a large number of routine CEMR follow-up exams. Gadolinium may not be needed for these exams to inform management decisions. A non-contrast MRI follow-up protocol appears to be non-inferior to a contrast-enhanced protocol for assessment of changes in tumor size based on interim analysis. Secondary benefits of a non-contrast follow-up protocol include decreased cost and risk to the patient.

INTRODUCTION

Gadolinium based contrast agents (GBCAs) are currently employed in approximately 40% of the MRI studies performed in the United States in pediatric patients [1]. After recognition of the association between GBCA administration and nephrogenic systemic fibrosis (NSF) prompted avoidance of use in the setting of renal failure, GBCAs have generally been well tolerated in the adult as well as pediatric population [2]. However, there has been recent recognition of long-term, dose dependent gadolinium (Gd) retention in both children and adults without severe renal dysfunction, including accumulation in the central nervous system (CNS) [2-6], bone [7], skin [8], and iron-overloaded liver [9]. The clinical significance of its accumulation in the human body in the context of normal renal function is currently unclear. Excluding NSF, there have been rare clinical case reports of Gd-induced toxicity (all in adult patients) that include various symptoms [10, 11]. Presently, however, there is no definitive evidence that accumulated Gd in patients with normal renal function results in clinical pathology. In pre-clinical research studies, multiple Gd-associated toxicities have been observed [10]. These data have led to increased caution in the medical community regarding potential unknown long-term clinical consequences that might result from accumulated Gd and exploration of imaging protocols designed to limit Gd exposure over long term follow up [12-17].

In the interest of diminishing possible risks associated with the administration of GBCAs and avoiding unnecessary cost related to GBCA administration, we sought to examine the routine use of GBCAs at our tertiary pediatric referral center for high-exposure populations such as patients undergoing multiple CEMR exams for brain tumor surveillance. Here, we present results from our

retrospective study of CEMR exams for follow-up of pediatric patients with optic pathway gliomas (OPG).

An OPG is most commonly a pilocytic astrocytoma involving any combination of the optic nerve, chiasm, optic tracts and radiations, and comprises approximately 5% of all childhood tumors [18], most frequently presenting between 2 to 8 years of age [19]. Up to 30% are associated with neurofibromatosis type 1 (NF1) [20]. The evolution of these tumors is unpredictable and progression can result in permanent vision loss and other neurologic and endocrine symptoms [20-22]. Diagnosis is typically made on CEMR prompted by clinical symptoms (e.g. unexplained vision loss, nystagmus, ataxia, proptosis, precocious puberty, absence of pain/inflammatory signs). Biopsy is generally not performed for lesions with classic imaging appearance (optic pathway involvement, solid or mixed solid/cystic, T1 isointense, avidly and heterogeneously T2 hyperintense, with variable contrast enhancement) [21, 23].

Although OPGs have approximately 90% overall survival, average progression-free survival is variable, ranging from approximately 40% at 10 years without treatment in NF1-associated cases [20] to 24% at 5 years with treatment in sporadic cases [22]. The general prognosis (less visual impairment at diagnosis, less likely to progress) and typical MR imaging features (e.g. smaller tumor size, decreased cystic components) of NF1-associated OPG versus sporadic OPG may be partly attributable to detection bias [24-27]. There are no universally accepted policies for MRI-based OPG screening in asymptomatic NF1 patients; however, some authors have advocated for this practice [28], and children with NF1, by nature of their disease and frequent evaluations by multiple clinical specialists, are more likely to undergo baseline diagnostic brain imaging than children without NF1.

Once identified, NF1-associated and sporadic OPGs are typically followed in similar fashion, using serial MRI and neuro-opthalmologic exam.

Treatment of OPGs is typically directed by a multi-disciplinary team. Upon presentation, "watchful waiting," including observation with serial imaging and neuro-ophthalmologic exams, is generally instituted [21]. Tumor directed therapy is initiated with progression, either based on exam (usually changes in vision), or tumor growth on imaging. Although radiation therapy is generally accepted to be a more definitive treatment, in order to avoid deleterious effects in children, particularly neurodevelopmental/intellectual effects in younger children, chemotherapy is usually attempted as a first-line therapy. Surgical resection is also selectively employed, but the diffuse nature and potential morbidity (e.g. additional vision loss) limit its utility [21].

The imaging modality that is typically employed for follow-up is CEMR of the brain. Although intervals of neuro-ophthalmological follow-up are generally standardized (every 3 months for the first year, followed by increasing intervals), there is no consensus on the optimal frequency of neuroimaging for follow-up of OPG [29]. Proposed imaging intervals range from 3 to 24 months, and may be influenced by the location of the tumor, the degree of visual impairment, and evidence of tumor progression [29-33].

Given the frequent CEMR monitoring and the excellent overall life-expectancy of children with OPGs, these patients will be at higher risk of Gd tissue deposition related to repeated GBCA administration and, theoretically, to potential complications associated with Gd accumulation. We hypothesized that GBCA administration was not necessary for routine MR surveillance of isolated OPGs to direct tumor management, obviating the potential risk and cost of repeated GBCA

administration. To test this hypothesis we undertook a retrospective review of our institutional experience, and conducted a multi-reader-multi-case non-inferiority trial [34, 35] of a non-contrast MR follow-up protocol versus the standard contrast-enhanced protocol.

MATERIALS AND METHODS

Exploratory data collection – Case identification

The study was performed at a tertiary pediatric hospital which is a referral center for pediatric OPGs. Following Institutional Review Board approval, a search of radiology reports was performed using an internal electronic patient-imaging-report database (zVision software, version 1.4.65, Clario Medical Imaging). Using the key phrase "optic pathway glioma", spanning the dates of June 1, 2002 through December 30, 2016, we identified 43 patients with OPGs who were treated at our institution and had a minimum of 3 CEMR brain examinations on our institutional Picture Archiving and Communication System (PACS) available for review. To minimize the impact of non-OPG-related imaging follow-up practices on our study, only isolated OPG cases were included. Four patients who had suspected or pathologically proven additional CNS tumors (all with NF1), as well as one patient with concurrent Moya Moya syndrome (also with NF1) were excluded. The electronic medical record (EMR) and available MR brain reports for the 38 patients with isolated OPG were reviewed by one of five of the investigators to identify instances of coincident changes on CEMR and changes in tumor-directed therapies. Specifically, we examined consensus tumor board notes to identify instances of initiation of chemotherapy or clinical-trial medications, surgery, or radiation therapy that were attributed, at least in part, to changes in tumor attributes observed on CEMR exams (e.g., changes in size, composition, or enhancement characteristics). We then verified the change in the attributed exam's report, with review of imaging on PACS. The relevant CEMR exams in all cases were performed at our institution. Per protocol during the study period at our

institution, post-Gd sequences were always obtained immediately following GBCA administration. No specific time-linked cut-off was used for "coincident" changes, only specific attribution of the change in therapy to a change on a specific CEMR exam.

Additional data including age, sex, and weight at OPG diagnosis, NF1 diagnosis status, insurance carrier, and duration of clinical follow-up for OPG was obtained from the EMR. The latter was defined as the time between diagnosis and the most recent relevant clinical note. Tumor board notes, incorporating consensus rationale from multiple subspecialty physicians (e.g., ophthalmologists, neuroradiologists, neuro-oncologists) were further evaluated to assess all of the underlying motivations of changes in tumor-directed therapies. This process included review with a pediatric neuro-oncology specialist to determine any influence that OPG enhancement characteristics might have had on clinical decisions. In some cases, the EMR identified additional CEMR brain exams performed before PACS inception or at outside institutions that were not present on our institutional PACS; the total number of CEMR brain exams for each patient included these additional outside institutional exams dating back to 1990. To establish the "denominator" of patients with OPG seen clinically at our institution over the same date range with any duration of follow up, a separate search of a regional cancer registry inclusive of all patients treated or diagnosed by our hematology/oncology department was performed, specifying our institution, ICD-O histology codes for glioma (9380-9489), and all available site codes involving the optic pathways (C72.30-C72.32).

Exploratory Data Collection – Blinded Review

To preliminarily assess the need for post-Gd sequences to identify tumor progression, a subspecialty-board-certified neuroradiologist conducted blinded review of the 8 cases where CEMR findings alone had resulted in a change in tumor-directed therapy. For each case, two exam pairs

were compiled. The two exams in each pair were always consecutive, obtained within 1 year of one another, and were performed at our institution. The first pair was comprised of the exam and relevant comparison previously identified in tumor board notes where a change in the tumor size had motivated a change in tumor directed therapy. The second pair was selected from later time point, when consensus tumor board notes stated that there was no significant difference in tumor size between exams. This second pair served as an internal control for each patient. Magnet field strength at 1.5T or 3T was not necessarily matched between studies and there was some slight variability in sequence slice thickness and specific sequences parameters in some cases.

All exam pairs were reviewed by the neuroradiologist, blind to the pair type and report contents, and the product of the greatest axial plane whole-tumor perpendicular diameters was calculated for each exam on T1 weighted post-Gd, T2 weighted, and fluid attenuated inversion recovery (FLAIR) sequences. Percent change in cross product was assessed for each pair. This method was adapted from previous studies assessing changes in OPG size on traditional MRI [36-39], and consensus guidelines for size assessments of low grade gliomas [40, 41]. Paired student's t-tests were used to compare the results obtained for measurements made on the different imaging sequences for both the size-change group and the control group.

Preliminary cost-efficacy analysis

Preliminary cost efficacy analysis was performed with attention to: 1) imaging exam-related costs, including relative value units (RVUs) and cumulative imaging time for a contrast-enhanced MR brain protocol model versus non-contrast model (eliminating only the post contrast sequences), 2) per-GBCA-administration patient risks, including incidence of moderate-to-severe allergic-like contrast reaction and cumulative Gd deposition. Statistical analysis of the relationship between cumulative

contrast enhanced examination burden and clinical follow-up time was performed using the statistical software package R [42]. The Healthcare Common Procedure Coding System codes 70551 and 70553 were used to identify appropriate global RVU values, combining the professional and technical components, for non-contrast (6.43) and contrast-enhanced (10.51) MR brain exams from the Centers for Medicare and Medicaid Services (CMS) website [43]. To translate RVU's to "generalizable" dollar amounts, we: 1) Used published national payment indices for private insurer-to-Medicare of 1.30 [44], and Medicaid-to-Medicare of 0.82 [45], 2) Assumed no Geographic Practice Cost Index adjustments that are typically applied to RVUs when calculating Medicare payments, 3) Used only the current Medicare conversion factor of $35.8887.

Powering and building the multi-reader-multi-case set

To generate pilot data for power calculations in the multi-reader-multi-case non-inferiority trial, I reviewed the same 16 exam pairs previously reviewed by the single, blinded neuroradiologist, and measured tumor size in similar fashion. A set of custom formulas were derived for the standard deviation of the main statistic, with the final formulas as follows:

- Variance of Di = $2p(1 - p)(1 - \kappa)$
- Variance of Di averaged over readers 1,2,…,R = $2p(1 - p)(1 - \kappa) \times (1 + (R-1)\beta) / R$ = $Var(Di) \times (1 + (R-1)\beta) / R$

Where variables are defined as: D = difference between 1) intra-reader assessment of tumor size change between non-contrast and contrast-enhanced protocols and 2) intra-reader assessment of tumor size change between two independent reads of the contrast-enhanced protocol; p = probability of agreement; κ = intra-reader correlation between A and B; β = inter-reader correlation between Di and Dj for readers i and j; R = number of readers.

These formulas, in combination with the pilot data, were then used to approximate power using t-test formulas. Power calculations revealed that a ≥ 80% probability of rejecting, at the $\alpha = 0.05$ level, the null hypothesis that there is -12% difference (non-inferiority margin) in intra-reader measurements beyond background variability between the non-contrast and contrast-enhanced protocols when the alternative hypothesis was a 0% difference could reasonably be achieved with an "n" of 60 exam pairs when each pair was interpreted by at least 4 readers. The non-inferiority threshold of 12% (rather than a more commonly employed threshold of 10%) was chosen because: 1) sufficient power would be achievable under a broader range of possible standard deviations of Di (see formulas above), which was unknown a priori; 2) the pilot data suggested that intra-reader agreement was likely to be ≥90% and inter-reader correlation was likely to be 0.6-0.7; 3) the overall rarity of MRI findings alone motivating a change in clinical management; 4) the relative frequency of repeated MRI exams and clinical assessments for our patient population; 5) the cumulative deposition of Gd with each administration; 6) the limited availability of exam pairs meeting including criteria for this relatively rare disease process.

The powered case set was derived from the "denominator" population of OPG patients previously identified in the above-described search of a regional cancer registry. I evaluated consensus tumor board notes for each of these patients. Any case where tumor board notes discussed tumor size change, or size stability, between two specific points in time on MRI was further reviewed. Cases where a change in tumor size was described were included in the case set if: 1) the relevant MRI exams were available in PACS and included axial and coronal (or sagittal) plane T1W images pre- and post-contrast, as well as pre-contrast axial and coronal (or sagittal) plane T2W images, and pre-contrast axial FLAIR images; 2) the patient had no additional intracranial tumors or

pathology at the time of imaging (either biopsy proven or suspected on imaging); 3) there were no other sources of significant intracranial signal alterations or degradation (e.g. significant post-operative changes, metallic susceptibility artifact, significant motion during image acquisition, sub-optimal contrast bolus delivery per the technologist notes). These criteria yielded 8 "size change" exam pairs from 7 NF1 patients and 20 "size change" exam pairs from 18 non-NF1 patients that were included in the case set. The remaining cases of "stable size" tumors described in tumor board consensus notes were stratified by NF1 status, and reviewed in order of most recent clinical follow up date. The same inclusion criteria described above for the "size change" exam pairs was employed, and attention was paid to the balance of NF1 and non-NF1 exam pairs in the case set, alternating between lists to achieve an optimal proportion of each patient type. Since the "denominator" population of OPG patients at our institution was comprised of approximately 50% NF1 patients, we built the case set to be 50% NF1 patients as well. After all patients in each category had been reviewed once, repeated instances of "size stable" exam pairs from already included patients were added. Twenty-two "stable size" exam pairs were added from 19 NF1 patients and 10 "stable size" exam pairs were added from 10 non-NF1 patients. These ratios were consistent with the known history of OPG tumors in NF1 and non-NF1 patients, with a higher frequency of progression in non-NF1 patients [24-27]. In total, exams from 23 NF1 and 19 non-NF1 patients were used to build the 60 exam pair case set, with no more than 2 exam pairs contributed by a single patient. For all patients who contributed two exam pairs, there were at least two years of separation between the later exam in the first pair, and the earlier exam in the second. Patient demographics, MRI exam pair characteristics (magnet strength, minimum slice thickness of sequences), and the anatomic distribution of the OPG according to the modified dodge criteria [46] were recorded for each included exam pair.

The selected exam pairs were exported from PACS to a research workstation, where a computer algorithm was used to curate and standardize the included MRI sequences for each pair of exams, anonymize the exam pairs, and remove and replace the dates of acquisition with standardized dates ("01/01/2001" for the "prior" exam and "02/02/2002" for the more recent, "current" exam). Two versions of each exam pair were built – one with the "minimum" sequences described above using the minimum slice thickness available, and one with the same pre-contrast sequences but without any post-gadolinium sequences. In all selected cases where the tumor involved an optic nerve, fat saturated T1W dedicated orbital sequences had been acquired pre and post-contrast administration for both referenced points in time, and these sequences were also included for these cases, in addition to the "minimum" sequences. Curated and anonymized exam pairs were then re-uploaded to a research PACS server. Exam pairs were randomly assigned to the 7 members of the reader panel (all fellowship-trained pediatric radiologists, 2 with additional subspecialty expertise in neuroradiology) such that each case would be interpreted by 4 different readers -- consistent with the so-called "mixed" multi-reader-multi-case study design [47]. This resulted in a randomly ordered "work list" of 34 to 35 anonymous exam pairs for each reader to interpret. Readers were initially granted access to the non-contrast version of this work list, which contained only the versions of each exam pair that did not include post-contrast sequences. After completing this task, access to data entry was withheld for 1 week to allow memory washout, and then access was granted to the contrast-enhanced version of the work list, containing the same exam pairs but now with the post-contrast sequences included as well. After this second interpretive session, access to data entry was again withheld for 1 week, and then access was granted to a new, randomly ordered and re-named version of the contrast-enhanced work list, for the final interpretive session. Of note, 4 members of the 7 reader panel had participated in the initial exploratory phases of the study and were aware of

the study premise, but agreed to approach each case with an open mind when evaluating the potential utility of the gadolinium-enhanced sequences.

Multi-reader-multi-case study data were collected and managed using Research Electronic Data Capture (REDCap) tools hosted at the University of Washington [48]. REDCap is a secure, web-based application designed to support data capture for research studies, providing: 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources. Standardized data entry for each MRI brain exam pair included the following measurements from the "prior" and "current" examination: 1) greatest perpendicular diameters of the tumor on a single slice in the axial plane (2 separate measurements), 2) greatest craniocaudal diameter of the tumor from a single slice in the coronal or sagittal plane. These measurement techniques were adapted from previous studies assessing changes in OPG size on traditional MRI [36-39], and consensus guidelines for size assessments of low grade gliomas [40, 41]. In cases where there was a significant cystic component of the tumor, separate measurements were provided for this component. For tumors with more diffuse involvement through the optic nerves and tracts, where the tumor board had discussed a specific part of the tumor, readers were directed to this portion of the tumor for making measurements. Readers were asked to identify the sequence that they chose to make each measurement. Finally, readers were asked if the post-contrast sequences were helpful in any way to assess change in tumor size, with a free-text entry box to describe reasons why this might have been true.

Statistical analysis of the multi-reader-multi-case set data

Non-inferiority hypothesis testing for multi-reader-multi-case studies has been described previously

[34, 35, 47, 49, 50]. Intra-reader agreement in assessing the difference in tumor size between exams in terms of greatest perpendicular diameters was the primary focus of our analysis. Secondary endpoints included: 1) Assessment of intra-reader agreement in determination of change in tumor volume, 2) Assessment of the correlation between axial tumor size versus tumor volume measurements and consensus-tumor board determinations of tumor enlargement. Variability parameter assessment and continuous-measurement analysis were performed using the software package R [42].

We defined "D" as the difference in intra-reader agreement on tumor size change between the non-contrast and contrast-enhanced versions of the exam pair protocols (A) and the reader's two independent interpretations of the contrast-enhanced protocol (B) (D = A minus B). For non-inferiority testing, the null hypothesis is that $D \leq -12\%$ (i.e. the non-contrast protocol is inferior to the contrast-enhanced) and the alternative hypothesis is $D = 0\%$ (i.e. the non-contrast protocol has the same performance). To perform the non-inferiority test, we calculated a bootstrap-based 95% two-sided confidence interval for the estimate of "D" using the percentile method and rejected the null hypothesis of inferiority if the confidence interval excluded -12%. This corresponds to a two-sided significance level of $\alpha = 0.05$. The bootstrap resampling was performed by exam pair, clustering the repeated reads (4 per pair), to account for the non-independence of those reads.

Data collection for the multi-reader-multi-case data set from the 7 reader panel is ongoing. An interim analysis was performed based on available data from the 2 readers who had completed the most interpretations. This truncated data set only included a single set of non-contrast, and contrast-enhanced interpretations, and so different statistical methods were employed to account for this difference. The mean difference in tumor size fold-changes measured with and without contrast-enhanced images available was assessed using generalized estimating equations (GEE) to account for non-independence of the multiple timepoints and reads per patient. Agreement in

measurements made with and without contrast-enhanced images was summarized using the intraclass correlation coefficient (ICC) and Cohen's kappa. Confidence intervals were calculated using the non-parametric bootstrap with resampling by patient to account for repeated measurements per patient.

RESULTS

38 patients with isolated OPG and at least 3 CEMR exams obtained between 2002 and 2016 were identified by imaging-record search within a population of 104 patients with OPG who were seen clinically at our institution over the same date range with any duration of follow up. Selected characteristics of the 38 patients meeting our search criteria are presented in **Table 1**. The number of CEMR brain exams performed per year of clinical follow-up for OPGs is presented in **Figure 1A**. On average, NF1 patients tended to have fewer exams per year of clinical follow up. Linear regression analysis with adjustment for NF1 status and age at diagnosis demonstrated a first-order linear trend between years of clinical follow-up for OPG and the cumulative number of CEMR brain examinations undergone by the child (**Figure 1B**). Eighteen patients were identified to have at least one instance of change in tumor-directed therapy motivated at least in part by change on MRI. Of these 18 cases, 8 therapy changes were made based on MRI findings alone (i.e. no significant changes were noted in the patient's neuroophthalmologic exam/testing or other reported symptoms), and 10 were made in the context of changes occurring both on MRI and in the patient's clinical symptoms. For all 18 cases, increase in tumor size was the only imaging feature that drove therapy decision making (see **Figure 2** for a representative case). Consensus tumor board notes in the 8 cases where MRI findings alone motivated change in management cited only the increase in tumor size as evidence of tumor progression, and most commonly stated "preservation of tenuous remaining vision" as the accompanying clinical rationale for initiation of tumor-directed therapy.

Discussion of clinical management of OPG patients and changes in CEMR findings with a neuro-oncology specialist revealed that in all 18 cases, changes in the contrast-enhancement characteristics of the OPG did not: influence tumor-directed therapies, impact assessment of treatment response, or alter frequency of follow-up imaging.

Selected patient demographics and MRI exam details from our preliminary blinded review of 8 patients (1 NF1 patient and 7 non-NF1 patients) with isolated OPG are presented in **Table 2**. Percent changes in tumor size derived from measurements on T2 weighted, FLAIR and T1 weighted post-Gd sequences are presented in **Figure 3**. For exam pairs where size increase had motivated change in tumor directed therapy, all size changes based on T2/FLAIR sequence measurements exceeded 25% (range: 28-106%), a commonly employed threshold for assessing meaningful size increase beyond measurement variability on neuroimaging [36, 40]. Size changes based on Gd-enhanced sequence measurements showed greater variability, and in two instances were below 25% (range: 6-133%). For control exam pairs, percent size changes based on T2/FLAIR sequences (range: -20-11%) and Gd-enhanced sequences (range: -12-14%) were similar. Paired t-tests revealed no significant difference between sequences within the tumor-board-defined size-increase or control groups.

To retrospectively estimate the potential impact of a non-contrast MR surveillance strategy on risk and cost in our cohort, the median values for 17 NF1 / 21 non-NF1 patients with isolated OPGs of 2.2/3.0 yearly MR brain examinations over 7.7/8.1 years of clinical follow-up were used to estimate median cumulative MR brain examination totals of 16.9/24.3. For a non-contrast follow-up protocol model, we assumed the patient would receive a single contrast-enhanced MRI at diagnosis and 15.9/23.3 subsequent non-contrast MRI surveillance studies, resulting in 15.9/23.3 fewer GBCA

administrations. Assuming a constant moderate-to-severe allergic-like reaction incidence of 18 per 100,000 exams [51], for the median NF1/non-NF1 patient with isolated OPG undergoing 16.9/24.3 CEMRs, the cumulative incidence would be 0.30%/0.44%, versus a non-contrast follow-up protocol where the single diagnostic exam exposure would lead to a cumulative incidence of 0.018% for both patient subpopulations. Yearly and cumulative RVU burdens derived from contrast-enhanced, and non-contrast MR follow-up protocol models are depicted in **Figure 4**. Employing the assumptions described in the methods section to the commercial insurance coverage distribution in our study population, the 39% reduction in median yearly RVU burden achieved for both the NF1 and non-NF1 patient cohorts with a non-contrast MR follow-up protocol equates to median-yearly-dollars-saved of $383 per NF1 patient and $481 per non-NF1 patient. Over a median 7.7/8.1 years of clinical follow-up, this translates to $2950 and $3897 saved for each NF1 and non-NF1 patient, respectively.

There was no significant difference in mean age at the time of imaging for 23 NF1/ 19 non-NF1 patients included in the multi-reader multi-case study (7.7/7.6 years with interquartile range of 5.0-18.0 and 3.0-17.0 years, respectively, $p = 0.95$). The sex of these patients was also not significantly different (52.1% male NF1 / 47.4% male non-NF1, $p = 0.76$; overall 50% male). The anatomic distributions of the included OPGs are presented in **Figure 5**. There were significantly higher proportions of bilateral optic nerve (modified dodge 1b), and cisternal segment (modified dodge 1c) optic nerve involvement in NF1 versus non-NF1 patients (30% NF1 versus 0% non-NF1, $p = 8.3 \times 10^{-4}$ and 63.3% NF1 versus 30% non-NF1, $p = 9.1 \times 10^{-3}$, respectively). There were significantly more instances of mis-matched magnet strength (1.5 vs 3.0 T) within exam pairs for the non-NF1 patients (43% non-NF1 versus 17% NF1, $p = 0.02$), but the overall number of exams performed at 3.0 T was identical for the two groups (48%), and the overall proportion of mis-matched magnet

strength for exam pairs was only 30%. The differences in minimum slice thicknesses for both the T1 and T2 weighted sequences for each exam pair were always ≤ 2 mm, and slice thickness was not statistically significantly different between baseline and follow up exams. The thinnest acquired T2 weighted sequences averaged 3.5 mm (1-4 mm range) at baseline exam, and 3.5 mm (1-4 mm range) at follow up ($p = 0.90$). The thinnest acquired T1 weighted, post-contrast sequences averaged 1.4 mm (1-4mm range) at baseline, and 1.6 mm (1-4 mm range) at follow up ($p = 0.22$).

For the interim analysis of the multi-reader-multi-case study, one reader completed 34 cases (both without contrast-enhanced images and with) and another reader completed 26 cases. Of these 60 sets of reads, there were 46 distinct cases. Fourteen cases were read by both readers. Data from these interpretations is presented in **Figure 6**.

The average fold-change in tumor size was 1.36 ± 0.71 when contrast-enhanced images were available and 1.35 ± 0.70 when they were not (difference: -0.01, 95% CI: -0.11 to 0.09, $p=0.86$). The intraclass correlation coefficient (ICC) between the fold-changes in tumor size measured with and without the contrast-enhanced images available was 0.82 (95% CI: 0.44-0.92). After dichotomizing the changes into ≥ 1.25 (i.e., ≥25% increase) and <1.25, there were 26 (43%) size increases when contrast-enhanced images were available and 24 (40%) when there were not. Cohen's kappa for the agreement in determining a ≥ 25% increase between the reviews with and without contrast-enhanced images was 0.73 (95% CI: 0.53-0.89), indicating moderate agreement [52].

DISCUSSION

The definition of "progression" of OPGs on follow-up CEMR has been variably reported to involve an increase in tumor size/extension or a change in the pattern of enhancement [29, 31-33, 46, 53-55]. Heterogeneity within the literature reflects the lack of consensus imaging assessment criteria for OPGs. The Dodge and modified Dodge criteria have been used to qualitatively classify tumors by anatomic distribution, as involvement of specific locations has distinct prognostic significance [46, 54, 56, 57]. Quantitative evaluation of most gliomas is typically performed using the Response Assessment in Neuro-Oncology (RANO) criteria [40, 41, 55]. However, no well-accepted guidelines that combine these two criteria currently exist, and the complex radiological features of OPGs make them difficult to characterize using the RANO criteria. For low grade gliomas, RANO classification emphasizes size measurements (the product of perpendicular diameters) on the T2/FLAIR sequence for assessment, rather than the size of contrast enhancing portions (since fibrillary astrocytomas rarely enhance), as well as avoidance of cystic component measurement [40, 41]. OPGs, primarily pilocytic astrocytomas, however, often demonstrate heterogeneous foci of enhancement in addition to T2/FLAIR hyperintensity of solid components and their cystic components may contribute to clinically-relevant mass effect on adjacent nerve fibers [23]. Such features make the use of RANO guidelines for therapeutic response or surveillance of untreated tumor progression difficult to apply for OPGs. Some groups are currently working toward development of an OPG-specific imaging classification/monitoring system, emphasizing both tumor location and changes in tumor size, with sub-segmentation of tumor components (solid enhancing, solid non-enhancing, and cystic) [58-61]. The results of this study indicate that at our institution changes in OPG size, uninfluenced by contrast enhancement characteristics, drive current management decisions that are based solely on follow-up MR imaging. In our preliminary retrospective review, changes in size were reliably apparent on T2/FLAIR sequences, as the slow growth and anatomic distribution typical of OPGs results in little or no surrounding edema that

might obscure tumor borders [23]. Significant peritumoral edema effacing tumor borders has also not been described in MRI follow-up studies of chemo/radiation treatment response for OPGs [36, 38, 39].

Our study suggests that routine surveillance MR exams for isolated OPGs do not require Gd to guide management decisions. Untreated OPGs are now recognized to have variable contrast enhancement patterns over time without associated clinical significance [23]. Similar to the experience at our institution, Gaudino *et al.* observed "spontaneous fluctuations" (random appearance, increase, decrease or disappearance) in conventional MR contrast enhancement foci of optic-hypothalamic pilocytic astrocytomas over 3-7 years of follow-up in patients not receiving tumor directed therapies [62]. In their study, no correlation was observed between enhancement characteristics and clinical symptoms or tumor size (the latter based on largest diameter and orthogonal line, including cystic and solid regions, measured on T2 weighted fast-spin-echo sequence). Theoretically, inflammatory reactions/degenerative changes (e.g. those induced by successful treatments) could increase blood-brain barrier permeability resulting in cyst formation and increased contrast enhancement, but this has not been demonstrated in clinical imaging studies [63, 64]. Studies of reliable means to identify OPG treatment response with conventional, clinical MRI are limited and difficult to perform given tumor heterogeneity and the rarity of the disease. There have been mixed results when correlating MRI findings with visual acuity changes in the context of variable study designs and patient populations [36, 38]. However, one common feature implicit in the designs of the studies performed is the minimal value attributed to Gd-enhanced sequences as criteria for OPG response to treatment. A recent study of CEMR-based response assessment for 1st-line chemotherapy agents in 15 patients with isolated OPG revealed a mean decrease in the solid tumor volume (9.7%+/-23%), and a mean increase in the tumor cystic component (35% +/-

100%), but no significant difference in tumor enhancement [39]. Fisher *et al.* employed only T2-weighted sequences for imaging-based assessment of OPG response to initial chemotherapy treatment in 71 NF1 patients [36]. Kelly *et al.* measured tumor volumes exclusively on T2/FLAIR sequences for imaging-based assessment of OPG response to initial therapy with chemotherapy or radiation therapy in 21 patients [38]. In addition, a recent large clinical trial assessing chemotherapy efficacy for treatment of low-grade gliomas, including OPGs, relied primarily on measurements from T2 weighted sequences to evaluate treatment response, rather than contrast enhanced MRI sequences [37]. These studies, combined with the findings in our preliminary patient series call into question any added benefit of GBCAs administration for patients with isolated OPG on routine follow-up exams, particularly given the large number of exams these patients typically undergo.

Gd administration can, however, add critical information to the initial diagnostic MR brain/orbits evaluation for a child presenting with visual symptoms, at a time when the differential diagnosis is broad. In addition, at initial diagnosis, CEMR can be helpful in identifying OPGs that are more likely to progress. There have been mixed results regarding the prognostic significance of the degree of conventional contrast enhancement at the time of OPG diagnosis [65, 66]; however, advanced quantitative techniques (such as dynamic contrast enhanced imaging), have been shown to more reliably predict subsequent clinical progression [66, 67].

OPGs are the most common CNS tumor in NF1 patients; however, additional, predominantly benign and indolent CNS tumors are also more common in NF1 patients than the general population [29, 68, 69]. In one large series (n=104) that included 84 NF1 patients with OPG, 20% of all patients had multiple CNS tumors, most commonly a combination of an OPG and a low grade brainstem tumor, equating to approximately 25% of the OPG cohort [70]. When all brain tumors

are taken into account, previous studies have found contrast administration in the NF1 population to be "useful" and "necessary" in documenting tumor stability [71, 72]. In our own study, additional CNS tumors, either pathologically-proven or suspected on imaging, were seen in 4 of 22 cases (18%) of OPG for NF1 patients meeting initial search criteria, and in none of the sporadic cases. Non-isolated OPG cases were excluded from our analysis to avoid confounding influence on neuroimaging frequency and treatment decisions, and our results are only relevant to the majority of NF1 patients with OPG who do not have concurrent CNS tumors.

Fisher *et al.* identified initial chemotherapy treatment rationale from retrospective chart review of 115 NF1 patients treated at 10 referral centers between 1997 and 2007 [36]. In their article, "tumor enhancement" was identified as an indication for initiation of the first course of chemotherapy treatment for 23 patients, and in 16 as a "primary" indication (NB this study allowed multiple indications to be listed simultaneously as the "primary" indication, with a median of 2 "primary" indications per case). Two critical points should be considered when synthesizing these results with our study: 1) Fisher *et al.* did not distinguish between information derived from diagnostic versus follow up MRI exams. "Tumor enhancement" might be a feature of the initial diagnostic MRI that helped to establish the diagnosis and direct therapy while the differential remained broad, or a change in the enhancement characteristics on follow up for a known tumor; 2) Despite the high degree of variability in rationale for initiating treatment between the participating institutions and the relatively large number of cases analyzed, tumor enhancement characteristics were never cited as an *isolated* indication for initiation of therapy, only in combination with other imaging and clinical considerations. The only isolated indications for initiation of therapy in this study were visual acuity loss (n=18) and tumor growth (n=10). The latter is consistent with the results of our study, and has reasonable causal inference in the complex context of medical treatment directives. The variable

significance attributed to OPG enhancement by the diverse set of clinicians in Fisher *et al.*'s study may be attributable to evolving practice patterns at the time. This study was published 5 months prior to the paper by Gaudino *et al.* that described variable enhancement characteristics of untreated OPGs over time without clinical significance [62].

In the same study by Fisher *et al.*, the authors mention that the non-neoplastic T2 hyperintense white matter foci commonly observed in the brains of NF1 patients might influence tumor size assessments, but thought this to be unlikely when exams were interpreted by subspecialty neuroradiologists using a minimum 25% change as threshold for size increase. In our study, only 1 of the 8 cases of therapy initiated based solely on change in MRI occurred in an NF1 patient with isolated OPG – the remaining cases were sporadic OPGs. For this single case, OPG size increase on follow up CEMR was readily apparent on T2/FLAIR sequences, was the only change identified, and motivated initiation of chemotherapy. These findings are consistent with the growing body of literature relying on T2 weighted sequences for size assessments at follow up of both NF1 and non-NF1 OPGs [36-39].

The ability of various GBCA's to release free Gd and result in tissue deposition varies but is lowest among macrocyclic agents [73, 74]. A recent survey revealed that 80% of North American pediatric hospitals currently use macrocyclic Gd-based contrast agents (m-GBCAs) [5], and for hospitals who recently switched agents, the most common rationale was concern regarding Gd tissue deposition. In the era of m-GBCAs, some might voice skepticism as to the utility of tailoring tumor follow-up protocols to exclude Gd-enhanced sequences. This stems, in part, from the radiologic community's focus on surrogate markers of Gd deposition. For example, a recent study by Radruch *et al.* demonstrated no significant signal intensity increase in the dentate nucleus on non-contrast MRI in

pediatric patients who had previously received up to 23 doses of m-GBCA (a phenomenon demonstrated in children exposed to linear agents) [75, 76]. Brain deposition with m-GBCAs has, however, been demonstrated pathologically in adults [7]. Furthermore, deposition is greater in some other tissues, such as bone, than in the brain [7]. Some have suggested that Gd in the bone matrix could function as a long-term systemic reservoir, slowly releasing Gd into the circulation and redistributing it to other organs over a long period of time [6, 16, 77]. Precise estimates of cumulative Gd tissue deposition in children are difficult given the incomplete understanding of specific patient and contrast-agent derived factors that contribute; however, given the known dose-dependence of Gd deposition in children [2], and the high number of follow-up CEMR exams OPG patients undergo, implementation of a non-contrast MRI follow-up protocol for OPG patients would be expected to result in a substantial reduction in accumulated Gd over time. Moreover, the unknown long-term risks of Gd deposition accompany the per-dose-risk of allergic-like contrast reaction (likely underestimated in our current study, as we limited our analysis to moderate-to-severe reactions that have relatively constant incidence, versus the increasing-per-dose incidence of mild reactions), the per-dose risk of contrast extravasation, and the minor trauma of intravenous administration [78, 79]. In this context, for the purposes of guiding clinical follow-up policy decisions, the risk reduction that a non-contrast MR follow-up protocol provides to the patient is likely best quantified in absolute terms. In our study and proposed non-contrast protocol model, this was a median 15.9 and 23.3 fewer GBCA administrations over 7.7 and 8.1 years of clinical follow-up for NF1 and non-NF1 patients, respectively.

Our relatively simplistic, preliminary estimations of cost efficacy are intended to provide a general impression of the impact that implementation of a non-contrast MRI follow-up protocol might have on the value of care provided to the average child with OPG. Such considerations are particularly

important in the evolving value-focused healthcare remuneration system [80]. CMS global RVUs were used to provide a standardized benchmark for the cost of follow-up MR imaging to the healthcare system and have become the reimbursement basis for Medicaid programs and many commercial payers covering the vast majority of children in the United States [81]. RVUs are robust to between-institution imaging variability and this strengthens the generalizability of our results. Assessing the cost, in dollars as well as medical risk, per patient examined is only a preliminary step in cost-effectiveness analysis. Assessment of the diagnostic performance and impact of a test requires knowledge of its overall accuracy for a given application, which we attempted to discern in the multi-reader-multi-case phase of our study [82, 83].

To our knowledge, there have been no prior, large comparative studies of a non-contrast versus contrast-enhanced MRI protocol for assessment of isolated OPG at follow-up. The wide range of anatomic distribution of OPGs included in this phase of our study was similar to prior studies including both non-NF1 and NF1 patients with this tumor [36, 46]. The significantly increased proportion of bilateral and cisternal optic nerve involvement in the NF1 population is also consistent with prior studies, and is a known tendency within this patient population, who have loss of tumor suppression [25, 27, 36, 46, 84, 85].

The interim analysis of our multi-reader-multi-case study suggests that when measurements are considered on a continuous scale for a diverse OPG case set with and without Gd-enhanced sequences, there is no significant difference in assessments of size change and excellent ICC. Dichotomization at precisely a 25% increase in the cross product of the greatest perpendicular tumor diameters to determine "size increase" generated a kappa value indicating moderate agreement, with reclassification of only 3% of "size increase" cases when Gd-enhanced sequences

were omitted. This small difference may be attributable to normal intra-reader measurement variability, and we will verify this when our data acquisition is complete.

CONCLUSION

Pediatric patients with isolated OPG undergo a large number of routine follow up MRI exams. Gd may not be needed on these exams to inform tumor management decisions when driven by increasing OPG size on MRI. Given the lack of clinical benefit, and the potential reduction in Gd deposition, adverse contrast reactions, and monetary cost to the healthcare system, it is prudent to reassess Gd use in MR imaging follow-up of isolated OPGs. Although these are preliminary results from a single center, they are consistent with a growing body of literature from multiple OPG referral centers. Interim analysis from our multi-reader-multi-case study suggests that a non-contrast protocol might be non-inferior to a contrast-enhanced protocol for assessment of change in tumor size on routine follow up MRI exams.

References

1. Gale EM, Caravan P, Rao AG, et al. Gadolinium-based contrast agents in pediatric magnetic resonance imaging. *Pediatr Radiol*. 2017;47(5):507-21.
2. McDonald JS, McDonald RJ, Jentoft ME, et al. Intracranial Gadolinium Deposition Following Gadodiamide-Enhanced Magnetic Resonance Imaging in Pediatric Patients: A Case-Control Study. *JAMA Pediatr*. 2017;171(7):705-7.
3. Kanda T, Fukusato T, Matsuda M, et al. Gadolinium-based Contrast Agent Accumulates in the Brain Even in Subjects without Severe Renal Dysfunction: Evaluation of Autopsy Brain Specimens with Inductively Coupled Plasma Mass Spectroscopy. *Radiology*. 2015;276(1):228-32.
4. Miller JH, Hu HH, Pokorney A, Cornejo P, Towbin R. MRI Brain Signal Intensity Changes of a Child During the Course of 35 Gadolinium Contrast Examinations. *Pediatrics*. 2015;136(6):e1637-40.
5. Mithal LB, Patel PS, Mithal D, Palac HL, Rozenfeld MN. Use of gadolinium-based magnetic resonance imaging contrast agents and awareness of brain gadolinium deposition among pediatric providers in North America. *Pediatr Radiol*. 2017;47(6):657-64.
6. Roberts DR, Chatterjee AR, Yazdani M, et al. Pediatric Patients Demonstrate Progressive T1-Weighted Hyperintensity in the Dentate Nucleus following Multiple Doses of Gadolinium-Based Contrast Agent. *AJNR Am J Neuroradiol*. 2016;37(12):2340-7.
7. Murata N, Gonzalez-Cuyar LF, Murata K, et al. Macrocyclic and Other Non-Group 1 Gadolinium Contrast Agents Deposit Low Levels of Gadolinium in Brain and Bone Tissue: Preliminary Results From 9 Patients With Normal Renal Function. *Invest Radiol*. 2016;51(7):447-53.
8. Roberts DR, Lindhorst SM, Welsh CT, et al. High Levels of Gadolinium Deposition in the Skin of a Patient With Normal Renal Function. *Invest Radiol*. 2016;51(5):280-9.
9. Maximova N, Gregori M, Zennaro F, et al. Hepatic Gadolinium Deposition and Reversibility after Contrast Agent-enhanced MR Imaging of Pediatric Hematopoietic Stem Cell Transplant Recipients. *Radiology*. 2016;281(2):418-26.
10. Rogosnitzky M, Branch S. Gadolinium-based contrast agent toxicity: a review of known and proposed mechanisms. *Biometals*. 2016;29(3):365-76.
11. Semelka RC, Ramalho M, AlObaidy M, Ramalho J. Gadolinium in Humans: A Family of Disorders. *AJR Am J Roentgenol*. 2016;207(2):229-33.
12. U. S. Food and Drug Administration (2017) Drug safety commu- nications. FDA evaluating the risk of brain deposits with repeated use of gadolinium-based contrast agents for magnetic resonance imaging. https://www.fda.gov/downloads/Drugs/DrugSafety/UCM455390.pdf. Accessed 17 April 2017.
13. European Medicines Agency (2017) PRAC concludes assessment of gadolinium agents used in body scans and recommends regula- tory actions, including suspension for some marketing authorisations. http://www.ema.europa.eu/docs/en_GB/document_library/Press_release/2017/03/WC500223209.pdf. Accessed 17 April 2017.
14. American College of Radiology (2017) ACR manual on contrast media version 10.3 (78). https://www.acr.org/-/media/ACR/Files/Clinical-Resources/Contrast_Media.pdf. Accessed 12 April 2018.

15. American College of Radiology (2017) News releases: ACR response to the European PRAC recommendations (April 4, 2017). http://www.publicnow.com/view/DA714236788989CE92F38E18EA3CCB6FCC9DFDE3?2017-04-04-17:31:30+01:00-xxx935. Accessed 12 April 2018.
16. Goischke HK. Safety assessment of gadolinium-based contrast agents (GBCAs) requires consideration of long-term adverse effects in all human tissues. *Mult Scler J Exp Transl Clin*. 2017;3(2):2055217317704450.
17. Gupta A, Al-Dasuqi K, Xia F, et al. The Use of Noncontrast Quantitative MRI to Detect Gadolinium-Enhancing Multiple Sclerosis Brain Lesions: A Systematic Review and Meta-Analysis. *AJNR Am J Neuroradiol*. 2017;38(7):1317-22.
18. Heideman RL, et al. Tumors of the central nervous system. In Pizzo PA and Poplack DG eds. Principles and Practice of Pediatric Oncology, Philadelphia: JB Lippincott 1993:633–681.
19. Ostrom QT, Gittleman H, Fulop J, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro Oncol*. 2015;17 Suppl 4:iv1-iv62.
20. Albers AC, Gutmann DH. Gliomas in patients with neurofibromatosis type 1. *Expert Rev Neurother*. 2009;9(4):535-9.
21. Recht LD (2017) Optic pathway glioma. In: UpToDate. https://www.uptodate.com/contents/optic-pathway-glioma. Accessed 12 April 2018.
22. Wan MJ, Ullrich NJ, Manley PE, et al. Long-term visual outcomes of optic pathway gliomas in pediatric patients without neurofibromatosis type 1. *J Neurooncol*. 2016;129(1):173-8.
23. Gaudino S, Martucci M, Russo R, et al. MR imaging of brain pilocytic astrocytoma: beyond the stereotype of benign astrocytoma. *Childs Nerv Syst*. 2017;33(1):35-54.
24. Hernaiz Driever P, von Hornstein S, Pietsch T, et al. Natural history and management of low-grade glioma in NF-1 children. *J Neurooncol*. 2010;100(2):199-207.
25. Kornreich L, Blaser S, Schwarz M, et al. Optic pathway glioma: correlation of imaging findings with the presence of neurofibromatosis. *AJNR Am J Neuroradiol*. 2001;22(10):1963-9.
26. Kerrison JB. Chapter 38: Phacomatoses. In: Miller NR (ed) Walsh & Hoyt's clinical neuro-ophthalmology, 6th edn, Philadelphia: Lippincott Williams & Wilkins 2005:1823-1898. .
27. Chateil JF, Soussotte C, Pedespan JM, et al. MRI and clinical differences between optic pathway tumours in children with and without neurofibromatosis. *Br J Radiol*. 2001;74(877):24-31.
28. Prada CE, Hufnagel RB, Hummel TR, et al. The Use of Magnetic Resonance Imaging Screening for Optic Pathway Gliomas in Children with Neurofibromatosis Type 1. *J Pediatr*. 2015;167(4):851-6 e1.
29. Listernick R, Ferner RE, Liu GT, Gutmann DH. Optic pathway gliomas in neurofibromatosis-1: controversies and recommendations. *Ann Neurol*. 2007;61(3):189-98.
30. Balcer LJ, Liu GT, Heller G, et al. Visual loss in children with neurofibromatosis type 1 and optic pathway gliomas: relation to tumor location by magnetic resonance imaging. *Am J Ophthalmol*. 2001;131(4):442-5.
31. Listernick R, Louis DN, Packer RJ, Gutmann DH. Optic pathway gliomas in children with neurofibromatosis 1: consensus statement from the NF1 Optic Pathway Glioma Task Force. *Ann Neurol*. 1997;41(2):143-9.
32. Liu GT. Optic gliomas of the anterior visual pathway. *Curr Opin Ophthalmol*. 2006;17(5):427-31.

33. Thiagalingam S, Flaherty M, Billson F, North K. Neurofibromatosis type 1 and optic pathway gliomas: follow-up of 54 patients. *Ophthalmology*. 2004;111(3):568-77.
34. Obuchowski NA, Gallas BD, Hillis SL. Multi-reader ROC studies with split-plot designs: a comparison of statistical methods. *Acad Radiol*. 2012;19(12):1508-17.
35. Obuchowski NA, Beiden SV, Berbaum KS, et al. Multireader, multicase receiver operating characteristic analysis: an empirical comparison of five methods. *Acad Radiol*. 2004;11(9):980-95.
36. Fisher MJ, Loguidice M, Gutmann DH, et al. Visual outcomes in children with neurofibromatosis type 1-associated optic pathway glioma following chemotherapy: a multicenter retrospective analysis. *Neuro Oncol*. 2012;14(6):790-7.
37. 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):2641-7.
38. Kelly JP, Leary S, Khanna P, Weiss AH. Longitudinal measures of visual function, tumor volume, and prediction of visual outcomes after treatment of optic pathway gliomas. *Ophthalmology*. 2012;119(6):1231-7.
39. Shofty B, Mauda-Havakuk M, Weizman L, et al. The effect of chemotherapy on optic pathway gliomas and their sub-components: A volumetric MR analysis study. *Pediatr Blood Cancer*. 2015;62(8):1353-9.
40. van den Bent MJ, Wefel JS, Schiff D, et al. Response assessment in neuro-oncology (a report of the RANO group): assessment of outcome in trials of diffuse low-grade gliomas. *Lancet Oncol*. 2011;12(6):583-93.
41. Wen PY, Chang SM, Van den Bent MJ, et al. Response Assessment in Neuro-Oncology Clinical Trials. *J Clin Oncol*. 2017;35(21):2439-49.
42. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. < URL: https://www.R-project.org/ >.
43. Centers for Medicare and Medicaid Services. Searchable Medicare Physician Fee Schedule. Accessed 09/12/2017. Available from: http://www.cms.gov/apps/physician-fee-schedule/overview.aspx
44. Clemens J, Gottlieb JD. In the Shadow of a Giant: Medicare's Influence on Private Physician Payments. *J Polit Econ*. 2017;125(1):1-39.
45. The Henry J. Kaiser Family Foundation. State Health Facts - Medicaid-to-Medicare fee index (2016). Accessed 09/15/2017. Available from: http://www.kff.org/state-category/medicaid-chip/medicaid-physician-fees/.
46. Taylor T, Jaspan T, Milano G, et al. Radiological classification of optic pathway gliomas: experience of a modified functional classification system. *Br J Radiol*. 2008;81(970):761-6.
47. Obuchowski NA. Reducing the number of reader interpretations in MRMC studies. *Acad Radiol*. 2009;16(2):209-17.
48. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-81.
49. Chen YH, Gutmann DH. The molecular and cell biology of pediatric low-grade gliomas. *Oncogene*. 2014;33(16):2019-26.
50. Chen W, Petrick NA, Sahiner B. Hypothesis testing in noninferiority and equivalence MRMC ROC studies. *Acad Radiol*. 2012;19(9):1158-65.

51. Dillman JR, Ellis JH, Cohan RH, Strouse PJ, Jan SC. Frequency and severity of acute allergic-like reactions to gadolinium-containing i.v. contrast media in children and adults. *AJR Am J Roentgenol*. 2007;189(6):1533-8.
52. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276-82.
53. Astrup J. Natural history and clinical management of optic pathway glioma. *Br J Neurosurg*. 2003;17(4):327-35.
54. Dodge HW, Jr., Love JG, Craig WM, et al. Gliomas of the optic nerves. *AMA Arch Neurol Psychiatry*. 1958;79(6):607-21.
55. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J Clin Oncol*. 2010;28(11):1963-72.
56. Pollack IF, Hoffman HJ, Humphreys RP, Becker L. The long-term outcome after surgical treatment of dorsally exophytic brain-stem gliomas. *J Neurosurg*. 1993;78(6):859-63.
57. Lee AG. Neuroophthalmological management of optic pathway gliomas. *Neurosurg Focus*. 2007;23(5):E1.
58. Lambron J, Rakotonjanahary J, Loisel D, et al. Can we improve accuracy and reliability of MRI interpretation in children with optic pathway glioma? Proposal for a reproducible imaging classification. *Neuroradiology*. 2016;58(2):197-208.
59. Weizman L, Sira LB, Joskowicz L, et al. Semiautomatic segmentation and follow-up of multicomponent low-grade tumors in longitudinal brain MRI studies. *Med Phys*. 2014;41(5):052303.
60. Weizman L, Ben Sira L, Joskowicz L, et al. Automatic segmentation, internal classification, and follow-up of optic pathway gliomas in MRI. *Med Image Anal*. 2012;16(1):177-88.
61. Shofty B, Weizman L, Joskowicz L, et al. MRI internal segmentation of optic pathway gliomas: clinical implementation of a novel algorithm. *Childs Nerv Syst*. 2011;27(8):1265-72.
62. Gaudino S, Quaglio F, Schiarelli C, et al. Spontaneous modifications of contrast enhancement in childhood non-cerebellar pilocytic astrocytomas. *Neuroradiology*. 2012;54(9):989-95.
63. Beni-Adani L, Gomori M, Spektor S, Constantini S. Cyst wall enhancement in pilocytic astrocytoma: neoplastic or reactive phenomena. *Pediatr Neurosurg*. 2000;32(5):234-9.
64. Takeuchi H, Kubota T, Sato K, Arishima H. Ultrastructure of capillary endothelium in pilocytic astrocytomas. *Brain Tumor Pathol*. 2004;21(1):23-6.
65. Strong JA, Hatten HP, Jr., Brown MT, et al. Pilocytic astrocytoma: correlation between the initial imaging features and clinical aggressiveness. *AJR Am J Roentgenol*. 1993;161(2):369-72.
66. Jittapiromsak N, Hou P, Liu HL, et al. Prognostic Role of Conventional and Dynamic Contrast-Enhanced MRI in Optic Pathway Gliomas. *J Neuroimaging*. 2017.
67. Jost SC, Ackerman JW, Garbow JR, et al. Diffusion-weighted and dynamic contrast-enhanced imaging as markers of clinical behavior in children with optic pathway glioma. *Pediatr Radiol*. 2008;38(12):1293-9.
68. Singhal S, Birch JM, Kerr B, Lashford L, Evans DG. Neurofibromatosis type 1 and sporadic optic gliomas. *Arch Dis Child*. 2002;87(1):65-70.
69. Campian J, Gutmann DH. CNS Tumors in Neurofibromatosis. *J Clin Oncol*. 2017;35(21):2378-85.
70. Guillamo JS, Creange A, Kalifa C, et al. Prognostic factors of CNS tumours in Neurofibromatosis 1 (NF1): a retrospective study of 104 patients. *Brain*. 2003;126(Pt 1):152-60.

71. Mentzel HJ, Seidel J, Fitzek C, et al. Pediatric brain MRI in neurofibromatosis type I. *Eur Radiol.* 2005;15(4):814-22.
72. Bonawitz C, Castillo M, Chin CT, Mukherji SK, Barkovich AJ. Usefulness of contrast material in MR of patients with neurofibromatosis type 1. *AJNR Am J Neuroradiol.* 1998;19(3):541-6.
73. Jost G, Lenhard DC, Sieber MA, et al. Signal Increase on Unenhanced T1-Weighted Images in the Rat Brain After Repeated, Extended Doses of Gadolinium-Based Contrast Agents: Comparison of Linear and Macrocyclic Agents. *Invest Radiol.* 2016;51(2):83-9.
74. Sieber MA, Steger-Hartmann T, Lengsfeld P, Pietsch H. Gadolinium-based contrast agents and NSF: evidence from animal experience. *J Magn Reson Imaging.* 2009;30(6):1268-76.
75. Radbruch A, Haase R, Kickingereder P, et al. Pediatric Brain: No Increased Signal Intensity in the Dentate Nucleus on Unenhanced T1-weighted MR Images after Consecutive Exposure to a Macrocyclic Gadolinium-based Contrast Agent. *Radiology.* 2017;283(3):828-36.
76. Flood TF, Stence NV, Maloney JA, Mirsky DM. Pediatric Brain: Repeated Exposure to Linear Gadolinium-based Contrast Material Is Associated with Increased Signal Intensity at Unenhanced T1-weighted MR Imaging. *Radiology.* 2017;282(1):222-8.
77. Abraham JL, Thakral C. Tissue distribution and kinetics of gadolinium and nephrogenic systemic fibrosis. *Eur J Radiol.* 2008;66(2):200-7.
78. Costello JR, Kalb B, Martin DR. Incidence and Risk Factors for Gadolinium-Based Contrast Agent Immediate Reactions. *Top Magn Reson Imaging.* 2016;25(6):257-63.
79. Jung JW, Kang HR, Kim MH, et al. Immediate hypersensitivity reaction to gadolinium-based MR contrast media. *Radiology.* 2012;264(2):414-22.
80. Stecker EC, Schroeder SA. Adding value to relative-value units. *N Engl J Med.* 2013;369(23):2176-9.
81. The Henry J. Kaiser Family Foundation. Fact Sheet - Next steps for CHIP: what is at stake for children (June, 2017). Accessed 09/14/2017. Available from: http://files.kff.org/attachment/Fact-Sheet-Next-Steps-for-CHIP-What-is-at-Stake-for-Children.
82. Patton DD, Woolfenden JM. A utility-based model for comparing the cost-effectiveness of diagnostic studies. *Invest Radiol.* 1989;24(4):263-71.
83. Hollingworth W, Jarvik JG. Technology assessment in radiology: putting the evidence in evidence-based radiology. *Radiology.* 2007;244(1):31-8.
84. Beres SJ, Avery RA. Optic Pathway Gliomas Secondary to Neurofibromatosis Type 1. *Semin Pediatr Neurol.* 2017;24(2):92-9.
85. Liu GT, Brodsky MC, Phillips PC, et al. Optic radiation involvement in optic pathway gliomas in neurofibromatosis. *Am J Ophthalmol.* 2004;137(3):407-14.

Figures

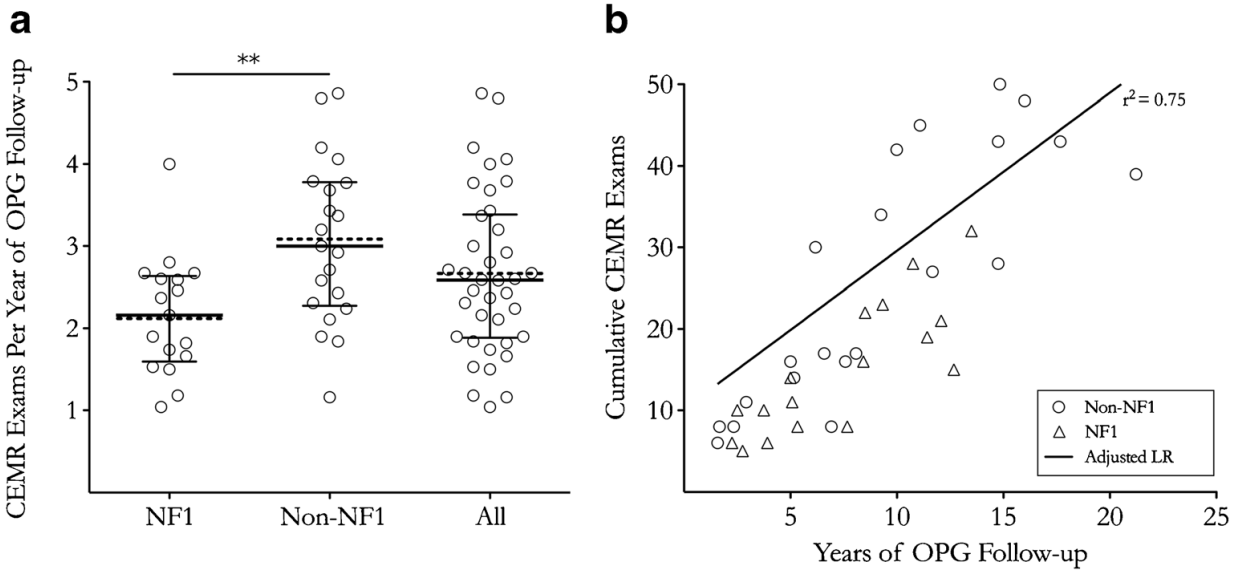

Figure 1 – **Cumulative contrast enhanced magnetic resonance (CEMR) brain examination burden for patients with isolated optic pathway glioma (OPG). A)** The mean (dotted line) / median (solid line) and interquartile range (whiskers) of CEMR brain examinations performed per year of clinical follow-up for OPG were: 2.2/2.2 (1.6-2.6) for neurofibromatosis type-1 (NF1) patients, 3.1/3.0 (2.3-3.8) for non-NF1 patients, and 2.7/2.6 (1.9-3.4) overall. Comparing the average values across the two subpopulations, the associated p-value was 0.002 (**). **B)** A first-order linear trend was demonstrated between years of clinical OPG follow-up and the cumulative number of CEMR brain examinations the patient underwent. Linear regression (LR) analysis with adjustment for NF1 status and age of diagnosis using robust standard errors generated a slope value of 1.9 (95% confidence interval 1.4-2.5) and an associated p-value of 3x10^-9.

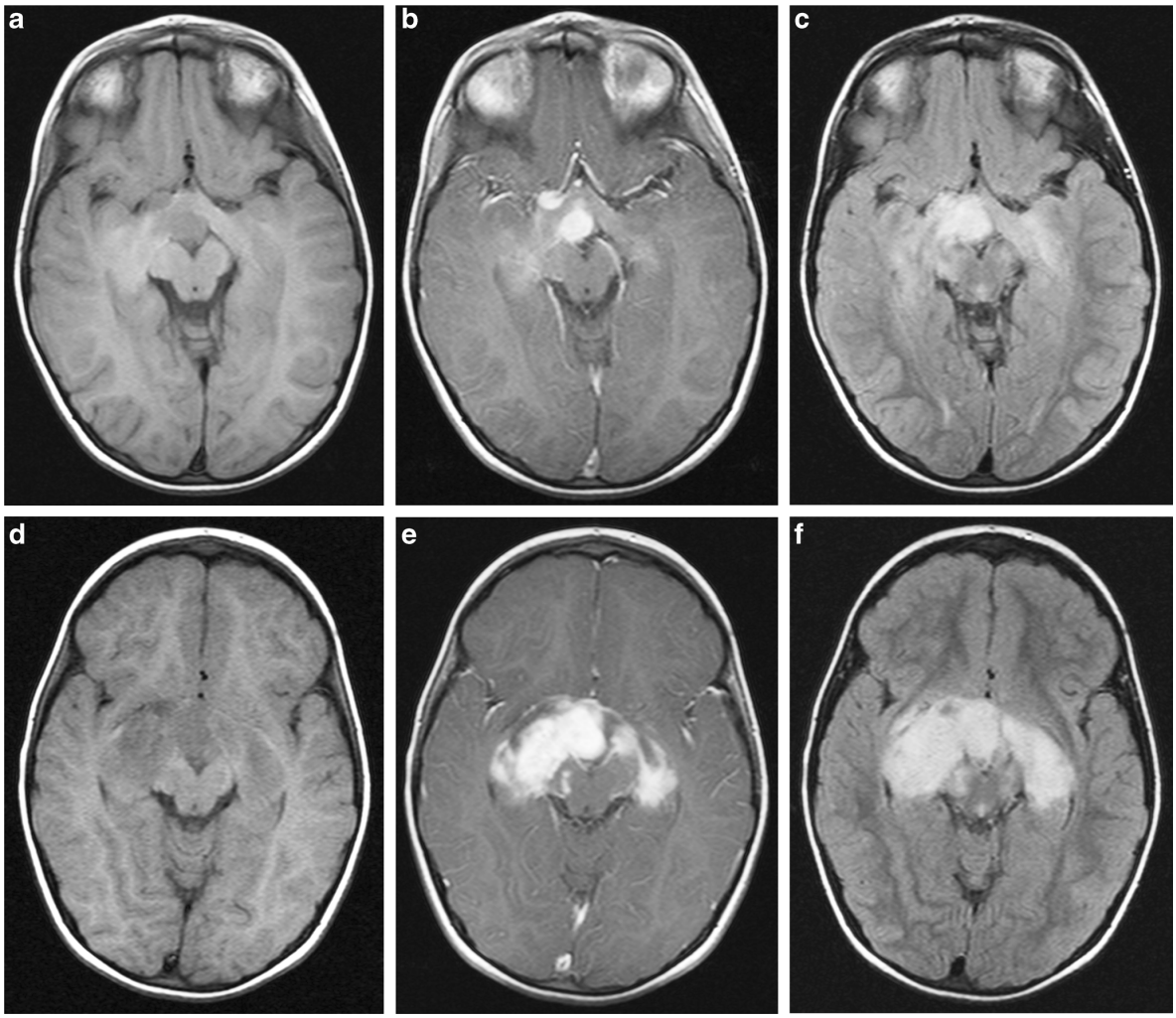

Figure 2 – **Example case.** 3 year-old male undergoing routine 3-month interval follow-up of a sporadic isolated OPG diagnosed 2 years previously. Axial T1 weighted (A & D), post-contrast (B&E), and non-contrast fluid attenuated inversion recovery (FLAIR; C&F) images at the level of the midbrain are shown from two consecutive CEMR exams. The CEMR from baseline (A-C) showed stable disease, and ophthalmologic exam revealed stable visual testing abnormalities. The CEMR from 3 months later (D-F) showed increase in tumor size that was readily apparent on both the axial post-contrast T1 and FLAIR sequences, although visual testing abnormalities / exam

remained unchanged. The patient was subsequently started on chemotherapy with good response in 2004. To date, he has undergone 50 CEMR brain exams over 14.8 years of surveillance.

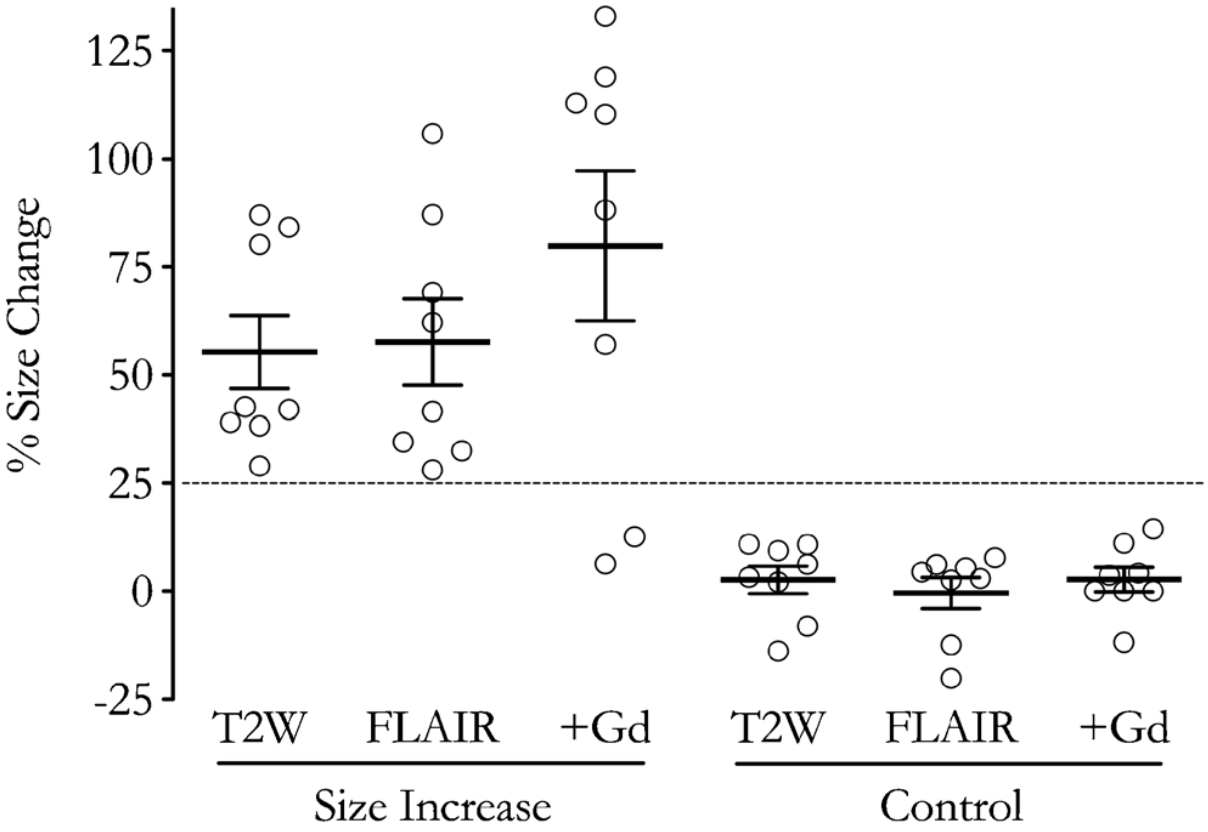

Figure 3 – MRI sequence-specific assessments of changes in tumor size on routine follow up exam pairs from 8 patients with isolated OPG. The "Size Increase" group were exam pairs identified by consensus tumor board notes where an isolated increase in tumor size on MRI had lead

to a change in tumor directed therapy. The "Control" group were exam pairs from the same patients where no significant change in tumor size was identified. In the "Size Increase" group, all size change assessments on T2W and FLAIR sequences met a commonly used threshold of 25%, reflecting meaningful size increase. In two instances for measurements made on the post-gadolinium sequence, this threshold was not met. FLAIR = fluid attenuated inversion recovery; +Gd = T1 weighted post-gadolinium sequence; T2W = T2 weighted. Solid lines represent the mean values for each group. Error bars indicate standard error from the mean.

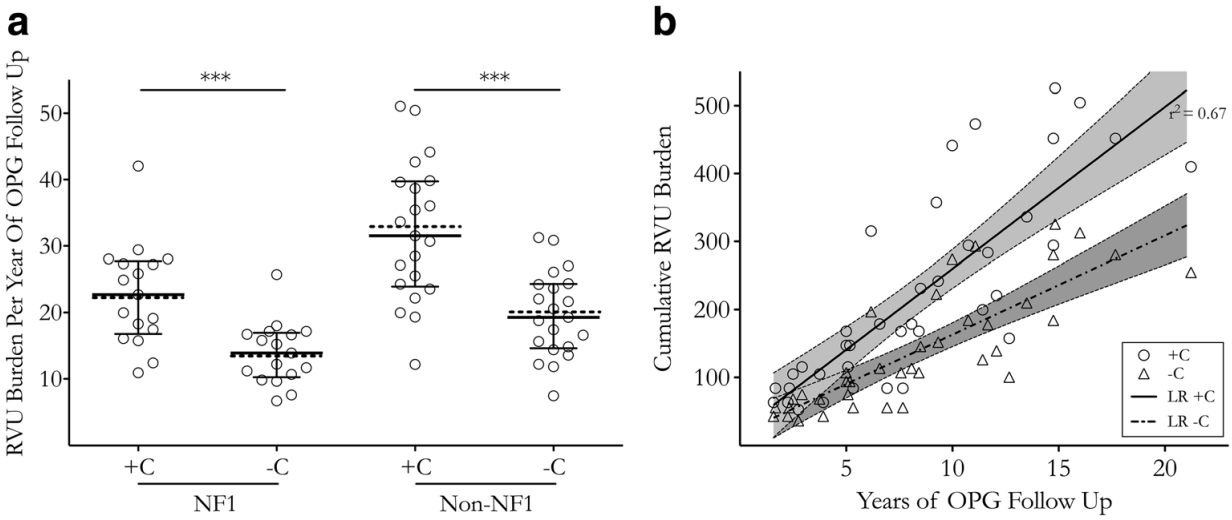

Figure 4 – Potential cost impact of a non-contrast MRI follow-up protocol for patients with isolated OPG. A) Mean (dotted line) / median (solid line) and interquartile range (whiskers) for global Relative Value Units (RVUs) generated by MR brain examinations per year of isolated OPG

follow-up with a contrast-enhanced protocol (+C; 22.7/22.7, 16.8-27.7 for NF1 patients, 32.2/31.5, 23.9-39.7 for non-NF1 patients), versus a non-contrast protocol model (-C; 13.9/13.9, 10.3-16.9 for NF1 patients, 19.7/19.3, 14.6-24.3 for non-NF1 patients). Yearly RVU burden was significantly reduced for both NF1 and non-NF1 patients under the non-contrast model, with associated p values <0.001 (***). **B)** Cumulative RVU burden for OPG MR follow-up protocols. Linear regressions (LR) fit to contrast enhanced (+C) and non-contrast enhanced (–C) models for the combined NF1 and non-NF1 patient populations generated slope values and 95% confidence intervals of 23.8 (18.2-29.5) versus 14.6 (11.1-18.0), respectively. The associated p-value was 0.006. Note that after 5 years of follow-up the 95% confidence intervals for the best fit lines (shaded areas) no longer overlap.

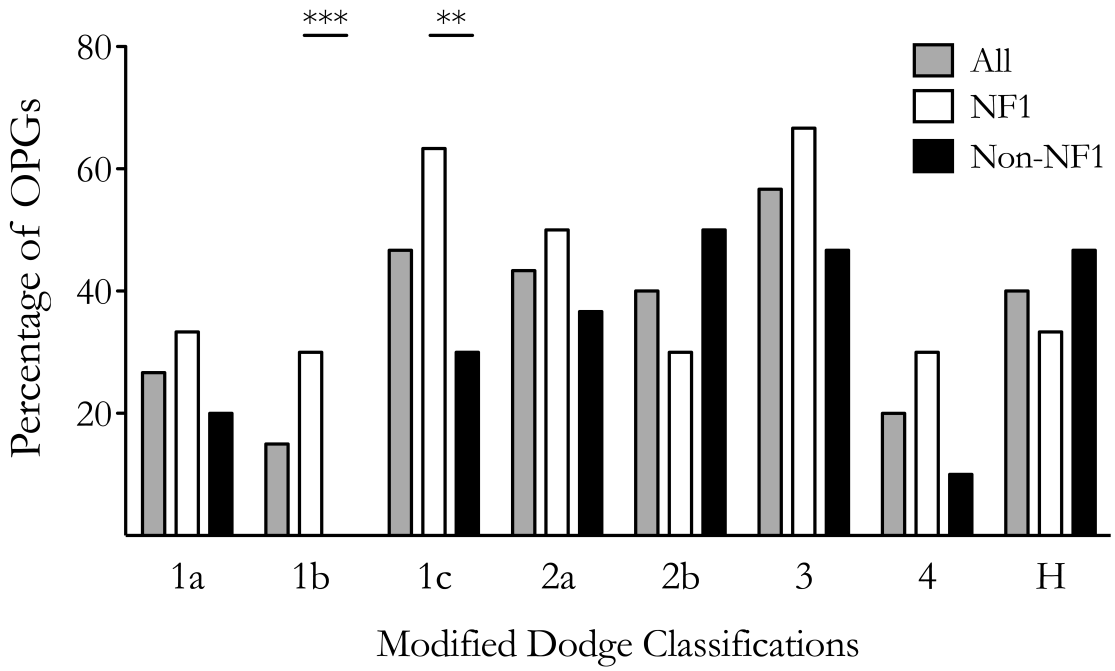

Figure 5 – Anatomic distribution of optic pathway gliomas included in the multi-reader-multi-case study. ***: p<0.001; **:p<0.01. See Taylor T et al 2008 for further detail regarding the modified dodge classification system.

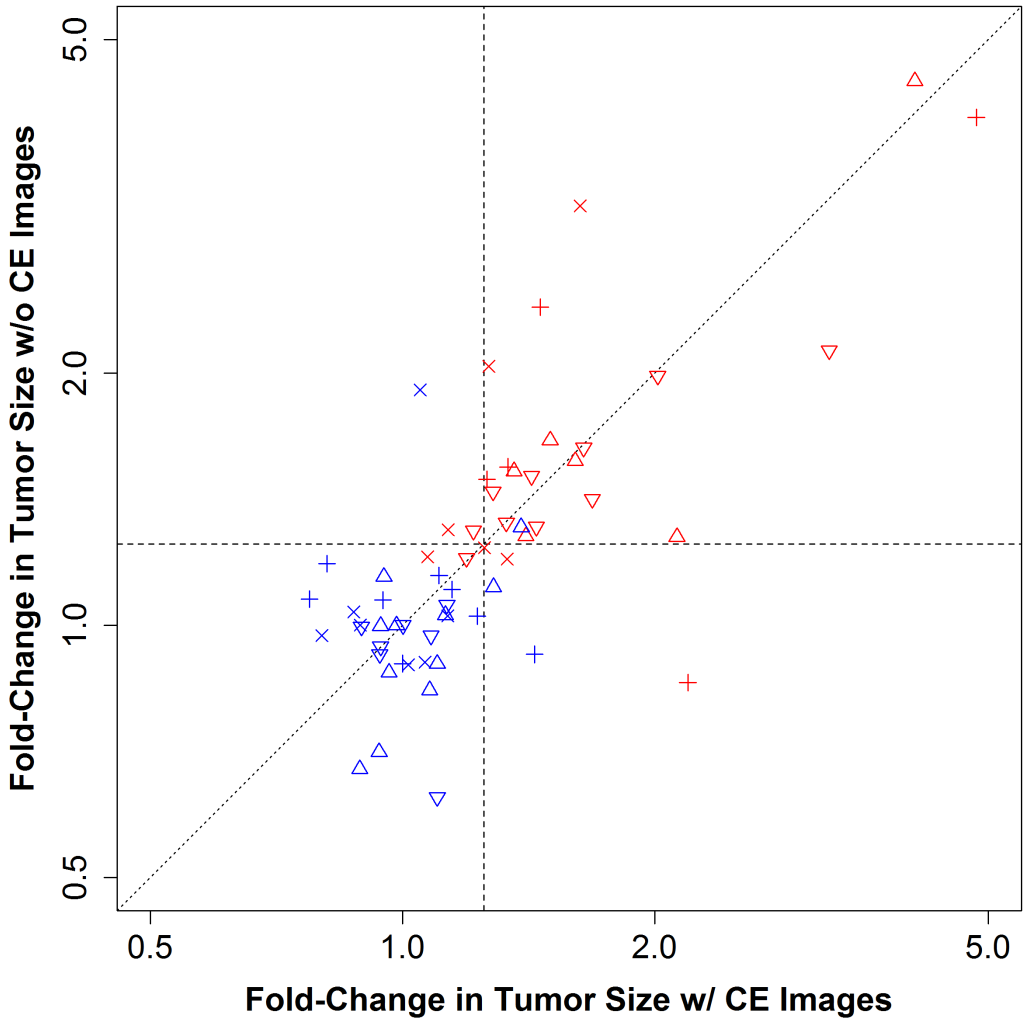

Figure 6 – Fold-changes in the tumor cross-product between the prior and current exam with and without contrast-enhanced (CE) images. The line of unity is denoted by the dotted line. The dashed vertical and horizontal lines indicate a 25% increase (fold-change of 1.25). Both axes are on the log-scale. The cases where the tumor board noted a size increase are colored in red, the remainder colored in blue. The measurements by reader 1 are indicated by the up-triangle (NF1 positive cases)

and the down-triangle (NF1 negative cases). The measurements by reader 2 are indicated by the plus symbol (NF1 positive cases) and the x-symbol (NF1 negative cases).

Table 1 – Selected characteristics of patients with isolated optic pathway glioma and ≥3 contrast-enhanced MRI brain examinations

Parameter	**Non-NF1 (n = 21)**	**NF1 (n = 17)**	**All**
Sex (% male)	52.3	65.0	57.5
Mean/Median age at diagnosis in years (IQR)	4.0/2.3 (0.8-6.3)	6.0/5.7 (3.6-7.7)	4.9/4.9 (2.1-6.4)
Mean/Median weight at diagnosis in kilograms (IQR)*	21.6/16.3 (11.1-26.6)	27.9/27.0 (17.1-37.5)	24.4/22.3 (13.2-34.5)
Mean/Median years of follow-up for OPG (IQR)	9.3/8.1 (5.1-14.8)	7.3/7.7 (3.8-11.1)	8.4/7.9 (4.7-11.8)
Private/Medicaid insurance (%)	57.1/42.9	76.6/23.4	65.8/34.2

*Weight within 1 month of diagnosis was missing for 2 (12%) of the NF1 patients, and 2 (10%) of the non-NF1 patients. Data was otherwise complete. For all variables listed, there was no significant difference between the NF1 and non-NF1 data distributions.

IQR = interquartile range; MRI = magnetic resonance imaging; NF1 = neurofibromatosis type 1.

Table 2 – Selected patient information and MRI exam-pair details from our preliminary isolated optic pathway glioma case review

Patient number	**Age at size-change MRI exams / Control exams (years)**	**Sex**	**NF1 status**	**Modified Dodge tumor classification[a]**	**Size-change exam pair axial sequences and slice thickness**	**Control exam pair axial sequences and slice thickness**
1	12 / 17	M	No	2bR 3bR	Baseline sequences (1.5T) • T1W: 5mm • T2W & FLAIR: 5 mm • T1W post-Gd: 5 mm Follow-up sequences (1.5T) • Same as above	Baseline sequences (1.5T) • T1W: 4 mm • T2W & FLAIR: 4 mm • T1W post-Gd: 4mm Follow-up sequences (1.5T) • Same as above
2	6 / 14	F	No	2a 3B	Baseline sequences (1.5T) • T1W: 5 mm	Baseline sequences (3T) • T1W: 1 mm

					• T2W & FLAIR: 5 mm • T1W post-Gd: 5 mm <u>Follow-up sequences</u> (1.5T) • Same as above	• T2W & FLAIR: 4 mm • T1W post-Gd: 1 mm <u>Follow-up sequences</u> (3T) • Same as above
3	3 / 16	M	No	2bR 3B 4B	<u>Baseline sequences</u> (1.5T) • T1W: 5mm, 3 mm FS • T2W & FLAIR: 5 mm • T1W post-Gd: 5 mm, 3 mm FS <u>Follow-up sequences</u> (1.5T) • Same as above	<u>Baseline sequences</u> (3T) • T1W: 1 mm, 2.5 mm FS orbit • T2W & FLAIR: 3 mm • T1W post-Gd: 1 mm, 2.5 mm FS orbit <u>Follow-up sequences</u> (3T) • Same as above
4	1 / 9	F	No	2bR 3B	<u>Baseline sequences</u> (3T) • T1W: 2 mm • T2W & FLAIR:	<u>Baseline sequences</u> (3T) • T1W: 1 mm • T2W & FLAIR:

					4 mm • T1W post-Gd: 1 mm, 3 mm FS <u>Follow-up sequences</u> (3T) • Same as above	3 mm • T1W post-Gd: 3 mm <u>Follow-up sequences</u> (3T) • Same as above
5	6 / 8	M	No	1cB 2bR 3bL 4L	<u>Baseline sequences</u> (1.5T) • T1W: 4 mm • T2W & FLAIR : 4 mm • T1W post-Gd: 4 mm <u>Follow-up sequences</u> (1.5T) • Same as above	<u>Baseline sequences</u> (3T) • T1W: 2 mm • T2W & FLAIR: 4 mm • T1W post-Gd: 2 mm <u>Follow-up sequences</u> (3T) • Same as above
6	6 / 9	M	Yes	1cbR 2bR 3B	<u>Baseline sequences</u> (3T) • T1W: 2 mm • T2W & FLAIR: 4 mm • T1W post-Gd: 2 mm	<u>Baseline sequences</u> (3T) • T1W: 3 mm, 3 mm FS • T2W & FLAIR: 3 mm • T1W post-Gd: 2

					Follow-up sequences (1.5T) • T1W: 4 mm • T2W & FLAIR: 4 mm • T1W post-Gd: 4 mm	mm, 3 mm FS Follow-up sequences (3T) • Same as above
7	2 / 6	M	No	1cR 2bR	Baseline sequences (3T) • T1W: 2 mm • T2W & FLAIR: 4 mm • T1W post-Gd: 2 mm Follow-up sequences (1.5T) • T1W: 4 mm • T2W & FLAIR: 4 mm • T1W post-Gd: 4 mm	Baseline sequences (3T) • T1W: 1 mm • T2W & FLAIR: 3 mm • T1W post-Gd: 1 mm Follow-up sequences (1.5T) • T2W & FLAIR: 4 mm • Otherwise the same
8	9 / 17	M	No	1aL 2a 3bR	Baseline sequences (1.5T)	Baseline sequences (3T)

	• T1W: 5 mm • T2W & FLAIR: 5 mm • T1W post-Gd: 3 mm <u>Follow-up sequences</u> (1.5T) • T1W post-Gd: 5 mm, 3 mm FS • Otherwise the same	• T1W: 2 mm, 3 mm FS • T2W & FLAIR: 4 mm • T1W post-Gd: 2 mm, 3 mm FS <u>Follow-up sequences</u> (3T) • T1W: 1 mm, 3 mm FS orbit • T2W & FLAIR: 3 mm • T1W post-Gd: 1 mm, 3 mm FS orbit

[a] See reference Taylor T, 2008.

FLAIR = fluid attenuated inversion recovery; F = female; FS = fat saturated; Gd = gadolinium; mm = millimeter; M = male; MRI = magnetic resonance imaging; NF1 = neurofibromatosis type-1; "orbit" = orbital-specific sequence; T = tesla; T1W = T1 weighted; T2W = T2 weighted.