



Published in final edited form as:

Clin Cancer Res. 2023 June 13; 29(12): 2194–2198. doi:10.1158/1078-0432.CCR-22-3524.

Accessible Data Collections for Improved Decision Making in Neuro-Oncology Clinical Trials

Rifaquat Rahman¹, Steffen Ventz², Robert Redd³, Timothy Cloughesy⁴, Brian M. Alexander⁵, Patrick Y. Wen⁶, Lorenzo Trippa⁷

¹Rifaquat Rahman, MD. Department of Radiation Oncology, Dana-Farber/Brigham and Women's Cancer Center, Harvard Medical School, Boston, MA, USA.

²Steffen Ventz, PhD. Division of Biostatistics, University of Minnesota, Minnesota, MN, USA.

³Robert Redd, MS Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, USA.

⁴Timothy Cloughesy, MD. Neuro-Oncology Program and Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA.

⁵Brian M. Alexander, MD. Department of Radiation Oncology, Dana-Farber/Brigham and Women's Cancer Center, Harvard Medical School, Boston, MA, USA; Foundation Medicine, Inc., Cambridge, MA, USA

⁶Patrick Y. Wen, MD. Center for Neuro-Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA.

⁷Lorenzo Trippa, PhD. Department of Data Sciences, Dana-Farber Cancer Institute, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

Abstract

Drug development can be associated with slow timelines, particularly for rare or difficult-to-treat solid tumors such as glioblastoma. The use of external data in the design and analysis of trials has attracted significant interest since it has the potential to improve the efficiency and precision of drug development. A recurring challenge in the use of external data for clinical trials, however, is the difficulty in accessing high-quality patient-level data. Academic research groups generally do not have access to suitable datasets to effectively leverage external data for planning and analyses of new clinical trials. Given the need for resources to enable investigators to benefit from

Corresponding author: Rifaquat Rahman, MD, Dana-Farber/Brigham and Women's Cancer Center, CNS Radiation Oncology, Harvard Medical School, rrahman@bwh.harvard.edu, Phone: 857-215-0462, Twitter: @RifRahmanMD.

Related links: Link to GBM External Data Platform website: <https://rconnect.dfci.harvard.edu/gbmdata/>

Competing interests: TC reports personal fees from Roche, Trizel, Medscape, Bayer, Amgen, Odonate Therapeutics, Pascal Biosciences, Del Mar, Tocagen, Karyopharm, GW Pharma, Kiyatec, AbbVie, Boehringer Ingelheim, VBI, Dicephera, VBL, Agios, Merck, Genocoea, Puma, Lilly, BMS, Cortice, Wellcome Trust, other from Notable Labs; outside the submitted work. BA reports employment at Foundation Medicine. BA reports personal fees from AbbVie, Bristol-Myers Squibb, Precision Health Economics, and Schlesinger Associates, outside of submitted work. He reports research support from Puma, Eli Lilly, Celgene, outside of submitted work. PW reports personal fees from Abbvie, Agios, Astra Zeneca, Blue Earth Diagnostics, Eli Lilly, Genentech/Roche, Immunomic Therapeutics, Kadmon, Kiyatec, Merck, Puma, Vascular Biogenics, Taiho, Tocagen, Deciphera, VBI Vaccines and research support from Agios, Astra Zeneca, Beigene, Eli Lilly, Genentech/Roche, Karyopharm, Kazia, MediciNova, Merck, Novartis, Oncocutics, Sanofi-Aventis, VBI Vaccines, outside the submitted work. SV, RR, LT declare no competing interests.

existing data assets, we have developed the Glioblastoma External (GBM-X) Data Platform which will allow investigators in neuro-oncology to leverage our data collection and obtain analyses. GBM-X strives to provide an uncomplicated process to use external data, contextualize single arm trials, and improve inference on treatment effects early in drug development. The platform is designed to welcome interested collaborators and integrate new data into the platform, with the expectation that the data collection can continue to grow and remain updated. With such features, the GBM-X Data Platform is designed to help to accelerate evaluation of therapies, to grow with collaborations, and to serve as a model to improve drug discovery for rare and difficult-to-treat tumors in oncology.

Introduction

In oncology, drug development is associated with slow timelines, particularly for rare or difficult-to-treat solid tumors such as glioblastoma. Randomized controlled trials (RCTs) constitute the gold standard for therapeutic testing, but they can have limitations, particularly in the early stages of development, including high cost and prolonged timelines. For disease settings such as gliomas, single arm trials are commonly used but associated with risks of bias and inaccurate decision making in early phase trials.¹ There is therefore a need for novel approaches and statistical designs to evaluate candidate therapies and for decision making throughout the development pathway of novel treatments.

Leveraging external data with patient-level information has been proposed as a potential strategy to accelerate the development of new therapies, particularly for rare or difficult-to-treat cancers.² *Externally augmented clinical trial* (EACT) designs incorporate pre-specified external data with patient-level information in the analysis plan of a clinical trial. Several feasibility studies have shown that external datasets could accelerate the evaluation of new therapies,^{2,3} and this has garnered interest from regulatory agencies.^{4,5} The most direct application is to use an external control arm to contextualize the results of single arm trials, and this has been explored across several indications in oncology.^{2,3,5} The use of statistically valid adjustment methods to remove the influence of confounding covariates, which differ in the trial and external populations, can facilitate the evaluation of novel treatments with greater precision compared to conventional single arm trial analyses. While other uses of external data are largely unexplored, external control data has the potential to bolster decision making in clinical trials, including possible applications in interim analyses⁶, subgroup analyses, hybrid randomized trials, and selection of the sample size and other design characteristics.

Applications of External Data in Neuro-Oncology

The use of external data has attracted significant interest in neuro-oncology.⁷ Initial efforts have focused on glioblastoma, a disease with a dismal prognosis and in need of therapeutic advances. Of note, despite a long history of poor decision making leading to repeated failed phase 3 trials, single arm studies continue to be common in glioblastoma early phase trials.¹ Retrospective analyses evaluating the use of external control arms have demonstrated potential value in glioblastoma, with reduction of false positive results compared to standard analyses of single arm trials.² Other applications of external control data in neuro-oncology

clinical trials, ranging from interim decisions to the analysis of randomized trials, have been previously discussed.^{2,6,7} A prospective phase 2 single-arm trial analyzed with an external control dataset has recently been reported in recurrent glioblastoma, and a subsequent registrational trial that will leverage external data in a hybrid randomized design has been announced.⁸

To develop designs that utilize external controls for newly diagnosed glioblastoma trials, a recurring challenge we have encountered is the difficulty in overcoming barriers to access high-quality patient-level data. These datasets ideally come from previously completed clinical trials or alternatively from *real-world* repositories such as institutional libraries of well-annotated datasets. Academic research groups generally will not have access to suitable datasets to effectively leverage external data for planning and analyses of new clinical trials. Data sharing has long been difficult, and barriers to data sharing have been discussed at length, including concerns related to patient privacy, academic credit, data sharing infrastructures, costs, data standards and inappropriate secondary analyses.⁹ Datasets of previously completed clinical studies, including trials conducted with support from federal agencies, remain difficult to access for the outlined purposes. In our experience, years can elapse when requesting data from investigators, industry, cooperative-groups and other organizations. Moreover, after pursuing the process to request data, it is difficult to predict if approval will be granted or all requested data elements will be provided. These challenges affect the balance between costs -including time and dedicated personnel to utilize these data sources- and potential efficiencies that could be achieved integrating external patient-level data in future clinical trials. Of note, in the literature¹⁰ and in our discussions with patient advocacy groups in neuro-oncology, there is support for the use of data from completed trials for further clinical research.⁷

Development of a Glioblastoma Data Platform

In light of the potential of external data in oncology research and preferences of patients to have their data used outside of a single study, we need practical solutions to reduce barriers raised by industry and academia. Thus, we recognized the need for mechanisms to enable investigators to benefit from existing data assets. With this in mind, we have developed the Glioblastoma External (GBM-X) Data Platform (<https://rconnect.dfci.harvard.edu/gbmdata/>) that will allow investigators in neuro-oncology to leverage our data collection. Current datasets include deidentified individual patient-level data of over 1,200 patients with newly diagnosed glioblastoma with relevant pre-treatment covariates, extracted from 6 clinical trials and 2 institutional databases (Table 1). All these datasets include patients that received the standard of care therapy of radiation therapy and temozolomide for newly diagnosed glioblastoma.

GBM-X strives to provide an uncomplicated process to use external data, contextualize single arm trials, and improve inference on treatment effects early in drug development. The proposed workflow for the data platform is illustrated in Figure 1. Investigators who have an ongoing or completed glioblastoma clinical trial can request an analysis to integrate information from the GBM-X Data Platform. We have structured the platform to allow users at various institutions to obtain these analyses without direct sharing of deidentified

patient-level data from our data collection. We offer a library of standardized analyses that integrate data from new trials with patient-level information from our data collection. Once standardized agreements are achieved and patient privacy is assured via institutional review board requirements, we will run the analysis and provide results, such as treatment effect estimates, by using established statistical methods. The investigators that request the analyses will receive (i) data dictionaries of our datasets (ii) simulated datasets that resemble major characteristics of the actual GBM-X datasets (unit of measures, names of the variables, etc.) and (iii) the R code used for the analyses. These are key components that will provide the users transparent information on the statistical procedures underlying the data analysis, and investigators can ask the GBM-X team for additional details as needed. The platform will be initially geared towards investigator-initiated trials, and we will limit the service to ten clinical studies for an initial one-year pilot period. We will then expand beyond these constraints based on the experience during the pilot period and feedback from the community and users.

The GBM-X Data Platform builds on the experience of existing data-sharing platforms (e.g. Project Data Sphere, YODA, Vivli) that are generally broad in scope and have an assortment of trial datasets across different indications, including common and rare cancers. While these efforts have shown tremendous potential, investigators hoping to leverage external data for trials will generally benefit from as much disease-specific datasets as possible.² GBM-X seeks to unify data assets for a single disease with the explicit purpose of serving future analyses and interim decisions of early-stage trials in glioblastoma.

Definitions of all variables in the current GBM-X database and study populations have been compared to identify potential differences across studies. We used data dictionaries and publications to examine the definitions and representations of populations, outcomes, and pre-treatment variables across datasets. Consistency of formats and definitions of variable are necessary to effectively leverage external datasets in the analysis of clinical trials.

Future Directions

Having a data hub is essential to effectively leverage external data in future clinical trials, but the data platform must be dynamic and up to date to stay relevant. Indeed, it is necessary to account for changes such as improvements in treatments, supportive care, or the identification of novel biomarkers. Moving forward, GBM-X can potentially expand in multiple directions. While we have deidentified patient-level datasets for newly diagnosed glioblastoma, we envision future growth of the data collection by (1) incorporating additional datasets, (2) integrating additional patient-level information, such as imaging, next generation sequencing data, and toxicity data, and (3) broadening to other disease populations (recurrent glioblastoma, IDH-mutant gliomas, and H3K27M-mutated gliomas). We will prioritize extensions that can translate into more efficient development of new treatments.

The platform is designed to welcome interested collaborators that want to share data, with the expectation that the data collection can continue to grow and remain temporally relevant. In accepting data from other research groups, we are allowing for flexibility to accept data through two collaborative data-sharing models. The first model is (1) centralized

patient-level data sharing, where deidentified patient-level data is housed on the GBM-X server (see Figure 1). The second model is (2) data-private collaborative learning without direct patient-level data sharing (i.e., contributors submit data summaries to GBM-X). In this latter model of data sharing, we will incorporate data summaries, such as regression models, which will add information from completed studies to GBM-X. This approach facilitates data sharing and provides an option for investigators who want to contribute without the complexities of sharing patient-level records. Standard meta-analytic methods can be used to summarize the regression models that will be provided by different study groups into a single regression function,¹¹ and this summary can be used in the analysis of single-arm trials to infer treatment effects. Any data provided to the GBM-X Data Platform will be used only for its prespecified purpose, as specified by data contributors.

Conclusion

RCTs will remain the indisputable gold standard for the evaluation of treatments, but external datasets can supplement information gleaned from RCTs and single arm studies. As data access remains the greatest barrier to studying and implementing EACTs, data sharing platforms such as GBM-X attempt to break down these barriers and allow for more efficient treatment development in neuro-oncology. The GBM-X Data Platform can help accelerate the evaluation of new therapies in neuro-oncology and can serve as a model to improve drug discovery for rare and difficult-to-treat tumors in oncology.

Acknowledgements:

The authors thank Jon McDunn and Bill Louv for input on this manuscript and Project Data Sphere for research support. RR is supported by the Dana-Farber Cancer Institute Early Career Faculty Innovation Fund and the Joint Center for Radiation Therapy Foundation Grant. LT supported by NIH grant R01LM013352.

References

1. Vanderbeek AM et al. The clinical trials landscape for glioblastoma: is it adequate to develop new treatments? *Neuro-oncology* (2018) doi:10.1093/neuonc/noy027.
2. Venz S. et al. Design and Evaluation of an External Control Arm Using Prior Clinical Trials and Real-World Data. *Clin. Cancer Res* 25, 4993–5001 (2019). [PubMed: 31175098]
3. Carrigan G. et al. Using Electronic Health Records to Derive Control Arms for Early Phase Single-Arm Lung Cancer Trials: Proof-of-Concept in Randomized Controlled Trials. *Clin Pharmacol Ther* 107, 369–377 (2020). [PubMed: 31350853]
4. Mishra-Kalyani PS et al. External control arms in oncology: current use and future directions. *Ann Oncol* 33, 376–383 (2022). [PubMed: 35026413]
5. Amiri-Kordestani L. et al. A Food and Drug Administration analysis of survival outcomes comparing the Adjuvant Paclitaxel and Trastuzumab trial with an external control from historical clinical trials. *Ann Oncol* (2020) doi:10.1016/j.annonc.2020.08.2106.
6. Venz S. et al. The Use of External Control Data for Predictions and Futility Interim Analyses in Clinical Trials. *Neuro-Oncology* (2021) doi:10.1093/neuonc/noab141.
7. Rahman R. et al. Leveraging external data in the design and analysis of clinical trials in neuro-oncology. *Lancet Oncol* 22, e456–e465 (2021). [PubMed: 34592195]
8. Sampson JH et al. Targeting the IL4 Receptor with MDNA55 in Patients with Recurrent Glioblastoma: Results of a Phase 2b Trial. *Neuro Oncol* noac285 (2023) doi:10.1093/neuonc/noac285.

9. Rockhold F, Nisen P. & Freeman A. Data Sharing at a Crossroads. *N Engl J Med* 375, 1115–1117 (2016). [PubMed: 27653563]
10. Mello MM, Lieou V. & Goodman SN Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing. *N Engl J Med* 378, 2202–2211 (2018). [PubMed: 29874542]
11. DerSimonian R. & Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 7, 177–188 (1986). [PubMed: 3802833]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Statement of Translational Relevance:

External data in the design and analysis of trials has the potential to improve the efficiency and precision of drug development. Given a need for mechanisms to enable investigators to benefit from high-quality patient-level data of newly diagnosed glioblastoma patients, the Glioblastoma External (GBM-X) Data Platform is a tool designed to allow neuro-oncology investigators to leverage our data collection and obtain analyses. GBM-X strives to provide an uncomplicated process to use external data, contextualize single arm trials, and improve inference on treatment effects early in drug development. The GBM-X Data Platform can serve as a model to improve drug discovery for rare and difficult-to-treat tumors in oncology.

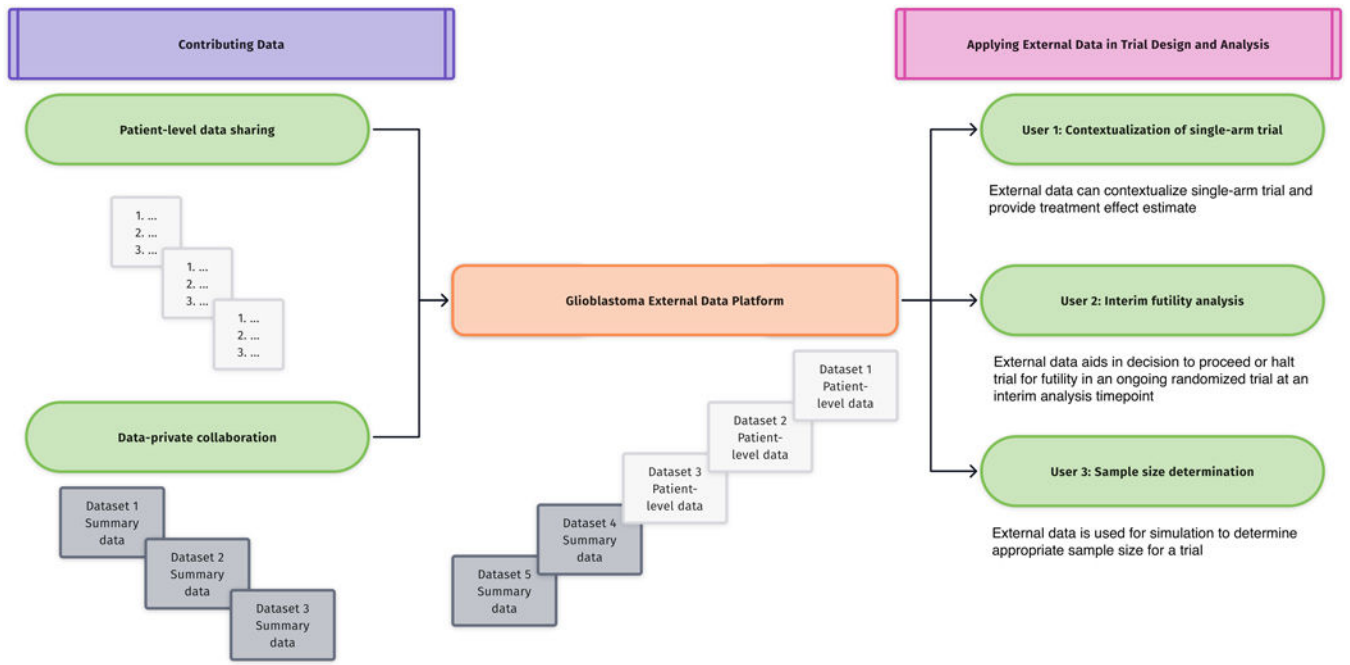


Figure 1:
A schematic representation of the Glioblastoma External Data Platform.

Table 1:

Initial patient-level data collections in the Glioblastoma External Data Platform

	Source							
	Cho et al. 2011 (PMID: 22001862)	DFCI GBM	NCT00441142	NCT00689221	NCT00813943	NCT00943826 #	NCT02977780	UCLA/KP GBM
Total	n = 1,662 (%)	n = 625 (38)	n = 27 (2)	n = 272 (16)	n = 88 (5)	n = 459 (28)	n = 69 (4)	n = 106 (6)
Age								
Median (range)	58 (17 – 94)	60 (17 – 94)	55 (26 – 73)	57 (21 – 78)	57 (21 – 74)	56 (18 – 79)	59 (24 – 75)	58 (20 – 79)
< 65	1,277 (77)	442 (71)	25 (93)	219 (81)	77 (88)	378 (82)	54 (78)	70 (66)
>= 65	385 (23)	183 (29)	2 (7)	53 (19)	11 (12)	81 (18)	15 (22)	36 (34)
Sex								
Female	684 (41)	270 (43)	12 (44)	129 (47)	34 (39)	165 (36)	28 (41)	38 (36)
Male	978 (59)	355 (57)	15 (56)	143 (53)	54 (61)	294 (64)	41 (59)	68 (64)
KPS								
<90	481 (29)	245 (39)	5 (19)	120 (44)	49 (56)	-	24 (35)	31 (29)
90–100	646 (39)	304 (49)	22 (81)	152 (56)	39 (44)	-	45 (65)	75 (71)
Unknown	535 (32)	76 (12)	-	-	-	459 (100)	-	-
RPA								
3	234 (14)	65 (10)	-	46 (17)	16 (18)	78 (17)	-	27 (25)
4–5	1,309 (79)	537 (86)	-	226 (83)	72 (82)	381 (83)	-	79 (75)
Unknown	119 (7)	23 (4)	27 (100)	-	-	-	69 (100)	-
MGMT promoter methylation status								
Unmethylated	768 (46)	313 (50)	15 (56)	-	88 (100)	236 (51)	69 (100)	40 (38)
Methylated	687 (41)	254 (41)	6 (22)	272 (100)	-	116 (25)	-	30 (28)
Unknown	207 (12)	58 (9)	6 (22)	-	-	107 (23)	-	36 (34)
Extent of Surgical Resection								
Biopsy	130 (8)	55 (9)	5 (19)	-	-	42 (9)	6 (9)	22 (21)
Gross total resection	771 (46)	293 (47)	9 (33)	137 (50)	46 (52)	192 (42)	37 (54)	46 (43)
Subtotal resection	746 (45)	277 (44)	13 (48)	126 (46)	36 (41)	225 (49)	26 (38)	38 (36)
Unknown	15 (1)	-	-	9 (3)	6 (7)	-	-	-
IDH mutation status								

	Source							
	Cho et al. 2011 (PMID: 22001862)	DFCI GBM	NCT00441142	NCT00689221	NCT00813943	NCT00943826 #	NCT02977780	UCLA/KP GBM
Total	n = 1,662 (%)	n = 625 (38)	n = 27 (2)	n = 272 (16)	n = 88 (5)	n = 459 (28)	n = 69 (4)	n = 106 (6)
Wildtype	760 (46)	620 (99)	17 (63)	-	-	-	69 (100)	54 (51)
Mutant	6 (0)	1 (0)	4 (15)	-	-	-	-	1 (1)
Unknown	896 (54)	4 (1)	6 (22)	272 (100)	88 (100)	459 (100)	-	51 (48)
Time-to-event in months								
Follow-up time								
Median (95% CI)	33 (32 – 35)	NR (NR - NR)	38 (22 - NR)	30 (28 – 32)	22 (21 - NR)	31 (29 – 32)	19 (15 – 23)	40 (37 – 47)
Overall survival								
Median (95% CI)	19 (19 – 20)	21 (20 – 22)	16 (11 – 32)	27 (24 – 33)	14 (13 – 15)	17 (15 – 18)	15 (13 – 17)	21 (19 – 26)
Progression-free survival								
Median (95% CI)	9 (8 – 10)	10 (10 – 11)	8 (5 – 23)	15 (12 – 19)	8 (6 – 8)	6 (6 – 8)	6 (6 – 8)	8 (6 – 11)

KPS = Karnofsky Performance Status

RPA = Recursive Partitioning Analysis

MGMT = O⁶-methylguanaine-DNA-methyltransferase

IDH = isocitrate dehydrogenase

Limited data access, used in prior analysis³