



November 6, 2001

VIA HAND DELIVERY

Advisors and Consultants Staff
Karen M. Templeton-Somers, Ph.D.
Executive Secretary
Oncology Drugs Advisory Committee
Center for Drug Evaluation and Research
Food and Drug Administration
HFD-21, Room 1093
5630 Fishers Lane
Rockville, MD 20852-1734

**RE: NDA 20-637
GLIADEL[®] Wafer (polifeprosan 20 with carmustine implant)
Briefing Document – December 6, 2001 ODAC Meeting**

Dear Ms. Somers:

Please find enclosed forty copies of the GLIADEL[®] Wafer Briefing Document which summarizes the clinical trials supporting the new labeling.

This document consists of the following information:

- **Executive Summary**
- **Proposed Product Insert**
- **Study 8802**
 - Placebo-controlled trial of safety and efficacy of interoperative controlled delivery by biodegradable polymers of chemotherapy for recurrent gliomas
 - Study Synopsis
- **Study 0190**
 - Interstitial Chemotherapy with Carmustine-loaded Polymers for High-grade Gliomas: A Randomized Double-Blind Study
 - Study Synopsis

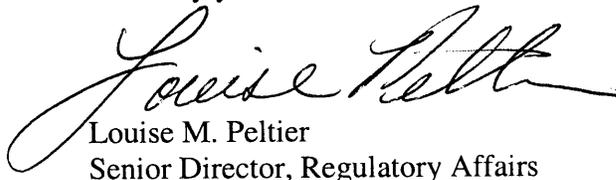
- **Study T-301**
 - Study Synopsis

- **Statistical References**
 - Adjustments for Center in Multicenter Studies: An Overview

 - Analysis of Data from Multicenter Trials

Your assistance has been greatly appreciated. Should you have any questions about this submission, please contact me at (410) 631-6356.

Sincerely yours,



Louise M. Peltier
Senior Director, Regulatory Affairs

LP/lkb

Enclosures



GLIADEL[®] WAFER

**(polifeprosan 20 with carmustine implant)
NDA No. 20-637, Supplement # 016
First-Line Treatment for Malignant Glioma Indication**

BRIEFING DOCUMENT

**Oncologic Drugs
Advisory Committee Meeting**

December 6, 2001

Available For Public Disclosure Without Redaction

Guilford Pharmaceuticals Inc., Baltimore, Maryland 21224

I. Background

GLIADEL® Wafer currently has marketing authorization in the United States and in Europe (including France, Germany, Austria, Greece, Ireland, Luxembourg, Portugal, Spain, United Kingdom and The Netherlands), for the treatment of patients with recurrent glioblastoma multiforme in whom surgical resection is indicated. In Canada GLIADEL® is approved for recurrent and newly diagnosed malignant glioma.

The approval of GLIADEL® for recurrent GBM was based upon the results obtained from four clinical studies (Table 1).

TABLE 1:

<u>Clinical Trials: Recurrent Malignant Glioma</u>		
Clinical Trial	Patients Enrolled	Study Design
Study 8701	21	Multicenter, open-label, dose escalation Phase I/II
Study 9115	40	Multicenter, open-label Phase III
Study 8802	222	Multicenter, randomized, double-blind, placebo controlled Phase III
Study 9501	<u>349</u>	Treatment protocol
Total	632	

Three additional clinical trials were conducted to assess the safety and efficacy of GLIADEL® Wafer in the treatment of patients with primary malignant glioma. (Table 2).

TABLE 2:

<u>Clinical Trials: Newly Diagnosed Malignant Glioma</u>		
Clinical Trial	Patients Enrolled	Study Design
Study 9003	22	Multicenter, open-label Phase I/II
Study 0190	32	Multicenter, randomized, double-blind, placebo controlled Phase III
Study T-301	<u>240</u>	Multicenter, randomized, double-blind, placebo controlled Phase III
Total	294	

The results of the two Phase III trials (Studies 0190 and T-301) in patients with newly diagnosed malignant glioma and the trials of GLIADEL® Wafer in recurrent disease support the proposed new indication for GLIADEL® wafer as a treatment to significantly prolong survival and maintain overall function (as measured by preservation of Karnofsky Performance Status) and neurological function in patients with malignant glioma undergoing primary and/or recurrent surgical resection.

II. Overview of Primary Malignant Glioma

Patients with malignant glioma typically present with symptoms and signs of neurological dysfunction including headache, seizure, or a new neurologic deficit. The average age of onset is 55-60 years old. There are approximately 17,500 new cases of primary malignant glioma diagnosed each year with more than 10,000 deaths attributed to this condition. Approximately 75% of the cases are glioblastoma multiforme (or Grade IV malignant glioma) which is the most malignant form of the disease.

The initial provisional diagnosis of malignant glioma is made after an imaging study such as a CT or MRI scan. While the surgeon may have a high index of suspicion that the patient has a high-grade glioma, the tentative diagnosis of malignant glioma cannot be confirmed until pathological examination has been completed. At the time of initial craniotomy for tumor biopsy and resection a provisional pathological diagnosis is made based on an intraoperative tissue sample that the neuropathologist examines by frozen section or squash prep. This allows the pathologist to inform the surgeon that the patient likely has a 'malignant glioma'. The exact histological diagnosis cannot be rendered until the final pathological assessment has been completed requiring fixed tissue examination. Therefore, the neurosurgeon proceeds with a provisional diagnosis in the operating room.

Certain clinical factors have repeatedly been shown to significantly influence survival in patients with malignant glioma. Advanced age (≥ 60 years of age), poor Karnofsky Performance Score (≤ 70), and tumor histology (glioblastoma multiforme or Grade IV glioma) are factors conferring the worst prognosis. Tumor size and extent (fraction) of tumor resection have been proposed to be important prognostic factors but are not universally accepted.

Therapy of Malignant Glioma

The standard treatment for patients with malignant glioma is palliative in nature and typically consists of primary surgical resection followed by radiotherapy and in some cases chemotherapy. Complete resection of the primary malignant glioma is virtually impossible due to the infiltrative nature of the lesion and the limitation that wide resection margins are impossible in the brain. Although clinical observational studies have suggested that optimal tumor resection improves survival, this point is not universally accepted. Local tumor recurrence is frequent, usually within 2 cm of the original resection margin.¹ This tumor recurrence leads to reoperation in many patients.

The standard treatments for malignant glioma are only modestly effective. With surgery alone the median survival is approximately 4 months.² Two clinical studies from the Brain Tumor Study Group (currently called the Brain Tumor Cooperative Group – BTCCG) showed that adding postoperative external beam radiation therapy to surgery alone increased median survival by approximately 18 to 22 weeks from 4-6 months to approximately 9-11 months. Adding systemic chemotherapy to surgery and radiation therapy was of very limited additional benefit in these studies.

Chemotherapy has been extensively examined in clinical studies. Carmustine or BCNU is the most widely studied agent. Other chemotherapeutic regimens include PCV (procarbazine, lomustine, and vincristine) and temozolomide although these therapies have not been conclusively shown to increase survival in patients with GBM or malignant glioma in randomized controlled trials. In the largest study conducted to date, the efficacy of PCV chemotherapy has been studied in 674 patients with the diagnosis of malignant glioma.³ Patients aged ≤ 70 years old with the diagnosis of WHO Grade 3 or 4 malignant glioma were randomized to receive either surgery and radiotherapy (RT Group) (n= 339) or surgery, radiotherapy, and PCV chemotherapy (RT-PCV Group) (n= 335). The groups were well matched for baseline prognostic indicators including tumor type/histology, age, and KPS: 76% of the patients in this study had Grade IV gliomas (GBM). The median survival for the RT Group was 9.5 months versus 10 months for the RT-PCV Group. The authors of this study concluded that ‘the results of this large randomized trial have failed to demonstrate a routine place for adjuvant chemotherapy with PCV in the treatment of high-grade glioma’. This study also confirms that the projected median survival for patients with high grade malignant glioma receiving surgery and radiotherapy is approximately 9-10 months.

However, Fine, *et al.*, conducted a meta-analysis of all available chemotherapy trials (only studies through 1989 were included in this analysis) that suggested that the proportion of patients diagnosed as having malignant glioma surviving for 12 months increased from 43% to 53% with the addition of systemic chemotherapy.⁴ Therefore, the role of systemic chemotherapy for patients with high grade malignant glioma remains unclear.

Carmustine or BCNU, a chemotherapeutic agent active both in vitro and in vivo against glioma cells, has been approved for use in the treatment of malignant brain tumors in the United States since 1979. However, intravenous BCNU has several limitations. Although it is lipophilic and crosses the blood-brain barrier, its half-life in the circulation after intravenous administration is only approximately 15 minutes.^{5,6} Furthermore, Carmustine doses used are often associated with systemic toxicities such as delayed myelosuppression⁷ and, less frequently, pulmonary fibrosis.^{8,9}

Local Therapy of Malignant Glioma

Because most malignant gliomas recur within two centimeters of their initial boundaries,¹⁰ local (regional) therapy for malignant gliomas is a logical approach to treatment. Local (regional) therapy affords an opportunity to increase tumor exposure to a chemotherapeutic agent by increasing the local concentration of the chemotherapeutic agent or the duration of contact with the tumor, or both. In addition, local therapy decreases systemic toxicity. Local (regional) therapy in the treatment of malignant gliomas has taken several approaches, including targeted intra-arterial infusions,¹¹ infusion through implanted catheters,¹² reservoirs,¹³ or pumps,¹⁴ and targeted disruption of the blood-brain barrier followed by systemic chemotherapy.¹⁵ A new approach to local (regional) therapy for malignant gliomas is the use of implanted polymers containing chemotherapeutic agents.¹⁶

GLIADEL® Wafer (Polifeprosan 20 with carmustine implant) is a biodegradable wafer, composed of a copolymer matrix containing 3.85% BCNU. The GLIADEL® Wafer is designed to deliver carmustine directly into the surgical cavity created when a brain tumor is resected, and thereby to increase local BCNU concentrations in adjacent brain tissue. The drug delivery effectiveness of polifeprosan 20 with BCNU implant is based on the diffusion release of carmustine from the copolymer matrix as the wafer degrades following intracerebral implantation. Each polifeprosan 20 with carmustine implant (chemical name poly[bis(p-carboxyphenoxy)propane:sebacic acid 20:80] with 3.85% 1,3-bis(2-chloroethyl)-1-nitrosourea) weighs 200 mg.

Preclinical data on BCNU released from intracerebrally-implanted polymers composed of ethylene vinyl acetate, (EVAC), or poly[bis(p-carboxyphenoxypropane):sebacic acid], (PCPP:SA), have shown sustained release, producing high local BCNU concentrations,^{17,18} and survival has been shown to be extended when compared with controls in a model of established intracranial 9L gliosarcoma.¹⁹

The release of BCNU from GLIADEL® Wafer implants has been studied in vitro and in vivo.²⁰ The GLIADEL® copolymer matrix degrades by hydrolysis in vivo. Results from these studies show that BCNU is delivered for approximately 2-3 weeks. Concentrations at the tumor site are approximately 100-fold greater than levels achieved after intravenous BCNU administration. Several studies have compared the survival of animals treated with parenteral BCNU using the 9L gliosarcoma model to those treated with GLIADEL® Wafer polymer implants and have shown a significant increase in survival with local therapy. There has been no evidence of systemic toxicity,²¹ and only minimal local inflammatory changes around the implant in a primate model.¹⁸

Clinically, there are no detectable plasma BCNU levels in patients treated with GLIADEL® wafers containing 3.85% BCNU. Because the systemic levels are so low, no systemic toxicities are attributable to the use of GLIADEL® wafer.

Virtually all patients receive primary resection at the time of initial diagnosis and treatment for malignant glioma. The use of GLIADEL® wafer does not necessitate any additional

surgery or intervention. The wafers can be simply implanted at the time of surgical resection. (Figure 1).

Dosage And Administration

Each GLIADEL wafer contains 7.7 mg of carmustine, resulting in a BCNU dose of 61.6 mg when eight wafers are implanted. It is recommended that eight wafers be placed in the resection cavity if the size and shape of it allows. Should the size and shape of the tumor resection cavity not accommodate eight wafers, the maximum number of wafers should be placed. Since there is no clinical experience with a dose of >8 wafers, therefore no more than eight wafers should be used in each surgical procedure.

Figure 1:
Implantation of GLIADEL® Wafer after Tumor Resection



The clinical experience to date with GLIADEL® wafer is significant with more than 6,000 patients having been treated with GLIADEL® wafers. GLIADEL® is generally well tolerated with attention by the surgeon to post operative management of cerebral edema with corticosteroids, a watertight dural closure to decrease the likelihood of CSF leak, and use of postoperative anti-convulsant medications. The postoperative use of corticosteroids and anticonvulsants and securing a watertight dural closure are standards of care in this patient population regardless of GLIADEL® wafer use. Therefore, no extraordinary measures are necessary when using GLIADEL® wafers.

Efficacy of GLIADEL® Wafer in Recurrent Glioma

Importantly, GLIADEL® wafer has been previously shown to be effective in improving survival in two separate randomized, placebo-controlled, double-blinded studies in patients with malignant glioma. In the clinical setting of reoperation for recurrent GBM (Study 8802) GLIADEL® wafer was demonstrated to be safe and effective in prolonging survival.

This study was a multicenter, double-blind, randomized, placebo-controlled, Phase III clinical trial. Patients underwent intraoperative, surgical implantation of GLIADEL or PLACEBO wafers for treatment of recurrent malignant glioma.

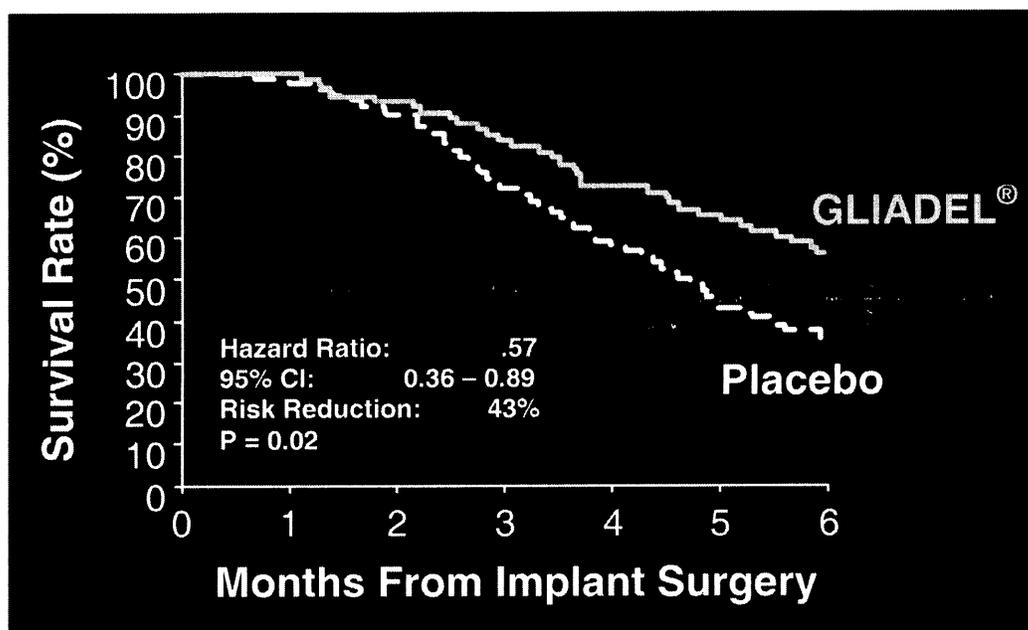
The primary efficacy measures were cumulative mortality and mortality rates through six months after wafer implantation surgery. The secondary efficacy measures were mortality through the end of the post-surgery observation period, Karnofsky Performance Status (KPS) scores and Mini-Mental State Examinations (MMSE).

In 222 patients with recurrent malignant glioma who had failed initial surgery and radiation therapy, the six-month survival rate after surgery increased from 47% (53/112) for patients receiving placebo to 60% (66/110) for patients treated with GLIADEL®. Median survival increased by 33%, from 24 weeks with placebo to 32 weeks with GLIADEL® treatment. In patients with GBM, the six-month survival rate increased from 36% (26/73) with placebo to 56% (40/72) with GLIADEL® treatment. Median survival of GBM patients increased by 41% from 20 weeks with placebo to 28 weeks with GLIADEL® treatment. In patients with pathologic diagnoses other than GBM at the time of surgery for tumor recurrence, GLIADEL® produced no survival prolongation.

Ninety-five percent of the patients treated with GLIADEL® had 7-8 wafers implanted.

In the 8802 study there was a risk reduction of death of 43% over 6 months (95% CIs: 11% - 64%) associated with GLIADEL® wafer use versus placebo wafer. This effect was statistically significant and led to the present indication for GLIADEL® wafer. (Figure 2)

FIGURE 2:
6-Month Survival Primary Endpoint – Recurrent GBM (8802)



The use of GLIADEL was associated with an increased frequency of convulsions in the immediate post-operative period (0-5 days) but did not cause the serious adverse effects frequently seen with systemic BCNU. Thus, the risk/benefit ratio for GLIADEL in this study was large, and favors the use of GLIADEL as palliative therapy in patients undergoing reoperation for recurrent GBM.

Differences in the local tissue environment between recurrent disease and newly diagnosed gliomas as well as the concomitant treatment of primary malignant glioma patients with radiotherapy warranted a separate evaluation of local therapy in primary malignant glioma patients. Therefore, two Phase III efficacy trials (Study 0190 and T-301) have been conducted to evaluate the overall survival benefit and safety of Gliadel® wafers at the time of initial surgery for malignant glioma.

Efficacy in Primary Malignant Glioma: Study 0190

Study 0190 was a Phase III, multicenter (4 centers in Finland and Norway), randomized, double-blind, placebo-controlled trial to determine the safety and efficacy of polifeprosan 20 with carmustine 3.85% (GLIADEL® Wafer) implants plus surgery and external beam radiation therapy, compared to placebo implants plus surgery and limited field radiation therapy, in patients with newly-diagnosed malignant glioma. Patients with initially diagnosed malignant glioma, and without prior surgical, radiotherapeutic, or chemotherapeutic treatment, were eligible for the study. Patients had to have an intra-operative pathological diagnosis of malignant glioma (or high grade glioma, per amended protocol).

The primary efficacy parameters were 12-month and 24-month survival rates, median survival duration and time-to-treatment failure. Secondary efficacy parameters included KPS scores, Mini-Mental State Examination (MMSE) scores, and results of neurological examinations. Safety parameters included collections of adverse events and laboratory testing (hematology and serum chemistry).

Survival was assessed by two methods: survival rate 12 months after wafer implantation surgery and by the Kaplan-Meier method at 12 and 24 months after wafer implantation surgery. In addition, the treatment effect on both 12-month and overall survival (24 months) was estimated using a proportional hazards multiple regression method. Time-to-treatment failure was analyzed using the Kaplan-Meier technique over the entire 24 month follow-up period.

The baseline characteristics of the patients in the 0190 study are shown in Table 3.

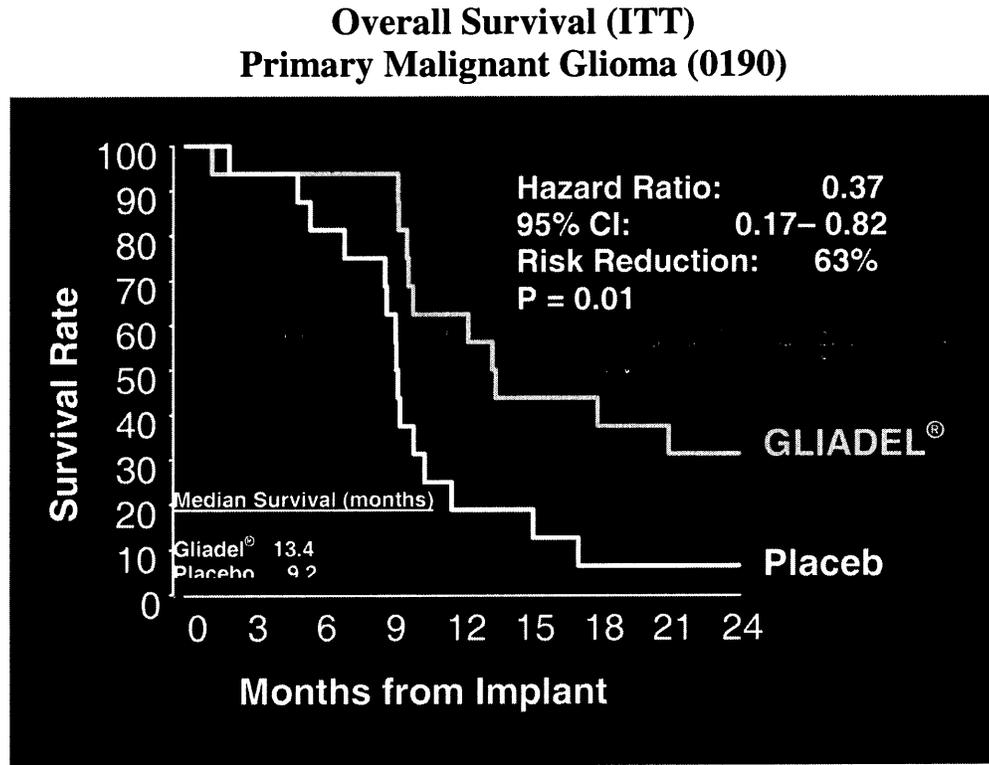
TABLE 3:

Study 0190 Baseline Patient Characteristics		
Characteristics	GLIADEL[®] Wafer (n=16)	Placebo Wafers (n=16)
Median age (years)	56	54
Median Mini-Mental Score	24.5	24.5
Median Karnofsky Performance Score	75	90
Median No. of Wafers Implanted	8	8
GBM Tumor Histology	11	16

There were 16 patients per treatment group and the characteristics of these patients were well balanced as far as age and the median Mini-Mental score. However, the KPS of the placebo group was higher (i.e., the patients had a better performance level) with a mean KPS of 90 vs. 75 for the GLIADEL[®] group. There were more patients with a GBM tumor histology in the placebo group (16) than the GLIADEL[®] group (11).

The effectiveness of GLIADEL[®] Wafer in the treatment of initially diagnosed malignant glioma was demonstrated in the 0190 Study by the statistically significant improvement in one-year survival rate and survival over the 12- and 24-month period after implant surgery in the GLIADEL[®] Wafer treatment group when compared to placebo. The median survival of the Gliadel[®] group is 13.4 months vs. 9.2 months in the placebo group with a risk reduction of 63% (95% CIs: 18 – 83%, p=0.01). At two years, approximately 33% of the Gliadel[®] wafer treated group vs. 6% of the placebo wafer treated group were still alive. (Figure 3)

FIGURE 3:



Ten of 16 GLIADEL[®] Wafer patients (63%) compared to 3 of 16 placebo patients (19%) survived to one year (52 weeks) ($p = 0.029$). Overall, 11 of 16 (69%) GLIADEL[®] Wafer patients and 15 of 16 placebo patients (94%) died during the two year study conduct period.

When the data were adjusted for important prognostic factors (age and MMSE), whether stratified by tumor type or not, a significant GLIADEL[®] wafer treatment effect was observed. For the 12-month period after implantation the adjusted risk reduction for GLIADEL[®] Wafer vs. placebo treatment was 85% for all patients by nonstratified analysis ($p = 0.004$) and 82% for all patients stratified by tumor type ($p = 0.006$). For the 24-month (overall) period after study surgery the adjusted risk reduction for GLIADEL[®] Wafer vs. placebo treatment was 82% for all patients by nonstratified analysis ($p = 0.0005$) and 79% for all patients stratified by tumor type ($p = 0.003$).

Six of 11 GBM patients (55%) in the GLIADEL[®] Wafer treatment group and 3 of 16 GBM patients (19%) in the placebo treatment group survived to one year ($p = 0.1$). In the GLIADEL[®] Wafer group the median post implantation survival duration for GBM patients was 53.3 weeks compared with 39.9 weeks in the placebo treatment group ($p = 0.093$ for overall survival).

Overall, 9 of 11 (82%) patients with GBM in the GLIADEL[®] Wafer group died compared with 15 of 16 (94%) patients with GBM in the placebo group. After accounting for the effect of prognostic factors (age and KPS), GLIADEL[®] Wafer produced a statistically significant reduction in the risk of dying relative to placebo in GBM patients for both the 12- and 24-month periods after wafer implantation surgery. The adjusted risk reductions were 85% (95% CIs: 53% to 95%) for 12 months and 82% (95% CIs: 53% to 93%) for 24 months, with p-values of 0.004 and 0.0005, respectively.

In the 0190 Study, there were no significant differences in Time-to-Treatment failure or change in KPS scores between the two treatment groups.

Of the 11 parameters evaluated in neurological examinations, improvements in mean scores were noted in four parameters for the GLIADEL[®] Wafer treatment group patients. In the GLIADEL[®] Wafer treatment group, the greatest improvements in the mean changes from Baseline to the Final Visit were seen in the following parameters: visual change, fundus (papilledema), cranial nerves III, IV, VI, and cerebellar signs.

In Study 0190, 12 of 16 patients (75%) in the GLIADEL[®] treatment group and 9 of 16 patients (56%) in the placebo treatment group experienced at least one treatment-emergent adverse event during the study period. The most frequently documented treatment-emergent adverse events in the GLIADEL[®] treatment group were hemiplegia (38%) followed by convulsion (19%), aphasia (13%), and visual field defect (13%). In the placebo treatment group, the most frequently reported treatment-emergent adverse events were hemiplegia (25%) and convulsion (13%). In both the GLIADEL[®] treatment group and the placebo treatment group, most events were considered by the investigator to have no relationship to study drug [22 of 31 events (71%) in the GLIADEL[®] treatment group and 10 of 16 events (63%) in the placebo treatment group]. No event was considered to be definitely or probably related to GLIADEL[®] or placebo wafers by the investigator. In the GLIADEL[®] treatment group, 3 of 31 treatment-emergent adverse events (10%) were considered to be possibly related. One of 16 treatment-emergent adverse events (6%) in the placebo treatment group was considered by the investigator to be possibly related to study medication.

The adverse event profile observed in Study 0190 in both treatment groups was typical of patients in the post-operative period following resection for malignant glioma. Systemic toxicities were not noted in evaluation of laboratory parameters.

In conclusion, the 0190 study demonstrated that GLIADEL[®] Wafer treatment prolongs survival in patients with primary malignant glioma both at 12 and 24 months. GLIADEL[®] Wafer was found to be safe and well tolerated. However, study 0190 was a small study of 32 patients, only 16 of whom were treated with GLIADEL[®] Wafer. At the time of the original approval of GLIADEL[®] Wafer for recurrent GBM patients, Study 0190 was deemed by the ODAC Panel to be an “adequate and well controlled” study. However, the small number of patients treated with GLIADEL[®] Wafer precluded any firm conclusions of its safety in this clinical setting. Therefore, an additional study was required to confirm efficacy and establish

safety of GLIADEL® Wafer treatment in a larger group of primary malignant glioma patients at the time of initial surgery. Trial T-301 fulfills these requirements.

STUDY T-301

Study T-301 was designed to confirm the efficacy and safety of GLIADEL® in the treatment of newly diagnosed malignant glioma patients and to determine the potential benefit of GLIADEL® in maintenance of overall function (as measured by the Karnofsky Performance Score), maintenance of neurological function, Time-to-Disease Progression and Quality of Life.

T-301 was a Phase III, multicenter (42 centers in 14 countries), randomized, double-blind, placebo-controlled trial to determine the safety and efficacy of polifeprosan 20 with carmustine 3.85% (GLIADEL® Wafer) implants plus surgery and limited field radiation therapy, compared to placebo implants plus surgery and limited field radiation therapy, for improving survival in 240 patients undergoing initial surgery for newly-diagnosed malignant glioma (Table 4). Male and Female patients between the ages of 18 and 65 years who had radiographic evidence on cranial magnetic resonance imaging (MRI) of a single contrast-enhancing unilateral supratentorial cerebral tumor for whom surgical treatment within two weeks of the baseline MRI scan was indicated were eligible for the study. Patients had to have an intra-operative pathological diagnosis of malignant glioma and have a Karnofsky Performance Score ≥ 60 . Patients who had received prior cytoreductive surgery, prior radiotherapy to the brain or chemotherapy, or who had more than one focus of the tumor or a tumor crossing the midline, or concomitant life-threatening disease, were excluded from the study. Patients were prohibited from receiving additional systemic chemotherapy until documented recurrence of disease progression.

TABLE 4:

Study T-301: Trial Design	
<ul style="list-style-type: none"> • Randomized, double-blind, placebo-controlled study 	
<ul style="list-style-type: none"> • Primary Efficacy Endpoint <ul style="list-style-type: none"> • Overall Survival – All Patients Randomized (ITT) by the Kaplan-Meier method 12 months after final patient was enrolled • Secondary Efficacy Endpoints <ul style="list-style-type: none"> • Overall Survival – GBM Patients • Karnofsky Performance Decline, Neuroperformance Decline, Progression-Free Survival, and Quality of Life Evaluation 	

A total of 42 sites in 14 countries actively recruited patients for the study. (Table 5)

TABLE 5:

Study T-301: Clinical Sites			
Australia	3 sites	Italy	3 sites
Austria	1 site	The Netherlands	2 sites
Belgium	2 sites	New Zealand	1 site
France	7 sites	Spain	3 sites
Germany	5 sites	Switzerland	2 sites
Greece	1 site	United Kingdom	4 sites
Israel	3 sites	United States	5 sites

Safety was assessed by tabulation of treatment emergent adverse events and laboratory testing.

The primary efficacy endpoint was overall survival in the ITT population 12 months after enrollment of the last patient (Table 4). The secondary efficacy endpoints were overall survival in a subgroup of patients with GBM, survival to 12 months, progression-free survival, Quality of Life (QOL), Karnofsky Performance Status (KPS) scores, neurological evaluation, survival censoring patients with reoperation for disease progression, and safety parameters.

T-301 Trial Design and Analytical Methods:

There were several features of the design of trial T-301 that reduced bias in the assessment of Gliadel's® treatment effect.

First, patients were assigned to treatment using a randomized treatment allocation procedure. The randomization schedule was balanced in blocks of four treatments (two GLIADEL and two placebo) at each participating center, ensuring that an equal or very similar number of patients would be treated at each center in each treatment group. Since any one treatment center is located in only one country, the randomization scheduled was also blocked by country. The randomization scheme was stratified by country and center because it was expected that the treatment and subsequent survival of patients with malignant glioma would vary from center to center and from country to country. For example, the effects of the magnitude of tumor resection, end-of-life care practices, and the proclivity of clinicians to reoperate on patients at the time of tumor recurrence all may influence a primary outcome measure such as survival. Stratifying the randomization by center and country was intended to minimize the bias that would ensue from imbalance in the two treatment groups in enrollment in a given center or country. The expectation of a country effect was explicitly pre-specified in the statistical analytical plan for trial T-301 as stated below:

In order to analyze the potential effects of covariates on treatment effect, after checking for proportionality of hazards, the following covariates will be entered in a Cox proportional hazard model:

treatment, Karnofsky Performance Status, age, final histopathological diagnosis (tumor type), country, and eventually additional covariates with an imbalance at baseline.

Second, a placebo control group was employed. The placebo used was polifeprosan 20 wafers without BCNU. Polymer wafers without BCNU do not have any tumoricidal activity against malignant glioma cells (as measured by in vivo activity in the 9L gliosarcoma model) and do not cause any adverse effects in the central nervous system (as measured by toxicology studies in multiple species, including primates). All patients received best conventional therapy as determined by their treating physicians and included protocol specified optimum radiation treatment. Patients could also receive parenteral chemotherapy and re-operation for disease recurrence at the discretion of their treating physician but only at the time of tumor recurrence.

Third, the use of placebo wafers allowed the assessment of the primary and secondary outcome measures in a double-masked (or double-blind) fashion. Thus, bias in the assessment of the outcome or in co-administration of other potentially effective treatments was controlled for.

Finally, the analytic plan for the trial was pre-specified and provided to FDA prior to unmasking the results of the study. This approach minimizes any bias due to post-hoc analysis and controls the type I error of the study.

Pre-specified Analytic Plan:

All of the analyses conducted in trial T-301 were rigorously pre-specified. The analytical plan pre-specified the methods to be used to determine if GLIADEL® is safe and effective. For the primary efficacy analysis, survival was to be estimated by the Kaplan-Meier method and treatment differences assessed by the log-rank test. Several patient characteristics were known to be sources of variability in assessing survival (age, Karnofsky score, tumor histology, and country of treatment) and were pre-specified in the analytical plan as covariates to be used in determining if an apparent GLIADEL® treatment effect was independent of known prognostic factors.

The study was conducted using a stratified blocked randomization by clinical center and by country as is typical in multicenter, multinational studies. This procedure explicitly acknowledges center and country as a source of variation, and requires the use of a stratified log rank test. Countries in which a small number of patients were enrolled were pooled together, as specified in the analytical plan, in order to avoid over stratification, which can reduce the power of a study and increase type II error^(22,23). The use of stratification in this trial was consistent with standard statistical practice^(22,23).

Thus, by design, trial T-301 provided an unbiased, precise estimate of the treatment effect of GLIADEL® wafer in the primary surgery setting.

STUDY T-301: RESULTS

T-301 BASELINE CHARACTERISTICS:

A total of 120 patients were enrolled in each treatment group of the T-301 study (Table 6). The age (mean and range) were very similar for the two treatment groups as was the sex distribution. The tumor types (histology) were also very similar in the two treatment groups with the GBM subtype comprising approximately 85% of both treatment groups.

TABLE 6:

Study T-301: Baseline Characteristics			
Characteristic		GLIADEL [®] Wafer (n=120)	Placebo Wafer (n=120)
Age (years)	Mean	52.6	53.6
	Range	21-72	30-67
Sex	Male (n)	76	84
	Female (n)	44	36
Tumor Type	Anaplastic astrocytoma	1	1
	Anaplastic oligodendroglioma	5	4
	Anaplastic oligoastrocytoma	7	3
	Glioblastoma multiforme	101	106
	Metastasis/Brain Metastasis	2	1
	Other	4	5

There were no statistically significant differences in the Karnofsky scores between the two treatment groups (Table 7). However, there were more patients in the GLIADEL[®] treatment group that had KPS \leq 70 versus the placebo group (37 vs. 33) and fewer patients in the GLIADEL[®] group with KPS \geq 90 (56 vs. 63).

TABLE 7:

Study T-301: Baseline Characteristics		
Karnofsky Performance Status	<u>Karnofsky Score</u>	
	GLIADEL [®] Wafer (n=120)	Placebo Wafer (n=120)
60	16	16
70	21	17
80	25	24
85	2	0
90	31	40
95	0	1
100	25	22

The tumor size at baseline was significantly different in the two treatment groups with the GLIADEL[®] treatment group having larger tumors than the placebo treatment group (Table 8).

TABLE 8:

Study T-301: Tumor Size*		
	GLIADEL [®] Wafer (n=120)	Placebo Wafer (n=120)
Number reported	83	76
Mean (cm ²)	66.8	50.8
Median (cm ²)	60.0	34.0

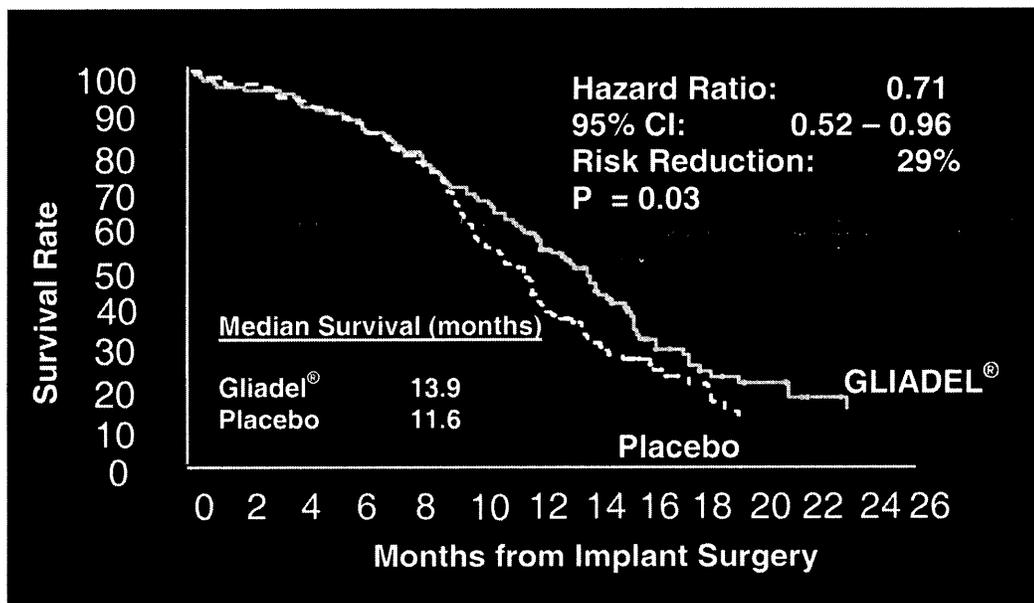
* Difference in tumor size at baseline; p-value <0.05

Efficacy Results from Study T-301

The primary pre-specified efficacy endpoint in Study T-301 was overall survival in the ITT population at the time the final patient had 12 months of follow up. The Kaplan Meier analysis of the survival shows a statistically significant treatment effect with a risk reduction of 29% (95% CIs: 4% - 48%) (Figure 4). The survival benefit is statistically significant (p=0.03, log rank statistic stratified by country).

FIGURE 4:

Study T-301: Overall Survival Analysis (ITT)



Baseline KPS score, age, final histopathological diagnosis and the number of wafers implanted are known to be statistically important predictors of survival in the overall population. Table 9 shows for the overall population the effects of these variables on survival (irrespective of treatment group). From these data it is clear that a number of factors significantly influence survival. (Note that number of wafers implanted instead of tumor size was used as a univariable prognostic factor due to availability of data). Therefore, one must examine the trial results taking these factors into account. Such an analysis was prespecified in the Statistical Analysis Plan.

TABLE 9:

Univariable Prognostic Factors			
Prognostic Factor	Hazard Ratio	95% CI	P-Value
Karnofsky Score ≤ 70 vs. KPS > 70	1.9	1.4 - 2.4	<0.001
Age ≥ 60 vs. <60	1.6	1.2 - 2.2	0.03
Number of wafers implanted <8 vs. 8	1.4	1.0 - 1.9	0.02
GBM Patients vs. Non-GBM Patients	1.8	1.1 - 2.9	0.02

Cox Model stratified by country with single covariates

None of the baseline factors differed significantly between the two treatment groups. (Tables 6 and 7). However, to control for chance imbalances in these factors, adjusted analyses were performed using the Cox Proportional Hazards Model (Table 10).

TABLE 10:

Overall Survival – Adjusted for Prognostic Factors – (ITT)			
	Hazard Ratio	95% CI	P-Value ¹
GLIADEL® Wafer vs. Placebo	0.72	0.53 – 0.98	0.03
KPS ≤70 vs. KPS >70	1.93	1.37 – 2.72	0.0002
Age ≥60 vs. <60	1.73	1.24 – 2.42	0.001

¹ Stratified by country. Tumor histology was not significant in the final model.

After adjusting for age, tumor histology, country of treatment and KPS the treatment effect remains significant with a risk reduction of 28% (95% CIs: 2% to 47%) (p=0.03).

GLIADEL® Wafer administration produces a clinically significant increase in survival (risk reduction = 29%) in malignant glioma patients undergoing primary surgery. This treatment effect is maintained after accounting for the effect of prognostic factors (risk reduction = 28%). Therefore, the conclusion is that in the T-301 study GLIADEL® wafer treatment had a positive effect on survival whether or not one adjusts for impact of baseline prognostic factors.

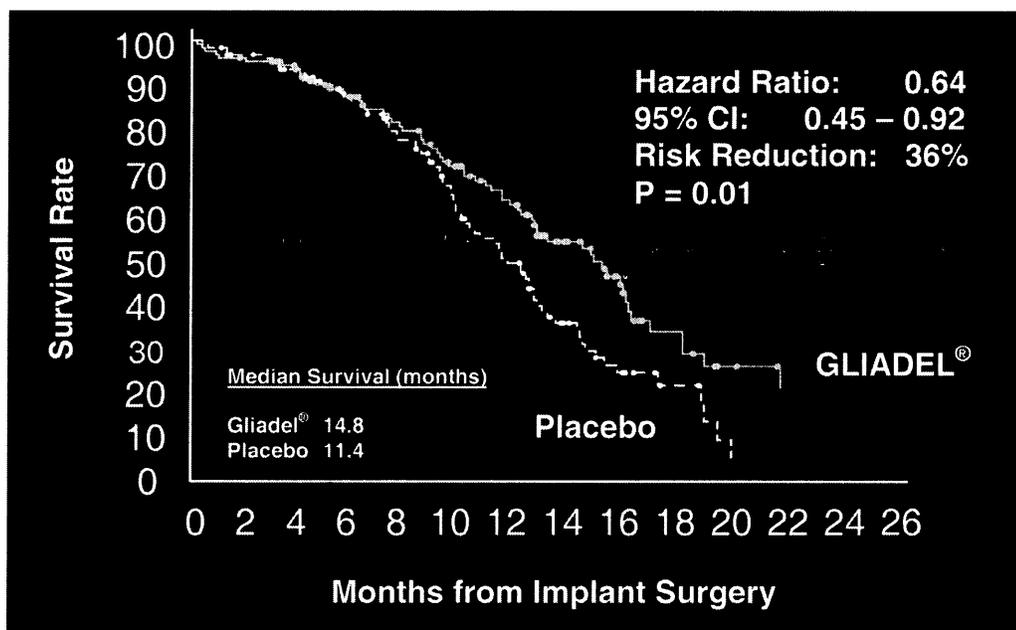
EFFECT OF REOPERATION FOR DISEASE PROGRESSION ON SURVIVAL

The SAP specified that a sensitivity analysis be conducted to adjust for additional therapies administered to patients at the time of relapse or progression of disease as these therapies may have influenced survival. It was noted that a much higher percentage of patients underwent reoperation for disease progression in the T-301 study than originally projected based on the 0190 study. Sixty-six patients had re-operation for disease recurrence or progression in the T-301 study versus only one in the 0190 study.

Physicians re-operate due to disease recurrence and/or to relieve symptoms and to prolong survival. The sensitivity analysis was performed to adjust the results in the survival endpoint by censoring patients alive at the time of re-operation. Such an analysis would adjust for the confounding factor of re-operation and provide a more precise measurement of the GLIADEL® Wafer treatment effect.

FIGURE 5:

**Study T-301: Overall Survival Analysis -
Reoperation for Disease Progression (ITT)**



The Kaplan Meier survival analysis in the intent-to-treat population censoring patients alive at the time of re-operation, shows a statistically significant survival benefit ($p=0.01$) with GLIADEL[®] wafer treatment (Figure 5). This treatment effect represents a risk reduction of 36% (95% CIs: 8% to 55%, $p = 0.01$). This analysis most closely approximates the conditions of the 0190 study where only one patient underwent reoperation for disease recurrence/progression. This analysis most accurately demonstrates the treatment effect that is conferred by GLIADEL[®] Wafer treatment alone compared to placebo wafer.

STUDY T-301: SECONDARY EFFICACY ENDPOINTS

The prespecified endpoints in the study included overall survival in the GBM population of patients. In the ITT population of patients the Time-to-KPS decline, Neuroperformance Score decline, Progression-Free survival, and a Quality of Life Evaluation were prespecified as secondary endpoints.

GLIADEL[®] wafer treatment produced an increase in survival versus placebo wafer treatment in the GBM subpopulation of patients. The magnitude of the treatment effect was similar to that seen in the ITT population and represented a risk reduction of 24% (95% CIs: 5% to 45%). The p-value for this effect was 0.10 (Figure 6). The treatment effect was statistically significant when adjusted for the prognostic factors of age, KPS and country (with age falling

out of the final model) with a p value of 0.04 (Table 11). The risk reduction was 31% (95% CIs: 3% to 51%).

FIGURE 6:

Study T-301: Overall Survival (GBM Patients)

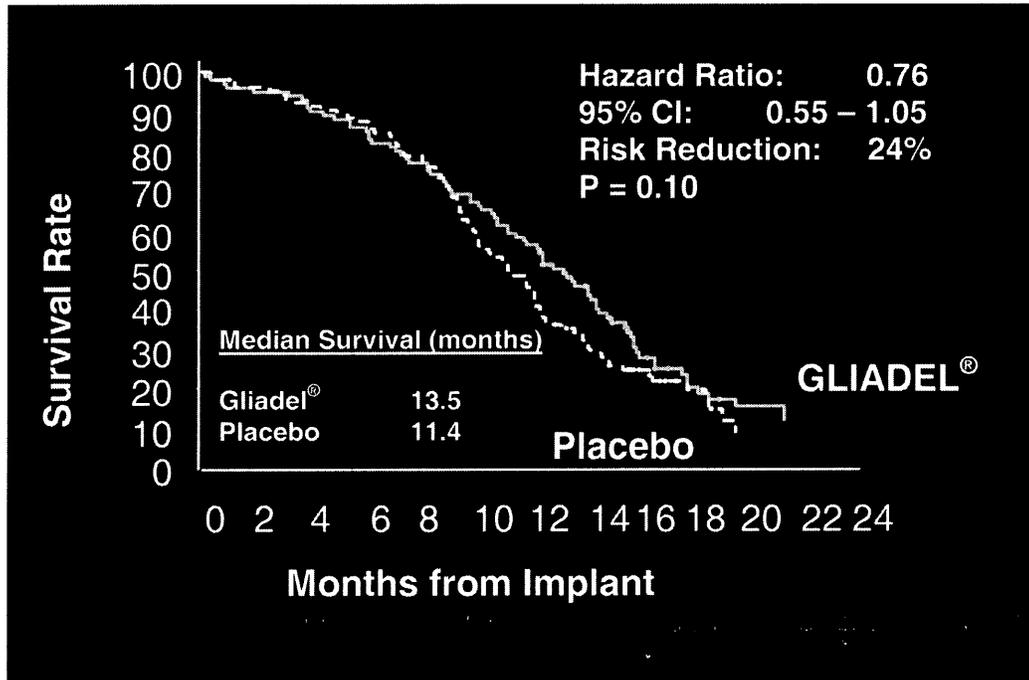


TABLE 11:

**Study T-301: Overall Survival Adjusted for Prognostic Factors¹
(GBM Patients)**

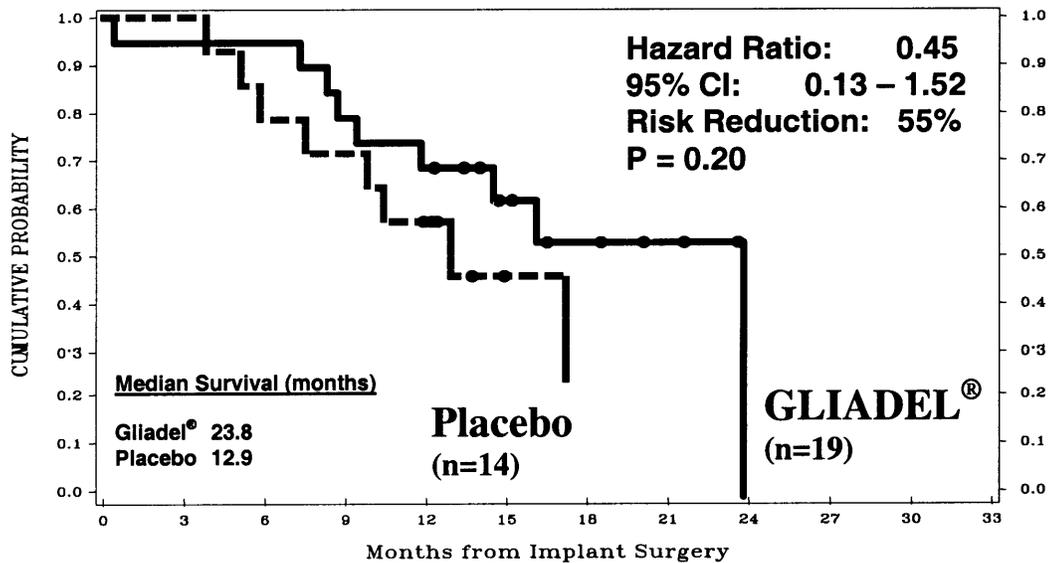
	Hazard Ratio	95% CI	P-Value ²
GLIADEL® vs. Placebo	0.69	0.49 - 0.97	0.04
KPS ≤70 vs. KPS >70	2.04	1.38 - 3.01	<0.001

¹ Adjusted for age and Karnofsky Score
² Cox Proportional Hazard model stratified by country and number of wafers (8,8) implanted

The overall survival of the non-GBM patients with malignant glioma is shown in Figure 7. The GLIADEL® wafer-treated patients have improved survival compared to the placebo control wafer treated patients (Figure 7). These data suggest that the treatment effect of GLIADEL® wafer treatment is present in both the GBM and non-GBM patient populations.

Figure 7 :

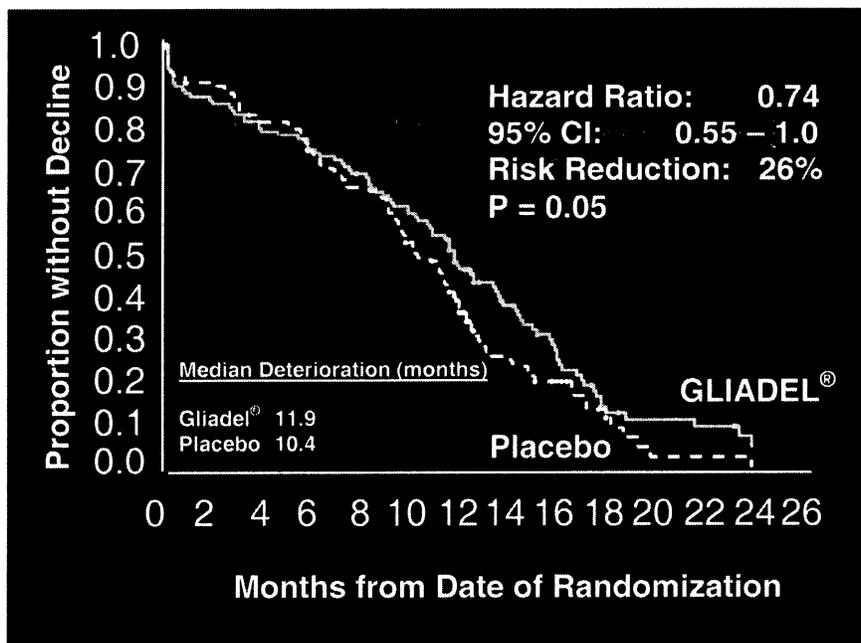
Study T-301: Overall Survival (Non-GBM Patients)



The Kaplan Meier analysis of the Time-To-KPS decline also was significantly better for the GLIADEL® wafer treated group vs. the placebo wafer treated group (p=0.05 with a risk reduction of 26%; 95% CIs: 0% to 45%) (Figure 8). This finding suggests that the survival benefits of GLIADEL® wafer treatment are accompanied by maintenance of overall function.

FIGURE 8:

Study T-301: Karnofsky Performance Decline (ITT)



Maintenance of neurological function was measured in the T-301 study by 11 prespecified neurological assessments. If patients demonstrated a clinically significant progression they were determined to have ‘progressed’, therefore, these neurological measures assess how long patients can maintain a higher level of neurological function. The neuroperformance measures did not differ between the two treatment groups at baseline. GLIADEL® wafer treated patients had a significantly longer time to neurological progression in 10 of 11 measures of neurologic function (Table 12). In the one measure where there was not a statistically significant treatment effect (Visual Status), a treatment trend ($p=0.09$) in favor of GLIADEL® wafer treatment was demonstrated.

TABLE 12:

Study T-301: Neuroperformance Decline (ITT)			
Neuroperformance Measure	Median Time Without Deterioration (weeks)		P-Value
	GLIADEL® Wafer (n=120)	Placebo Wafer (n=120)	
Vital Signs	54.9	49.1	0.01
Level of Consciousness	52.1	45.4	0.02
Personality	51.7	40.0	0.008
Speech	49.6	36.7	0.003
Visual Status	44.0	42.4	0.09
Fundus	55.1	46.3	0.007
Cranial Nerves III, IV, VI	54.9	49.1	0.02
Cranial Nerves, Other	54.3	46.3	0.003
Motor Status	45.4	31.4	0.01
Sensory Status	51.6	44.1	0.02
Cerebellar Status	54.1	46.7	0.01

A number of specific examples of neuroperformance maintenance are illustrated below and include Speech Function, Cranial Nerve Function, Motor Function, and Cerebellar Function (Figures 9, 10, 11, and 12).

FIGURE 9:

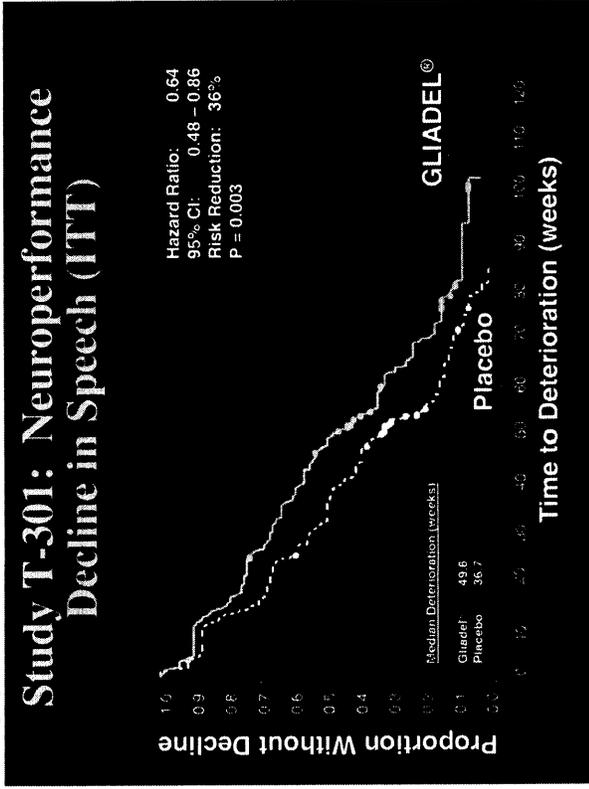


FIGURE 11:

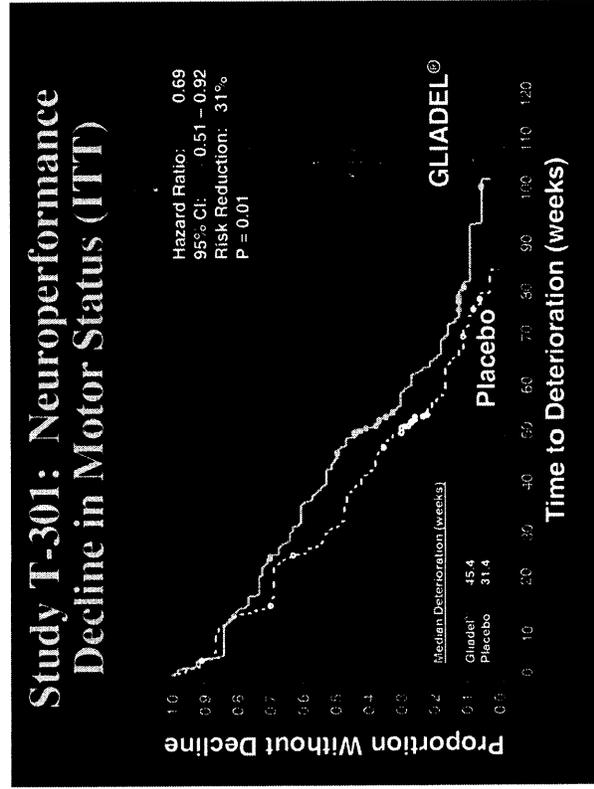


FIGURE 10:

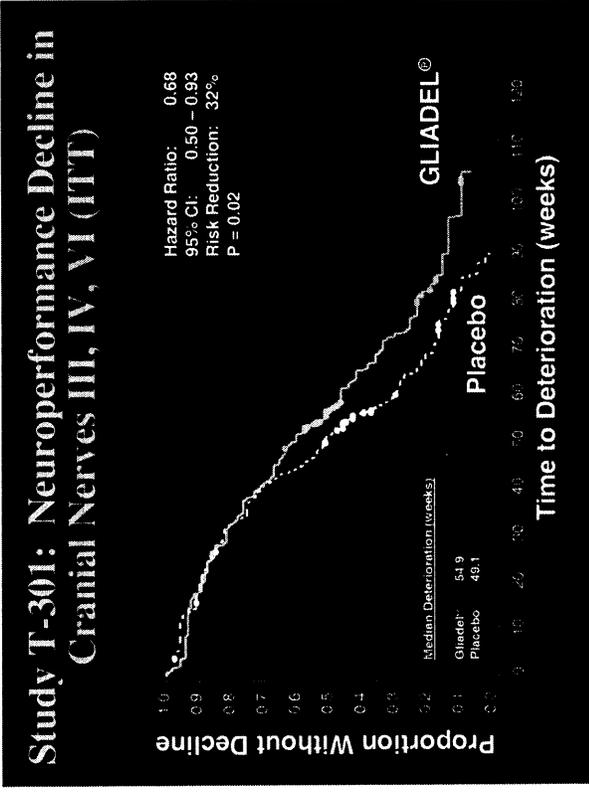
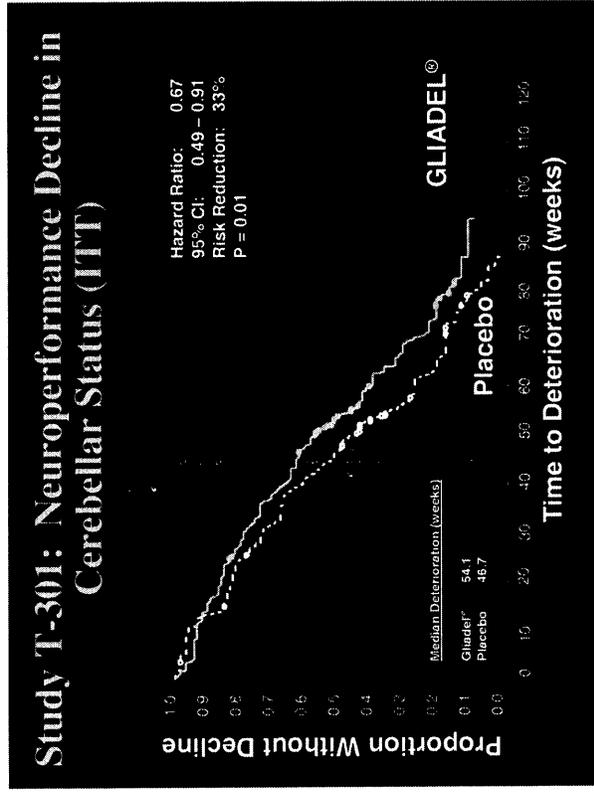


FIGURE 12:



SAFETY

No new or major safety issues concerning treatment with GLIADEL® Wafer implants were raised in Study T-301. Safety results were comparable between the treatment groups and generally consistent with those expected in patients undergoing major surgery for resection of malignant glioma, and those seen in previous clinical studies and described in the package insert. Early post operative convulsions, infections, and healing abnormalities reported with GLIADEL® wafer use in the recurrent GBM clinical setting did not occur in the treatment of patients with primary malignant glioma with GLIADEL®.

Three general conclusions can be made about the safety of GLIADEL® use in primary surgery for malignant glioma:

1. Careful monitoring of GLIADEL®-treated patients for cerebral edema and/or intracranial hypertension with aggressive consequent steroid use is warranted.
2. CSF leak, though uncommon, may be more frequent in GLIADEL®-treated patients. Attention to a water tight dural closure and local wound care are indicated.
3. The safety profile of GLIADEL® appears to be more benign in the primary surgical setting than in the recurrent surgery setting in patients with malignant glioma.

A summary of the safety findings of Study T-301 is indicated in the table below (Table 13). This summary indicates that CSF leak was the only adverse event convincingly associated with GLIADEL® wafer use (6% in the GLIADEL® group versus 0.8% in the placebo group). Intracranial hypertension was observed more frequently in the GLIADEL® wafer treated group (9.2% vs 1.7%). However, in 9 of 11 patients treated with GLIADEL® in whom intracranial hypertension was diagnosed this adverse event occurred late (200 days or greater) after surgical implantation and was, therefore, not likely associated with GLIADEL® therapy. In addition, even though there was an increased frequency of intracranial hypertension associated with GLIADEL® there was not an increase in cerebral edema associated with GLIADEL® use.

TABLE 13:

Safety Summary GLIADEL[®] Wafer in Primary Surgery

- Intracranial hypertension was present in more GLIADEL[®] wafer treated patients vs. Placebo patients (9.2% vs. 1.7%). However, there is no difference in brain edema.
- Intracranial hypertension was typically observed late, at the time of tumor recurrence, and was not likely associated with GLIADEL[®] use.
- Careful monitoring of GLIADEL[®]-treated patients for cerebral edema/intracranial hypertension with consequent steroid use is warranted.
- CSF leak (5% vs. 0.8%) was more common in GLIADEL[®] -treated patients. However, intracranial infections and other healing abnormalities were not increased. Attention to a water tight dural closure and local wound care is indicated.
- Convulsions are not more common in GLIADEL[®]-treated vs. placebo-treated patients.
- The safety profile of GLIADEL[®] appears more benign in the primary surgery setting vs. recurrent disease.

CSF leak was more frequently reported in the GLIADEL[®]-treated group vs. the placebo-treated group (5% vs. 0.8%). However, CNS infection, a possible sequelae of CSF leak, was not more frequently reported in the GLIADEL[®] -treated patients. Convulsions, intracranial infections, and other healing abnormalities were not more common in the GLIADEL[®]-treated patients vs. the placebo-treated patients.

Table 14 is a detailed listing of the neurological adverse events occurring in 5% or more in the GLIADEL[®] or placebo groups. There are no differences between the two treatment groups with the exception of intracranial hypertension as previously noted.

TABLE 14:

Study T-301: Neurologic Adverse Events Occurring in \geq 5% of Patients		
Adverse Event	GLIADEL® Wafer (n=120)	Placebo Wafer (n=120)
Abnormal gait	6 (5.0)	6 (5.0)
Amnesia	11 (9.2)	12 (10.0)
Anxiety	8 (6.7)	5 (4.2)
Aphasia	21 (17.5)	22 (18.3)
Ataxia	7 (5.8)	5 (4.2)
Brain edema	27 (22.5)	23 (19.2)
Coma	5 (4.2)	6 (5.0)
Confusion	28 (23.3)	25 (20.8)
Convulsion	40 (33.3)	45 (37.5)
Depression	19 (15.8)	12 (10.0)
Dizziness	6 (5.0)	11 (9.2)
Facial paralysis	8 (6.7)	5 (4.2)
Grand mal convulsion	6 (5.0)	5 (4.2)
Hallucinations	6 (5.0)	4 (3.3)
Hemiplegia	49 (40.8)	53 (44.2)
Hypesthesia	7 (5.8)	6 (5.0)
Hypokinesia	2 (1.7)	8 (6.7)
Incoordination	3 (2.5)	8 (6.7)
Insomnia	6 (5.0)	7 (5.8)
Intracranial hypertension	11 (9.2)	2 (1.7)
Neuropathy	8 (6.7)	12 (10.0)
Paresthesia	7 (5.8)	10 (8.3)
Personality disorder	10 (8.3)	9 (7.5)
Somnolence	13 (10.8)	18 (15.0)
Speech disorder	13 (10.8)	10 (8.3)
Thinking abnormal	7 (5.8)	10 (8.3)
Tremor	6 (5.0)	8 (6.7)

There were no differences in systemic adverse events or laboratory abnormalities between the two treatment groups.

‘Aggravation Reaction’ was the most common, general or non-neurological adverse event reported in both treatment groups. In the GLIADEL[®] wafer treated group it was reported in 98 (81.7%) patients and in the placebo wafer treated group it was reported in 95 (79.2%) patients. Adverse events coded to the COSTART Term ‘Aggravation Reaction’ were mainly tumor progression/disease progression or general deterioration in the patient’s condition associated with progression of disease. ‘Aggravation Reactions’ were only considered to be treatment-related for two patients (1.7%) in the GLIADEL[®] group and three patients (2.5%) in the placebo group.

The frequency of convulsions including ‘severe’ convulsions, Grand Mal convulsions, or the frequency of convulsions occurring within the first 5 days after surgery did not differ between the two treatment groups (TABLE 15). An additional analysis which determined the time-to-first convulsion was performed and similarly showed no difference between the two treatment groups.

TABLE 15:

Study T-301: Convulsions		
	GLIADEL [®] Wafer (n=120)	Placebo Wafer (n=120)
Number of patients (5)	40 (33.3)	45 (37.5)
Convulsions – severe	14 (11.7)	24 (20)
Grand Mal convulsions	6 (5.0)	5 (4.2)
Convulsions (≤ 5 days)	3 (2.5)	5 (4.2)
Time-to-First Seizure did not differ in the two treatments groups		

Specific wound healing data were collected in the T-301 study based on the safety profile of GLIADEL[®] in the recurrent GBM treatment setting (Study 8802). In the 8802 study patients with recurrent GBM treated with GLIADEL[®] had an increased frequency in CNS infections and wound healing abnormalities. In the T-301 study in primary malignant glioma patients data on ‘CSF Leak’, ‘CSF or Subdural Collections’, ‘Wound Dehiscence, Breakdown, or Poor Healing’ and ‘Subgaleal or Wound Effusion’ were collected. As shown in Table 16 the only difference between the two treatment groups is in the frequency of CSF Leak (as noted above).

TABLE 16:

Study T-301: Healing Abnormalities:		
	GLIADEL® Wafer n (%)	Placebo Wafer n (%)
Fluid, CSF or Subdural Collections		
Number of Patients (%)	5 (4.2)	6 (5.0)
Median Duration (days)	15	10
Range for Duration (days)	12-60	1-68
CSF Leak		
Number of Patients (%)	6 (5.0)	1 (0.8)
Median Duration (days)	9	3
Range for Duration (days)	2-211	3
Wound Dehiscence, Breakdown or Poor Healing		
Number of Patients (%)	6 (5.0)	6 (5.0)
Median Duration (days)	10	6
Range for Duration (days)	2-281	2-172
Subgaleal or Wound Effusion		
Number of Patients (%)	4 (3.3)	5 (4.2)
Median Duration (days)	3	10
Range for Duration (days)	3-30	2-26

The frequency of serious CNS infections including meningitis and abscess were determined in the T-301 study. As indicated in Table 17 the frequency of these adverse events did not appear to differ in the two treatment groups.

TABLE 17:

Serious Intracranial Infections (T-301)		
	GLIADEL® Wafer n (%)	Placebo n (%)
Abscess	4 (3)	5 (4)
Meningitis	2 (2)	2 (2)
Total	6 (5)	7 (6)

To summarize the safety profile of GLIADEL® wafer use in primary malignant glioma patients:

1. There was no evidence of earlier onset, frequency, or severity of seizures in primary malignant glioma patients treated with GLIADEL® wafers.
2. CSF Leak was more common in GLIADEL® wafer treated patients.
3. There was no evidence of increased frequency of intracranial infections or other healing abnormalities in GLIADEL® Wafer treated patients.

SUMMARY OF GLIADEL® BENEFITS AND RISKS IN PRIMARY MALIGNANT GLIOMA

To summarize the benefits of GLIADEL® wafer treatment: The use of GLIADEL® Wafer in patients with newly diagnosed malignant glioma demonstrates an increase in survival in patients compared to placebo wafers. This effect is statistically significant and clinically meaningful as demonstrated by the results of two separate clinical studies: the 0190 study and the T-301 study. Importantly, this survival increase is accompanied by a maintenance of function in patients. There is a delayed time to overall functional decline as measured by the Karnofsky Performance Score (p=0.05). The increase in survival is also accompanied by maintenance of good neurological function. In 10/11 neuroperformance measures GLIADEL® wafer treatment was superior to placebo wafer treatment in delaying decline (p<0.05).

The consistency of the Phase III GLIADEL® wafer trial results (both Study 0190 and T-301 are shown in Table 18. The results of the 0190 and T-301 trials in primary malignant glioma and the 8802 trial in recurrent malignant glioma demonstrate the overall consistent efficacy of GLIADEL® Wafer treatment in this patient population. The risk reduction and confidence intervals for both the 0190 and T-301 studies are shown in this Table along with the same data from the 8802 study. These data show that in three separate Phase III randomized, placebo-controlled, double-blinded studies that GLIADEL® wafers have significant activity and have produced a significant clinical benefit in prolonging survival. In the ITT population in three separate trials GLIADEL® has demonstrated a treatment benefit with a risk reduction of death of 31%, 29%, and 63% for studies 8802, T-301, and 0190 respectively.

TABLE 18:

Summary of Efficacy Results from Randomized Controlled Trials of GLIADEL® Wafer (ITT)			
Study	Hazard Ratio	95% CI	P-Value
8802	0.69	0.47 – 1.02	0.06
T-301	0.71	0.52 – 0.96	0.03
0190	0.37	0.17 – 0.82	0.01

The same analyses are shown for the GBM group of patients in Table 19. All studies have demonstrated a benefit. Therefore, GLIADEL® wafers have been shown to have significant efficacy in three randomized, placebo controlled double-blind studies in patients with malignant glioma.

TABLE 19:

Summary of Efficacy Results from Randomized Controlled Trials of GLIADEL® Wafer (GBM)			
Study	Hazard Ratio¹	95% CI	P-Value
8802	0.57	.036 – 0.89	0.02
T-301	0.69	0.49 – 0.97	0.04
0190	0.21	0.08 – 0.60	<0.01

¹ Hazard Ratio adjusted for prognostic factors

Therefore, the conclusion can be made that:

The benefit to risk ratio for GLIADEL® Wafers in patients with primary malignant glioma is favorable.

We therefore feel that the data support the following new indication for GLIADEL® Wafer:

‘GLIADEL® Wafer is indicated for use as a treatment to significantly prolong survival and maintain overall function (as measured by preservation of Karnofsky Performance Status) and neurological function in patients with malignant glioma undergoing primary and/or recurrent surgical resection.’

References

1. Hochberg FH, et.al. Assumptions in the radiotherapy of glioblastoma. *Neurology* 1980; 30: 907-11
2. McDonald JD, Rosenblum ML: In: Rengachary SS, Wilkin RH, eds. *Principles of Neurosurgery*. St Louis, MO: Mosby-Wolfe; 1994: chap 26.
3. The Medical Research Council Brain Tumour Working Party. Randomized Trial Of Procarbazine, Lomustine, And Vincristine In The Adjuvant Treatment Of High-Grade Astrocytoma: A Medical Research Council Trial. *J Clin Oncol* 2001 Jan 15;19(2):509-18
4. Fine HA, Dear KBG, Loeffler JS, Black PMcL, Canellos, GP. Meta-analysis of radiation therapy with and without adjuvant chemotherapy for malignant gliomas in adults. *Cancer* 1993; 71: 2585-2597.
5. DeVita VT, Denham C, Davidson JD, Oliverio VT. The physiological disposition of the carcinostatic 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU) in man and animals. *Clinical Pharmacology and Therapeutics* 1967; 8:566-577.
6. Levin VA, Hoffman W, Weinkam RJ. Pharmacokinetics of BCNU in man: a preliminary study of 20 patients. *Cancer Treatment Reports* 1978; 62: 1305-13 12.
7. Hochberg FH, Pruitt AA, Beck DO, DeBrun G, Davis K. The rationale and methodology for intra-arterial chemotherapy with BCNU as treatment for glioblastoma. *Journal of Neurosurgery* 1985; 63:876-880.
8. Deutsch M, Green SB, Strike TA, Burger PC, Robertson JT, Selker RG, Shapiro WR, Mealey J, Ransohoff J, Paoletti P, Smith KR, Odom GL, Hunt WE, Young B, Alexander E, Walker MD, Pistenmaa DA. Results of a randomized trial comparing BCNU plus radiotherapy, streptozotocin plus radiotherapy, BCNU plus hyperfractionated radiotherapy, and BCNU following misonidazole plus radiotherapy in the postoperative treatment of malignant glioma. *International Journal of Radiation Oncology. Biology. Physics* 1989; 16: 13 89-1396.
9. Shapiro WR, Green SB, Burger PC, Mahaley MS, Selker RG, VanGilder JC, Robertson JT, Ransohoff J, Mealey J, Strike TA, Pistenmaa DA. Randomized trial of three chemotherapy regimens and two radiotherapy regimens in postoperative treatment of malignant glioma. *Journal of Neurosurgery* 1989; 71: 1-9.
10. Hochberg FH, Pruitt A. Assumptions in the radiotherapy of glioblastoma. *Neurology* 1980; 30: 907-911.
11. Fine HA, Dear KBG, Loeffler JS, Black PMcL, Canellos, GP. Meta-analysis of radiation therapy with and without adjuvant chemotherapy for malignant gliomas in

- adults. *Cancer* 1993; 71: 2585-2597.
12. Garfield J, Dayan AD, Weller RO. Postoperative intracavitary chemotherapy of malignant supratentorial astrocytomas using BCNU. *Clinical Oncology* 1975; 1:213-222.
 13. Morantz RA, Kimler BF, Vats TS, Henderson SD. Bleomycin and brain tumors. *Journal of Neuro-Oncology* 1983; 1:249-255.
 14. Nierenberg D, Harbaugh R, Maurer H, Reeder T, Scott G, Fratkin J, Newman B. Continuous intratumoral infusion of methotrexate for recurrent glioblastoma: a pilot study. *Neurosurgery* 1991; 28:752-761.
 15. Inamura T, Tomojiro N, Bartus RT, Black KL. Intracarotid infusion of RMP-7, a bradykinin analog: a method of selective drug delivery to brain tumors. *Journal of Neurosurgery* 1994;81:752-758.
 16. Brem H, Walter KA, Langer R. Polymers as controlled drug delivery devices for the treatment of malignant brain tumors. *European J Pharm Biopharm* 1993; 39:2-7.
 17. Grossman SA Reinhard C, Colvin OM, Chasm M, Brundrett R, Tamargo RJ, Brem H. The intracerebral distribution of BCNU delivered by surgically implanted biodegradable polymers. *Journal of Neurosurgery* 1992; 76:640-647.
 18. Yang MB, Tamargo RJ, Brein H. Controlled delivery of 1,3-Bis(2-chloroethyl)-1-nitrosourea from ethylene-vinyl acetate copolymer. *Cancer Research* 1989; 49:5103-5107.
 19. Tamargo RJ, Myseros JS, Epstein JI, Yang MB, Chasm M, Brem H. Interstitial chemotherapy of the 9L gliosarcoma: controlled release polymers for drug delivery in the brain. *Cancer Research* 1993; 53:329-333.
 20. Grossman SA, Reinhard C, Colvin OM, Chasin M, Brundrett R, Tamargo RJ, Brem H. The Intracerebral Distribution of BCNU Delivered by Surgically Implanted Biodegradable Polymers. *Journal of Neurosurgery* 1992; 76:640-647.
 21. Brem H, Tamargo RJ, Olivi A, Pinn M, Wharam M, Epstein JI. Biodegradable polymers for controlled delivery of chemotherapy with and without radiation therapy in the monkey brain. *Journal of Neurosurgery* 1994; 80:283-290.
 22. Fleiss, JL (1986). Analysis of data from multiclinic trials. *Controlled Clinical Trials* 7:267-275.
 23. Localio, AR et al (2001). Adjustments for center in multicenter studies: An Overview. *Annals of Internal Medicine* 135:112-123.

Guilford Pharmaceuticals Inc.
6611 Tributary Street
Baltimore, Maryland 21224

GLIADEL[®] WAFER

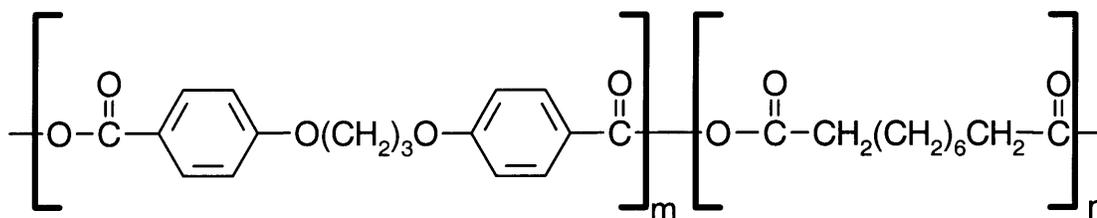
(polifeprosan 20 with carmustine implant)

Rx only

DESCRIPTION

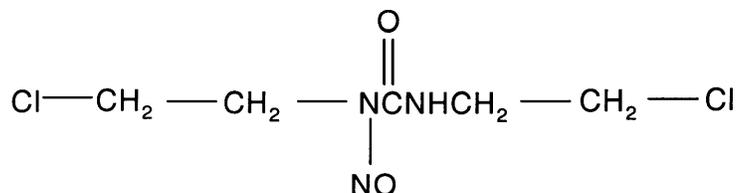
GLIADEL[®] Wafer (polifeprosan 20 with carmustine implant) is a sterile, off-white to pale yellow wafer approximately 1.45 cm in diameter and 1 mm thick. Each wafer contains 192.3 mg of a biodegradable polyanhydride copolymer and 7.7 mg of carmustine [1,3-bis(2-chloroethyl)-1-nitrosourea, or BCNU]. Carmustine is a nitrosourea oncolytic agent. The copolymer, polifeprosan 20, consists of poly[bis(p-carboxyphenoxy) propane: sebacic acid] in a 20:80 molar ratio and is used to control the local delivery of carmustine. Carmustine is homogeneously distributed in the copolymer matrix.

The structural formula for polifeprosan 20 is:



Ratio m:n = 20:80; random copolymer

The structural formula for carmustine is:



CLINICAL PHARMACOLOGY

GLIADEL is designed to deliver carmustine directly into the surgical cavity created when a brain tumor is resected. On exposure to the aqueous environment of the resection cavity, the anhydride bonds in the copolymer are hydrolyzed, releasing carmustine, carboxyphenoxypropane, and sebacic acid. The carmustine released from GLIADEL diffuses into the surrounding brain tissue and produces an antineoplastic effect by alkylating DNA and RNA.

Carmustine has been shown to degrade both spontaneously and metabolically. The production of an alkylating moiety, hypothesized to be chloroethyl carbonium ion, leads to the formation of DNA cross-links.

The tumoricidal activity of GLIADEL is dependent on release of carmustine to the tumor cavity in concentrations sufficient for effective cytotoxicity.

More than 70% of the copolymer degrades by three weeks. The metabolic disposition and excretion of the monomers differ. Carboxyphenoxypropane is eliminated by the kidney and sebacic acid, an endogenous fatty acid, is metabolized by the liver and expired as CO₂ in animals.

The absorption, distribution, metabolism, and excretion of the copolymer in humans is unknown. Carmustine concentrations delivered by GLIADEL in human brain tissue have not been determined. Plasma levels of carmustine after GLIADEL wafer implant were not determined. In rabbits implanted with wafers containing 3.85% carmustine, no detectible levels of carmustine were found in the plasma or cerebrospinal fluid.

Following an intravenous infusion of carmustine at doses ranging from 30 to 170 mg/m², the average terminal half-life, clearance, and steady-state volume of distribution were 22 minutes, 56 mL/min/kg, and 3.25 L/kg, respectively. Approximately 60% of the intravenous 200 mg/m² dose of ¹⁴C-carmustine was excreted in the urine over 96 hours and 6% was expired as CO₂.

GLIADEL wafers are biodegradable in human brain when implanted into the cavity after tumor resection. The rate of biodegradation is variable from patient to patient. During the biodegradation process, a wafer remnant may be observed on brain imaging scans or at re-operation even though extensive degradation of all components has occurred. Data obtained from review of CT scans obtained 49 days after implantation of GLIADEL demonstrated that images consistent with wafers were visible to varying degrees in the scans of 11 of 18 patients. Data obtained at re-operation and autopsies have demonstrated wafer remnants up to 232 days after GLIADEL implantation.

Wafer remnants removed at re-operation from two patients with recurrent malignant glioma, one at 64 days and the second at 92 days after implantation, were analyzed for content. The following table presents the results of analyses completed on these remnants.

COMPOSITION OF WAFER REMNANTS REMOVED FROM
TWO PATIENTS ON RE-OPERATION

<u>Component</u>	<u>Patient A</u>	<u>Patient B</u>
Days After GLIADEL Implantation	64	92
Anhydride Bonds	None detected	None detected
Water Content (% of wafer remnant weight)	95-97%	74-86%
Carmustine Content (% of initial)	<0.0004%	0.034%
Carboxyphenoxypropane Content (% of initial)	9%	14%
Sebacic Acid Content (% of initial)	4%	3%

The wafer remnants consisted mostly of water and monomeric components with minimal detectable carmustine present.

CLINICAL STUDIES

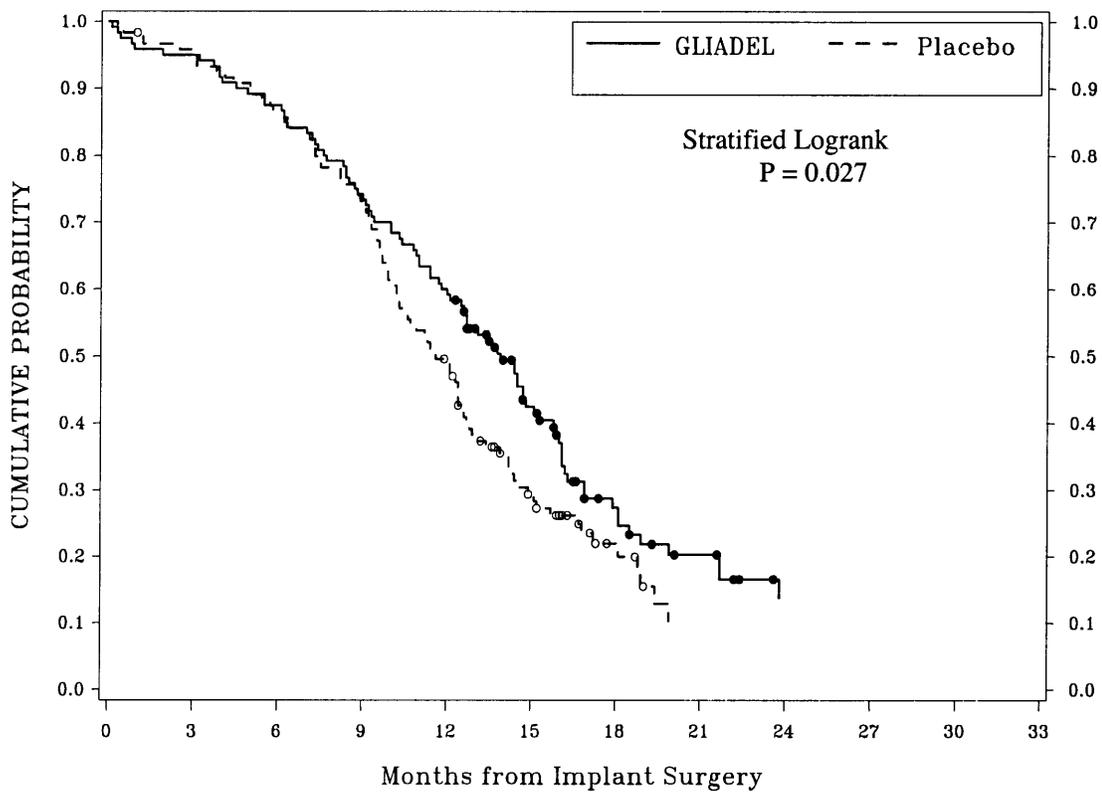
Primary Surgery

In two randomized, double-blind, placebo-controlled clinical trials, GLIADEL prolonged survival in adults with newly diagnosed malignant glioma.

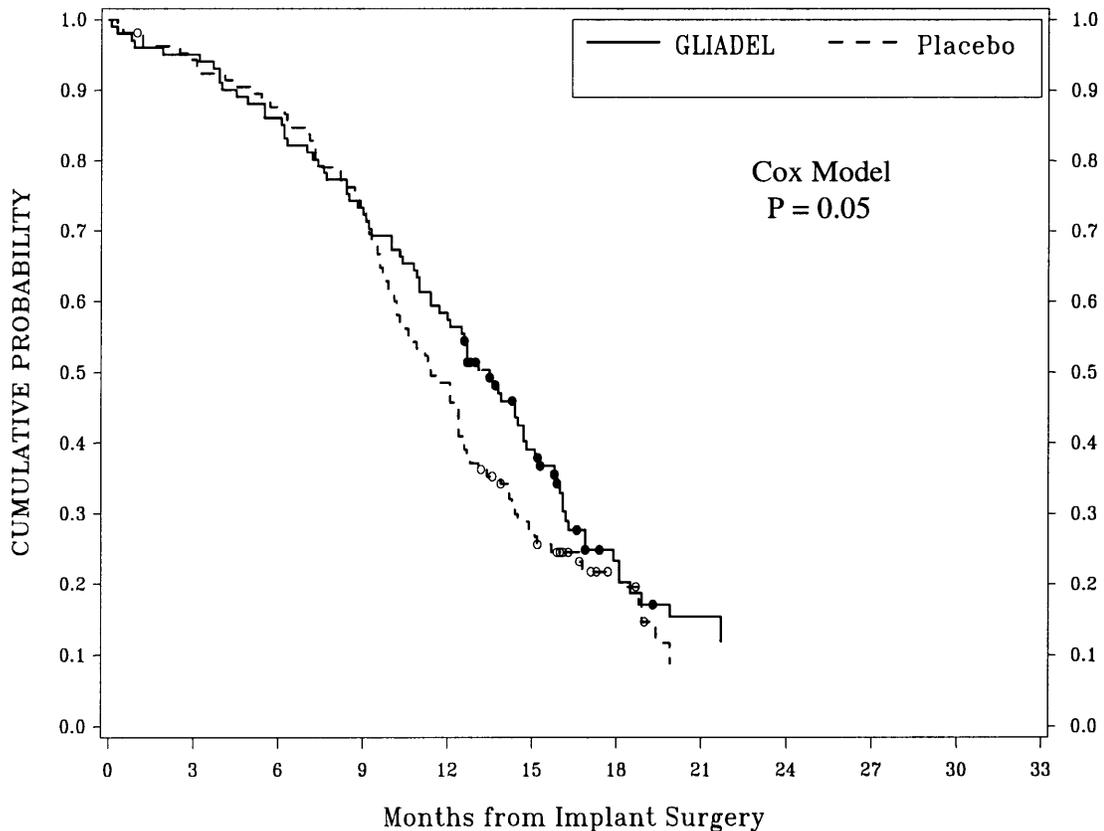
In a large 240 patient study with patients undergoing initial surgery for malignant glioma, the median survival after surgery increased from 11.6 months for patients receiving placebo to 13.9 months for patients treated with GLIADEL ($p = 0.027$). In patients with glioblastoma multiforme, the median survival was increased from 11.4 months with placebo to 13.5 months with GLIADEL treatment ($p = 0.05$ after adjustment for prognostic factors). In addition, one year survival was 59.2% in the GLIADEL group and 49.6% in the placebo group.

Fifty- six percent of the patients treated with GLIADEL had 7-8 wafers implanted.

OVERALL KAPLAN-MEIER SURVIVAL CURVE FOR PATIENTS UNDERGOING SURGERY FOR NEWLY DIAGNOSED MALIGNANT GLIOMA (ITT POPULATION)



OVERALL KAPLAN-MEIER SURVIVAL CURVE FOR PATIENTS UNDERGOING SURGERY FOR NEWLY DIAGNOSED GLIOBLASTOMA MULTIFORME MALIGNANT GLIOMA (GBM)



Approximately 27% of patients in this study underwent reoperation for disease progression. Since the primary endpoint was survival, reoperation for tumor progression may have confounded this endpoint. Censoring patients at the date of reoperation for tumor progression, the median survival was 14.8 months in the GLIADEL group compared to 11.4 months in the placebo group ($p = 0.014$ stratified logrank test).

The median time to deterioration of overall function (as measured by Karnofsky Performance Status score determination) was longer in the GLIADEL[®] group compared to the placebo group. The median time to deterioration was 11.9 months (95% CI: 10.4 to 13.7 months) in the GLIADEL[®] group and 10.4 months (95% CI: 9.5 to 11.9 months) in the placebo group. This difference was statistically significant ($p=0.050$, stratified logrank test)

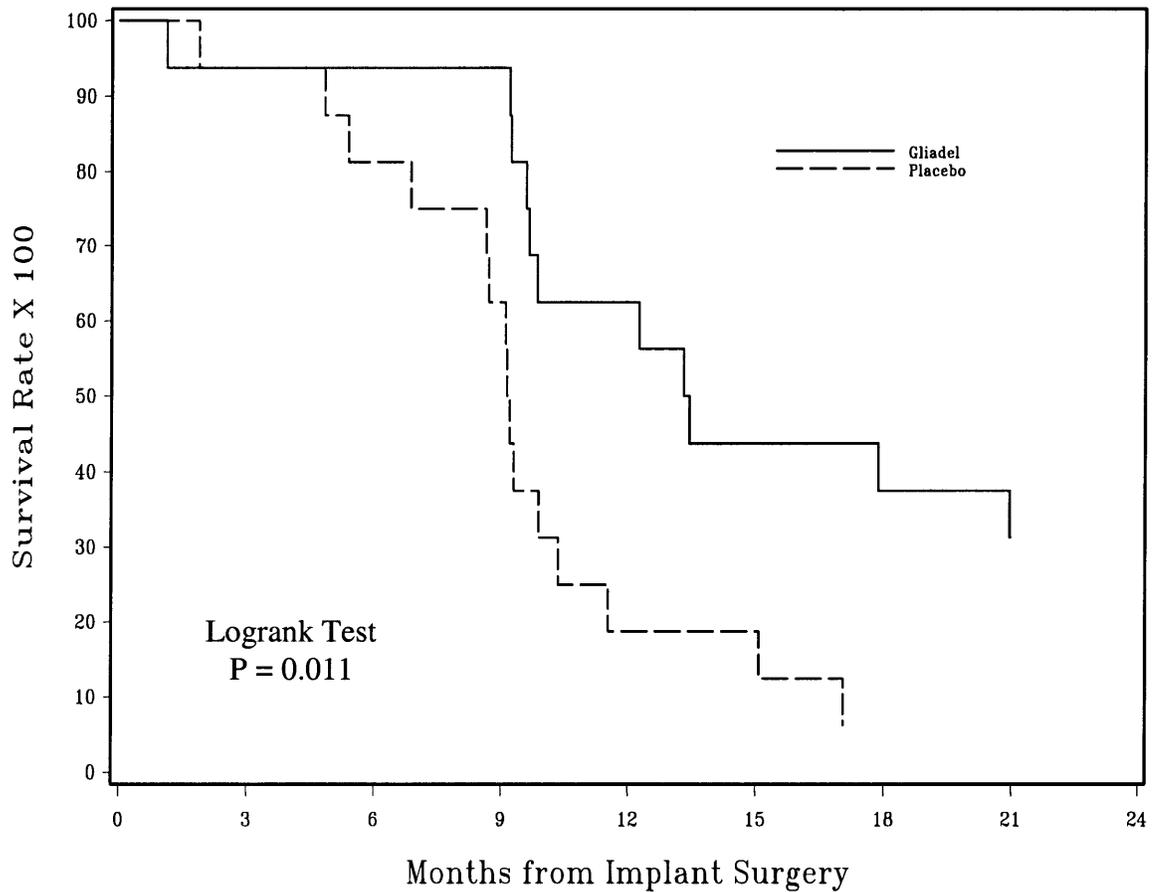
The difference between the GLIADEL[®] group and the placebo group in time to deterioration was statistically significant ($p<0.05$, stratified logrank test) for all but one of eleven neuroperformance measures (Table 1). The exception was visual status ($p=0.087$), although the median time to deterioration was longer in the GLIADEL[®] group (44.0 weeks) compared to the placebo group (42.4 weeks).

Table 1: Time to Neuroperformance Measures Deterioration, in Weeks

Neuroperformance measure	Median time to deterioration (weeks)		p-value
	GLIADEL® n=120	Placebo n=120	
Vital signs	54.9	49.1	0.010
Level of consciousness	52.1	45.4	0.016
Personality	51.7	40.0	0.008
Speech	49.6	36.7	0.003
Visual status	44.0	42.4	0.087
Fundus	55.1	46.3	0.007
Cranial nerves II, IV, VI	54.9	49.1	0.016
Cranial nerves, other	54.3	46.3	0.003
Motor status	45.4	31.4	0.013
Sensory status	51.6	44.1	0.024
Cerebellar status	54.1	46.7	0.011

In a smaller study of 32 patients undergoing initial surgery for malignant glioma (16 treated with GLIADEL and 16 placebo treated), the median overall survival after surgery increased from 39.9 weeks with placebo to 58.1 weeks with GLIADEL treatment (P = 0.011).

Overall Kaplan-Meier Survival Curves -- All Patients

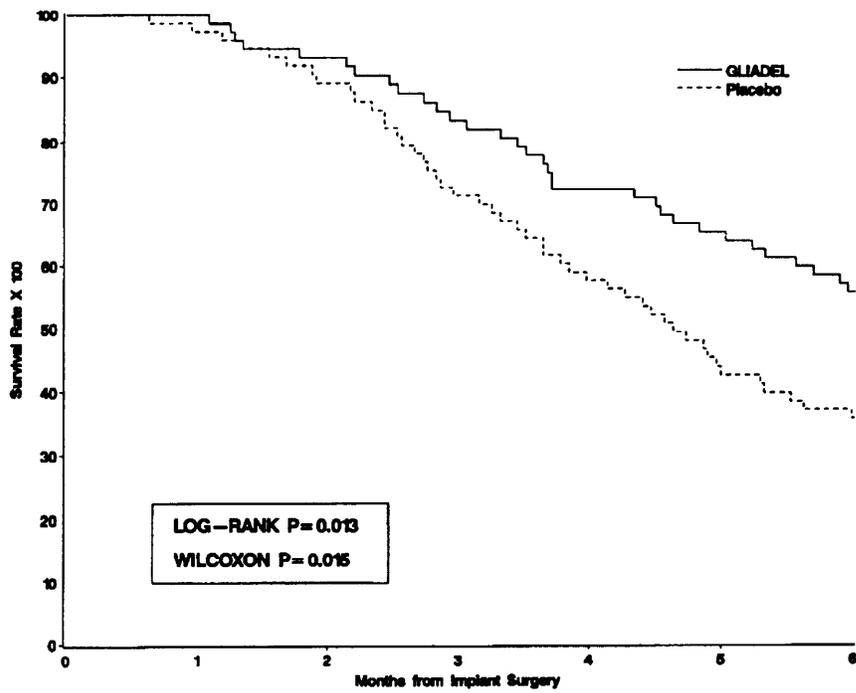


Recurrent Surgery

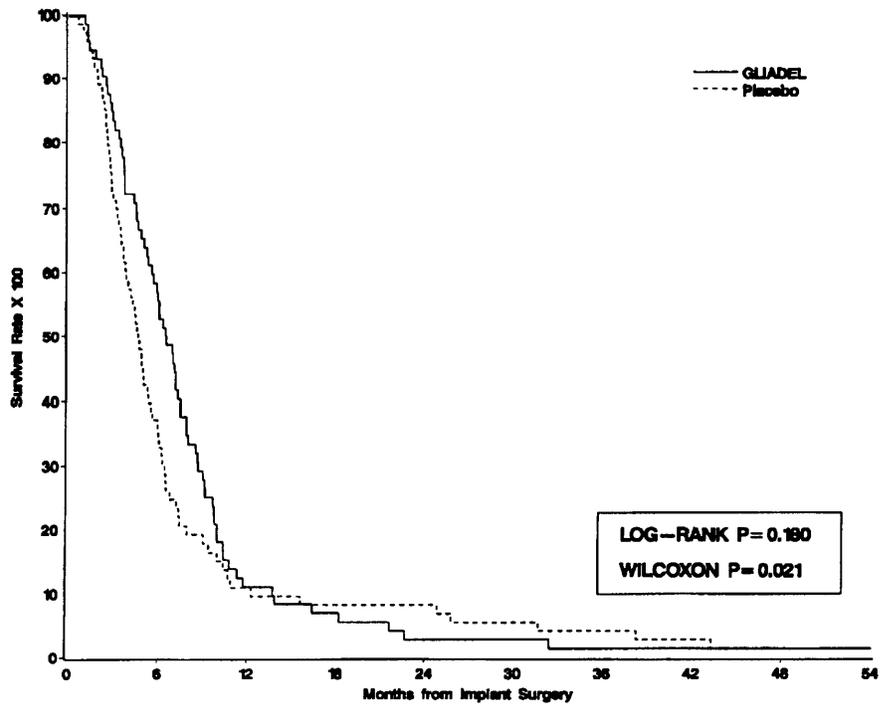
In 222 patients with recurrent malignant glioma who had failed initial surgery and radiation therapy, the six-month survival rate after surgery increased from 47% (53/112) for patients receiving placebo to 60% (66/110) for patients treated with GLIADEL. Median survival increased by 33%, from 24 weeks with placebo to 32 weeks with GLIADEL treatment. In patients with GBM, the six-month survival rate increased from 36% (26/73) with placebo to 56% (40/72) with GLIADEL treatment. Median survival of GBM patients increased by 41% from 20 weeks with placebo to 28 weeks with GLIADEL treatment. In patients with pathologic diagnoses other than GBM at the time of surgery for tumor recurrence, GLIADEL produced no survival prolongation.

Ninety-five percent of the patients treated with GLIADEL had 7-8 wafers implanted.

6-MONTH KAPLAN-MEIER SURVIVAL CURVES FOR PATIENTS UNDERGOING SURGERY FOR RECURRENT GBM



OVERALL KAPLAN-MEIER SURVIVAL CURVES FOR PATIENTS UNDERGOING SURGERY FOR RECURRENT GBM



INDICATIONS AND USAGE

GLIADEL wafer is indicated for use as a treatment to significantly prolong survival and maintain overall function (as measured by preservation of Karnofsky Performance Status) and neurological function in patients with malignant glioma undergoing primary and/or recurrent surgical resection.

CONTRAINDICATIONS

GLIADEL contains carmustine. GLIADEL should not be given to individuals who have demonstrated a previous hypersensitivity to carmustine or any of the components of GLIADEL.

WARNINGS

Patients undergoing craniotomy for malignant glioma and implantation of GLIADEL should be monitored closely for known complications of craniotomy, including seizures, intracranial infections, abnormal wound healing, and brain edema. Cases of intracerebral mass effect unresponsive to corticosteroids have been described in patients treated with GLIADEL, including one case leading to brain herniation.

Pregnancy: There are no studies assessing the reproductive toxicity of GLIADEL. Carmustine, the active component of GLIADEL, can cause fetal harm when administered to a pregnant woman. Carmustine has been shown to be embryotoxic and teratogenic in rats at i.p. doses of 0.5, 1, 2, 4, or 8 mg/kg/day when given on gestation days 6 through 15. Carmustine caused fetal malformations (anophthalmia, micrognathia, omphalocele) at 1.0 mg/kg/day (about 1/6 the recommended human dose (eight wafers of 7.7 mg carmustine/wafer) on a mg/m² basis). Carmustine was embryotoxic in rabbits at i.v. doses of 4.0 mg/kg/day (about 1.2 times the recommended human dose on a mg/m² basis). Embryotoxicity was characterized by increased embryo-fetal deaths, reduced numbers of litters, and reduced litter sizes.

There are no studies of GLIADEL in pregnant women. If GLIADEL is used during pregnancy, or if the patient becomes pregnant after GLIADEL implantation, the patient must be warned of the potential hazard to the fetus.

PRECAUTIONS

General: Communication between the surgical resection cavity and the ventricular system should be avoided to prevent the wafers from migrating into the ventricular system and causing obstructive hydrocephalus. If a communication larger than the diameter of a wafer exists, it should be closed prior to wafer implantation.

Imaging Studies: Computed tomography and magnetic resonance imaging of the head may demonstrate enhancement in the brain tissue surrounding the resection cavity after implantation of GLIADEL wafers. This enhancement may represent edema and inflammation caused by GLIADEL or tumor progression.

Therapeutic Interactions: Interactions of GLIADEL with other drugs have not been formally evaluated

Primary Surgery

No chemotherapeutic agents or other anti-neoplastic therapy were permitted for patients following initial surgery and GLIADEL until tumor progression or recurrence. In clinical trials, few patients received systemic chemotherapy. Only 17 patients (14.2%) in the GLIADEL group and 12 patients (10.0%) in the placebo group received systemic chemotherapy during the study.

The majority of patients (93/120, 77.5% in the GLIADEL group and 98/120, 81.7% in the placebo group) with newly diagnosed malignant glioma received a standard course of radiotherapy typically starting 3 weeks after surgery.

Recurrent Surgery

Chemotherapy was withheld at least four weeks (six weeks for nitrosoureas) prior to and two weeks after surgery in patients undergoing re-operation for malignant glioma.

The short-term and long-term toxicity profiles of GLIADEL when given in conjunction with chemotherapy have not been fully explored. GLIADEL, when given in conjunction with radiotherapy does not appear to have any short-term or chronic toxicities.

Carcinogenesis, Mutagenesis, Impairment of Fertility: No carcinogenicity, mutagenicity or impairment of fertility studies have been conducted with GLIADEL. Carcinogenicity, mutagenicity and impairment of fertility studies have been conducted with carmustine, the active component of GLIADEL. Carmustine was given three times a week for six months, followed by 12 months observation, to Swiss mice at i.p. doses of 2.5 and 5.0 mg/kg (about 1/5 and 1/3 the recommended human dose (eight wafers of 7.7 mg carmustine/wafer) on a mg/m² basis) and to SD rats at i.p. dose of 1.5 mg/kg (about 1/4 the recommended human dose on a mg/m² basis). There were increases in tumor incidence in all treated animals, predominantly subcutaneous and lung neoplasms. *Mutagenesis:* Carmustine was mutagenic *in vitro* (Ames assay, human lymphoblast HGPRT assay) and clastogenic both *in vitro* (V79 hamster cell micronucleus assay) and *in vivo* (SCE assay in rodent brain tumors, mouse bone marrow micronucleus assay). *Impairment of Fertility:* Carmustine caused testicular degeneration at i.p. doses of 8 mg/kg/week for eight weeks (about 1.3 times the recommended human dose on a mg/m² basis) in male rats.

Pregnancy: Pregnancy Category D: see **WARNINGS**.

Nursing Mothers: It is not known if either carmustine, carboxyphenoxypropane, or sebacic acid is excreted in human milk. Because many drugs are excreted in human milk and because of the potential for serious adverse reactions from carmustine in nursing infants, it is recommended that patients receiving GLIADEL discontinue nursing.

Pediatric Use: The safety and effectiveness of GLIADEL in pediatric patients have not been established.

ADVERSE REACTIONS

The spectrum of adverse events observed in patients who received GLIADEL or placebo in clinical studies was consistent with that encountered in patients undergoing craniotomy for malignant gliomas.

GLIADEL was not reported to be the cause of death in any of the GLIADEL clinical trials.

Primary Surgery

The following data are the most frequently occurring adverse events observed in 5% or more of the patients in the large study (240 patients) of newly diagnosed malignant glioma.

**COMMON ADVERSE EVENTS OBSERVED IN \geq 5% OF PATIENTS
RECEIVING GLIADEL AT INITIAL SURGERY**

Body System Adverse event	GLIADEL [®] N=120 n (%)	Placebo N=120 n (%)
Body as a whole		
Abdominal pain	10 (8.3)	2 (1.7)
Abscess	6 (5.0)	3 (2.5)
Accidental injury	6 (5.0)	8 (6.7)
Aggravation reaction*	98 (81.7)	95 (79.2)
Allergic reaction	2 (1.7)	6 (5.0)
Asthenia	26 (21.7)	18 (15.0)
Back pain	8 (6.7)	4 (3.3)
Chest pain	6 (5.0)	0
Face edema	7 (5.8)	6 (5.0)
Fever	21 (17.5)	21 (17.5)
Headache	33 (27.5)	44 (36.7)
Infection	22 (18.3)	24 (20.0)
Pain	16 (13.3)	18 (15.0)
Cardiovascular system		
Deep thrombophlebitis	12 (10.0)	11 (9.2)
Hemorrhage	8 (6.7)	7 (5.8)
Pulmonary embolus	10 (8.3)	10 (8.3)
Digestive system		
Constipation	23 (19.2)	14 (11.7)
Diarrhea	6 (5.0)	5 (4.2)
Liver function tests abnormal	1 (0.8)	6 (5.0)
Nausea	26 (21.7)	20 (16.7)
Vomiting	25 (20.8)	19 (15.8)
Endocrine system		
Cushings syndrome	4 (3.3)	6 (5.0)
Diabetes mellitus	6 (5.0)	5 (4.2)
Metabolic and nutritional disorders		
Healing Abnormal	19 (15.8)	14 (11.7)
Peripheral edema	11 (9.2)	11 (9.2)
Musculoskeletal system		
Myasthenia	5 (4.2)	6 (5.0)
Nervous system		
Abnormal gait	6 (5.0)	6 (5.0)
Amnesia	11 (9.2)	12 (10.0)
Anxiety	8 (6.7)	5 (4.2)
Aphasia	21 (17.5)	22 (18.3)
Ataxia	7 (5.8)	5 (4.2)
Brain edema	27 (22.5)	23 (19.2)
Coma	5 (4.2)	6 (5.0)
Confusion	28 (23.3)	25 (20.8)

*Adverse events coded to the COSTART term "aggravation reaction" were usually events involving tumor/disease progression or general deterioration of condition (e.g. condition/health/Karnofsky/neurological/physical deterioration).

**COMMON ADVERSE EVENTS OBSERVED IN \geq 5% OF PATIENTS
RECEIVING GLIADEL AT INITIAL SURGERY**

Body System Adverse event	GLIADEL [®] N=120 n (%)	Placebo N=120 n (%)
Nervous system (continued)		
Convulsion	40 (33.3)	45 (37.5)
Depression	19 (15.8)	12 (10.0)
Dizziness	6 (5.0)	11 (9.2)
Facial paralysis	8 (6.7)	5 (4.2)
Grand mal convulsion	6 (5.0)	5 (4.2)
Hallucinations	6 (5.0)	4 (3.3)
Hemiplegia	49 (40.8)	53 (44.2)
Hypesthesia	7 (5.8)	6 (5.0)
Hypokinesia	2 (1.7)	8 (6.7)
Incoordination	3 (2.5)	8 (6.7)
Insomnia	6 (5.0)	7 (5.8)
Intracranial hypertension	11 (9.2)	2 (1.7)
Neuropathy	8 (6.7)	12 (10.0)
Paresthesia	7 (5.8)	10 (8.3)
Personality disorder	10 (8.3)	9 (7.5)
Somnolence	13 (10.8)	18 (15.0)
Speech disorder	13 (10.8)	10 (8.3)
Thinking abnormal	7 (5.8)	10 (8.3)
Tremor	6 (5.0)	8 (6.7)
Respiratory system		
Dyspnea	4 (3.3)	8 (6.7)
Pneumonia	10 (8.3)	9 (7.5)
Skin and appendages		
Alopecia	12 (10.0)	14 (11.7)
Rash	14 (11.7)	13 (10.8)
Special senses		
Abnormal vision	7 (5.8)	7 (5.8)
Conjunctival edema	8 (6.7)	8 (6.7)
Diplopia	1 (0.8)	6 (5.0)
Eye disorder	3 (2.5)	6 (5.0)
Visual field defect	6 (5.0)	8 (6.7)
Urogenital system		
Urinary incontinence	9 (7.5)	9 (7.5)
Urinary tract infection	10 (8.3)	13 (10.8)

In the smaller 32 patient study, hemiplegia was the most frequently reported adverse event [GLIADEL: 6 (38%); placebo: 4 (25%)], followed by convulsions [GLIADEL: 3 (19%); placebo: 2 (13%)].

Recurrent Surgery

The following post-operative adverse events were observed in 4% or more of the patients receiving GLIADEL at recurrent surgery. Except for nervous system effects, where there is a possibility that the placebo wafers could have been responsible, only events more common in the GLIADEL group are listed. These adverse events were either not present

pre-operatively or worsened post-operatively during the follow-up period. The follow-up period was up to 71 months.

COMMON ADVERSE EVENTS OBSERVED IN $\geq 4\%$ OF PATIENTS
RECEIVING GLIADEL AT RECURRENT SURGERY

Body System Adverse Event	GLIADEL Wafer with Carmustine [N=110] n (%)	PLACEBO Wafer without Carmustine [N=112] n (%)
Body as a Whole		
Fever	13 (12)	9 (8)
Pain*	8 (7)	1 (1)
Digestive System		
Nausea and Vomiting	9 (8)	7 (6)
Metabolic and Nutritional Disorders		
Healing Abnormal*	15 (14)	6 (5)
Nervous System		
Aphasia	10 (9)	12 (11)
Brain Edema	4 (4)	1 (1)
Confusion	11 (10)	9 (8)
Convulsion	21 (19)	21 (19)
Headache	16 (15)	14 (13)
Hemiplegia	21 (19)	22 (20)
Intracranial Hypertension	4 (4)	7 (6)
Meningitis or Abscess	4 (4)	1 (1)
Somnolence	15 (14)	12 (11)
Stupor	7 (6)	7 (6)
Skin and Appendages		
Rash	6 (5)	4 (4)
Urogenital System		
Urinary Tract Infection	23 (21)	19 (17)

*p < 0.05 for comparison of GLIADEL versus placebo groups in the randomized trial (two-sided Fisher's Exact Test)

Recurrent Surgery

The following adverse events were also reported in 4-9% of GLIADEL patients but were at least as frequent in the placebo group as in GLIADEL-treated patients: infection, deep thrombophlebitis, pulmonary embolism, nausea, oral moniliasis, anemia, hyponatremia, pneumonia.

The following four categories of adverse events are possibly related to treatment with GLIADEL. The frequency with which they occurred in the randomized trial along with descriptive detail are provided below.

1. Seizures: In the randomized study, the majority of seizures in the placebo and GLIADEL groups were mild or moderate in severity. The incidence of new or worsened seizures was 19% in patients treated with GLIADEL and 19% in patients receiving

placebo. Of the patients with new or worsened seizures post-operatively, 12/22 (54%) of patients treated with GLIADEL and 2/22 (9%) of placebo patients experienced the first new or worsened seizure within the first five post-operative days. The median time to onset of the first new or worsened post-operative seizure was 3.5 days in patients treated with GLIADEL and 61 days in placebo patients. The occurrence of seizures did not reduce the survival benefit of GLIADEL.

2. Brain Edema: In the randomized trial, brain edema was noted in 4% of patients treated with GLIADEL and in 1% of patients treated with placebo. Development of brain edema with mass effect (due to tumor recurrence, intracranial infection, or necrosis) may necessitate re-operation and, in some cases, removal of wafer or its remnants.

3. Healing Abnormalities: The majority of these events were mild to moderate in severity. Healing abnormalities occurred in 14% of GLIADEL-treated patients compared to 5% of placebo recipients. These events included cerebrospinal fluid leaks, subdural fluid collections, subgaleal or wound effusions, and wound breakdown.

4. Intracranial Infection: In the randomized trial, intracranial infection (meningitis or abscess) occurred in 4% of patients treated with GLIADEL and in 1% of patients receiving placebo. In GLIADEL-treated patients, there were two cases of bacterial meningitis, one case of chemical meningitis, and one case of meningitis which was not further specified. A brain abscess developed in one placebo-treated patient. The rate of deep wound infection (infection of subgaleal space, bone, meninges, or neural parenchyma) was 6% in both GLIADEL and placebo treated patients.

The following adverse events, not listed in the table above, were reported in less than 4% but at least 1% of patients treated with GLIADEL in all studies (n=273). The events listed were either not present pre-operatively or worsened post-operatively. Whether GLIADEL caused these events cannot be determined.

Body as a Whole: peripheral edema (2%); neck pain (2%); accidental injury (1%); back pain (1%); allergic reaction (1%); asthenia (1%); chest pain (1%); sepsis (1%)

Cardiovascular System: hypertension (3%); hypotension (1%)

Digestive System: diarrhea (2%); constipation (2%); dysphagia (1%); gastrointestinal hemorrhage (1%); fecal incontinence (1%)

Hemic and Lymphatic System: thrombocytopenia (1%); leukocytosis (1%)

Metabolic and Nutritional Disorders: hyponatremia (3%); hyperglycemia (3%); hypokalemia (1%)

Musculoskeletal System: infection (1%)

Nervous System: hydrocephalus (3%); depression (3%); abnormal thinking (2%); ataxia (2%); dizziness (2%); insomnia (2%); monoplegia (2%); coma (1%); amnesia (1%); diplopia (1%); paranoid reaction (1%). In addition, cerebral hemorrhage and cerebral

infarct were each reported in less than 1% of patients treated with GLIADEL.

Respiratory System: infection (2%); aspiration pneumonia (1%)

Skin and Appendages: rash (2%)

Special Senses: visual field defect (2%); eye pain (1%)

Urogenital System: urinary incontinence (2%)

OVERDOSAGE

There is no clinical experience with use of more than eight GLIADEL wafers per surgical procedure.

DOSAGE AND ADMINISTRATION

Each GLIADEL wafer contains 7.7 mg of carmustine, resulting in a dose of 61.6 mg when eight wafers are implanted. It is recommended that eight wafers be placed in the resection cavity if the size and shape of it allows. Should the size and shape not accommodate eight wafers, the maximum number of wafers as allowed should be placed. Since there is no clinical experience, no more than eight wafers should be used per surgical procedure.

Handling and Disposal¹⁻⁷: Wafers should only be handled by personnel wearing surgical gloves because exposure to carmustine can cause severe burning and hyperpigmentation of the skin. Use of double gloves is recommended and the outer gloves should be discarded into a biohazard waste container after use. A surgical instrument dedicated to the handling of the wafers should be used for wafer implantation. If repeat neurosurgical intervention is indicated, any wafer or wafer remnant should be handled as a potentially cytotoxic agent.

GLIADEL wafers should be handled with care. The aluminum foil laminate pouches containing GLIADEL should be delivered to the operating room and remain unopened until ready to implant the wafers. **The outside surface of the outer foil pouch is not sterile.**

Instructions for Opening Pouch Containing GLIADEL

Figure 1: To remove the sterile inner pouch from the outer pouch, locate the folded corner and slowly pull in an outward motion.

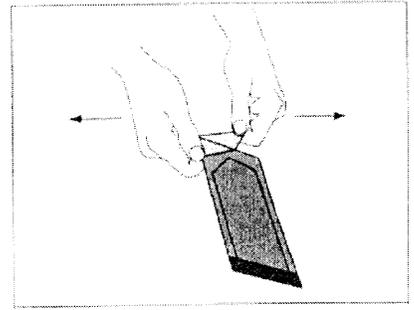


Figure 2: Do NOT pull in a downward motion rolling knuckles over the pouch. This may exert pressure on the wafer and cause it to break.

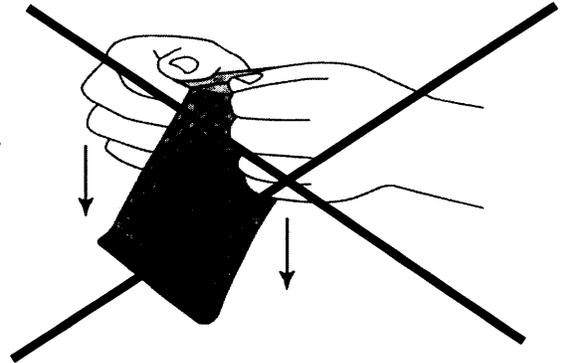


Figure 3: Remove the inner pouch by grabbing hold of the crimped edge and pulling upward.

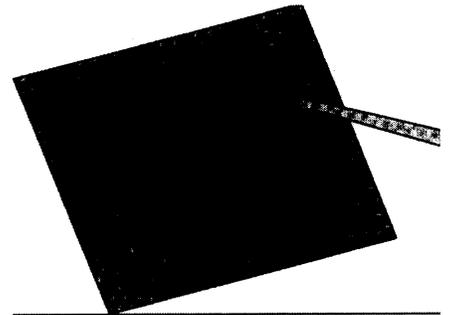


Figure 4: To open the inner pouch, gently hold the crimped edge and cut in an arc-like fashion around the wafer.

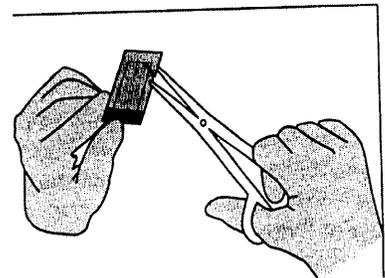
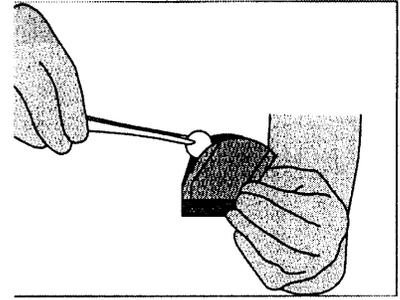


Figure 5: To remove the GLIADEL wafer, gently grasp the wafer with the aid of forceps and place it onto a designated sterile field.



Once the tumor is resected, tumor pathology is confirmed, and hemostasis is obtained, up to eight GLIADEL[®] Wafers (polifeprosan 20 with carmustine implant) may be placed to cover as much of the resection cavity as possible. Slight overlapping of the wafers is acceptable. Wafers broken in half may be used, but wafers broken in more than two pieces should be discarded in a biohazard container. Oxidized regenerated cellulose (Surgicel[®]) may be placed over the wafers to secure them against the cavity surface. After placement of the wafers, the resection cavity should be irrigated and the dura closed in a water tight fashion.

Unopened foil pouches may be kept at ambient room temperature for a maximum of six hours at a time.

HOW SUPPLIED

GLIADEL is available in a single dose treatment box containing eight individually pouched wafers. Each wafer contains 7.7 mg of carmustine and is packaged in two aluminum foil laminate pouches. The inner pouch is sterile and is designed to maintain product sterility and protect the product from moisture. The outer pouch is a peelable overwrap. **The outside surface of the outer pouch is not sterile.**

GLIADEL must be stored at or below -20°C (-4°F).

REFERENCES

1. Recommendations for the Safe Handling of Parenteral Antineoplastic Drugs, NIH Publication No. 83-2621. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402.
2. AMA Council Report, Guidelines for Handling Parenteral Antineoplastics. JAMA, 1985; 253(11):1590-1592.
3. National Study Commission on Cytotoxic Exposure -- Recommendations for Handling Cytotoxic Agents. Available from Louis P. Jeffrey, ScD., Chairman, National Study Commission on Cytotoxic Exposure, Massachusetts College of Pharmacy and Allied Health Sciences, 179 Longwood Avenue, Boston, Massachusetts 02115.
4. Clinical Oncological Society of Australia, Guidelines and Recommendations for Safe Handling of Antineoplastic Agents. Med J Australia, 1983; 1:426-428.

5. Jones RB, et al: Safe Handling of Chemotherapeutic Agents: A Report from the Mount Sinai Medical Center. CA -- A Cancer Journal for Clinicians, 1983; (Sept/Oct) 258-263.
6. American Society of Hospital Pharmacists Technical Assistance Bulletin on Handling Cytotoxic and Hazardous Drugs. Am J. Hosp Pharm, 1990; 47:1033-1049.
7. OSHA Work-Practice Guidelines for Personnel Dealing with Cytotoxic (Antineoplastic) Drugs. Am J Hosp Pharm, 1986; 43:1193-1204.

NDC: 61379-0100-1

CAUTION: FEDERAL LAW PROHIBITS DISPENSING WITHOUT PRESCRIPTION.

U.S. Patent Nos. 4,789,724 and 5,179,189.

Manufactured by
Guilford Pharmaceuticals Inc.
Baltimore, MD 21224

Rev. DRAFT

IN-1000

Placebo-controlled trial of safety and efficacy of intraoperative controlled delivery by biodegradable polymers of chemotherapy for recurrent gliomas

Henry Brem, Steven Piantadosi, Peter C Burger, Michael Walker, Robert Selker, Nicholas A Vick, Keith Black, Michael Sisti, Steven Brem, Gerard Mohr, Paul Muller, Richard Morawetz, S Clifford Schold, for the Polymer-Brain Tumor Treatment Group*

Summary

Chemotherapy for brain tumours has been limited because of difficulty in achieving adequate exposure to the tumour without systemic toxicity. We have developed a method for local sustained release of chemotherapeutic agents by their incorporation into biodegradable polymers. Implantation of the drug-impregnated polymer at the tumour site allows prolonged local exposure with minimal systemic exposure. We conducted a randomised, placebo-controlled, prospective study to evaluate the effectiveness of biodegradable polymers impregnated with carmustine to treat recurrent malignant gliomas.

In 27 medical centres, 222 patients with recurrent malignant brain tumours requiring re-operation were randomly assigned to receive surgically implanted biodegradable polymer discs with or without 3.85% carmustine. Randomisation balanced the treatment groups for all of the prognostic factors examined. Median survival of the 110 patients who received carmustine polymers was 31 weeks compared with 23 weeks for the 112 patients who received only placebo polymers (hazard ratio=0.67, $p=0.006$, after accounting for the effects of prognostic factors). Among patients with glioblastoma, 6-month survival in those treated with carmustine-polymer discs

was 50% greater than in those treated with placebo (mortality=32 of 72 [44%] vs 47 of 73 [64%], $p=0.02$). There were no clinically important adverse reactions related to the carmustine polymer, either in the brain or systemically.

Interstitial chemotherapy delivered with polymers directly to brain tumours at the time of surgery seems to be a safe and effective treatment for recurrent malignant gliomas.

Lancet 1995; **345**: 1008–12

Introduction

In view of the poor outlook of patients with malignant gliomas,^{1,2} we investigated the direct introduction of chemotherapeutic agents by controlled-release polymers. Our rationale behind this approach was based on the high local recurrence rate of primary brain tumours,³ the restrictions to systemic drug delivery imposed by the blood-brain barrier, and the severe complications from systemic exposure to drugs targeted for the brain.¹ A biodegradable polymer capable of sustained local delivery of a drug might circumvent the restrictions imposed by the blood-brain barrier and allow more effective direct treatment of the tumour.

The polymer consists of poly(carboxyphenoxypropane/sebacic acid) anhydride.⁴ Carmustine (BCNU), the most effective chemotherapeutic drug for brain tumour,^{1,5} can be incorporated into this hydrophobic matrix which protects the active agent from hydrolysis. We established the biocompatibility of the polymer, the kinetics of its degradation, and the pattern of drug release and distribution in animals.^{6,7} Carmustine incorporated into the polymer and released over a 2 to 3 week period was more effective than systemic administration in controlling growth of experimental brain tumours.⁸

A phase I trial established the safety of implanting polymers impregnated with carmustine at the time of surgery for recurrent gliomas.⁹ That study also determined the effective dose, with some patients displaying prolonged survival.⁹ To determine the effectiveness and safety of this new approach to treating brain tumours, we began a multicentre, prospective, randomised, double-blind, placebo-controlled study.

Patients and methods

Patients

222 patients were enrolled at 27 clinical centres. Patients were randomly assigned to receive either polymer discs containing carmustine or empty polymer implants.

Patients with recurrent malignant glioma were candidates for enrolment if they met the following criteria: presence of

*Participants are listed at the end of the article

Departments of Neurological Surgery (Prof H Brem MD), **Oncology** (Prof H Brem, S Piantadosi MD), **and Pathology** (Prof P C Burger MD), **Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; Department of Biostatistics, Johns Hopkins University School of Hygiene and Public Health, Baltimore** (S Piantadosi); **the National Institute of Neurological Diseases and Stroke of the National Institutes of Health, Bethesda, Maryland** (Prof M Walker MD); **Division of Neurosurgery, Western Pennsylvania Hospital, Pittsburgh, Pennsylvania** (Prof R Selker MD); **Department of Neurology, Evanston Hospital, Evanston, Illinois** (Prof N A Vick MD); **Department of Neurological Surgery, University of California, Los Angeles** (Prof K Black MD); **Department of Neurosurgery, Neurological Institute of New York** (M Sisti MD); **Department of Neurological Surgery, Northwestern University, Chicago, Illinois** (Prof S Brem MD); **Division of Neurosurgery, McGill University, Montreal, Quebec, Canada** (G Mohr MD); **Division of Neurosurgery, University of Toronto, Toronto, Ontario, Canada** (Prof P Muller MD); **Department of Neurosurgery, University of Alabama, Birmingham, Alabama** (Prof R Morawetz MD); **and Department of Neurology, University of Texas Southwestern Medical Center, Dallas, Texas** (Prof S C Schold MD)

Correspondence to: Prof Henry Brem, Department of Neurological Surgery and Oncology, Hunterian 817, Johns Hopkins University School of Medicine, 725 N Wolfe Street, Baltimore, MD 21205, USA

unilateral single focus of tumour in the cerebrum showing at least 1.0 cm³ enhancing volume on computed tomography scan or magnetic resonance imaging; a Karnofsky performance score of at least 60 (ie, ability to function independently); completion of external beam radiation therapy; and no nitrosoureas for 6 weeks and no other systemic chemotherapeutic agent for 4 weeks before enrolment. In addition, patients' surgeons made an independent determination that another tumour resection would be done irrespective of the study.

Carmustine discs

BIODEL, the polyanhydride polymer used, is a copolymer of poly-carboxyphenoxypyrone and sebacic acid prepared in a 20/80 ratio.⁹ Briefly, polymer and carmustine were co-dissolved in methylene chloride and spray dried into microspheres, which were compressed into discs of 1.4 cm diameter and 1.0 mm thickness, and sterilised by 2.2×10^4 Gy gamma irradiation.¹⁰ Loading with 50 µg carmustine/mm² of polymer (3.85% carmustine loading) yielded 7.7 mg of carmustine per wafer for a maximum patient dose of 62 mg. This dose was chosen as a result of previous experiments⁸ and a phase I clinical trial.⁹

Trial design

Patients underwent a craniotomy for maximum resection of tumour. The final admission criterion for the study was either the pathologist's report of malignant glioma or the report of recurrent tumour in a patient with a previously established malignant glioma. Randomisation was stratified by institution. Investigators and study monitors did not have access to the treatment assignments. After removal of the tumour, up to eight discs were applied to the resection cavity surface. Sheets of oxidised regenerated cellulose (Surgicel, Johnson & Johnson, New Brunswick, NJ, USA) were used occasionally to secure the polymers against the brain. All patients were clinically and radiologically reassessed at least once every 2 months. Patients were eligible to receive systemic chemotherapy 2 weeks after the implant surgery.

222 patients were enrolled between March 1, 1989, and January 17, 1992. An interim analysis to assess safety was done midway through the study by an outside reviewer (MW). The first analysis of all endpoints was done after all enrolled patients had passed the 6-month post-operative point (July 17, 1992). At the time of the analysis reported here (September 4, 1993), 93% of the enrolled patients had died.

Pathological evaluation

The tissue sections of the recurrent tumours were reviewed by one of us (PCB) without any knowledge of patients' treatment or outcome. Fibrillary astrocytic tumours were classified by a modified Ringertz system.¹¹ As part of the study, but not as a determinant of treatment, malignant gliomas (largely glioblastomas) were further subdivided into those that were clearly actively proliferating tumours and those that showed the effects of treatment. The "active" or "recurrent" neoplasms were cellular, mitotically active tumours resembling glioblastomas as encountered routinely before radiotherapy or chemotherapy. The "quiescent" or "persistent" tumours were generally extensively necrotic, but without peripheral pseudopalisading. These tumours were paucicellular neoplasms that often contained pleomorphic cells.¹²

To study the histological effects of the polymer implants with and without incorporated carmustine, 11 brains were evaluated at necropsy: 7 were from patients who had received carmustine polymers, and 4 from patients who had received placebo polymers. Postmortem magnetic resonance images were obtained for the brains of 8 patients.¹³

Statistical methods

The primary endpoint of this trial was survival from the time of polymer implantation. Secondary endpoints included rates of complications, and toxicity and quality of life measurements. The primary efficacy analysis included all the patients randomised,

and all analyses classified patients according to treatment assigned (intention-to-treat). Event times were censored if the patient was still alive on September 4, 1993. The primary endpoint represents time to death from any cause.

Event-time distributions were estimated by the product-limit method¹⁴ and compared by the log-rank statistic.¹⁵ To control for the effects of strong prognostic factors on outcome due to chance imbalances in the treatment groups,^{16,17} adjusted analyses were done with the proportional hazards regression model.¹⁸ Prognostic factors such as pathological type, Karnofsky performance score, extent of previous surgery, age, and previous use of nitrosoureas were thought to be important a priori.¹⁶ In practice, we included these and other statistically significant predictors in multiple regression models to examine their influence on the estimated treatment effect. Because of inter-correlations, some factors did not remain significant and were removed from the multiple regression. The estimated hazard ratio for carmustine polymers was not affected by these factors.

Differences in complication and toxicity rates between treatment groups were tested for statistical significance by the chi-squared or *t* tests. All *p* values reported are two-sided.

Results

Patients

Carmustine polymer discs were implanted in 110 patients and placebo polymer discs in 112. Table 1 shows that no significant differences were found between patient groups. Half the patients entered into the study had received previous systemic chemotherapy. Treatment with the carmustine polymer did not lower the performance status or neurological condition of patients compared with those who did not receive carmustine. Within 6 months of the polymer implantation, 11.8% of the carmustine group and 11.6% of the placebo group underwent re-operation.

Characteristics	Carmustine polymer (n=110)	Placebo polymer (n=112)	<i>p</i>
Mean (SD) age (years)	48.1 (12.3)	47.6 (13.6)	0.75
Sex (male)	74 (67%)	69 (62%)	0.38
Race (white)	100 (91%)	103 (92%)	0.78
Mean (SD) Karnofsky performance score	77.0 (13.1)	74.6 (12.1)	0.17
Mean (SD) mini-mental state exam score	24.1 (7.2)	22.6 (8.5)*	0.16
Previous treatment			
Operations			
1	83 (75.5%)	79 (70.5%)	
2	20 (18.2%)	30 (26.8%)	
≥3	7 (6.4%)	3 (2.7%)	0.17
Median interval from first operation	12.9 mo	11.3 mo	0.19
Amount of radiation therapy			
≥45 Gy	108 (98.2%)	110 (98.2%)	
<45 Gy	2 (1.8%)	2 (1.8%)	
None	0 (0.0%)	0 (0.0%)	0.99
Type of radiation therapy			
Local	53 (48.2%)	54 (48.2%)	
Whole brain	28 (25.5%)	23 (20.5%)	
Local and whole brain	29 (26.4%)	34 (30.4%)	
Unknown	0 (0.0%)	1 (0.9%)	0.60
Chemotherapy	58 (52.7%)	54 (48.2%)	0.50
Immunotherapy	7 (6.4%)	5 (4.5%)	0.53
Brachytherapy	2 (1.8%)	5 (4.5%)	0.45
Tumour histopathology at implantation			
Glioblastoma	72 (65.5%)	73 (65.2%)	
Astrocytoma (anaplastic)	15 (13.6%)	16 (14.3%)	
Oligodendroglioma (anaplastic)	4 (3.6%)	5 (4.5%)	
Oligodendroglioma	2 (1.8%)	2 (1.8%)	
Other glial tumours	16 (14.5%)	16 (14.5%)	
Necrosis	1 (0.9%)	0 (0.0%)	
>75% resection at reoperation	88 (79.9%)	87 (78.0%)	0.54

*n=108, scores were missing for 4 patients.

Table 1: Patient characteristics by treatment group

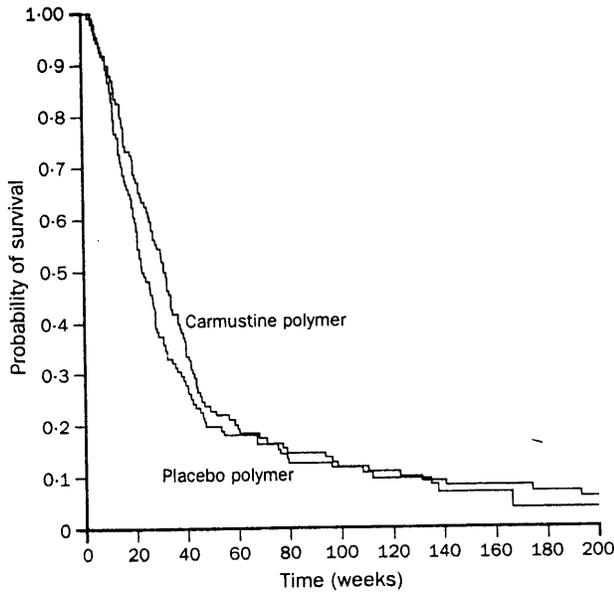


Figure 1: Overall survival by treatment group (Kaplan-Meier curve)

During this initial 6-month period, 25.5% of patients in the carmustine group and 18.8% of placebo patients had systemic chemotherapy.

Laboratory analyses

Neither significant reductions in blood cell counts nor abnormalities in blood chemistry or urinalysis were found even though these frequently occur with systemic exposure to carmustine. Hyperglycaemia and glycosuria were observed in both patient groups, but these could be attributed to the high doses of corticosteroid used routinely to reduce cerebral swelling in these patients.

Statistical analysis

Median survival was 31 weeks in the carmustine group and 23 weeks in the placebo group. 59 (53%) of 112 patients treated with placebo implants were dead at 6 months compared with 44 (40%) of 110 patients treated with carmustine implants (p=0.061). Among patients with glioblastoma, treatment with placebo polymer led to 64% (47 of 73 patients) mortality at 6 months compared with 44% (32 of 72 patients) mortality for those treated with the carmustine implants (p=0.020).

The overall treatment effect favoured the carmustine polymer (estimated hazard ratio 0.83, p=0.19, log rank,

Variable	Hazard ratio (95% CI)	p*
Carmustine polymer vs placebo polymer	0.83 (0.63-1.10)	0.19
>75% tumour resection vs <75% resection	0.56 (0.41-0.76)	<0.001
Age (per decade)	1.24 (1.11-1.39)	<0.001
White vs other races	1.83 (1.10-3.06)	0.02
Male vs female	0.80 (0.61-1.07)	0.14
Interval from first surgery to index surgery (per year)	0.90 (0.84-0.96)	0.001
Karnofsky ≥70 vs <70	0.53 (0.40-0.70)	<0.001
Local radiation vs whole brain	0.76 (0.55-1.06)	0.10
Previous chemotherapy vs none	1.58 (1.20-2.09)	<0.001
Previous nitrosoureas vs none	1.61 (1.22-2.12)	<0.001
Previous immunotherapy vs none	1.18 (0.66-2.12)	0.57
Active recurrent vs *quiescent* tumour at implant surgery	1.25 (0.76-2.05)	0.38
Anaplastic astrocytoma vs glioblastoma	0.60 (0.40-0.90)	0.01
Oligodendroglioma vs glioblastoma	0.39 (0.26-0.59)	<0.001
All other diagnoses vs glioblastoma	0.31 (0.13-0.70)	0.005

*Tests the hypothesis that hazard ratio=1.0.

Table 2: Estimated hazard ratios and 95% CIs for survival for prognostic factors (univariate regressions)

figure 1 and table 2). Although treatment groups were balanced with respect to prognostic factors, several of these were very strong predictors of outcome. For example, resecting 75% or more of the tumour, a Karnofsky performance score greater than 70, and pathological type were all strong predictors of survival irrespective of treatment with the carmustine implants (table 2). When accounting for the effects of treatment and prognostic factors simultaneously, the estimated hazard ratio for treatment (0.67) was statistically significant (p=0.006; table 3, model A). Similar effects were seen in a multiple regression model that stratified for the effect of pathology and adjusted for the other factors (table 3, model B). These different methods of evaluating prognostic factors yielded quantitatively consistent estimates of the beneficial effect of carmustine polymer.

Because the overall survival curves (figure 1) reflect both the treatment effect and influential differences in prognostic factors, we calculated survival curves adjusted by the proportional hazards regression model for the factors listed in table 3. Adjusted survival curves (figure 2) showed an increased median survival of 9 weeks attributable to carmustine, and slightly higher long-term survival.

The clinically most important subset of patients are those with glioblastoma. In these 145 patients, carmustine polymer lowered the risk of death with an estimated hazard ratio of 0.81 (p=0.22), a finding similar to the overall effect. Factors that were significant predictors of outcome in patients with glioblastoma included age (p=0.004), interval from previous surgery (p<0.001),

Variable	Model A (all patients)		Model B (all patients, stratified for pathology)		Model C (glioblastoma patients only n=145)	
	Hazard ratio (95% CI)	p	Hazard ratio (95% CI)	p	Hazard ratio (95% CI)	p
Carmustine polymer vs placebo polymer	0.67 (0.51-0.90)	0.006	0.69 (0.52-0.91)	0.01	0.67 (0.48-0.95)	0.02
Karnofsky >70 vs ≤70	0.65 (0.48-0.89)	0.007	0.66 (0.49-0.91)	0.01	0.62 (0.44-0.89)	0.009
Local radiation vs whole brain	0.60 (0.43-0.84)	0.003	0.59 (0.42-0.83)	0.003	0.64 (0.43-0.96)	0.03
Active vs *quiescent*	1.95 (1.13-3.35)	0.02	1.93 (1.26-3.78)	0.02	2.37 (1.20-4.66)	0.01
Previous nitrosoureas vs none	1.49 (1.11-2.01)	0.009	1.53 (1.13-2.08)	0.006	1.60 (1.12-2.28)	0.009
White vs other races	1.78 (1.04-3.03)	0.03	1.75 (1.03-2.99)	0.04	2.39 (1.15-4.99)	0.02
>75% resection vs <75% resection	0.66 (0.48-0.92)	0.01	0.67 (0.49-0.93)	0.02
Age (per decade)	1.24 (1.10-1.39)	<0.001	1.25 (1.11-1.40)	<0.001
Interval from previous operation (per year)	0.82 (0.73-0.92)	<0.001
Anaplastic astrocytoma vs glioblastoma	0.63 (0.42-0.95)	0.03
Oligodendroglioma vs glioblastoma	0.43 (0.28-0.67)	<0.001
All other diagnoses vs glioblastoma	0.46 (0.20-1.07)	0.07

Table 3: Effect of carmustine polymer adjusted for prognostic factors (multiple regressions)

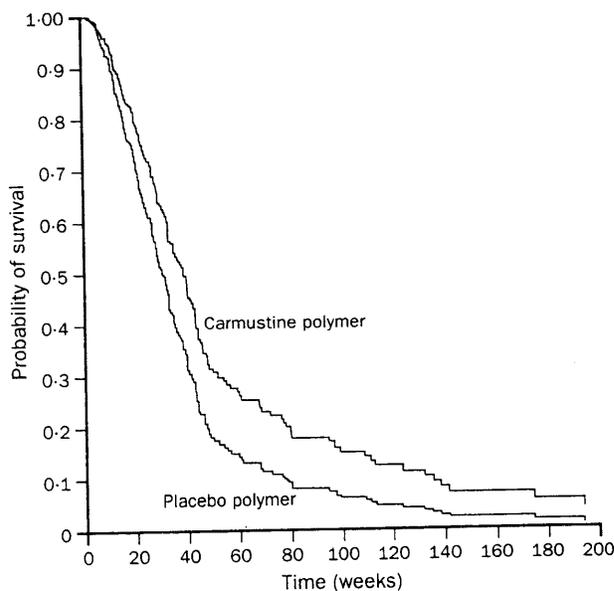


Figure 2: Overall survival by treatment group after adjustment for prognostic factors

The curves illustrate the treatment effect expected if all patients were about age 48, white, had performance status >70, underwent >75% resection, had local irradiation, had not previously been exposed to nitrosoureas, and had glioblastomas pathologically classified as active.

Karnofsky performance score ($p=0.02$), race ($p=0.06$), and previous nitrosourea chemotherapy ($p=0.03$). When treatment group and prognostic factors were considered simultaneously in a multiple regression analysis stratified by pathological type, carmustine polymer showed a significant beneficial effect in glioblastoma patients (hazard ratio 0.67, $p=0.02$; table 3, model C). Also, glioblastoma patients classified as having recurrent active tumours had significantly increased risk compared with those classified as quiescent (hazard ratio 2.37, $p=0.01$; table 3, model C). With the regression model in table 3, model C, there was no statistically significant interaction between use of carmustine polymer and active versus quiescent tumour, which indicates that the treatment benefit was not restricted only to patients with active recurrent tumour.

Adverse events

During postoperative follow-up, no deleterious effects occurred as a result of polymer implants. Anaemia occurred postoperatively in 7% of patients treated with carmustine polymers and in 11% of placebo controls; 2% of each group had thrombocytopenia and 1% of the carmustine polymer group had leukopenia. 73 patients had seizures postoperatively (41 carmustine, 32 placebo, $p=0.199$), which was within the expected frequency for postoperative seizures.¹⁹

Overall incidence of serious intracranial infection was low (5/222, 2.2%) but was more common with carmustine treatment (4/110) than placebo (1/112). This difference was not statistically significant and well within the reported range (9–13%) for recurrent glioma surgery.^{2,20} Other minor infections included urinary-tract infections, pneumonia, and conjunctivitis, which were equally common in the two treatment groups and were consistent with the expected general infection rate for patients on steroids who had undergone multiple craniotomies.³ All patients experienced cerebral oedema

during the study, as is typical for postoperative craniotomy patients, and were treated with corticosteroids. There were no apparent differences between the groups in requirement for steroids.

Postmortem studies

11 brains were examined after death. The brains of 9 of the 11 patients contained large disseminated glioblastomas; in no case was there extensive necrosis. Fibrous membranes were evident in the tumour bed in several specimens. In 2 of the 11 patients, the small amount of tumour did not explain the patient's death. 1 of these patients succumbed to disseminated colon cancer and the other died after a 3-week clinical deterioration that was unrelated to the intracranial disease. Postmortem magnetic resonance scans revealed the expected increased T2 signal in the region of the tumour, which often crossed the corpus callosum. In no case did the extent of abnormal magnetic resonance signal seem remarkable or unusually large for a recurrent glioblastoma, nor were there any changes directly attributable to the implants.

Discussion

Use of biodegradable polymers to deliver prolonged, high doses of chemotherapy directly to a tumour, thereby sparing the patient from systemic exposure to the drug, represents a new tool in the armamentarium against cancer. In this study, carmustine polymer implants significantly prolonged survival. By contrast with systemic carmustine therapy, no notable untoward events were associated with the treatment.

The polyanhydride polymer used in the present trial is hydrophobic and therefore protects carmustine from decomposition until it is released into the tumour environment. Compared with systemic delivery, intracranial implantation of a carmustine-containing polymer in animals increases brain exposure to the drug 113-fold.⁷

The study was designed to isolate the effect of drug-impregnated polymer from previous treatments, so that the efficacy of implantation of the polymer-drug could be stringently evaluated. Although the study design controlled for large imbalances by randomisation, we increased the precision of the evaluation of treatment effect with adjusted analyses. Consistency in the estimated hazard ratios in favour of carmustine polymer—irrespective of the method of analysis—and the control of bias and imbalance afforded by the study design, strongly support the efficacy of this drug-delivery system. Curran et al¹⁶ used a recursive partitioning technique to refine the stratification and design of malignant glioma trials. They observed an impact on survival of age, performance status, and tumour histopathology, independent of treatment method. Florell and colleagues¹⁷ have emphasised the selection bias of uncontrolled trials for assessing treatment of brain tumours. The benefits of interstitial radiation implants reported in previous studies could be obtained simply by prospectively applying the entry criteria, and did not depend on the actual treatment.¹⁷ In view of the modest but significant improvement in survival of carmustine-polymer-treated patients with recurrent gliomas, future studies will evaluate the effectiveness of higher doses of carmustine and the use of the polymer implants as the initial therapy for brain tumours.

The present results suggest that biodegradable polymers can assist delivery of other drugs. Brain tumour therapy might now be approached with agents that do not pass the blood-brain barrier. We have found that carboplatin,²¹ 4-hydroperoxycyclophosphamide,²² camptothecin,²³ and paclitaxel²⁴ can be effectively delivered intracranially to improve treatment of brain tumours in rats. Steroids²⁵ and immunotoxins such as the transforming growth factor alpha pseudomonas exotoxin fusion protein²⁶ may be more safely delivered by polymers. Peptides and polynucleotides including inhibitors of angiogenesis²⁷ and antisense oligonucleotides²⁸ might also be more effective when delivered locally.

Demonstration of effective polymeric delivery of carmustine directly into the brain opens the door to treatment of other diseases requiring central nervous system delivery. Solid tumours in other locations also might be treated with polymeric delivery of radiosensitisers or chemotherapeutic drugs. We suggest that, whenever local approaches such as surgery or radiation therapy are being used, consideration be given to development of biodegradable polymer delivery systems to maximize the benefit of such treatments.

The Polymer Brain Tumor Treatment Group consists of the following investigators (number of patients enrolled at each institution): Henry Brem, Eileen Bohan, Alessandro Olivi, Rafael Tamargo, Haring Nauta, Donlin M Long, and O Michael Colvin, Johns Hopkins Hospital, Baltimore, MD, USA (35); Robert Selker and Linda Kirnick, Western Pennsylvania Hospital, Pittsburgh, PA (26); Nicholas A Vick, Ivan S Ciric, Theodore W Eller, Jeffrey W Cozzens, Nina Palaologos, and Annette Walsh, Evanston Hospital, Evanston, IL (20); Keith L Black and Joe Ciacci, UCLA, Los Angeles, CA (16); Michael Sisti and Lisette Abad, Neurological Institute of New York, Columbia Presbyterian Medical Center, New York, NY (14); Steven Brem, Gerard Mohr, and Shirley Entis, Jewish General Hospital, McGill University, Montreal, Quebec, Canada (11); Paul Muller and Dina Evans, University of Toronto, Toronto, Canada (11); Richard Morawetz and Diane Taylor, University of Alabama, Birmingham, AL, USA (9); S Clifford Schold, Allan H Friedman, and Amy Gentry, Duke University Medical Center, Durham, NC (9); Kevin Lillehei and Betty Owens, University of Colorado, Denver, CO (8); Joseph Ransohoff and Arlene Wise, New York University Medical Center, New York, NY (7); Gaston Acosta-Rua and Cyndie Trogden, Jacksonville, FL (7); Stephen Bloomfield and Sandi Cline, West Virginia University, Morgantown, WV (6); Jeffrey Olsen and Karen Blakely, Emory University, Atlanta, GA (6); Scott Shapiro and Sandy Kay, Indiana University Wishard Hospital, Indianapolis, IN (6); Spiridon Koulouris and Karn Jorstad, Kaiser Permanente, Anaheim, CA (5); Lucy Love and Teresa Quisenberry, HarborSide Medical Tower, Tampa, FL (5); Rual Abad and Luanne Procyk, Easton, PA (3); W Craig Clark and Pat Coleman, Semme's-Murphey Clinic, Memphis, TN (3); Deborah Heros and Deborah Derby, Park Nicollet Medical Center, St Lewis Park, MN (3); Stephen Saris and Chris Robbins, Tufts New England Medical Center, Boston, MA (3); Thomas Trautmann and Sandra Heineken, Charlotte Medical Center, Charlotte, NC (3); Lynne Taylor and Claude Wetzell, Virginia Mason Clinic, Seattle, WA (2); Edward Connolly and Rosemary Dunn, Ochsner Clinic, New Orleans, LA (1); Greg Criscuolo and Debbie Pantalena, Yale University School of Medicine, New Haven, CT (1); Peter Heilbrun and Peter Sunderland, University of Utah, Salt Lake City, UT (1); and Patrick LaSala and Ada Korn, Albert Einstein Medical Center, Montefiore, Bronx, NY (1).

We thank our medical advisory committee of Dr Darell D Bigner, Dr Mark Chasin, Dr O Michael Colvin, Dr Victor Levin, and Dr Richard Trout. We also thank Dr Pamela Talalay for her advice. This study was supported by Guilford Pharmaceuticals Inc, Baltimore, MD; Scios-Nova Corporation, Mountain View, CA; and by the National Cooperative Drug Discovery Group (U01-CA52857 and U01-CA62474) of the National Cancer Institute of the National Institutes of Health, Bethesda, MD.

References

- Black PM. Brain tumors. *N Engl J Med* 1991; 324: 1471-76.
- Dirks P, Bernstein M, Muller PJ, Tucker WS. The value of reoperation for recurrent glioblastoma. *Can J Surg* 1993; 36: 271-75.
- Hochberg FH, Pruitt AA, Beck DO, DeBrun G, Davis K. The rationale and methodology for intra-arterial chemotherapy with BCNU as treatment for glioblastoma. *J Neurosurg* 1985; 63: 876-80.
- Mathiowitz E, Saltzman M, Domb A, Dor P, Langer R. Polyanhydride microspheres as drug carriers. II Microencapsulation by solvent removal. *J Appl Polymer Sci* 1988; 35: 755-74.
- Walker MD, Green SB, Byar DP, et al. Randomized comparisons of radiotherapy and nitrosoureas for the treatment of malignant glioma after surgery. *N Engl J Med* 1980; 303: 1323-29.
- Brem H, Tamargo RJ, Olivi A, et al. Biodegradable polymers for controlled delivery of chemotherapy with and without radiation therapy in the monkey brain. *J Neurosurg* 1994; 80: 283-90.
- Grossman SA, Reinhard CS, Colvin OM, et al. The intracerebral distribution of BCNU delivery by surgically implanted biodegradable polymers. *J Neurosurg* 1992; 76: 640-47.
- Tamargo RJ, Myseros JS, Epstein JI, Yang MB, Chasin M, Brem H. Interstitial chemotherapy of the 9L gliosarcoma: controlled release polymers for drug delivery in the brain. *Cancer Res* 1993; 53: 329-33.
- Brem H, Mahaley MS, Vick NA, et al. Interstitial chemotherapy with drug polymer implants for the treatment of recurrent gliomas. *J Neurosurg* 1991; 74: 441-46.
- Chasin M, Lewis D, Langer R. Polyanhydrides for controlled drug delivery. *Biopharm Manufact* 1988; 1: 33-46.
- Burger PC, Scheithauer BW, Vogel FS. Surgical pathology of the nervous system and its coverings, 3rd ed. New York: Churchill-Livingstone, 1991.
- Burger PC, DuBois PJ, Schold SC Jr, et al. Computerized tomographic and pathologic studies of the untreated, quiescent, and recurrent glioblastoma multiforme. *J Neurosurg* 1983; 58: 159-69.
- Boyko OB, Alston SR, Fuller GN, Hulett CM, Johnson GA, Burger PC. Utility of postmortem magnetic resonance imaging in clinical neuropathology. *Arch Pathol Lab Med* 1994; 118: 219-25.
- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53: 457-81.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; 22: 719-48.
- Curran WJ, Scott CB, Horton J, et al. Recursive partitioning analysis of prognostic factors in three radiation therapy oncology group malignant glioma trials. *J Natl Cancer Inst* 1993; 85: 704-10.
- Florell RC, Macdonald DR, Irish WD, et al. Selection bias, survival, and brachytherapy for glioma. *J Neurosurg* 1992; 76: 179-83.
- Cox DR. Regression models and life tables (with discussion). *J R Stat Soc B* 1972; 34: 187-220.
- North JB, Penhall RK, Hanieh A, Frewin DB, Taylor WB. Phenytoin and postoperative epilepsy: a double-blind study. *J Neurosurg* 1983; 58: 672-77.
- Tenney JH, Vlahov D, Salzman M, Ducker TB. Wide variation in risk of wound infection following clean neurosurgery. *J Neurosurg* 1985; 62: 243-47.
- Olivi A, Gilbert M, Duncan KL, Corden B, Lenartz D, Brem H. Direct delivery of platinum-based antineoplastics to the central nervous system: a toxicity and ultrastructural study. *Cancer Chemother Pharmacol* 1993; 31: 449-54.
- Judy KD, Olivi A, Buahin KG, et al. Controlled release of a cyclophosphamide derivative with polymers is effective against rat gliomas. *J Neurosurg* 1995; 82: 103-08.
- Weingart J, Tyler B, Colvin OM, Brem H. Local delivery of camptothecin as a new treatment for malignant brain tumors. Third NIH Drug Discovery and Development Symposium; San Diego, California; 1993.
- Walter KA, Cahan MA, Gur A, et al. Interstitial taxol delivered from a biodegradable polymer implant against experimental malignant glioma. *Cancer Res* 1994; 54: 2207-12.
- Tamargo RJ, Sills AK, Reinhard CS, Pinn ML, Long DM, Brem H. Interstitial delivery of dexamethasone in the brain for the reduction of peritumoral edema. *J Neurosurg* 1991; 74: 956-61.
- Phillips PC, Levow C, Catterall M, Colvin OM, Pastan I, Brem H. Transforming growth factor alpha pseudomonas exotoxin fusion protein (TGF alpha-PE38) treatment of subcutaneous and intracranial human medulloblastoma and glioma xenografts in athymic mice. *Cancer Res* 1994; 54: 1008-15.
- Weingart J, Sipos EP, Brem H. The role of minocycline in treatment of intracranial 9L glioma. *J Neurosurg* 1995; 82: 635-40.
- Simons M, Edelman ER, DeKeyser JL, Langer R, Rosenberg RD. Antisense c-myc oligonucleotides inhibit intimal arterial smooth muscle cell accumulation in vivo. *Nature* 1992; 359: 67-70.

STUDY SYNOPSIS

Compound:

A biodegradable, polyanhydride copolymer consisting of polycarboxyphenoxypropane (PCPP) and sebacic acid (SA) in a 20:80 molar ratio with BCNU (carmustine) incorporated into the matrix

Names	
Generic Name	Polifeprosan 20 with carmustine
Trade Name	GLIADEL® Wafer

Report Type:

Integrated Clinical and Statistical Report

Study Phase and IND Number:

Phase III, IND #30,237

Protocol Number/Title:

8802 / A Double-Blind, Placebo-Controlled Study of BCNU Delivered from BIODERL®, a Biodegradable, Surgically Implanted Polymer for the Treatment of Recurrent, Malignant Glioma.

Protocol Review:

The protocol, amendment and Informed Consent Form for this study were approved by an Institutional Review Board (IRB) at each investigational site before initiation as per regulation 21 CFR 56, and each patient signed an IRB-approved written Informed Consent Form as per regulation 21 CFR 50.

Dates of Study:

First patient enrolled - March 1, 1989

Last patient enrolled - January 17, 1992

Last observation on last patient - November 10, 1995

End of follow-up for survival and mortality analyses - November 10, 1995

Objectives of Study:

The objectives of the study were to determine the efficacy of BCNU delivered by surgically implanted polymer wafers for improving six-month survival, and to measure the side effects associated with this treatment.

Number of Patients:

Two hundred twenty-two patients between the ages of 19 and 80 years old were enrolled into Study 8802. A total of 110 patients (50%) were enrolled into the GLIADEL treatment group and 112 patients (50%) were enrolled into the PLACEBO treatment group.

Number of Clinical Centers:

Multicenter - 27 centers (25 in the United States and two in Canada)

Study Design:

This study was a multicenter, double-blind, randomized, placebo-controlled, Phase III clinical trial. Patients underwent intraoperative, surgical implantation of GLIADEL or PLACEBO wafers for treatment of recurrent malignant glioma. Patients received either GLIADEL (7.7 mg BCNU, 3.85% BCNU by weight) or blank wafers (PLACEBO) and all were evaluable for safety and efficacy analyses (intent-to-treat analyses).

The primary efficacy measures were cumulative mortality and mortality rates through six months after wafer implantation surgery. The secondary efficacy measures were mortality through the end of the post-surgery observation period (November 10, 1995), Karnofsky Performance Status (KPS) scores and Mini-Mental State Examinations (MMSE). Safety parameters included death, discontinuation from the study,

treatment-emergent adverse events, neurological examination, level of consciousness assessment, clinical laboratory test evaluation, post-baseline chemotherapy, and concomitant medications.

Indication:

Recurrent, malignant Grade III or IV astrocytoma in patients at least 18 years of age.

Investigational New Drug and Dosage:

Up to 8 wafers of GLIADEL (7.7 mg BCNU, 3.85% BCNU by weight) or blank wafers (PLACEBO) were implanted - Batches SRO42-49-1 through SRO42-49-8 were used in this trial. The number of wafers implanted was determined by the size of the resection cavity.

Duration of Administration:

GLIADEL or PLACEBO wafers were implanted once; repeat administration was not permitted.

Total Study Period:

The study continued until the time of patient death or to November 10, 1995, the end of the post-surgery follow-up period.

Efficacy Results:

The primary efficacy measure for this trial was six-month mortality.

Six-Month Mortality: All Patients by Treatment Group

A total of 44 of the 110 patients (40%) in the GLIADEL treatment group and 59 of the 112 patients (53%) in the PLACEBO treatment group died by six months after wafer implantation surgery (P = 0.061, Fisher's Exact Test). Cumulative mortality through six months after wafer implantation was compared using the Kaplan-Meier method and demonstrated lower cumulative mortality with GLIADEL (Log-Rank P = 0.063, Wilcoxon P = 0.077).

Prognostic Factors for Six-Month Mortality: All Patients

Using a univariate Cox regression, the following factors were found to be associated (P < 0.15) with increased six-month mortality: glioblastoma multiforme (GBM) tumor type, KPS score \leq 70, white race, < 75% tumor resection, advanced age, prior chemotherapy, years from first surgery to index surgery, MMSE < 26, and > 6 wafers implanted.

Six-Month Mortality Adjusted for Prognostic Factors: All Patients

After adjustment for prognostic factors, GLIADEL produced a statistically significant reduction in mortality compared to PLACEBO (risk ratio = 0.58; 95% CI: 0.39, 0.875; P = 0.009). Stratifying patients by tumor type and using the Cox regression model gave the same results (risk ratio = 0.058; 95% CI: 0.39, 0.86; P = 0.007).

Subgroup Analysis of Six-Month Mortality Rates: GBM and Non-GBM Patients

Among patients with GBM tumors, fewer GLIADEL-treated patients [32 of 72 (44%)] than PLACEBO-treated patients [47 of 73 (64%)] died by six months after surgery (P = 0.020, Fisher's Exact Test). Among patients with non-GBM tumors, a similar percentage of GLIADEL- and PLACEBO-treated patients died by six months after surgery (32% and 31%, respectively).

Controlling for tumor type using the Cochran-Mantel-Haenszel test showed that GLIADEL produced a statistically significant reduction in mortality during six months after wafer implantation (P = 0.052). There was no statistically significant treatment-by-tumor interaction (P = 0.153).

Compared to PLACEBO, GLIADEL produced a statistically significant decrease in cumulative mortality (Kaplan-Meier method) (Log-Rank Test: P = 0.013; Wilcoxon Test: P = 0.015) in patients with GBM, but had no statistically significant effect on cumulative mortality (Log-Rank Test: P = 0.849; Wilcoxon Test: P = 0.775) in patients with non-GBM tumors. The results of a Cox regression model containing treatment, tumor type, and their interaction, showed that GLIADEL reduced 6 month mortality (P = 0.04) but that there was no statistically significant treatment-by-tumor type interaction (P = 0.177).

Six-Month Mortality Adjusted for Prognostic Factors in GBM Patients

After adjustment for prognostic factors in a Cox regression model, GLIADEL produced a statistically significant reduction in cumulative mortality compared to PLACEBO in patients with GBM (risk ratio = 0.53, P = 0.005; 95% CI: 0.33, 0.825).

Analyses of Six-Month Mortality by Age, Gender, and Race

For six-month mortality, the results of the Cox regression model with treatment, gender, and their interaction as prognostic factors, showed that there were no statistically significant treatment-by-age, treatment-by-gender, or treatment-by-race interactions.

Secondary Efficacy Analyses

The median duration of survival through the 71 month post-surgery observation period was 7.24 months (95% CI: 6.05, 8.54 months) for GLIADEL and 5.42 months (95% CI: 4.73, 6.44 months) for PLACEBO treated patients (P = 0.297 Log Rank test, P = 0.106 Wilcoxon Rank Sum test). After adjustment for tumor type, Karnofsky score, location of radiation therapy, race, extent of resection, age, gender, prior chemotherapy, years from first surgery to index surgery, mini-mental state exam score, and number of wafers implanted, GLIADEL treatment reduced the risk of death compared to PLACEBO (risk ratio = 0.752; 95% CI: 0.57, 0.99; P = 0.045). Similar results were obtained if patients were stratified by tumor type (risk ratio = 0.70; 95% CI: 0.53, 0.94; P = 0.017).

No consistent differences between GLIADEL and PLACEBO treated patients were noted in Karnofsky or MMSE score changes over time.

Safety Results:

Patient Death or Trial Discontinuation:

One patient in the PLACEBO treatment group was lost to follow-up; all other study patients were followed to the time of their death or November 10, 1995. A total of 105 patients (95%) in the GLIADEL treatment group and 107 patients (96%) in the PLACEBO treatment group died during the follow-up period. None of the deaths was considered by the investigator to be related to GLIADEL or the PLACEBO. None of the study patients was discontinued from the trial for an adverse event.

Serious and Unexpected Adverse Events that Were Associated with Study Drug:

There were no treatment-emergent adverse events that were considered by the investigators to be serious, unexpected, and reasonably associated with the use of GLIADEL.

Adverse Events:

In the study, 100 of 110 (91%) patients in the GLIADEL treatment group and 100 of 112 (89%) patients in the PLACEBO treatment group experienced at least one treatment-emergent adverse event during the study period. A total of 1031 adverse events, of which 862 were treatment-emergent adverse events, were reported--428 in the GLIADEL treatment group and 434 in the PLACEBO treatment group. The most frequently reported adverse events were urinary tract infection (GLIADEL: 21%; PLACEBO: 17%) followed by hemiplegia (GLIADEL: 19%; PLACEBO: 20%), convulsion (GLIADEL: 19%; PLACEBO: 19%) and headache (GLIADEL: 15%; PLACEBO: 13%). Over 50% of all treatment-emergent adverse events were considered by the investigator to be moderate in severity and over 30% were considered to be severe. Most events were not considered to be related to study drug -- 238 of 428 (56%) in the GLIADEL treatment group and 252 of 434 (58%) in the PLACEBO treatment group. Seven treatment-emergent adverse events were considered to be probably related to wafer implantation by the investigator; 5 of 428 (1%) events in the GLIADEL treatment group and 2 of 434 (0.5%) events in the PLACEBO treatment group. The five events in the GLIADEL group were chemical meningitis, thrombocytopenia, neutropenia, right III nerve palsy, and headache. In the PLACEBO group hydrocephalitis and left hemiplegia were reported to be probably related.

Clinically Significant Adverse Events:

Abnormal Healing

Fifteen patients (14%) in the GLIADEL treatment group and six patients (5%) in the PLACEBO treatment group experienced abnormal healing (P = 0.040). The abnormal healing events included cerebrospinal

fluid leaks, localized fluid collections, wound dehiscence or poor healing, and subgaleal or wound effusions. All of these events were mild or moderate in severity, except one wound breakdown in the GLIADEL treatment group which occurred 175 days after surgery and was classified as severe. Three patients in the GLIADEL treatment group had documented abnormal healing events considered by the investigator to be possibly related to GLIADEL implantation. Two patients in the PLACEBO treatment group had abnormal healing considered by the investigator to be possibly related to study drug. For 6 of 15 (40%) abnormal healing events in the GLIADEL treatment group and 2 of 6 (33%) in the PLACEBO treatment group, the investigator considered the event to have no relationship to study drug. The relationship was unknown for the remaining abnormal healing events [6 of 15 (40%) in the GLIADEL treatment group and [2 of 6 (33%)] in the PLACEBO treatment group.

Convulsion (ALL Episodes in ALL Patients with Post-Baseline Convulsion)

The incidence of treatment-emergent convulsions was 19% in both the GLIADEL treatment group (21 of 110 patients) and the PLACEBO treatment group (21 of 112 patients) ($P = 1.000$). In the first five days after surgery there were 12 treatment-emergent convulsions in the GLIADEL group compared with 2 in the PLACEBO group ($P = 0.025$).

There were 41 patients in the GLIADEL treatment group and 32 patients in the PLACEBO treatment group with any post-baseline convulsion. In the first five days after surgery 15 convulsions occurred in the GLIADEL treatment group compared with 6 convulsions in the PLACEBO group ($P = 0.624$).

Infections

A total of 65 treatment-emergent infections occurred in the GLIADEL treatment group and 69 in the PLACEBO group ($P = 0.784$). There were no statistically significant differences between treatment groups in the types of infections reported.

The overall incidence of treatment-emergent meningitis was 4% (four patients) in the GLIADEL treatment group; no patient in the PLACEBO treatment group had meningitis ($P = 0.059$). The meningitis was bacterial in two cases, chemical in one case, and unspecified in one case. The investigator considered one case of meningitis to be probably related to GLIADEL, one possibly related to GLIADEL, and two cases were considered to have an unknown relationship to study medication. One patient was diagnosed as having meningitis on Study Day 3 and on Study Day 4, the GLIADEL wafers were surgically removed. One patient in the PLACEBO group developed a brain abscess which was diagnosed 76 days after wafer implant surgery. After the abscess was drained and antibiotics were given, the patient recovered.

Leukopenia and Thrombocytopenia

One patient in the GLIADEL treatment group developed leukopenia and thrombocytopenia after implant surgery. This patient received concomitant lomustine therapy starting 4 days after implant surgery.

Three other patients (one in the GLIADEL treatment group and two in the PLACEBO treatment group) had treatment-emergent thrombocytopenia. All three patients had been administered chemotherapy prior to GLIADEL or PLACEBO wafer implantation and had platelet counts that were low or below the lower limit of normal at Baseline.

Neurological Examinations:

Serial neurological examinations included vital signs, level of consciousness, personality change, speech disorder, motor involvement, sensory changes and cerebellar signs. There were no statistically significant differences between GLIADEL and PLACEBO in the changes from Baseline to final visit for any of these examinations. There were no consistent differences between GLIADEL and PLACEBO in the changes in these examinations during the study, although most of the scores for these examinations deteriorated significantly over time in both treatment groups.

Level of Consciousness Assessment:

There were no statistically significant between-group differences in the results of the level of consciousness scale at any study visit.

Clinical Laboratory Evaluations:

Seventy-one of 110 (65%) patients in the GLIADEL treatment group and 66 of 112 (59%) patients in the PLACEBO treatment group had at least one clinically notable, post-baseline laboratory abnormality (P = 0.410). The greatest numbers of clinically notable abnormal laboratory values were documented in hematological parameters -- principally, low values for erythrocyte-related parameters and high values for leukocyte counts. In total, 42 of 110 patients (39%) in the GLIADEL treatment group and 41 of 112 (37%) in the PLACEBO treatment group had clinically notable low hematocrit values. Clinically notable leukocytosis was documented post-baseline in 26 of 110 patients (24%) in the GLIADEL treatment group and 26 of 112 (23%) in the PLACEBO treatment group (P = 1.000).

Other clinically notable abnormalities were less common and included low WBC count and neutrophil percentage, high eosinophil percentage, high BUN (but not creatinine), elevations of liver enzymes (SGPT, SGOT, and alkaline phosphatase), high LDH, high uric acid and abnormalities seen on urinalysis (including high urinary glucose, high urinary protein, and increased numbers of casts). For most of these parameters, there were no statistically significant between-group differences in the incidence of clinically notable abnormalities.

The changes from Baseline in laboratory parameters (usually transient and maximal on the Day of Surgery) were consistent with the changes frequently seen after major surgery (and its attendant blood loss), or treatment with high doses of corticosteroids.

Conclusions:

GLIADEL reduced six month mortality in patients who underwent reoperation for recurrent malignant glioma. The association of GLIADEL treatment with improved survival was strong and persistent after adjustment for the effect of Baseline prognostic factors. The use of GLIADEL was associated with an increased frequency of convulsions in the immediate post-operative period (0-5 days) but did not cause the serious adverse effects frequently seen with systemic BCNU. Thus, the risk/benefit ratio for GLIADEL in this study was large, and favors the use of GLIADEL as palliative therapy in patients undergoing reoperation for recurrent malignant glioma.

Interstitial Chemotherapy with Carmustine-loaded Polymers for High-grade Gliomas: A Randomized Double-blind Study

Simo Valtonen, M.D., Ulla Timonen, M.D.,
 Petri Toivanen, M.Sc., Hannu Kalimo, M.D.,
 Leena Kivipelto, M.D., Olli Heiskanen, M.D.Prof.,
 Geirmund Unsgaard, M.D.Prof., Timo Kuurne, M.D.

Departments of Neurosurgery (SV) and Pathology (HK), Turku University Central Hospital, Turku, Finland; Department of Neurosurgery (LK, OH), Helsinki University Central Hospital, Helsinki, Finland; Department of Neurosurgery (TK), Tampere University Hospital, Tampere, Finland; Department of Neurosurgery (GU), University Hospital of Trondheim, Trondheim, Norway; and Orion Pharma (UT, PT), Espoo, Finland

OBJECTIVE: To find out the effect of carmustine (bischloroethyl-nitrosourea) combined with a biodegradable polymer in the treatment of malignant (Grades III and IV) gliomas, applied locally, at the time of the primary operation.

METHODS: Prospective, randomized double-blind study of an active treatment group versus a placebo group. Conducted at the Departments of Neurosurgery of the University Hospitals of Helsinki, Tampere, and Turku in Finland and Trondheim in Norway. The study consisted of 32 patients (16 in each treatment group) enrolled between March 23, 1992, and March 19, 1993. The study was planned to include 100 patients but had to be terminated prematurely, because the drug that was being used had become unobtainable. The main outcome measures included the survival times of patients after the operations and the application of an active drug or placebo.

RESULTS: The median time from surgery to death was 58.1 weeks for the active treatment group versus 39.9 weeks for the placebo group ($P = 0.012$). For 27 patients with Grade IV tumors, the corresponding times were 39.9 weeks for the placebo group and 53.3 weeks for the active treatment group ($P = 0.008$). At the end of the study, six patients were still alive, five of whom belonged to the active treatment group.

CONCLUSION: Carmustine applied locally in a biodegradable polymer at the time of primary operation, seems to have a favorable effect on the life span of patients with high-grade gliomas. (Neurosurgery 41:44-49, 1997)

Key words: BCNU, Biodegradable polymer, Chemotherapy, Malignant glioma

Chemotherapy has not fulfilled its early promise in the treatment of high-grade gliomas. This is partly because the drugs that are used do not have a significant effect on the survival of patients with high-grade gliomas and partly because their systemic use is associated with a high degree of adverse reactions and requires long hospitalization (11).

After diagnosis and treatment with tumor resection and external radiotherapy, the median survival time of patients with high-grade gliomas is still less than 1 year (1). Because a better method of further treatment is currently not available (9), local application of cytostatic drugs has been in-

troduced as a means to improve treatment results (3, 4, 6, 12). Infusion pumps, catheters, Ommaya reservoirs, and other drug delivery systems have been studied as methods for the administration of cytostatic agents, primarily carmustine (bischloroethyl-nitrosourea [BCNU]), to the target site (12).

BCNU, when combined with a biodegradable polymer wafer and implanted in the tumor resection cavity during surgery, has been observed to be an easy, safe, and clinically effective method of local application in the treatment of patients with recurrent gliomas (2-4). The polymer wafer releases the drug slowly during a period of approximately 2 weeks, without remarkable adverse reactions (2, 10). Because

previous studies have been conducted in patients with recurrent gliomas (3, 4), no previous results of this type of therapy are available, in which the drug is applied at the first operation of a high-grade glioma. However, the safety of BCNU polymer implants used at the time of the first surgery has been established in a Phase I study (8).

This placebo-controlled double-blind study was designed to evaluate the safety and efficacy of the BCNU wafer (Gliadel; Guilford Pharmaceuticals Inc., Baltimore, MD) placed in the tumor resection cavity at the time of the first surgery. The study had to be terminated prematurely, because the drug was temporarily not available.

MATERIALS AND METHODS

Patient selection

Patients were enrolled in the study between March 23, 1992, and March 16, 1993. Three Finnish hospitals (Turku University Central Hospital, Helsinki University Central Hospital, and Tampere University Hospital) and one Norwegian hospital (University Hospital of Trondheim) participated in the study. According to the protocol, 100 patients were to be enrolled in the study. The protocol was submitted to the ethical committees at each study center and approved. The national regulatory authorities were informed about the study in accordance with the national regulations.

Patients included in the study met the following criteria: unilateral, unifocal intrinsic brain tumor not crossing the midline of at least 1.0 cm in diameter as determined by computed tomography or magnetic resonance imaging (MRI), 18 to 65 years of age, a Karnofsky Performance Score (KPS) of 60 or higher (indicating ability to function independently), ability to provide witnessed informed consent before surgery, and a histopathological diagnosis of high-grade glioma (Grade III or IV) on a frozen section during surgery.

If any of the five following exclusion criteria were met, the patient was excluded from the study: 1) evidence of significant renal or hepatic disease as determined by blood urea nitrogen, creatinine, serum glutamic-oxaloacetic transaminase (aspartate aminotransferase), serum glutamic-pyruvic transaminase (alanine aminotransferase), lactate dehydrogenase, or bilirubin levels two times higher than the upper limit of normal value; 2) other concomitant life-threatening disease; 3) fewer than 100×10^9 circulating platelets per liter or fewer than 4.0×10^9 leukocytes per liter; 4) pregnancy; and 5) hypersensitivity to contrast media used in computed tomographic and MRI studies.

Patient evaluation

Each patient underwent a thorough examination before surgery. These included medical history, physical examination, KPS determination, neurological examination, mini-

mental state examination, tumor imaging by computed tomography or MRI, and laboratory examinations.

Treatment and study groups

Patients received either Gliadel, the polyanhydride used, which contained 3.85% BCNU by weight, or a placebo. Eight wafers were available for each patient. Each BCNU wafer contained 7.7 mg BCNU, the maximal dose being 61.6 mg of BCNU.

The patients were randomized to receive either BCNU or placebo wafers. The randomization was conducted in blocks of four patients (two patients in the active group and two in the placebo group, in random order). The study was kept blinded for 2 years after the last patient was entered.

All patients underwent resection of the tumor mass. An intraoperative sample was sent to the local neuropathologist for confirmation of the diagnosis of high-grade glioma. Additional samples of the resected tumor were immediately fixed in buffered formaldehyde for routine paraffin preparations. The specimens were later reviewed by a referred pathologist. After maximal tumor resection was accomplished, meticulous hemostasis was achieved, and as many wafers as the space allowed were placed over the resection surface. Materials such as absorbable gelatin sponge were occasionally used to cover the polymers and keep them in place on the brain surface. The decompression cavity was filled with irrigation fluid, and the dura was closed in a normal fashion.

All patients underwent standard radiotherapy. The median cumulative dose was 54.03 Gy for the placebo group and 54.92 Gy for the group receiving Gliadel. Because of poor condition, one patient in the group receiving Gliadel received no radiotherapy. All patients were treated with perioperative corticosteroids to reduce brain swelling. Subsequent operations were allowed if considered necessary.

Follow-up studies

The patients underwent KPS determination, neurological examination, mini-mental state examination, tumor imaging by computed tomography or MRI, and laboratory examinations before discharge and at three monthly intervals until death or for up to 2 years.

Statistical methods

Time from surgery to death was analyzed as the primary end point of the study. First, the unadjusted association between treatment and time was assessed using the log-rank test. Furthermore, medians and respective 95% confidence intervals were calculated for both treatment groups, and Kaplan-Meier (7) estimates of survival function are presented. These analyses were performed for all 32 patients (Intention To Treat analysis) and also for the subgroup consisting of 27 Grade IV patients.

Second, to adjust for covariates that may have an impact on survival time, a sequence of Cox's proportional hazards models were fitted. In addition to treatment, the following covariates were taken into consideration: patient's age, sex, KPS, tumor size, tumor type, and total cumulative dose of radiotherapy received. The strategy adopted for model selection was similar to the strategy presented by Collett (5), except that treatment was always retained in the model. Moreover, hazard ratios and respective 95% confidence intervals were calculated for treatment, and each covariate was included in the model. These analyses were performed both for all patients (Intention To Treat analysis) and for the subgroup formed by Grade IV patients. The adequacy of the assumptions inherent in the Cox model was assessed visually by the examination of various plots (e.g., the proportional hazards assumption was checked by plotting the logarithm of cumulative hazard against logarithm; martingale residuals of the model were plotted). All statistical analyses were performed using Statistical Analysis System (SAS) software (Cary, NC).

RESULTS

Patients

Thirty-two patients were enrolled in the study (nine patients from Helsinki, Trondheim, and Turku each, and five from Tampere), and they all completed it. The study had to be terminated prematurely, because the manufacturer of the drug was not able to deliver more of the product. There were no scientific reasons for the premature termination. The differences between the study groups were also not the cause for termination, because the study was still blinded at this stage.

Regarding age, sex, and tumor size, the groups were well matched (Table 1). There was a slight difference in KPS in favor of the placebo group. Two patients in the group receiving Gliadel received 38.5 and 48.2 mg, respectively, which was less than the maximal amount of drug (61.6 mg).

TABLE 1. Baseline Characteristics of the Two Study Groups*

	Treatment	
	Placebo	Gliadel
Sex (n)		
Male	6	8
Female	10	8
Age (yr)		
Median	53.0	55.5
Range	(36-65)	(36-67)
Tumor size (mg)		
Median	20	20
Range	(6.25-28.0)	(12.0-38.5)
Karnofsky Performance Score		
Median	90	75
Range	(40-100)	(60-100)

* Tumor size refers to the maximal planar area of the tumor in the computed tomographic or magnetic resonance imaging study.

Pathological findings

The original pathological examinations revealed that 15 patients in the placebo group developed glioblastomas and that 1 patient developed an anaplastic astrocytoma. In the group receiving Gliadel, 11 patients sustained glioblastomas, 2 sustained anaplastic astrocytomas, 2 sustained malignant oligodendroglioma, and 1 sustained malignant ependymoma. The review by the referee pathologist indicated a diagnosis of glioblastoma for all 16 patients in the placebo group (1 patient who was originally diagnosed with anaplastic astrocytoma was confirmed to have sustained glioblastoma). In the group receiving Gliadel, all the original diagnoses were confirmed.

Efficacy

The time from surgery to death was analyzed as the primary efficacy variable of the study (see Table 2 and Fig. 1). Of the 32 patients (16 in each treatment group), 6 were alive at the end of the follow-up period of 2 years (104 wk). Five of the patients had been assigned to the group receiving Gliadel and one patient to the placebo group. Thus, the survival times of six patients were recorded at 104 weeks. A *P* value of 0.012 between the treatment groups suggests that treatment had an effect on survival time. Six covariates in addition to the treatment were chosen as possible explanatory variables, as follows: sex, age, KPS, tumor type, tumor size, and total cumulative dose of radiotherapy received. Results for age, KPS, and tumor type are presented in Table 3. All of these factors are significant for the outcome, as was the mini-mental score (*P* = 0.016), but do not explain the risk ratio of 0.269 in favor of Gliadel treatment versus placebo.

Three years after termination of the study, five patients were still alive; four of the patients belonged to the group receiving Gliadel, and one belonged to the placebo group.

Because there were no patients with tumor types other than Grade IV glioma in the placebo group, estimations of the effect based on tumor type varies. For this reason, a subgroup analysis of 27 patients with glioblastomas was performed (Table 4 and Fig. 2). The model used was the same as described above. The median time from surgery to death was 39.9 weeks for the placebo group and 53.3 weeks for the group receiving Gliadel (*P* = 0.008). To date, the significance of the treatment effect cannot be weakened by introducing other covariates. The estimate of the hazard ratio was 0.280, with 95% confidence limits (Table 5).

TABLE 2. Median and 95% Confidence Interval for Time from Surgery to Death in Weeks for the Two Treatment Groups*

Treatment	Median	Lower 95% CI Limit	Upper 95% CI Limit
Placebo	39.9	37.6	45.0
Gliadel	58.1	42.0	-

* CI, confidence interval; -, cannot be estimated from the data. The difference between the two treatment groups is significant (*P* = 0.012). All 32 patients are included, with both Grade III and Grade IV tumors, but the placebo group has no patients with Grade III tumors.

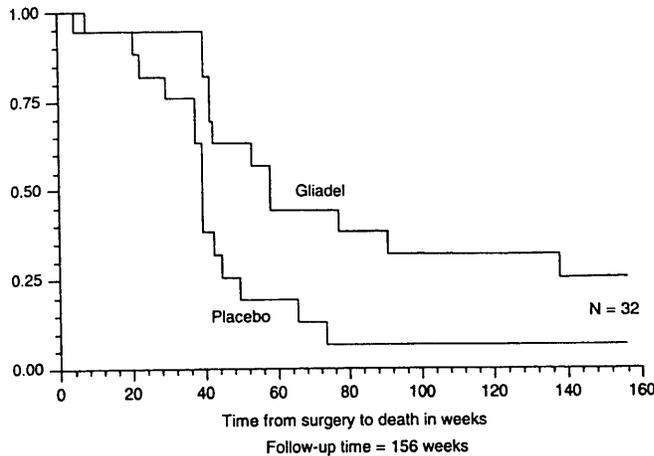


FIGURE 1. Kaplan-Meier survival curves for the both study groups, including patients both with Grade III and Grade IV tumors. All patients with Grade III tumors were included in the group receiving Gliadel.

TABLE 3. Estimates of the Cox Model for Time from Surgery to Death^a

Variable	Hazard Ratio	Lower 95% CI Limit	Upper 95% CI Limit	P Value
Treatment	0.27	0.11	0.68	0.006
KPS	0.96	0.93	0.99	0.010
Age (yr)	1.09	1.02	1.15	0.007
Tumor type	5.62	0.69	46.05	0.108

^a CI, confidence interval; KPS, Karnofsky Performance Score. Intention To Treat analysis (all patients included in the analysis, placebo group has no patients with Grade III tumors).

TABLE 4. Median and 95% Confidence Interval for Time from Surgery to Death in Weeks for Both Treatment Groups^a

Treatment	Median	Lower 95% CI Limit	Upper 95% CI Limit
Placebo	39.9	37.6	45.0
Gliadel	53.3	40.1	77.7

^a CI, confidence interval. Patients with only Grade IV tumors (n = 27) are included. The difference between the groups is significant.

Adverse events and complications

There were no deaths in the perioperative period. The total number of patients with adverse events during the study was 21 (9 in the placebo group and 12 in the group receiving Gliadel). Fifteen serious and unexpected adverse events were reported by nine patients. The group receiving Gliadel included 10 serious adverse events reported in five patients. These included wound infection, septic inflammation with meningismus, cerebrospinal fluid leukocytosis with hydrocephalus, deep venous thrombosis with pulmonary embolism, pneumonia with an increase in aphasia, visual disturbances, and hemiparesis. In the placebo group, five serious adverse events were reported in four patients, including pulmonary embolism, meningitis, wound infection, and deep

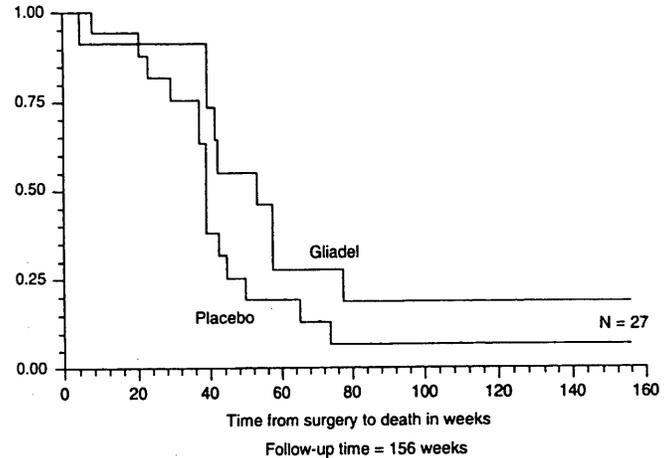


FIGURE 2. Kaplan-Meier survival curves for patients with Grade IV tumors.

TABLE 5. Estimates of the Cox Model for Time from Surgery to Death^a

Variable	Hazard Ratio	Lower 95% CI Limit	Upper 95% CI Limit	P Value
Treatment	0.27	0.10	0.71	0.008
KPS	0.96	0.93	0.99	0.019
Age (yr)	1.08	1.01	1.14	0.018

^a CI, confidence interval; KPS, Karnofsky Performance Score. Subgroup analysis for patients with Grade IV tumors (n = 27).

venous thrombosis with pulmonary embolism. The most frequently documented treatment-emergent adverse events in the group receiving Gliadel included hemiparesis (38%), convulsion (19%), aphasia (13%), and visual field defect (13%). In the placebo group, the most frequently reported treatment-emergent adverse events included hemiparesis (25%) and convulsions (13%). The treatment-emergent adverse events experienced were consistent with those expected in postoperative patients with malignant gliomas. No significant changes in blood chemistry or urinalysis were detected in the follow-up examinations. One patient in the group receiving Gliadel underwent subsequent surgery.

DISCUSSION

The effect of cytostatic drugs on the outcome of patients with high-grade gliomas was not overwhelming (8, 11). Their usefulness has also been limited by adverse reactions and the long hospitalization required. Because of these factors and based on the results of our study and previous investigations (3, 4), the local application of BCNU seems to have clear advantages compared with the traditional methods.

Gliadel-treated patients in our study had a longer median survival time than the patients in the placebo group. The difference is statistically significant, even if the number of patients in the study remained small (smaller than planned because of unforeseen circumstances). The difference of median survival times is perhaps not as significant as the difference in the number of patients surviving 2 years.

There was a bias in our study because of the lack of Grade IV tumors in the placebo group. Consequently, the significance of the difference regarding the life span between the two study groups is diminished. However, the difference between the study groups including only patients with Grade IV tumors is significant, although to a lesser degree.

An advantage of the local application of cytostatic drugs is their ease of use; the procedure itself only takes a few minutes, and the time of hospitalization is not prolonged. Regarding patients' comfort and hospital costs, this compares very favorably with traditional methods of drug administration. The administration of the drugs did not have any negative effect on the patients' quality of life.

The local application of BCNU combined with a biodegradable polymer is not associated with any additional drug-related adverse events, either in previous studies (4) or in our study. In our study, we used BCNU because it is the only cytostatic drug available in combination with a biodegradable polymer. Although other cytostatic drugs may have superior efficacy compared to BCNU, it has not been demonstrated to date (8).

Our high number of infectious complications can be explained. In one of the study centers, the instructions about the sterility of wafer packages were misunderstood and nonsterile packages were thought to be sterile. This probably contributed to two of the four infections recorded, which had no other apparent cause, and possibly to one other infection (the patient also had a cerebrospinal fluid leak through frontal sinus). Because these complications were evenly distributed between the study groups, they cannot be attributed to the active drug.

In conclusion, the administration of BCNU at first surgery to patients with high-grade gliomas seems to be beneficial. To fully establish the value of the therapy will require further studies with a higher number of patients and perhaps a randomized study comparing this type of drug administration with the traditional intravenous one.

ACKNOWLEDGMENTS

Both Gliadel and placebo were originally supplied by Nova Company, and the study was financially supported by the Orion Pharma Company.

Received, April 25, 1996.

Accepted, January 8, 1997.

Reprint requests: Simo Valtonen, M.D., Department of Neurosurgery, Turku University Central Hospital, 20520 Turku, Finland.

REFERENCES

1. Black PM: Brain tumors. *N Engl J Med* 324:1471-1476, 1991.
2. Brem H, Ewend MG, Piantadosi S, Greenhoot J, Burger PC, Sisti M: The safety of interstitial chemotherapy with BCNU-loaded polymer in the treatment of newly diagnosed malignant gliomas. *J Neurooncol* 26:111-123, 1995.
3. Brem H, Mahaley MS Jr, Vick NA, Black KL, Schold SC Jr, Ciric IS, Eller TW, Cozzens JW, Kenealy JN: Interstitial chemotherapy with drug polymer implants for the treatment of recurrent gliomas. *J Neurosurg* 74:441-446, 1991.
4. Brem H, Piantadosi S, Burger PC, Walker M, Selker R, Vick NA, Black K, Sisti M, Brem S, Mohr G, Muller P, Morawetz R, Schold SC, for the Polymer-Brain Tumor Treatment Group: Placebo-controlled trial of safety and efficacy of intraoperative controlled delivery by biodegradable polymers of chemotherapy for recurrent gliomas. *Lancet* 345:1008-1012, 1995.
5. Collett D: *Modelling Survival Data in Medical Research*. London, Chapman & Hall, 1994, pp 80-81.
6. Garfield J, Dyan AD: Postoperative intracavitary chemotherapy of malignant gliomas: A preliminary study using methotrexate. *J Neurosurg* 39:315-322, 1973.
7. Kaplan EL, Meier P: Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 53:457-481, 1958.
8. Kornblith PL, Walker M: Chemotherapy for malignant gliomas. *J Neurosurg* 68:1-17, 1988.
9. Lesser GJ, Grossman SA: The chemotherapy of adult brain tumors. *Cancer Treat Rev* 19:261-281, 1993.
10. Tamargo RJ, Myseros JS, Epstein JI, Yang MB, Chasin M, Brem H: Interstitial chemotherapy of the 9L gliosarcoma: Controlled release polymers for drug delivery in the brain. *Cancer Res* 53:329-333, 1993.
11. Walker MD, Green SB, Byar DP, Alexander E Jr, Batzdorf U, Brooks WH, Hunt WE, MacCarty CS, Mahaley MS Jr, Mealey J Jr, Owens G, Ransohoff J II, Robertson JT, Shapiro WR, Smith KR Jr, Wilson CB, Strike TA: Randomized comparisons of radiotherapy and nitrosoureas for the treatment of malignant glioma after surgery. *N Engl J Med* 303:1323-1329, 1980.
12. Walter KA, Tamargo RJ, Olivi A, Burger PC, Brem H: Intratumoral chemotherapy. *Neurosurgery* 37:1129-1145, 1995.

COMMENTS

This prospective randomized placebo-controlled study of the effectiveness of bischloroethyl-nitrosourea (BCNU) polymers as the initial therapy for malignant gliomas is an important addition to our understanding of the role of this new therapeutic modality. This study is based on the foundation of previous prospective multi-institutional studies demonstrating the safety of the BCNU polyanhydride polymers when implanted during the initial operation for malignant brain tumors (1), as well as during subsequent operations (2). We completed a prospective randomized placebo-controlled study involving 222 patients in 27 hospitals in North America who underwent polymer implantation for recurrent gliomas (3). These studies demonstrated that a chemotherapeutic drug, i.e., BCNU, could be safely and effectively administered directly to the tumor bed at the time of craniotomy, leading to a statistically significant prolongation of survival.

Using a similar study design, the authors have independently addressed the issue of whether the BCNU polymers were safe and effective as the initial treatment for malignant gliomas. Our previous laboratory work had suggested that the use of the BCNU polymer was optimal as the initial therapy. The study of the rat 9L glioma demonstrated a 30% long-term survival with Gliadel and no long-term survivors in the control rats treated with systemic BCNU (4). This figure is remarkably similar to the clinical findings reported by Valtonen et al., in which the 2-year survival for malignant glioma is 30% as compared to 6% in the placebo-controlled polymer patients. The small number of patients in the study does not diminish the quality of the study design or the

striking finding of prolonged survival. Although the authors planned a study involving 100 patients, it was unfortunately stopped after 32 patients for administrative and funding reasons and not for scientific reasons nor because of emerging treatment differences. Nonetheless, the study was kept blinded for the full 2 years after entry of the last patient. The 58-week median survival and 30% 2-year survival is a significant improvement compared with the randomized control group. The significance is reinforced by the authors' proportional hazards regression analysis and the analysis of patients with glioblastomas as a "separate group."

Of additional significance, the study by Valtonen et al. was analyzed by the Food and Drug Administration and presented to the Oncology Drug Advisory Committee on June 14, 1996, which recommended approval for the use in recurrent glioblastoma. This is the first approval in 22 years by the Food and Drug Administration of a new treatment for brain tumors.

I think that the BCNU polymer implants will serve as a "proof of principle" that controlled delivery of chemotherapy using biodegradable polymers is a valuable addition to the neurosurgical armamentarium. As newer treatments are brought from the laboratory to the clinic (5), I hope that they will be tested in as rigorous a fashion as by Valtonen et al.

Henry Brem
Baltimore, Maryland

1. Brem H, Ewend MG, Piantadosi S, Greenhoot J, Burger PC, Sisti M: The safety of interstitial chemotherapy with BCNU-loaded polymer followed by radiation therapy in the treatment of newly diagnosed malignant gliomas: Phase I trial. *J Neurooncol* 26:111-123, 1995.
2. Brem H, Mahaley MS Jr, Vick NA, Black KL, Schold SC Jr, Burger PC, Friedman AH, Ciric IS, Eller TW, Cozzens JW, Kenealy JN: Interstitial chemotherapy with drug polymer implants for the treatment of recurrent gliomas. *J Neurosurg* 74:441-446, 1991.
3. Brem H, Piantadosi S, Burger PC, Walker M, Selker R, Vick NA, Black K, Sisti M, Brem S, Mohr G, Muller P, Morawetz R, Schold SC, and the Polymer-Brain Tumor Group: Placebo-controlled trial of safety and efficacy of intraoperative controlled delivery by biodegradable polymers of chemotherapy for recurrent gliomas. *Lancet* 345:1008-1012, 1995.
4. Tamargo RJ, Myseros JS, Epstein JI, Yang MB, Chasin M, Brem H: Interstitial chemotherapy of the 9L gliosarcoma: Controlled release polymers for drug delivery in the brain. *Cancer Res* 53:329-333, 1993.
5. Walter KA, Tamargo RJ, Olivi A, Burger PC, Brem H. Intratumoral chemotherapy. *Neurosurgery* 37:1129-1145, 1995.

The goal of achieving a high local or regional dosage level of chemotherapy for a sustained interval is clearly desirable. Many approaches have been used with limited success. The concept of using wafers "impregnated" with chemotherapy provides an appealing route for local sustained delivery.

The results of this small initial trial are of interest. The data suggest that there is a prolongation of life as a result of the BCNU wafer placement.

The study has some serious limitations. The size is small. After a review of pathological findings, the grades are not

perfectly matched in the groups, and I have a concern about the relatively high rates of postoperative hemiplegia in both groups of patients. If this high rate reflects more aggressive tumor resection, this needs to be taken into account in interpreting the results.

In general, the concept of direct and sustained delivery of chemotherapy into a malignant glioma is very appealing. Although the results of this study are intriguing, they are not conclusive evidence of efficacy; however, the results may suggest that there would be value in a much larger, tighter, and better controlled randomized prospective trial.

Paul L. Kornblith
Pittsburgh, Pennsylvania

Valtonen et al. present the results of what was intended to be a larger placebo controlled Phase III study, which had to be stopped prematurely because of the unavailability of the study drug. Gliadel or placebo was placed in the surgical cavity during the first surgical resection of patients thought to have high-grade malignant tumors. After surgery and interstitial chemotherapy (or placebo), patients were treated with radiotherapy. The primary study end point was survival. A planned sample size of 100 patients was not achieved, and 32 patients were enrolled in four treatment centers. At final review of neuropathological findings, an imbalance in the placebo and treatment groups was observed, demonstrating more patients without glioblastomas in the treatment group. For the purposes of this analysis, the only valid study group is the 27 cases of centrally reviewed glioblastoma cases (16 in the placebo group, 11 in the treatment group). Three years after the study terminated, four patients with glioblastomas were alive (three in the treatment group, and one in the placebo group). Median survival for the placebo glioblastoma group was 39.9 weeks and for the treatment group was 53.3 weeks. Toxicity seemed to be greater in the group receiving Gliadel, with 10 serious adverse events in five patients versus 5 serious adverse events in four patients in the placebo group. A possible explanation for the toxicity is the use of nonsterile wafer packages at one of the study centers. The study population was fairly evenly matched, although at least one patient in the placebo group had an ineligible Karnofsky Performance Score of 40. The authors conclude that Gliadel improves survival in patients newly diagnosed with glioblastomas and other malignant tumors. This conclusion is premature and needs to be tested in an appropriately sized clinical trial. The current study results are really based on a very small number of cases, and the confidence intervals for the median survival estimates have a wide range. Although a small variability in patient selection in such a small patient group can greatly influence such results, a larger study with appropriate stratifications will be necessary to confirm the conclusion that this approach is superior to control groups. Appropriately, the authors agree that a larger study is needed to compare with the results observed with intravenous BCNU.

Michael D. Prados
Neuro-oncologist
San Francisco, California

STUDY SYNOPSIS

Compound:

A biodegradable, polyanhydride copolymer consisting of polycarboxyphenoxypropane (PCPP) and sebacic acid (SA) in a 20:80 molar ratio with 3.85% (by weight) 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU) incorporated into the polymer matrix.

Names

Generic name	Polifeprosan 20 with carmustine
Tradename	GLIADEL [®] Wafer

Report Type:

Integrated Clinical and Statistical Report

Study Phase and IND Number:

Phase III Clinical Trial conducted in Finland and Norway. Serious Adverse Event data have been submitted to IND #30,237.

Protocol Number:

F-GLI-CL-0190/Interstitial Chemotherapy for Malignant Glioma: a Phase III Placebo Controlled Study to Examine the Safety and Efficacy of GLIADEL[™] Placed at the Time of First Surgery

Protocol Review:

This study follows the recommendations for biomedical research involving human subjects (Declaration of Helsinki of the World Medical Association 1964 and Venice Revision 1983 and Hong Kong Revision 1989). Prior to the initiation of the study, the protocol and the informed consent form were submitted to and approved by the Ethics Committees of the study sites.

Dates of Study:

Date first patient enrolled - March 23, 1992

Date last patient enrolled - May 14, 1993

Date of last observation - May 14, 1995

Objective of Study:

To determine the safety and efficacy of using GLIADEL as adjunctive treatment with surgery and external beam radiotherapy in patients newly diagnosed with malignant glioma.

Number of Patients:

Thirty-two patients, 18 years of age and older with pathologically confirmed diagnosis of malignant glioma were enrolled into Study F-GLI-CL-0190. All patients received GLIADEL or PLACEBO wafers and were evaluable for safety and efficacy analyses.

Number of Clinical Centers:

Multicenter - 4 sites (3 Finnish, 1 Norwegian)

Study Design:

This was a multicenter randomized, double-blind, placebo-controlled, parallel group, Phase III study. Patients were enrolled into the study after pathological examination of the tumor during surgery established the presence of malignant glioma. After maximum tumor resection, the surgeon placed up to eight GLIADEL wafers, each containing approximately 7.7 mg BCNU or eight PLACEBO wafers, into the resection cavity. About three weeks after surgery, standard radiotherapy began. Patients were evaluated periodically for up to two years by neurological examination, Karnofsky Performance Status Score (KPS) evaluation, Mini-Mental State Examination (MMSE), clinical laboratory tests and computed tomography (CT) or magnetic resonance imaging (MRI) scan to measure time to treatment failure. The primary efficacy endpoints were 12 month survival rates, median survival duration, and time to treatment failure.

Indication:

Patients 18 to 65 years of age with initially diagnosed, malignant glioma, and without prior surgical, radiotherapeutic, or chemotherapeutic treatment.

Investigational New Drug and Dosage:

Each GLIADEL wafer contains approximately 7.7 mg (3.85% by weight) BCNU. Each PLACEBO wafer contained 200 mg of the polymer with no BCNU. Up to eight wafers may be implanted into each patient. The number of GLIADEL wafers placed in the tumor cavity varied depending on the exposed resection surface area.

Duration of Administration:

GLIADEL was administered only once for the treatment of newly diagnosed malignant glioma in this study. Patients were enrolled into the study for surgical implantation of up to eight GLIADEL wafers upon pathological confirmation of malignancy.

Total Study Period:

Approximately two to three weeks after tumor removal and wafer implantation, standard external beam radiotherapy began. Patients were followed for up to two years after wafer implantation.

Primary Efficacy Results:

All patients enrolled into the study and treated within the protocol were evaluated for efficacy. In total, 32 patients (16 GLIADEL and 16 PLACEBO) contributed to the efficacy assessments.

Survival and Time to Treatment Failure (All Patients):

The effectiveness of GLIADEL in the treatment of initially diagnosed malignant glioma was demonstrated by the statistically significant improvement in one-year survival rate compared to PLACEBO and the statistically significantly improved survival over the 12- and 24-month period after implant surgery in the GLIADEL treatment group when compared to PLACEBO. Statistically significantly more patients who were implanted with GLIADEL wafers survived to one year post-surgery. Ten of 16 GLIADEL patients (63%) compared to 3 of 16 PLACEBO patients (19%) survived to one year (52 weeks) ($P = 0.029$). Overall, 11 of 16 (69%) GLIADEL patients and 15 of 16 PLACEBO patients (94%) died during the two year study conduct. The overall median survival durations were 58.1 weeks (95% CI: 42.0 - inestimable) and 39.9 weeks (95% CI: 37.6 to 45.0 weeks) ($P = 0.011$) for GLIADEL and PLACEBO group patients, respectively. The results of the Log-Rank and Wilcoxon tests show that there were significant between-group differences in the effect on survival during both the 12-month interval after study surgery ($P = 0.0087$ and $P = 0.0105$, respectively) and the up-to-24-month interval after study surgery ($P = 0.0116$ and $P = 0.0106$, respectively).

When the data were adjusted for important prognostic factors (KPS and age), whether stratified by tumor type or not, a significant GLIADEL treatment effect was observed. For the 12-month period after study surgery the adjusted risk ratios for GLIADEL vs. PLACEBO treatment were 0.154 (95% CI: 0.051 - 0.467) for all patients by nonstratified analysis ($P = 0.0044$) and 0.179 (95% CI: 0.056 - 0.574) for all patients stratified by tumor type ($P = 0.0059$). For the 24-month (overall) period after study surgery the adjusted risk ratios for GLIADEL vs. PLACEBO treatment were 0.177 (95% CI: 0.067 - 0.468) for all patients by nonstratified analysis ($P = 0.0005$) and 0.214 (95% CI: 0.078 - 0.590) for all patients stratified by tumor type ($P = 0.0029$).

Twelve patients (75%) in the GLIADEL treatment group and 14 patients (88%) in the PLACEBO treatment group were considered to be treatment failures. The median time to treatment failure was 1.12 months (7.79 months vs. 6.67 months; log rank $P = 0.4668$ and Wilcoxon p -value = 0.9635).

Survival and Time to Treatment Failure (Glioblastoma Multiforme Patients):

In this study, 27 patients had a diagnosis of Glioblastoma multiforme (GBM) [11 of 16 patients (69%) in the GLIADEL treatment group and 16 of 16 patients (100%) in the PLACEBO treatment group]. Six of 11 GBM patients (55%) in the GLIADEL treatment group and 3 of 16 GBM patients (19%) in the PLACEBO treatment group survived to one year ($P = 0.097$). In the GLIADEL group the median post implantation

survival duration for GBM patients was 53.3 weeks (95% CI: 40.1 - 77.7 weeks) compared with 39.9 weeks (95% CI: 37.6 - 45.0 weeks) in the PLACEBO treatment group ($P = 0.093$) for overall survival. Overall, 9 of 11 (82%) patients with GBM in the GLIADEL group died compared with 15 of 16 (94%) patients with GBM in the PLACEBO group. After adjustment for prognostic factors, GLIADEL produced a statistically significant reduction in mortality relative to PLACEBO in GBM patients for both the 12- and 24-month periods after wafer implantation surgery. The adjusted risk ratios were 0.196 (95% CI: 0.060 to 0.642) for 12 months and 0.213 (95% CI: 0.076 to 0.601) for 24 months, with P values of 0.0072 and 0.0035, respectively.

Other Results:

Changes in KPS Scores:

Among patients in both treatment groups, the mean KPS Scores declined from Baseline [GLIADEL 79 (± 14) and PLACEBO 82 (± 15)] to the Final Visit [GLIADEL 52 (± 30) and PLACEBO 43 (± 24)]. The mean change from Baseline to the Final Visit [-27 (± 29) in the GLIADEL group, and -40 (± 27) in the PLACEBO group] was not statistically significant in between-treatment-group comparisons.

Using the Observed Cases (OC) method of the Generalized Estimating Equations (GEE) analyses for both the continuous outcome (mean values over time) and the categorical outcome (treatment frequencies by visit for patients with worsening from Baseline), the results of overall tests for both treatment effect and treatment-by-visit interaction effects were not statistically significant. The results of the overall tests for visit effect were statistically significant for the continuous outcome ($P=0.010$ for mean values over time) and there was a trend toward statistical significance for the categorical outcome ($P = 0.056$ for categorical analysis of worsening over time). There was no statistically significant overall treatment-by-visit interaction for either analysis.

Longitudinal assessments of both continuous and categorical variables evaluated by a last observation carried forward (LOCF) method of GEE analyses showed a statistically significant result in testing for overall visit effect ($P \leq 0.001$), but not for treatment effect or overall treatment-visit interaction.

Changes in MMSE Scores:

Among patients in both treatment groups, the mean MMSE Scores declined from Baseline to Final Visit for all parameters. The mean total score worsened by 6.1 (± 9.7) points in GLIADEL patients and by 4.9 (± 5.7) points in PLACEBO patients ($P = 0.683$). This change from Baseline to the Final Visit was not statistically significantly different in the two treatment groups.

Neurological Examination Changes:

Of the 11 parameters evaluated, improvements in mean scores were noted in only four parameters and only for the GLIADEL treatment group patients. In the GLIADEL treatment group, the greatest improvement in the mean change from Baseline to the Final Visit was seen in the following parameters: visual change, fundus (papilledema), cranial nerves III, IV, VI, and cerebellar signs.

Patients Who Died:

A total of 11 patients (69%) in the GLIADEL treatment group and 15 (94%) in the PLACEBO treatment group died during the study. None of the deaths in the GLIADEL treatment group were considered by the investigator to be related to study drug. One death in the PLACEBO treatment group was considered to be remotely related to wafer implantation; the cause of death for this patient was "not assessable". All other deaths in the PLACEBO treatment group were considered by the investigator to have no relationship to study drug.

Serious and Unexpected Adverse Events:

There were 15 serious and unexpected adverse events reported by nine patients. In the GLIADEL group there were 10 serious adverse events reported in five patients. These included: wound infection, septic inflammation with meningismus and CSF leukocytosis with hydrocephalus, DVT and pulmonary embolism, pneumonia and a increase in aphasia, visual disturbances and hemiplegia. In the PLACEBO group five serious adverse events were reported in four patients. These included: pulmonary embolism, meningitis, wound infection and DVT and pulmonary embolism. In the opinion of the investigator 12 of

the 15 serious and unexpected adverse events were not, unlikely or remotely related to the study drug. The two occurrences of wound infection and one occurrence of meningitis were noted by the investigator to be possibly related to the study drug. All three infection related events occurred at study site 01 where the principal investigator inadvertently placed the unsterile outer wafer pouch on the sterile operating field.

Adverse Events:

Twelve of 16 patients (75%) in the GLIADEL treatment group and 9 of 16 patients (56%) in the PLACEBO treatment group experienced at least one treatment-emergent adverse event during the study period. The most frequently documented treatment-emergent adverse events in the GLIADEL treatment group were hemiplegia (38%) followed by convulsion (19%), aphasia (13%) and visual field defect (13%). In the PLACEBO treatment group, the most frequently reported treatment-emergent adverse events were hemiplegia (25%) and convulsions (13%). Six percent (two events) of the treatment-emergent adverse events in the GLIADEL treatment group were considered by the investigator to be life-threatening in severity, 55% (17 events) were considered to be severe and 32% (10 events) were considered to be moderate in severity. In the PLACEBO treatment group, 44% (7 events) of the treatment-emergent adverse events were considered to be severe and 38% (6 events) were considered to be moderate. There were no life-threatening treatment-emergent adverse events reported by investigators in the PLACEBO treatment group. In both the GLIADEL treatment group and the PLACEBO treatment group, most events were considered by the investigator to have no relationship to study drug [22 of 31 (71%) events in the GLIADEL treatment group and 10 of 16 (63%) events in the PLACEBO treatment group]. No event was considered to be probably related to GLIADEL or PLACEBO wafers by the investigator. In the GLIADEL treatment group, 3 of 31 treatment-emergent adverse events (10%) were considered to be possibly related. One of 16 treatment-emergent adverse events (6%) in the PLACEBO treatment group was considered by the investigator to be possibly related to study medication.

Tumor Imaging:

Tumor areas were similar in the two treatment groups at Baseline and at Visits 3, 5, 6, and 7. At Visits 8 and 9, however, the mean tumor area for GLIADEL patients was statistically significantly smaller than for PLACEBO patients [P = 0.027 at Visit 8 (N = 8 GLIADEL and N = 2 PLACEBO) and P = 0.007 at Visit 9 (N = 6 GLIADEL and N = 1 PLACEBO)].

Clinical Laboratory Parameters:

The changes from Baseline values in laboratory parameters were usually transient and maximal on the Day of Surgery and consistent with the clinical context of major surgery, blood loss and use of high dose corticosteroids. Nine of 16 (56%) patients in the GLIADEL treatment group and 12 of 16 (75%) patients in the PLACEBO treatment group had a clinically notable, post-baseline laboratory abnormality. The highest percentage of clinically notable laboratory findings were abnormalities in hematological values (low hemoglobin or leukocytosis) for both treatment groups. Four of 16 patients in each treatment group (25%) had clinically notable elevations of serum glutamic pyruvic transaminase [SGPT (ALT)]. Clinically notable abnormalities were, in general, transient in nature. There were no laboratory abnormalities or trends in laboratory value changes that indicated a systemic toxicity associated with GLIADEL therapy.

Conclusions:

In this 32-patient, Phase III, multicenter randomized, placebo-controlled clinical study in patients with newly diagnosed malignant glioma, implantation of up to eight GLIADEL wafers (each 3.85% BCNU by weight, containing about 7.7 mg of BCNU) increased one-year survival rates by approximately 230% (63% of GLIADEL patients were alive compared to 19% of PLACEBO patients; P = 0.029). GLIADEL treatment produced statistically significant reductions in mortality relative to PLACEBO treatment over both the 12-month period (relative risk 0.154 [95% CI: 0.051 to 0.467; P = 0.0010) and the 24-month period (relative risk 0.177 [95% CI: 0.067 to 0.468]; P = 0.0005) after wafer implantation surgery. GLIADEL treatment increased median overall patient survival by more than 18 weeks (58.1 weeks vs. 39.9 weeks; P = 0.011). The adverse event profile was typical of patients in the post-operative period following resection for malignant glioma. Systemic toxicity was not noted in evaluation of laboratory parameters.

SYNOPSIS

Study RPR132596T - 301

Title of the study:

A PHASE III, MULTICENTER RANDOMIZED DOUBLE-BLIND, PLACEBO-CONTROLLED TRIAL OF POLIFEPROSAN 20 WITH CARMUSTINE 3.85% IMPLANT IN PATIENTS UNDERGOING INITIAL SURGERY FOR NEWLY-DIAGNOSED MALIGNANT GLIOMA.

Investigators:

This was a multicenter study conducted in 14 countries (number of centers) as follows: Australia (3), Austria (1), Belgium (2), France (7), Germany (5), Greece (1), Israel (3), Italy (3), The Netherlands (2), New Zealand (1), Spain (3), Switzerland (2), United Kingdom (4), United States (5).

Study center (s):

A total of 42 centers were initiated. A total of 38 centers enrolled patients.

Publications (reference):

None to date.

Study period:	19 December 1997 to 30 June 2000	Clinical phase:	III
----------------------	----------------------------------	------------------------	-----

Objectives:

The objective of the study was to determine the efficacy and safety of polifeprosan 20 with carmustine 3.85% (GLIADEL® wafer) implants plus surgery and limited field radiation therapy compared to placebo implants plus surgery and limited field radiation therapy for improving the survival in patients undergoing initial surgery for newly-diagnosed malignant glioma.

Methodology:

Multicenter, randomized, double-blind placebo-controlled Phase III study.

Number of Patients (total and for each treatment):

A total of 240 patients were randomly assigned to receive either surgery plus GLIADEL® wafer implants plus limited field radiation therapy or surgery plus placebo implants plus limited field radiation therapy in a 1:1 GLIADEL®:placebo ratio. The total number of 240 patients was to include at least 168 patients with a final histopathological diagnosis of glioblastoma multiforme.

Diagnosis & criteria for inclusion:

Males or females aged 18 to 65 years who had radiographic evidence on cranial magnetic resonance imaging (MRI) of a single contrast-enhancing unilateral supratentorial cerebral tumor for whom surgical treatment within two weeks of the baseline MRI scan was indicated. Patients had to have an intra-operative pathological diagnosis of malignant glioma. Patients who had received prior cytoreductive surgery, prior radiotherapy to the brain or chemotherapy, or who had more than one focus of the tumor or a tumor crossing the midline, or concomitant life-threatening disease, were excluded from the study.

Test product, dose and mode of administration, batch N°:

Polifeprosan 20 with carmustine 3.85% (GLIADEL® wafer) implants. Up to eight wafers were implanted into the tumor resection cavity after maximal tumor resection, in order to cover the entire resection surface. Each implant contained 7.7 mg carmustine. Batch numbers of GLIADEL® used in this study were: ST6050, ST6328E and A7078E.

Duration of treatment:

Duration of the study : 30 months (19 December 1997 to 30 June 2000).
Enrollment duration : 18 months (19 December 1997 to 30 June 1999). All patients were followed for a minimum of 12 months after study surgery or until death.

Reference therapy, dose and mode of administration, batch N°:

Placebo wafer implants. Up to eight wafers were implanted into the tumor resection cavity after maximal tumor resection, in order to cover the entire resection surface. Batch numbers of placebo implants used in this study were: ST6050, ST6328E and A7078E.

Criteria for evaluation:

Efficacy: Overall survival 12 months after enrollment of the last patient, overall survival in GBM subgroup, survival to 12 months, progression-free survival (time to tumor progression was assessed according to clinical and radiographic criteria), Quality of Life (QoL) [assessed using the European Organization for Research and Treatment of Cancer (EORTC) Core Quality of Life Questionnaire (QLQ-30) with the supplemental Brain Cancer Module (BCM-20)], Karnofsky Performance Status (KPS) scores (to assess functional status), neurological evaluation and survival censoring patients with reoperation for disease progression.

Safety: Incidence of adverse events, serious adverse events (SAEs), laboratory safety tests (hematology and biochemistry).

Statistical methods:

The primary efficacy parameter was overall survival in the ITT population 12 months after enrollment of the last patient. The secondary efficacy parameters were overall survival in the GBM subgroup, survival to 12 months, progression-free survival, QoL, KPS scores, neurological evaluation, survival censoring patients with reoperation for disease progression, and safety parameters.

Using a two-tailed logrank test with an α level of 0.05 and a power of $1-\beta=0.90$, the estimated sample size to detect an 18% difference in 12 month survival rates between the two treatment groups (based on survival rates of 68% on the GLIADEL[®] group and 50% in the placebo group, and assuming 18 months accrual, 12 months follow-up time and a 15% patient loss rate) was 240 patients (120 per treatment group).

For the efficacy analysis, all randomized patients (whether they were eligible or not) were included in the ITT population. The sub-group of patients with glioblastoma multiforme was also analyzed. All randomized patients who had at least one wafer implanted were evaluable for safety.

All statistical tests were two-sided and the level of statistical significance was fixed at 5%. Categorical data were presented in contingency tables. Continuous data were summarized with at least the following: frequency (n), median, mean, standard error of the mean (SEM), minimal and maximal values. Time to event analyses were performed using the Kaplan-Meier method and compared using a logrank test stratified by country as a primary comparison and the Wilcoxon test as a sensitivity comparison. For the survival analysis, the treatment effect was also examined after adjusting for prognostic factors.

Summary - Conclusions:

• **Efficacy results**

Primary efficacy parameter – overall survival: Median survival in the ITT population was increased by 20% in the GLIADEL[®] group (13.9 months, 95% CI: 12.1 to 15.3 months) compared to the placebo group (11.6 months, 95% CI: 10.2 to 12.6 months). The percentage of patients surviving to one year was approximately 10% higher in the GLIADEL[®] group (59.2%, 95% CI: 50.4% to 68.0%) compared to the placebo group (49.6%, 95% CI: 40.6% to 58.6%). The difference in overall survival between the treatment groups was statistically significant for both the stratified logrank test ($p=0.027$) and the stratified logrank test adjusted for prognostic factors ($p=0.020$). In the GBM subgroup there was a similar increase in median survival and the percentage of patients surviving to one year in the GLIADEL[®] group [13.5 months (95% CI: 11.4 to 14.8 months) and 57.4% (95% CI: 47.8% to 67.1%), respectively] compared to the placebo group [11.4 months (95% CI: 10.2 to 12.6 months) and 48.6% (95% CI: 39.0% to 58.1%), respectively]. The difference between the treatment groups was not statistically significant for the main stratified logrank test ($p=0.098$), but the treatment effect was statistically significant when the results were adjusted for prognostic factors ($p=0.050$).

The results for the supportive survival analysis, excluding the two patients who had undergone further surgery with GLIADEL[®] reimplantation, were similar to the overall results. There was no statistically significant difference between the treatment groups in survival up to 12 months after initial surgery (i.e. censoring survival data after 12 months) for either the ITT population or the GBM subgroup. An additional survival analysis censoring patients with reoperation for tumor progression was significant for the ITT population ($p=0.014$) but not the GBM subgroup ($p=0.131$). Baseline KPS score, age, final histopathological diagnosis and the number of wafers implanted were shown to be statistically important predictors of survival in the ITT population ($p<0.001$, $p=0.001$, $p=0.011$ and $p=0.037$, respectively). In the GBM subgroup, baseline KPS score ($p=0.001$), age ($p=0.040$) and the number of wafers implanted ($p=0.018$) were shown to be statistically important predictors of survival.

Secondary efficacy parameters: The results for the secondary efficacy parameters in the ITT population were also more favorable for patients in the GLIADEL[®] group compared to patients in the placebo group. The difference between the treatment groups was statistically significant and favored GLIADEL[®] for the time to KPS score deterioration ($p=0.050$) and time to deterioration of neuroperformance measures ($p<0.05$ for 10/11 neuroperformance measures assessed). The difference between the treatment groups in secondary efficacy parameter results in the GBM subgroup were smaller (although still favoring GLIADEL[®] over placebo), and not statistically significant for any of the parameters except 5 of the 11 neuroperformance measures.

The median progression-free survival was almost identical for the two treatment groups for both the ITT population (5.9 months and approximately 48% of patients progression-free at one year for both groups, $p=0.901$) and the GBM subgroup (5.8 months for

the GLIADEL® group and 5.7 months for the placebo group, and 47.6% of patients in the GLIADEL® group and 44.1% of patients in the placebo group progression-free at one year, $p=0.621$). The median time to deterioration of the KPS score and the percentage of patients deterioration-free after one year in the ITT population were both higher in the GLIADEL® group (11.9 months and 47.5%, respectively) than in the placebo group (10.4 months and 39.3%, respectively) ($p=0.050$). The median time to deterioration of the KPS score and the percentage of patients deterioration-free after one year in the GBM subgroup were both higher in the GLIADEL® group (11.7 months and 43.6%, respectively) than in the placebo group (10.3 months and 38.0%, respectively) ($p=0.189$). The difference between treatment groups in time to deterioration of neuroperformance measures was statistically significant and favored GLIADEL® for 10 out of 11 neuroperformance measures in the ITT population (the exception was visual status). In the GBM subgroup the time to deterioration favored GLIADEL® for all neuroperformance measures except visual status, but the treatment difference was only statistically significant for 5 of the 11 neuroperformance measures.

- **Safety results**

No new or major safety issues concerning treatment with GLIADEL® wafer implants were raised in this study. Safety results were comparable between the treatment groups and generally consistent with those expected in patients undergoing major surgery for resection of malignant glioma, and those seen in previous studies and described in the package insert.

Eighty-eight patients (73.3%) in the GLIADEL® group and 93 patients (77.5%) in the placebo group died before the study cut-off date and most deaths [75 patients (62.5%) in the GLIADEL® group and 84 patients (70.0%) in the placebo group] were due to malignant disease. Only three patients (all in the GLIADEL® group) died within 30 days of randomization; two died due to a complication of the initial surgery and one died due to a complication of subsequent surgery for tumor recurrence.

Overall, 1244 AEs were reported by 119 patients in the GLIADEL® group and 1224 AEs were reported by 120 patients in the placebo group. The AE profile was similar for both treatment groups, with no statistically significant difference between the treatment groups in frequency of any of the AEs tested ($p<0.05$) except intracranial hypertension, which was reported by 11 patients (9.2%) in the GLIADEL® group and two patients (1.7%) in the placebo group. More patients in the GLIADEL® group [6 (5.0%)] compared to the placebo group [1 (0.8%)] had cerebrospinal fluid (CSF) leaks. Less than 10% of AEs were considered to be treatment-related. Aggravation reaction was the most frequently reported AE in both groups [reported for 98 patients (81.7%) in the GLIADEL® group and 95 patients (79.2%) in the placebo group]. Aggravation reactions were mainly tumor/disease progression or general deterioration of condition. Other frequently reported AEs were nervous system AEs (hemiplegia, convulsion, confusion, brain edema and aphasia), digestive system AEs (nausea and vomiting) and body as a whole AEs (fever, headache and infection).

A total of 374 SAEs were reported by 112 patients in the GLIADEL® group and 370 SAEs were reported by 110 patients in the placebo group. The most frequently reported SAEs were aggravation reaction [85 patients (70.8%) in the GLIADEL® group and 83 patients (69.2%) in the placebo group] and the nervous system AEs convulsion [40 patients (33.3%) in the GLIADEL® group and 44 patients (36.7%) in the placebo group] and hemiplegia [19 patients (15.8%) in the GLIADEL® group and 18 patients (15.0%) in the placebo group]. Less than 20% of all SAEs were considered to be treatment-related.

Patterns of change from baseline in laboratory parameters were similar for both treatment groups and consistent with changes frequently seen after major surgery. There were no clinically significant patterns of change in laboratory parameters that could be associated with study treatment.

- **Quality of life results:** Quality of life results were comparable for the GLIADEL® group and the placebo group. Missing data due to attrition (death) and noncompliance with questionnaire completion were significant in this study, reducing the amount of data available for analysis and thus limiting any conclusions. GLIADEL® treated patients showed no decline in QoL due to GLIADEL® therapy compared to the placebo treated patients.

- **Conclusion**

GLIADEL® wafer implants increased overall survival in patients undergoing initial surgery for malignant glioma. In these patients, GLIADEL® increased the median survival from 11.6 months to 13.9 months, a 20% improvement, which was statistically significant ($p=0.020$, stratified logrank test adjusted for prognostic factors). GLIADEL® increased the one year survival rate from 49.6% to 59.2%. GLIADEL® wafer implants also increased overall survival in the GBM subgroup. GLIADEL® increased the median survival in this subgroup from 11.4 months to 13.5 months, an 18% improvement, which was statistically significant ($p=0.050$, stratified logrank test adjusted for prognostic factors). GLIADEL® increased the one year survival rate from 48.6% to 57.4%.

There were no safety concerns arising from the use of GLIADEL® wafer implants. The adverse event profile and laboratory safety test results were similar for GLIADEL® wafer implants compared to placebo wafer implants and characteristic of patients undergoing major surgery for resection of malignant glioma, although there was a higher frequency of CSF leaks and intracranial hypertension in the GLIADEL® group compared to the placebo group.

The overall risk:benefit ratio was positive and favors the use of GLIADEL® as an adjunct to surgery and limited field radiation therapy in patients undergoing initial surgery for malignant glioma.

- **Date of report** 07 March 2001

Adjustments for Center in Multicenter Studies: An Overview

A. Russell Localio, JD, MS; Jesse A. Berlin, ScD; Thomas R. Ten Have, PhD; and Stephen E. Kimmel, MD, MS

Increasingly, investigators rely on multicenter or multigroup studies to demonstrate effectiveness and generalizability. Authors too often overlook the analytic challenges in these study designs: the correlation of outcomes and exposures among patients within centers, confounding of associations by center, and effect modification of treatment or exposure across center. Correlation or clustering, resulting from the similarity of outcomes among patients within a center, requires an adjustment to confidence intervals and *P* values, especially in observational studies and in ran-

domized multicenter studies in which treatment is allocated by center rather than by individual patient. Multicenter designs also warrant testing and adjustment for the potential bias of confounding by center, and for the presence of effect modification or interaction by center. This paper uses examples from the recent biomedical literature to highlight the issues and analytic options.

Ann Intern Med. 2001;135:112-123.

www.annals.org

For author affiliations and current addresses, see end of text.

For a glossary of terms, see end of text.

Multicenter studies offer powerful methods for understanding the effects of treatment and exposure on patient outcomes. Although convenient and expedient, simple pooling of data as if they arose from a single population can produce incorrect results. More complex methods are needed for analyzing data characterized by patients clustered within centers. A hospital, ward, clinic, physician's office, neighborhood, housing complex, or family, in which patients are naturally grouped, might qualify as a *center* for the purposes of design and analysis. Typically, individuals within centers are more similar than those from different centers. Siblings share genes and environment more than do strangers. Patients in one hospital experience common treatment protocols delivered by shared clinicians. In statistical terms, observations within a center are correlated; those in different centers are independent. Although well known among statisticians and epidemiologists (1-8), this failure of a key methodologic assumption in standard statistics, the independence of observations, is overlooked in manuscripts and published articles.

Authors and readers recognize the general principle of confounding by covariates (age) in the association between treatments (hormone replacement therapy) and outcomes (stroke). Effect modification (smokers might be at greater risk from oral contraceptives than non-smokers) is also a well-known feature of clinical studies. Less appreciated is the potential for study center to produce confounding or effect modification when treatments are administered across several centers.

Failure to consider the center in an analysis might result in incorrect *P* values and confidence intervals (because of clustering), biased estimates (because of uncon-

trolled confounding), and unrecognized heterogeneity across centers (because of effect modification). For purposes of exposition, we use the common example of the patient as the observation and the hospital, clinic, or physician's practice as the center (Table 1).

HISTORICAL DEVELOPMENT

Early applications of statistical methods for clustered data arose from the specialized field of survey research (9, 10), where the practice of sampling clusters of individuals in entire households or census blocks necessitated new statistical methods. A parallel development occurred in the social sciences and education, where children are naturally grouped in classrooms (11). In 1978, Cornfield was credited with an early assessment of the implications of clustering in biomedical studies (3). Despite these developments, reviews document the inattention of published studies to departures from the assumption behind most statistical methods— independence of observations (1, 12-15). This inattention perhaps reflected the now-remedied dearth of commercially available software and the inadequate documentation of early pertinent findings from survey literature.

EFFECTS ON *P* VALUES AND CONFIDENCE INTERVALS

When patients are not independent observations, *P* values and confidence intervals generated from standard statistical methods can be incorrect. Without an adjustment for clustering, results might appear statistically significant when they are not, or vice versa. For that reason, investigators need to identify centers, incorporate the concept of centers into their designs, estimate the

Table 1. Key Points: Effects of Ignoring Center on Study Design and Analysis

<p>Underpowered study designs, especially when centers are few and large Incorrect <i>P</i> values from standard statistical formulae or software Incorrect confidence intervals, often too narrow Biased estimates from unrecognized confounding by center Unrecognized heterogeneity of effect of treatment or exposure across centers</p>
--

design effect, and adjust confidence intervals and *P* values appropriately.

Identifying the Center

In traditional randomized multicenter studies, the identity of the centers is obvious; in other designs, it is not. For example, Dexter and colleagues (16) chose physician as the center and the physician-patient pairs as the observation in their study of the effectiveness of computerized reminders on the quality of discussion about advance directives. This choice was reasonable because patients might have similar (correlated) success in their dialogues with their common physician. In an observational study of hand-washing behavior among clinicians at a single hospital (17), the authors selected a combination of period of observation and ward as the center and the opportunity to wash as the observation. Although the authors might have selected a different definition of center (and observation), it was reasonable to assume a common correlation of the probability of hand-washing within the ward during a shift, perhaps because of common personnel and shared time pressures.

Relating the Center to the Design

In traditional multicenter trials, patients are randomly assigned to different treatments within each center, and estimates of treatment effect are within-center comparisons. In a cluster-randomization design, by contrast, all the patients in a practice, hospital, organization, or entire community are assigned to the same treatment (1, 18, 19). Comparisons are effectively done among communities, although data are collected and analyzed on each person. This study design appears typically in physician-practice-based interventions, as in a comparison of individualized office systems to improve the rate of breast cancer screening (20). In many studies of computer-generated reminders, for example, the intervention group of physicians receives a reminder for all patients, whereas the control group of physicians re-

ceives no reminders for any patients (2). The intervention (the reminder) is therefore a center-level (among-center) comparison. Ethics, politics, logistics, or the need to avoid contamination of treatment arms often dictates this design. Authors should indicate the unit of randomization, patient or center, in reports of clinical trials (21). Design will therefore control whether key comparisons should be within or among centers.

The Design Effect

A measure of the degree of clustering is the design effect (10, 22, 23). When patients do not cluster, the design effect is 1.0 and standard statistical methods should produce correct confidence intervals. We shall first look at the design effect for estimates of the overall risk for an outcome, such as disease or adverse event. Then we shall consider the design effect in the presence of covariates, such as treatment (exposure) or other patient- or center-level factors.

When no covariates are present, the design effect for an estimate of the overall risk for an outcome usually increases as the within-center correlation and the number of patients per center increase. A study with a design effect of 2.0 has double the variance and half the effective sample size of a comparable study with no clustering. Without an adjustment for clustering, confidence intervals on estimates of patient risk are too narrow.

A striking example appears in the analysis by Laine and associates (24) of rates of combined (right and left) cardiac catheterization among 41 000 Medicare patients in 73 Pennsylvania hospitals. A simple estimate of the 95% confidence interval (33.7% to 34.7%) of the overall rate of 34.2% suggests great precision. The calculation assumes incorrectly that patients' outcomes are independent observations. The hospital-specific rates vary widely, however, from 2% to 98%. This variation provides strong evidence that patients at the same hospital are correlated. Two patients in one hospital are more likely to have (or not to have) a combined procedure than two patients at different hospitals. This clustering of large numbers of patients into centers inflates variance, as represented by a design effect of 370 in this example. The large variance reflects the impact of hospital selection on the result; a slightly different sample of hospitals might produce a substantially different overall rate. The study is far less powerful than a study with 41 000 independent patients would have been. The ef-

fective sample size (25) is only 111 (41 000/370). Adjusted for this correlation of patients within hospitals, the confidence interval for the overall rate becomes 25% to 43%. A simple analysis would overstate statistical precision substantially.

Design effects for treatment, exposure, or other covariates are more complex. A multivariable model, with several predictors of outcome, can have several design effects, one for each treatment effect and covariate. For example, in the preceding example, the design effect was 169 for the hospital location covariate but only between 25 and 57 for the hospital size categories. The size of the design effect depends in part on whether the covariate applies to the patient or to the center. If all patients in a center receive identical treatment, as in a computerized physician reminder study, the design effect for reminders (a center-level factor) can be substantial. By contrast, in the same study, if the investigator wants to examine the effect of patient sex on outcomes, the design effect might be much closer to 1.0, especially if the distribution of patients by sex is about the same across centers. The design effect for a treatment variable also can fall below 1.0 in some instances of perfectly balanced studies, with equal proportions of patients assigned to treatment and control groups at each center (6, 26). In that case, standard statistical methods can overstate the width of confidence intervals and mask statistically significant results. Therefore, study design, the comparison of interest, and the degree of correlation of outcomes affect confidence intervals and *P* values.

Details for computing the design effect in simple cases are given by Kerry and Bland (22, 23). One can test for the need to adjust for clustering by comparing the design effect with 1.0. In a study of the incidence of adverse drug reactions among 3137 patients admitted to 62 departments in 33 hospitals, Pouyanne and coworkers considered a design effect below 1.5 to be "negligible" (27). However, failure to adjust for a design effect of 1.5 means that the "95% confidence interval" as computed by standard statistical packages is in fact only a 90% interval. Whenever the design effect exceeds 1.0, the investigator should adjust confidence intervals.

Adjusting Confidence Intervals To Account for Center

As we have noted, investigators can moderate the design effect by ensuring that key comparisons involve within-center variables that are well balanced (the same

proportions) across centers. A complementary approach involves adding covariates to the regression model, subject to the usual cautions (28). Of more importance, residual correlation of patient outcomes within center will probably persist even after the investigator controls for all known and measurable patient and center characteristics. This was the finding of Fink and colleagues (29) in their assessment of the correlation of outcome (urea reduction ratios) among 6969 patients receiving hemodialysis at 154 facilities. Statistical methods have been developed to adjust for the effect of this residual correlation and produce correct *P* values and confidence intervals.

The choice of methods for this adjustment depends on how the comparison of interest relates to the center. Conventional randomized, controlled trials compare treatments allocated to like fractions of patients within each center (balanced design). The investigator can sometimes increase statistical power (as reflected by lower *P* values) through optimal statistical methods. By contrast, when treatments are not well balanced within centers, as when the studies are not randomized, the investigator must choose a method that will compensate for a design effect greater than 1.0. If the covariate of interest is a center-level factor, such as in a cluster-randomization design, the statistical method must be suitable for among-cluster comparisons.

We outline two sets of methods for adjusting confidence intervals: conditional and unconditional (Table 2). Each asks a different question, relies on separate assumptions, and therefore produces different results. Conditional methods estimate patient-level factors by conditioning (or stratifying) on center. Although conditional estimates average the effects of treatment over the centers, they reflect only the within-center component of treatment effect. They evaluate, for example, what the impact would be of switching a patient within a hospital from control to treatment. They are appropriate for conventional multicenter studies in which patients are randomly assigned to different treatments within each center. In that case, treatment is a patient-level, within-center comparison. By contrast, unconditional analyses estimate the average effect of treatment over the population of patients without regard to the center from which the patients were drawn. As such, unconditional estimates represent the joint impact of within- and among-center effects of treatment or exposure. For ex-

Table 2. Key Points: Alternative Approaches To Estimating Correct Confidence Intervals in Multicenter Data

Approach	Comments
Conditional methods	For within-center (patient-level) factors
Fixed-effects analyses	Centers represent themselves rather than the population of all centers
Mantel-Haenszel methods	Simple methods when there are no covariates other than treatment
Conditional regression*	Focuses on within-center factors when centers are neither large nor small
Fixed-effects regression	Appropriate when there are many patients per center
Random-effects or center-specific	Ideal for within-center factors but allows case mix-adjusted comparisons across centers
Unconditional methods	For population-averaged interpretation of patient- and center-level factors
Sample survey methods	Ideal for complex surveys, yet broadly applicable to other multicenter designs
Generalized estimating equations	Especially suited for designs involving many centers (at least 30)
Bootstrap resampling by center	Flexible but computer-intensive method for obtaining confidence intervals

* Classified by some authors as equivalent to random-effects analysis in some cases.

ample, to assess the association of patient ethnicity and outcome, conditional analyses compare ethnic groups treated at the same hospital. Unconditional analyses compare ethnic groups' outcomes without regard to the place of hospitalization and then adjust the confidence intervals for the effects of clustering of patients within hospitals. At present, we are assuming that treatment effects, such as relative risks, do not vary across centers, although patients' overall risk might vary because of intercenter differences in patients' health. Issues of variable treatment effect, as reflected by heterogeneity of relative risks across centers, are discussed in the section on effect modification.

Conditioning on Center

Methods that condition on center measure the association between treatment and outcome within each center, and then combine results across centers. These conditional analyses can result in greater statistical power, narrower confidence intervals, and smaller *P* values than pooled analyses because they take advantage of similarity of patients within clusters, just as would a crossover or longitudinal study in which a patient serves as his or her own control. Conditional methods can be subdivided into fixed- and random-effects models. Fixed-effects analyses consider a center to be fixed and

to represent only itself, while in random-effects analyses the centers represent the population of similar centers in the region from which the study sample was drawn.

Estimation under fixed-effects methods is limited to treatments or exposures that vary within center, as in traditional randomized, controlled trials. If some centers contain only patients of one treatment group, the entire center is lost because there are no comparison patients in the same center. Likewise, if outcomes are rare and centers are not large, some centers might experience only successful outcomes (or failures). These centers will also be dropped, and results will then reflect only the experience of the remaining centers. Whether results from the remaining centers are meaningful depends on the investigator's clinical question. We present three subclasses of fixed-effects methods: Mantel-Haenszel analysis, conditional logistic regression, and fixed-effects regression.

Mantel-Haenszel methods, discussed in elementary textbooks and available in most software packages, compute odds ratios, relative risks, or risk differences and their confidence intervals. These methods are suitable for studies with many centers and few patients per center, as well as for designs with few centers but many patients per center. In the context of data clustered by center, these methods are limited to estimating the association of a binary outcome and a single categorical covariate (the treatment or exposure variable). The stratification variable is the center. For multivariable analyses, more flexible tools are needed.

Conditional logistic regression is the multivariable analogue to Mantel-Haenszel methods. Increasingly available in standard statistical packages, conditional regression can estimate only patient-level factors. Center-level factors, such as hospital size, are removed as "nuisance" effects not in need of estimation. In terms of its underlying theory, this method has been classified by different authors as either a fixed- or random-effects analysis. The effects of patient-level variables, whether continuous or categorical, are easily estimated. Investigators and readers should regard with caution any data sets containing many variables relative to the number of events, because large biases from overly sparse data sometimes produce grossly inflated odds ratios (30). Having too many patients per center also presents problems because some statistical software packages will not

accommodate such large data structures (31). In the latter case, fixed-effects logistic regression is an alternative.

Like conditional logistic regression, fixed-effects logistic regression estimates only within-center effects. The regressions include a set of indicator variables, each representing a center, in addition to the patient-level variables of interest. The regression generates a separate estimate of the risk to the unexposed or untreated patient for each center. It works best when many patients are spread across few centers. For example, Marrazzo and associates (32) were able to use fixed-effects regression to assess the impact of screening for chlamydia trachomatis among 10 118 teenagers in 12 facilities (32). Severely biased estimates occur, however, when each center has few patients and few events (31, 33, 34). As with any logistic regression, investigators must always avoid having too few events for the number of covariates, including centers (35).

Center-specific or random-effects models assume that the centers are a random rather than a fixed sample from the population. These models estimate an overall treatment effect but allow for heterogeneity across centers in the form of random differences in patient populations at each center. For example, the use of cardiac stents might reduce the risk for death within 6 months by 20% (a relative risk of 0.80). Because of unmeasured differences in the patients across centers, however, the risk for death among patients without stents might vary substantially. As a result, the risk for death among the patients with stents might also vary by center, even when the relative benefit or harm of stents remains constant across centers. By assuming by convention that the log odds of patient risks across centers take on a typical bell-shaped (normal) distribution, the model can consider this variation in outcomes across 30 centers by means of a single parameter, the variance of the center-specific random effects. A comparable fixed-effects regression would require estimation of 29 variables to represent variation among the 30 centers. Fewer variables can lead to increased statistical power.

As with fixed-effects methods, the center-specific models can estimate the effect of treatment within a center. But this model relies on a key assumption: the absence of association between the random effect (the center) and the chances of being treated or exposed. In most multicenter randomized trials, this assumption is satisfied because patients are allocated in equal propor-

tions to treatment and control within each center. Readers must be alert to studies that do not meet this assumption, such as observational studies with different proportions of patients exposed by center. In addition, especially with low-risk outcomes, some centers might experience no events. While conditional logistic regression will explicitly drop observations from these centers, the center-specific model will simply ignore them. Looking at the data by center is therefore essential to determine which centers contributed to the results.

For continuous outcomes, such as blood pressure or many laboratory values, center-specific models have been used for two decades. These models are well described and reviewed (36, 37), and software is widely available. An example is the study by Smith and coworkers (38) of the average delay from symptoms to hospitalization of 1334 patients with acute stroke in 23 hospitals. For binary or count outcomes, however, model interpretation and fitting are more difficult. Patient-level factors (within-center effects) have straightforward interpretations. Estimates of patient sex, for example, reflect the average association of sex and outcome within centers. Interpretation of center-level factors becomes more challenging (8). For example, if some centers are for-profit, the interpretation of a binary "profit status" variable is the following: If patients could be treated at a single clinic that was for-profit for some patients and nonprofit for others, how would patients respond when exposed to the different business status (39)? This awkward interpretation of a center-level factor as a within-center variable leads some authors to avoid these models when center-level covariates are important.

As we have explained, patients' outcomes within center are often positively correlated; they are similar rather than independent. In the infrequent case of negatively correlated outcomes, however, some center-specific models assume zero correlation (independence), and their confidence intervals are too wide. Negatively correlated outcomes occur, for example, in birthweights of twins; when one infant is large, the other might be small. For this reason, center-specific models demand the input of a statistician who can offer alternative strategies (40, 41).

For an increasingly common application, comparisons of hospital or physician performance in the form of "report cards," "score cards," or "league tables," center-specific models are attractive but challenging. Fiscella

and Franks (42) used a center-specific model to assess whether patient satisfaction differed across 100 primary care practices (centers) after adjustment for disparities in the mix of patients. When used to compare performance by center, these methods pose severe challenges in model fitting and testing. Institutional performance rankings based on simple analyses can be extremely unreliable (43).

Authors refer to the spectrum of center-specific models alternatively as mixed-effects, random-effects, variance component, hierarchical, multistage, or empirical Bayes regressions and, in the context of survival (time-to-event) data, frailty models (44–46). Programs in both general and specialized statistical packages are readily available, although their proper use requires considerable expertise.

Unconditional Models

Marginal or population-averaged models are common in the medical literature, in part because of ease of interpretation and wide availability of software. Marginal models estimate the average impact of treatment (or exposure) without regard to the center, but then adjust confidence intervals for the correlation of patients within centers. They focus on the effect of treatment or exposure averaged over centers rather than on the corresponding estimate conditional on individual centers. This interpretation of the treatment effect in a population of patients (6) is often what authors desire when the experience of an individual center and prediction of patients' outcomes at that center are not important.

For continuous outcomes, estimates from population-average models will approximate those of the center-specific methods outlined previously. For binary outcome (mortality) or count data (numbers of infections), however, estimates of odds ratios or relative risk from a marginal model will usually be closer to 1.0 (no treatment effect) than comparable estimates from a center-specific analysis (47). Comparisons of effect size across studies with different analyses are therefore problematic.

We review three common population-averaged analyses: sample survey methods, generalized estimating equations, and bootstrap resampling.

Well-established survey-sampling methods are obvious choices for estimating means, rates and proportions,

their between-group differences, and parameters from multivariable regression models when data are collected from patient surveys (48). These methods are found routinely in reports from large national health surveys, as in the recent studies on the association between vitamin use and homocysteine levels (49), where responses of individuals will be clustered within areas or households. Survey analysis methods are equally useful in smaller, less complex surveys of patients clustered within centers, such as Bassuk and colleagues' examination of cognitive decline among elderly, community-dwelling persons (50). For survey data, the investigator should consider and the reader should expect an analysis that uses specialized software for clustered and weighted data.

Survey methods are generalizable to other types of clustered data (51). One example is the report of Shekelle and colleagues (52) on the appropriateness of clinical indications and choices of therapy in 1310 patients seen in 131 chiropractic offices at 6 sites. Such methods are also appropriate for randomized, controlled trials. If the design is well balanced, these methods will correctly report lower *P* values for within-center factors if patients' outcomes vary by center.

Generalized estimating equations represent an increasingly popular set of methods for estimating average or marginal effects of treatment or exposure. These methods do not model the center explicitly; they produce estimates comparable to those from ordinary logistic regression but then adjust confidence intervals for the correlation of outcomes within hospital, physician practice, or family. Statistical software is well documented (53), and examples abound. Robinson and Roter (54) recently examined the prevalence and factors associated with counseling on psychosocial problems among 308 patients in a sample of 69 primary care practices. Using the family as the center for analysis, Knox and associates (55) studied the association between cardiovascular events and hostility toward others. According to simulation studies, generalized estimating equations need large numbers of centers. No rules of thumb are available, but at least 30 and perhaps more centers are required for the underlying theory to apply. When centers are few, computed *P* values will be too small. An analogous "marginal" method of analysis applies to time-to-event outcomes (survival) among multiple patients per center, as well as to multiple events per patient (56).

Table 3. Key Points: Treatment or Exposure by Center

<p>Confounding</p> <p>Source of bias when both outcomes and proportions of exposed or treated patients vary by center</p> <p>Applies only to within-center treatments or factors</p> <p>Tested by confirming results with conditional regression methods</p> <p>Resolved by decomposition of within- and among-center components of treatment or exposure</p> <p>Effect modification (interaction)</p> <p>Always warrants investigation into the causes of variation across centers</p> <p>Analysis options include:</p> <ul style="list-style-type: none"> Using a random-effects model that allows for variation in effect of treatment across centers Using a fixed-effects regression with treatment-by-center interaction terms Explaining intercenter variation by interactions of patient-level factors with treatments

The bootstrap represents a class of flexible, computer-intensive methods applicable to a broad range of models for any type of outcome (57). When applied to clustered data, the bootstrap method uses the estimate from an ordinary regression model but adjusts the confidence interval by asking how much the results would change with different samples of centers. In so doing, this method follows the fundamental paradigm that patients are grouped within centers. Unlike the fixed-effects analyses, bootstrap resampling assumes that the centers are a random sample from a larger population of centers. To the extent that results vary by center, resampling by center detects that variability. For clustered data, the bootstrap is implemented easily by means of standard statistical software and yields estimates and confidence intervals approximately equal to those from generalized estimating equations or survey sampling methods.

Feldman and coworkers (58) used the bootstrap in their examination of the impact of dialyzer reuse on the survival of 28 000 patients treated at 1300 centers. If the facility elected to reuse dialyzers, all patients were treated with recycled equipment. The bootstrap method resamples the real data and creates many new data sets, each consisting of different combinations of the 1300 actual centers. Each of these created data sets differs because the algorithm can select the same center more than once and some centers not at all. The statistical analysis is applied to each of these created data sets, and the investigator then computes the middle 95% of these estimates. This range of estimates becomes the confidence interval for the estimate from the real data (59). Resampling by center, rather than by individual patient, pre-

serves the effect of the correlation of patients within center.

Because population-averaged methods do not exploit within-center contrasts between groups of patients (older vs. younger, for example), they are less efficient than the center-specific analyses when these within-center comparisons are of primary interest (60). The center-specific models will then offer more appropriate tools. On a more practical level, software sometimes behaves contrary to theory. Our experience suggests that population-averaged methods perform appropriately when the correlation of patients' outcomes within center are positive or negative. However, when within-study factors of interest occur with equal proportions by center (balanced design), we have found that population-averaged methods sometimes fail to reduce the width of confidence intervals and therefore understate statistical significance. Utmost caution is warranted, especially when the number of clusters falls below 30.

Although we have presented the paradigm of patients clustered within centers, the statistical issues we outline apply equally well to repeated measures or longitudinal studies, in which each patient (as the center) has multiple measurements of outcome and covariates over time. In these applications, the statistical issues and methods are well described in textbooks (61) and review articles (62).

BIAS: CENTER AS A CONFOUNDER

Less well recognized than the effect of center on confidence intervals and *P* values, but equally important, is the potential for confounding of within-center covariates by center (Table 3). A common example is the influence of patient ethnicity on outcome. The investigator wishes to determine whether good outcomes vary by patient ethnicity in a multihospital observational study. Ethnic mix varies across hospitals because they serve different populations. A significant association between ethnicity and outcome, even after accounting for clustering, might mean that 1) people of different ethnicities experience disparate outcomes at the same hospital because of variable quality of service to individual patients (within-center difference), or 2) hospitals that serve predominantly minority populations happen to be worse (or better) regardless of the ethnicity of the individual patient (among-center difference). Population-

averaged models make no attempt to distinguish the within- and among-center effects. Center-specific models estimate within-center differences, but only when a key assumption is satisfied: The random hospital effect is independent of the covariate of interest (ethnicity). This assumption equates to an absence of confounding due to center. Only when hospital ethnic mix is constant can we be certain that this assumption is met. Thus, both population-averaged and center-specific models are subject to confounding by center. Conditional logistic regression and its univariable counterpart, the Mantel-Haenszel methods, as well as fixed-effects regression, estimate solely the within-center effect and are immune to this form of bias.

One assessment of confounding by center, therefore, compares results from conditional logistic regression with those from an unconditional analysis. For example, in a recent examination of ethnic differences in the use of cardiovascular procedures among 4987 adults with renal disease who were treated at 303 facilities (63), clustering presented issues of confounding in addition to issues of correct confidence intervals. First, the authors used generalized estimating equations to confirm that their confidence intervals were not affected by clustering. Next, to estimate only the within-center effect of ethnicity, the authors used conditional logistic regression. These estimates being the same, the authors were able to rule out the possibility that observed differences resulted from variation in ethnic mix across hospitals.

Confounding can apply to any within-center treatment or exposure. Careful design of a randomized study can achieve balance (constant proportions) of the treatment across centers, at least initially. However, when this balance is absent at the completion of the study, as in many observational studies or in randomized studies in which patients are lost to follow-up, confounding by center can still cause biased estimates.

When interest lies not in the joint within- and among-center effect, as estimated by an unconditional regression, nor solely in the within-center effect, as from a conditional logistic regression, the investigator can use a modified center-specific or population-averaged regression model to distinguish within and between center components. The Appendix uses a clinical example to detail a method for decomposing within- and among-center effects (47, 64, 65). That decomposition bridges many of the differences we have outlined among the

conditional, center-specific, and population-averaged models.

EFFECT MODIFICATION BY CENTER

Our review has thus far assumed that the effect of treatment or exposure remains the same across centers, even if outcomes vary. Effect modification, or interaction (66), typically occurs when the association between exposure (asbestos) and outcome (lung cancer) varies with the level of another factor (cigarette consumption). Asbestos exposure might be more likely to lead to lung cancer in smokers than in nonsmokers. Smoking status is an effect modifier. Likewise, center can be an effect modifier. We are not referring simply to variations in outcomes among all patients across centers. Methods we have described previously address that issue. Rather, we refer to variation in the effect of treatment or exposure across centers (Table 3).

Multicenter clinical trials sometimes fail to address effect modification, perhaps because sample sizes are too small to detect treatment differences across centers. By the use of standard protocols for patient recruitment and treatment, the investigator assumes a single true treatment effect across all centers. However, this assumption does not always hold in practice. Populations at different centers might not react the same way, perhaps because of unmeasured population or environmental factors or incomplete adherence to protocols. This possibility increases in observational studies and meta-analyses, in which exposures or protocols are likely to vary across centers.

Conditional analyses, such as the Mantel-Haenszel methods or conditional logistic regression, assume the absence of effect modification. If, for example, 15 centers revealed beneficial treatment effects and 15 demonstrated harm, conditional analyses would mask the true differences across center by estimating a treatment effect somewhere between benefit and harm (for example, no effect). Because statistical tests have low power to uncover effect modification (67), looking at center-specific results is imperative. When effect modification is suspected or apparent, the investigator must consider methods that estimate treatment effects specific to each center.

Using fixed-effects logistic regression, the investigator can test for and estimate center-by-treatment interaction by comparing two models: one with and one

without an additional set of terms for these interactions. However, a 30-center study would require 29 interaction terms, and only in studies with large numbers of patients per center can the investigator estimate the treatment effects by center.

Center-specific models, as we described previously, are far more efficient in this application; they require only one additional parameter to model treatment interaction. These random-coefficient models (68) produce a separate estimate (coefficient) of treatment effect for each center, as well as an overall estimate (69, 70). Random-coefficient models, although flexible and powerful, are only beginning to surface in widely read medical journals. Yamaguchi and Ohashi (44) used this approach to examine the observed variation among centers in multicenter clinical trials of superficial bladder cancer. By contrast, population-averaged models by design do not estimate effect modification by center.

When treatment effects vary significantly across centers, the interpretation of an overall effect becomes problematic with any of the statistical models we have presented. Significant center-by-treatment interaction (effect modification) warrants further investigation of issues of omitted covariates, unmeasured population factors, or center-specific differences in study protocols (70, 71). Meta-analyses often assume interaction of treatment effect by center and use the methods we have reviewed (72).

DISCUSSION

Clustering, confounding, and effect modification should be considered in any study of patients grouped into centers. For simplicity, we concentrated on examples with binary outcomes. The same issues of clustering, confounding, and effect modification by center apply to continuous outcomes, count data (the number of infections), or time-to-event studies. This advice notwithstanding, readers should not conclude that all studies with centers must involve clustering. For example, case-control studies might remain unclustered and population-based although cases are identified through hospitals or clinics.

Because only a minority of journals regularly refer manuscripts for statistical review (73), readers must rely on their own knowledge of the analytic and interpretative issues of patients clustered within centers. Our

greatly simplified presentation downplays the complexity, uncertainty, and ongoing controversies with clustered data. Often, the skills required for optimal analysis will lie beyond those of most clinician-investigators. Multicenter studies will therefore benefit from the combined expertise of multidisciplinary teams.

APPENDIX: AN EXTENDED CLINICAL EXAMPLE OF CONFOUNDING BY CENTER

The decomposition of an effect into its within-center and among-center components can control for confounding. Our extended clinical example involves two strategies, staged and combined, for percutaneous transluminal coronary angioplasty in 18 500 patients at 49 clinical centers (69). The principal outcome, major complications, occurred in 339 patients (1.8%). The clinical question was whether the type of angioplasty influenced outcome.

The two necessary conditions for confounding by center were present: substantial variability across centers in both the rate of complications and the rate of combined procedures. To control confounding, the investigators had to adjust for the association between exposure (staged or combined) and center. They decomposed the procedure variable into within-center and among-center components, resulting in two odds ratio estimates. The usual binary indicator for the type of procedure received for each patient measured the within-center component, while a variable measuring the proportion of all patients who received the combined procedure in a center represented the among-center component. Because the data were clustered by center, the investigators used generalized estimating equations to adjust confidence intervals.

The unadjusted odds ratio of 1.57 (95% CI, 1.21 to 2.05; $P = 0.001$) for a binary variable representing cardiac procedure suggested a significant, increased risk for complications for patients with combined procedures. Adjusting for the design effect from clustering reduced the statistical significance but did not alter the odds ratio (1.57 [CI, 1.00 to 2.47]; $P = 0.053$). However, this analysis estimated the joint within- and among-center components of the effect of combined procedures. Conditional logistic regression, again using only a simple binary variable for cardiac procedure, suggested confounding by center; it produced an odds ratio estimate of 1.02 (CI, 0.75 to 1.37) for the within-center effect of angioplasty procedures.

Decomposing the variable for procedure assessed both the within- and among-center components of angioplasty procedure. A within-center odds ratio of 1.02 (CI, 0.73 to 1.41; $P > 0.2$) measured the association between procedure type and outcome on patients, conditional on center. The among-center odds ratio for an increase of 10 percentage points in the proportion of

combined procedures was 1.26 (CI, 1.16 to 1.38; $P < 0.001$). This component reflects the impact on patient outcome of the rate of performance of combined procedures at the hospital. The significant odds ratio suggests that centers with high proportions of combined procedures had more complications. The decomposition of the variable on type of cardiac procedure demonstrated that the choice of hospital, not the choice of cardiac procedure, influenced the risk for complications. The unadjusted odds ratio mixed these effects, one significant and the other not, into a single estimate of limited clinical meaning.

GLOSSARY

Balance: Equal fractions of patients are allocated to treatment (or are exposed) within each center, as in many conventional multicenter randomized, controlled trials. Not all centers must have equal numbers of patients to achieve balance. Only the ratios of treated to untreated patients must remain equal.

Bias: The systematic departure of an estimate (of the effect of a treatment or exposure) from the true value.

Bootstrap resampling: A method of computing confidence intervals that repeatedly draws samples of centers (such as hospitals) with replacement, creates numerous samples of the data set, and then computes numerous estimates of the statistic of interest.

Center-specific (cluster-specific) models: Statistical models that produce estimates of the average effect of treatment or exposure on outcome within centers and correct confidence intervals around these estimates by accounting for each center specifically in the model.

Clustering: Grouping of patients into centers so that they have correlated outcomes and no longer represent independent observations.

Cluster-randomization design: A designed study in which groups of patients in centers, rather than individual patients, are randomly assigned to treatments or interventions, and the same treatment (or intervention) is applied to all patients in a center.

Conditional analysis: In the context of patients clustered within centers, an analysis of the association of treatment (or exposure) and patient outcome within center.

Confidence interval: A range of values in which the true effect of treatment (or exposure) lies within a given frequency (for example, 95%) over repetitions of the study. A measure of the variance of an estimate.

Confounding factor (confounder): A variable that is related to both the treatment (exposure) of interest and the outcome so that it alters (biases) the measure of association.

Covariate: A factor of clinical interest other than the treatment or exposure being examined.

Design effect: The ratio of the actual (correctly computed) variance to the variance under conditions of simple random sampling from a population of independent observations.

Effect modification: Variation in the strength of association between a treatment (or exposure) and an outcome with the level of a third factor (such as center).

Effective sample size: The size of the study after adjustment for the impact of the design on the variance. Usually lower than the number of patients but greater than the number of centers.

Fixed-center effect: In the context of clustered data, the effect on patient outcome of being in a center that the investigator regards as representing only itself and not other centers in the population of all centers.

Frailty model: A center-specific model applied to survival (time-to-event) analysis.

Population-averaged (or marginal) models: Statistical models that produce estimates of the average effect of treatment or exposure on outcome across centers and corrected confidence intervals around these estimates, but without regard to a patient's membership in a particular center.

Random-center effect: In the context of clustered data, the effect on patient outcome of being treated at (or exposed to) a center that the investigator regards as a sample representing others in the population. The effects of these centers on outcome are often assumed to take on a normal (bell-shaped) distribution, with the overall average effect as the mean of this distribution.

Random-coefficient model: A center-specific model that allows for random variation in the treatment effect across centers.

Sample survey analysis: A set of statistical tools developed originally for the analysis of complex survey data characterized by the sampling of clusters of individuals.

Treatment effect: The estimate of the impact of a treatment or exposure on the patients; expressed commonly as an odds ratio, a relative risk, a risk difference, or a difference in means.

From University of Pennsylvania, Philadelphia, Pennsylvania.

Grant Support: In part by the Agency for Healthcare Research and Quality, Centers for Education and Research on Therapeutics, University of Pennsylvania School of Medicine (U18 HS10399).

Corresponding Author: A. Russell Localio, JD, MS, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 606 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021; e-mail, rlocalio@ceb.upenn.edu.

Current Author Addresses: Drs. Localio, Berlin, Ten Have, and Kimmel: Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 606 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021.

References

1. Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. The implications of adopting a cluster design are still largely being ignored

- [Editorial]. *BMJ*. 1998;317:1171-2. [PMID: 9794847]
2. Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol*. 1994;140:279-89; discussion 300-1. [PMID: 8030631]
 3. Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol*. 1978;108:100-2. [PMID: 707470]
 4. Donald A, Donner A. Adjustments to the Mantel-Haenszel chi-square statistic and odds ratio variance estimator when the data are clustered. *Stat Med*. 1987; 6:491-9. [PMID: 3629050]
 5. Ashby M, Neuhaus JM, Hauck WW, Bacchetti P, Heilbron DC, Jewell NP, et al. An annotated bibliography of methods for analysing correlated categorical data. *Stat Med*. 1992;11:67-99. [PMID: 1557577]
 6. Liang KY, Zeger SL. Regression analysis for correlated data. *Annu Rev Public Health*. 1993;14:43-68. [PMID: 8323597]
 7. Murray DM. Design and Analysis of Group Randomization Trials. New York: Oxford Univ Pr; 1998.
 8. Neuhaus JM. Statistical methods for longitudinal and clustered designs with binary responses. *Stat Methods Med Res*. 1992;1:249-73. [PMID: 1341660]
 9. Hansen MH, Hurwitz WN, Madow WG. Sample Survey Methods and Theory. Volume 1: Methods and Applications. New York: J Wiley; 1953.
 10. Kish L. Survey Sampling. New York: J Wiley; 1965:88.
 11. Goldstein H. Multilevel Statistical Models. London: Edward Arnold; 1995: 1-13.
 12. Andersen J. Methodological Errors in Medical Research: An Incomplete Catalogue. Oxford: Blackwell Scientific; 1990:147-56.
 13. Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979-1989. *Int J Epidemiol*. 1990;19:795-800. [PMID: 2084005]
 14. Simpson JM, Klar N, Donner A. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993. *Am J Public Health*. 1995;85:1378-83. [PMID: 7573621]
 15. Göttsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials*. 1989;10:31-56. [PMID: 2702836]
 16. Dexter PR, Wolinsky FD, Gramelspacher GP, Zhou XH, Eckert GJ, Waisburd M, et al. Effectiveness of computer-generated reminders for increasing discussions about advance directives and completion of advance directive forms. A randomized, controlled trial. *Ann Intern Med*. 1998;128:102-10. [PMID: 9441569]
 17. Pittet D, Mourouga P, Perneger TV. Compliance with handwashing in a teaching hospital. *Infection Control Program*. *Ann Intern Med*. 1999;130:126-30. [PMID: 10068358]
 18. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG, Donner A. Methods in health service research. Evaluation of health interventions at area and organisation level. *BMJ*. 1999;319:376-9. [PMID: 10435968]
 19. Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health Research. London: Arnold; 2000.
 20. Kinsinger LS, Harris R, Qaqish B, Strecher V, Kaluzny A. Using an office system intervention to increase breast cancer screening. *J Gen Intern Med*. 1998; 13:507-14. [PMID: 9734786]
 21. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276:637-9. [PMID: 8773637]
 22. Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. *BMJ*. 1998;316:1455-60. [PMID: 9572764]
 23. Kerry SM, Bland JM. Sample size in cluster randomisation. *BMJ*. 1998;316: 549. [PMID: 9501723]
 24. Laine C, Venditti L, Localio R, Wickenheiser L, Morris DL. Combined cardiac catheterization for uncomplicated ischemic heart disease in a Medicare population. *Am J Med*. 1998;105:373-9. [PMID: 9831420]
 25. Bailar JC 3rd, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. *Ann Intern Med*. 1988;108: 266-73. [PMID: 3341656]
 26. Neuhaus JM, Segal MR. Design effects for binary regression models fitted to dependent data. *Stat Med*. 1993;12:1259-68. [PMID: 8210825]
 27. Pouyanne P, Haramburu F, Imbs JL, Bégaud B. Admissions to hospital caused by adverse drug reactions: cross sectional incidence study. French Pharmacovigilance Centres. *BMJ*. 2000;320:1036. [PMID: 10764362]
 28. Katz MH. Multivariable Analysis: A Practical Guide for Clinicians. Cambridge: Cambridge Univ Pr; 1999.
 29. Fink JC, Gardner JF, Armistead NC, Turner MS, Light PD. Within-center correlation in dialysis adequacy. *J Clin Epidemiol*. 2000;53:79-85. [PMID: 10693907]
 30. Greenland S, Schwartzbaum JA, Finkle WD. Problems due to small samples and sparse data in conditional logistic regression analysis. *Am J Epidemiol*. 2000; 151:531-9. [PMID: 10707923]
 31. Breslow NE, Day NE. Statistical Methods in Cancer Research. Volume I: The Analysis of Case-Control Studies. Lyon: International Agency for Research on Cancer; 1980:248-9.
 32. Marrazzo JM, White CL, Krekler B, Celum CL, Lafferty WE, Stamm WE, et al. Community-based urine screening for *Chlamydia trachomatis* with a ligase chain reaction assay. *Ann Intern Med*. 1997;127:796-803. [PMID: 9382400]
 33. Cox DR, Hinkley DV. Theoretical Statistics. London: Chapman & Hall; 1974:292.
 34. Graubard BI, Korn EL. Regression analysis with clustered data. *Stat Med*. 1994;13:509-22. [PMID: 8023032]
 35. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49:1373-9. [PMID: 8970487]
 36. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med*. 1997;16: 2349-80. [PMID: 9351170]
 37. Sullivan LM, Dukas KA, Losina E. Tutorial in biostatistics. An introduction to hierarchical linear modelling. *Stat Med*. 1999;18:855-88. [PMID: 10327531]
 38. Smith MA, Doliszny KM, Shahar E, McGovern PG, Amett DK, Luepker RV. Delayed hospital arrival for acute stroke: the Minnesota Stroke Survey. *Ann Intern Med*. 1998;129:190-6. [PMID: 9696726]
 39. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988;44:1049-60. [PMID: 3233245]
 40. Ten Have TR, Landis JR, Weaver SL. Association models for periodontal disease progression: a comparison of methods for clustered binary data. *Stat Med*. 1995;14:413-29. [PMID: 7746980]
 41. Ten Have TR, Kunselman A, Zharichenko E. Accommodating negative intracluster correlation with a mixed effects logistic model for bivariate binary data. *J Biopharm Stat*. 1998;8:131-49. [PMID: 9547432]
 42. Fiscella K, Franks P. Influence of patient education on profiles of physician practices. *Ann Intern Med*. 1999;131:745-51. [PMID: 10577297]
 43. Marshall EC, Spiegelhalter DJ. Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. *BMJ*. 1998;316:1701-4; discussion 1705. [PMID: 9614016]
 44. Yamaguchi T, Ohashi Y. Investigating centre effects in a multi-centre clinical trial of superficial bladder cancer. *Stat Med*. 1999;18:1961-71. [PMID: 10440879]
 45. Andersen PK, Klein JP, Zhang MJ. Testing for centre effects in multi-centre

- survival studies: a Monte Carlo comparison of fixed and random effects tests. *Stat Med*. 1999;18:1489-500. [PMID: 10398287]
46. Hedeker D, Siddiqui O, Hu FB. Random-effects regression analysis of correlated grouped-time survival data. *Stat Methods Med Res*. 2000;9:161-79. [PMID: 10946432]
47. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics*. 1998;54:638-45. [PMID: 9629647]
48. Korn EL, Graubard BI. *Analysis of Health Surveys*. New York: J Wiley; 1999.
49. Selhub J, Jacques PF, Rosenberg IH, Rogers G, Bowman BA, Gunter EW, et al. Serum total homocysteine concentrations in the third National Health and Nutrition Examination Survey (1991-1994): population reference ranges and contribution of vitamin status to high serum concentrations. *Ann Intern Med*. 1999;131:331-9. [PMID: 10475885]
50. Bassuk SS, Glass TA, Berkman LF. Social disengagement and incident cognitive decline in community-dwelling elderly persons. *Ann Intern Med*. 1999;131:165-73. [PMID: 10428732]
51. Williams RL. A note on robust variance estimation for cluster-correlated data. *Biometrics*. 2000;56:645-6. [PMID: 10877330]
52. Shekelle PG, Coulter I, Hurwitz EL, Genovese B, Adams AH, Mior SA, et al. Congruence between decisions to initiate chiropractic spinal manipulation for low back pain and appropriateness criteria in North America. *Ann Intern Med*. 1998;129:9-17. [PMID: 9653012]
53. Horton NJ, Lipsitz SR. Review of software to fit generalized estimating equation regression models. *American Statistician*. 1999;53:160-9.
54. Robinson JW, Roter DL. Counseling by primary care physicians of patients who disclose psychosocial problems. *J Fam Pract*. 1999;48:698-705. [PMID: 10498076]
55. Knox SS, Siegmund KD, Weidner G, Ellison RC, Adelman A, Paton C. Hostility, social support, and coronary heart disease in the National Heart, Lung, and Blood Institute Family Heart Study. *Am J Cardiol*. 1998;82:1192-6. [PMID: 9832093]
56. Wei LJ, Glidden DV. An overview of statistical methods for multiple failure time data in clinical trials. *Stat Med*. 1997;16:833-9. [PMID: 9160483]
57. Efron B, Tibshirani R. *Statistical data analysis in the computer age*. Science. 1991;253:390-5.
58. Feldman HI, Kinoshian M, Bilker WB, Simmons C, Holmes JH, Pauly MV, et al. Effect of dialyzer reuse on survival of patients treated with hemodialysis. *JAMA*. 1996;276:620-5. [PMID: 8773634]
59. Diaconis P, Efron B. *Computer intensive methods in statistics*. Scientific American. 1983 May;248:116-30.
60. Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population averaged approaches for analyzing correlated binary data. *International Statistical Review*. 1991;59:25-35.
61. Diggle P, Liang KY, Zeger SL. *The Analysis of Longitudinal Data*. New York: Oxford Univ Pr; 1994.
62. Albert PS. Longitudinal data analysis (repeated measures) in clinical trials. *Stat Med*. 1999;18:1707-32. [PMID: 10407239]
63. Daumit GL, Hermann JA, Coresh J, Powe NR. Use of cardiovascular procedures among black persons and white persons: a 7-year nationwide study in patients with renal disease. *Ann Intern Med*. 1999;130:173-82. [PMID: 10049195]
64. Berlin JA, Kimmel SE, Ten Have TR, Sammel MD. An empirical comparison of several clustered data approaches under confounding due to cluster effects in the analysis of complications of coronary angiography. *Biometrics*. 1999;55:470-6.
65. Ten Have TR, Landis JR, Weaver S. Association models for periodontal disease progression: a comparison of methods for clustered binary data [Letter]. *Stat Med*. 1996;15:1227-9. [PMID: 8804150]
66. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology. The Essentials*. 3rd ed. Baltimore: Williams & Wilkins; 1996:235.
67. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven; 1998:276.
68. Longford NT. *Random coefficient models*. Oxford: Clarendon Press; 1993.
69. Agresti A, Hartzel J. Strategies for comparing treatments on a binary response with multi-centre data. *Stat Med*. 2000;19:1115-39. [PMID: 10790684]
70. Gould AL. Multi-centre trial analysis revisited. *Stat Med*. 1998;17:1779-97; discussion 1799-800. [PMID: 9749447]
71. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. *Stat Med*. 1999;18:1905-42. [PMID: 10532877]
72. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med*. 1999;18:321-59. [PMID: 10070677]
73. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: caveat lector? *J Gen Intern Med*. 1998;13:753-6. [PMID: 9824521]

Analysis of Data from Multiclinic Trials

Joseph L. Fleiss

Division of Biostatistics, Columbia University

ABSTRACT Because the clinics in a multiclinic randomized clinical trial represent neither fixed stratification effects nor random classificatory effects, the appropriate analysis of data from such a trial has been the subject of controversy and debate. The following are some of the elements of controversy that are discussed and for which some bases for resolution are proposed. Is it ever valid to ignore the effects of clinics in the analysis? Is it ever valid to drop clinics from the analysis? Is a multiclinic clinical trial similar in structure or not to a single-clinic clinical trial in which patients have been stratified on a classificatory factor? Assuming that clinics will be taken account of in the analysis, should it be the weighted or the unweighted average of within-clinic treatment differences that is to be taken as the best estimate of the overall difference between the treatments? How should the data be analyzed if there is evidence of treatment-by-clinic interaction?

KEY WORDS: *multiclinic clinical trial, pooling of data, treatment-by-clinic interaction*

INTRODUCTION

A multiclinic trial is one conducted in a number of participating clinics simultaneously. What makes such a trial a single study rather than a series of separate, unrelated studies is the adherence to a common protocol, in theory if not in actuality. Multiclinic trials are at least as popular today as they were in the 1940s and 1950s when Bradford Hill designed the classical controlled multiclinic studies of antihistamines, cortisone, and streptomycin [1]. The major reason for a multiclinic rather than a single-clinic trial is now, as it was then, the need to enroll sufficiently many patients so that the study has adequate power to detect a difference between the treatments being compared. The ability to generalize the study's results to more than one kind of patient and more than one kind of treatment facility is a second important reason.

Given the relatively long history of multiclinic trials, and given the relatively large literature on the planning and execution of such studies [2-9], there is a striking paucity of articles and chapters in books on the analysis of the

Address reprint requests to: Joseph L. Fleiss, Ph.D., Division of Biostatistics, Columbia University, 600 West 168th Street, New York, NY 10032.

Accepted June 18, 1986.

resulting data. This dearth in the literature is especially striking because uncertainty and controversy exist concerning most aspects of analysis. Several problematic areas of data analysis are identified in the article, and criteria are offered for resolving some of the disagreements and ambiguities that exist. Attention is restricted throughout the article to a measured response variable that is assumed to be normally distributed.

It is important to specify at the outset what is probably the key assumption underlying the analysis, criticisms, and recommendations that follow: clinics will vary one from another in the overall mean levels of response of their patients. One set of reasons includes the enrollment of demographically as well as medically different kinds of patients, and the existence of different treatment milieux. A second set of reasons includes different types and levels of departure from the study's protocol, and differing criteria for evaluating response to treatment. Both sets of factors are likely to produce clinic-to-clinic variation in the mean level of response. The second set of factors is, as well, likely to produce treatment-by-clinic interaction (i.e., nonconstancy across the clinics of the differences among the treatments' means), whereas the first set might not.

"POOLING" THE DATA

What may be the most important controversy concerns the meaning and the strategy of "pooling." Some notation will help here. Assume for simplicity that two treatments are being compared in each of C clinics and that the underlying variances are equal for all clinics and for both treatments. Further, assume for now that there is no treatment-by-clinic interaction. Two randomization procedures are considered, one employing separate and independent randomization schedules for the several clinics and the second ignoring the clinics in the random assignment of patients to treatment groups.

The results within a typical clinic, say clinic c , may be summarized as follows.

Treatment	Sample Size	Mean	Standard Deviation
1	n_{c1}	\bar{X}_{c1}	s_{c1}
2	n_{c2}	\bar{X}_{c2}	s_{c2}

To some, "pooling" means "averaging within-clinic differences" and is thus used in the same sense as "pooling variances." Specifically, pooling in this sense means taking

$$\bar{D}_w = \frac{\sum W_c (\bar{X}_{c1} - \bar{X}_{c2})}{\sum W_c} = \frac{\sum W_c D_c}{\sum W_c} \quad (1)$$

as the estimator of $\mu_1 - \mu_2$, the assumed common difference between the treatments' means, where $D_c = \bar{X}_{c1} - \bar{X}_{c2}$ and the W_c s are a set of positive weights; and taking

$$se(\bar{D}_w) = \frac{s_p}{\sum W_c} \sqrt{\sum W_c^2 \frac{n_{c1} + n_{c2}}{n_{c1}n_{c2}}} \quad (2)$$

as its estimated standard error, where $s_p^2 = \sum \sum (n_{ci} - 1) s_{ci}^2 / \sum \sum (n_{ci} - 1)$, the pooled variance with, say, $\nu = \sum \sum (n_{ci} - 1)$ degrees of freedom. The weighting systems most frequently employed either take W_c to be a constant for all c , or take W_c to be a function of n_{c1} and n_{c2} .

To others, however, "pool the data" is a euphemism for "throw together all the responses to a treatment, ignoring the clinics." Specifically, pooling in this second sense means taking

$$\bar{D} = \frac{\sum n_{c1} \bar{X}_{c1}}{n_{.1}} - \frac{\sum n_{c2} \bar{X}_{c2}}{n_{.2}} \quad (3)$$

as the estimator of the difference between the treatments' means, where $n_{.i} = \sum n_{ci}$ for $i = 1$ and 2 , and taking

$$se(\bar{D}) = s^* \sqrt{\frac{n_{.1} + n_{.2}}{n_{.1} n_{.2}}} \quad (4)$$

as its estimated standard error, where $(s^*)^2 = [(n_{.1} - 1)s_1^2 + (n_{.2} - 1)s_2^2] / (n_{.1} + n_{.2} - 2)$, with s_i^2 being the variance of the $n_{.i}$ responses to treatment i for $i = 1$ and 2 .

Under either of the randomization procedures considered here, randomization with regard or without regard to the clinics, both \bar{D}_w and \bar{D} are unbiased estimators of $\mu_1 - \mu_2$. The sampling variation of \bar{D} exceeds that of \bar{D}_w for appropriate choices of weights [10], however, and the standard deviation s^* that is used in analyses based on \bar{D} will tend to be greater than the standard deviation s_p that is appropriate when \bar{D}_w is the estimator (s^* will tend to be inflated relative to s_p because it will be affected by the differences among the clinics' means that are assumed always to exist).

Pooling in the first sense is the correct method of analysis whenever the random assignment of treatments to patients is carried out separately and independently within the clinics, for the analysis will then properly have been dictated by the design. Pooling in the second sense is theoretically valid only when the study design calls for randomizing without regard to clinic, so that it is theoretically possible for most or even all of a clinic's patients to be assigned the same treatment. Such randomization schemes are sometimes employed [3,11,12], mainly in cancer trials in which stratification on stage of disease is considered to be more important than control for the effects of clinics. For trials in diseases other than cancer, randomizing without regard to clinic seems rare. Even in such cases, it would be advisable to summarize the data first within clinics and then to average the treatment differences across the clinics. The strategy would therefore be analogous to post-stratification [13-15], that is, the control only in the analysis, not in the design, for a classificatory factor.

In summary, pooling in the sense of averaging within-clinic differences is almost always justified, and pooling in the sense of throwing together all the data is only rarely justified. When the design calls for separate and independent randomizations within the clinics, pooling in the latter sense is inappropriate on theoretical grounds. When the design calls for randomization without regard to clinics, it may be inefficient. The practice of pooling the data in the sense of lumping them together should generally be avoided, and

the word "pooling," given its two contradictory meanings, should also be avoided or, if not, defined whenever used.

DROPPING CLINICS FROM ANALYSIS

The practice of dropping entire clinics from analysis is not widespread, but it does occur. The protocol of one multiclinic study, for example, stipulated that "each center must have a minimum of 30 . . . enrollees to be included in a pooled analysis." A recent article on multiclinic trials suggested that "an entire institution can be dropped from the study if the percentage of non-valid patients . . . becomes too large or if the entry rate is not high enough" [16, p. 854]. In a personal communication, the senior author explained that the dropping referred both to withdrawal of permission to enroll future patients and to deleting from analysis the data collected on the patients already enrolled.

If enhanced prestige is a consequence of a clinic's having participated in a multiclinic trial, it is understandable that the threat of being dropped may provide an impetus to the enrollment of adequate numbers of patients, and to the enrollment of patients who satisfy the admission criteria enumerated in the protocol. Actually to carry out such a threat is to establish a precedent that may lead to bias. One may imagine a study whose sponsors are not totally blinded to treatment. Noting that the results are going in the "wrong" direction in a certain clinic, they may decide to drop that clinic with respect to recruiting future patients and thereby succeed in having it dropped from analysis because it failed to enroll the prespecified minimum number of patients.

The only valid reason I can think of for dropping a clinic from analysis is that all of its patients had been assigned the same treatment. Otherwise, the aphorism that describes the consensus view concerning which patients to analyze in a clinical trial, "if randomized, then analyzed," should apply to clinics as well.

TAKING THE CLINIC EFFECTS TO BE RANDOM.

A question exists as to just how the clinics are to be taken account of in the analysis, as fixed or as random effects. The most compelling presentation of a model that takes the clinic effects to be random is in a paper by Chakravorti and Grizzle [17]. The resulting statistical model for the obtained data is a mixed model, with the effects of treatments being fixed and the effects of clinics being random. Even though such a model is never strictly valid, inasmuch as the participating clinics are never sampled randomly or even haphazardly from either an infinite or a finite population of clinics [18], it might be tenable when there are relatively many clinics and relatively few patients in each [19]. Under this model, linear combinations of the pooled variance, s_p^2 , and the interaction mean square,

$$\text{IMS} = \frac{\sum_{c=1}^c W_c^* (D_c - \bar{D}_{W^*})^2}{C - 1} \quad (5)$$

with $C - 1$ degrees of freedom, must be used in making inferences about the difference between the treatments' means. In equation (5),

$$W_c^* = \frac{n_{c1}n_{c2}}{n_{c1} + n_{c2}} \quad (6)$$

and

$$\bar{D}_W^* = \frac{\sum W_c^* D_c}{\sum W_c^*} \quad (7)$$

Except under special circumstances, the analytical methods are complicated and the test procedures only approximate [17,18].

Generally, the clinics have more of the earmarks of a fixed factor than of a random factor, and thus analyses assuming a mixed model would be unnecessarily complicated. In fact, clinics are usually deliberately selected because of a reputation for having conducted clinical trials, because of an expectation of success in enrolling the requisite numbers of patients and in adhering to the study protocol, because of the unique kinds of patients they treat, and so on. Were the study to begin at a different time, most if not all of the participating clinics would be the same.

The consensus in the field seems to be that, except in certain circumstances [20], clinics are more appropriately modeled as representing levels of a fixed factor than levels of a random factor. The designation of a factor as fixed is not sufficient, however, to indicate how the data are to be analyzed and the results interpreted. Both classificatory and bona fide experimental factors are considered to represent levels of a fixed factor, after all, but how their effects are defined, estimated, and analyzed depends on whether assignment to one level or another is under the investigator's control [21, pp. 92-93]. Clinics obviously represent a classificatory, not an experimental factor.

ARE CLINICS LIKE STRATA?

Even identifying a factor as classificatory rather than experimental is insufficient for specifying how to conceive of and analyze its effects. Consider Greenberg's suggestion that clinics be considered similarly to such strata as stages of disease [22]. The analogy is an imperfect one, and it is especially weak when there is treatment-by-clinic interaction.

In the brief theoretical development given earlier, suppose that all the measurements were obtained on patients randomly sampled from a single clinic, with the patients having been stratified into C stages of disease. The classical method of analysis [23, pp. 264-265; 24, pp. 150-153] calls for the weighting factor W_c in equation (1) to be set equal to W_c^* in equation (6) for $c = 1, \dots, C$. The resulting test is based on the magnitude of the statistic

$$t = \frac{\sum W_c^* D_c}{s_p \sqrt{\sum W_c^*}} \quad (8)$$

with $\nu = \sum \sum (n_{ci} - 1)$ degrees of freedom.

When there is no treatment-by-stratum interaction, this weighting system is optimal in that the resulting weighted average in equation (1) is the min-

imum variance unbiased estimator of the common within-stratum difference between treatment means. When there is treatment-by-stratum interaction, this weighting system, although no longer optimal in any rigorous sense, is at least valid and defensible in that $W_c^*/\Sigma W_c^*$ is a consistent estimator of the proportion of patients in the study population who are from stratum c [24, p. 162], and thus \bar{D}_{W^*} in equation (7) is a consistent estimator of the difference between the two treatments' means in that population.

Consider, now, the apparently analogous multiclinic study, with C representing, as earlier, the number of clinics. When no treatment-by-clinic interaction exists, the weights given in equation (6) are optimal in the same sense as in a stratified study. When, on the other hand, such interaction exists, those weights can no longer be interpreted as they were in the case of a stratified study. In no easily stated sense is there a definable population of patients that was sampled from. A clinic's W_c^* may be informative about how many eligible patients sought treatment there, but it may be just as informative about how vigorous clinic c was in enrolling and retaining patients. The earlier informativeness of W_c^* about the relative size of stratum c is thus no longer pertinent.

THE DETECTION OF INTERACTION

The most challenging questions in the analysis of the data from a multiclinic trial are how to carry out the analysis when there is treatment-by-clinic interaction, and, prior to that, how to ascertain whether such interaction exists.

The U.S. Food and Drug Administration issued draft guidelines in June 1985 for the format and content of a new drug application's statistics section [25]. The importance it places on treatment-by-clinic interaction is seen in such recommendations as "The documentation of multi-investigator studies should include at a minimum an analysis of variance table with terms for investigators, treatments, (and) their interaction" (p. 19), "(present plots or graphs so as to) convey . . . information about consistency of drug effects across investigators" (p. 20), and "(as) rationale . . . for combining results across investigators, . . . suitable consistency of results across clinics should be documented" (p. 22). The test statistic for interaction in the analysis of variance is the ratio $F = IMS/s_p^2$ with $C - 1$ and ν degrees of freedom, with IMS as defined in equation (5). Unfortunately the draft guidelines provide neither warning nor reassurance as to the significance level at which the test should be performed, nor do they help determine the degree of inconsistency of results across clinics that would make the combination of results unsuitable.

Some statistical reviewers at the FDA have, allegedly, required applicants to consider the problem of interaction—and, perhaps, to take account of it in the analysis—if the F ratio for interaction is significant at the 0.10 level. Even though the chances are one in ten that there will be undue concern about interaction and perhaps even an inefficient analysis when interaction is not really there, the relatively high power that such a criterion provides when in fact there is interaction is reassuring to those who believe that clinical or demographic differences between the clinics' patients make interaction possible, and that loose controls and little or no monitoring of procedures at the

individual clinics make interaction inevitable [10]. I recommend basing the decision whether or not there is interaction on the test of significance described above, with a significance level of 0.10.

(Interactions due to factors inherent in the clinics—their patients or their treatment milieux—are impossible to eliminate, but the gratuitous interactions due to idiosyncratic interpretations or violations of the study's protocol may be eliminated or at least minimized by the adoption of certain procedures. Regular visits to or telephone contacts with each of the clinics, the consequent early detection of protocol violations, and thus the quick imposition of remedial controls are superior to leaving things out of control and thus running the risk of large interactions that may make the study's results uninterpretable. How to carry out the necessary monitoring, and what kinds of problems to look out for, are both known [2,4-6,9].)

ANALYSIS IN THE PRESENCE OF INTERACTION

Despite experimental prevention's theoretical superiority to statistical cure, it may be too expensive to apply or, if applied, ineffective in preventing interaction. Assuming that treatment-by-clinic interaction has been found to be significant, how should the data be analyzed in order to help answer the study's single primary question, Is there a difference between the two treatments' means? As was suggested by Yates over 50 years ago [26], the mean difference that makes sense when there is interaction is $\sum(\mu_{c1} - \mu_{c2})/C$, where μ_{ci} is the underlying mean response to treatment i within clinic c ($i = 1, 2; c = 1, \dots, C$). The hypothesis being tested is that the simple average of the differences $\mu_{c1} - \mu_{c2}$, with no clinic receiving greater or lesser weight than another, is zero. The pertinent test statistic when interaction has been found to exist is

$$t = \frac{\sum D_c}{s_p \sqrt{\frac{1}{W_c^*}}} \quad (9)$$

with ν degrees of freedom, which is equal to the ratio of the mean in equation (1) to the standard error in equation (2) for a weighting system that takes $W_c = 1$ for each value of c . Giving each clinic's estimated mean difference the same weight is easily explained to one's clinical colleagues when the C sample sizes are of comparable magnitudes but may be extremely difficult to defend when some of the sample sizes $n_{c1} + n_{c2}$ are small (2 or 3, for example) and others large (10 or more times greater). Nevertheless, the test based on the statistic in equation (9) is theoretically correct when interaction exists, regardless of the sample sizes.

As an alternative to performing this test when the F ratio for interaction is significant, Overall [27, pp. 69-70] suggested the following procedure: (1) identify and then eliminate from the analysis the clinic that contributes most to the mean square for interaction; (2) check whether the elimination of that clinic succeeds in reducing the magnitude of interaction to nonsignificance; and, if so, (3) test for the significance of the mean difference between treatments by applying the test statistic in equation (8) to the data from the clinics

that remain. If the interaction remains statistically significant, the process would be continued. The Type I error rate for Overall's suggested procedure would be improved by imposing control for multiple comparison artifacts, perhaps by comparing the absolute value of the test statistic in equation (8) to the Scheffé criterion $\sqrt{CF_{\alpha, C, v}}$. The fact that data had to be discarded in order for a consistent, statistically significant difference to be obtained, however, may damage the credibility of the final result.

Yet other procedures have doubtless been proposed or applied when interaction is present, but the relevant literature in which they are presented, analyzed, and criticized does not seem to exist. Perhaps this review article will provide an impetus for publication, debate, and, ultimately, consensus on how best to analyze the data from a multiclinic trial when treatment-by-clinic interaction exists.

This research was supported in part by grant DE 04068 from the National Institute of Dental Research. I benefited greatly from the criticisms of my colleagues Ralph D'Agostino, Edmund Gehan, Peter Imrey, Albert Kingman, Larry Laster, Kuo Lu, Shanna Swan, Sylvan Wallenstein, and Stuart Zimmerman and from two anonymous reviewers and the editors of this journal. Howard Hait drew my attention to reference numbers 25 and 27.

REFERENCES

1. Hill AB: Statistical Methods in Clinical and Preventive Medicine. New York: Oxford University Press, 1962
2. Friedman LM, Furberg CD, DeMets DL: Fundamentals of Clinical Trials, 2nd ed. Littleton, MA: PSG, 1985
3. Lange N, MacIntyre J: A computerized patient registration and treatment randomization system for multi-institutional clinical trials. *Controlled Clin Trials* 6:38-50, 1985
4. Pocock SJ: Clinical Trials: A Practical Approach. New York: Wiley, 1983
5. Schwartz D, Flamant R, Lellouch J: Clinical Trials. London: Academic, 1980
6. Stanley K, Stjernswärd J, Isley M: The Conduct of a Cooperative Clinical Trial. New York: Springer, 1981
7. Sylvester RJ, Staquet MJ: Operational methodology of E.O.R.T.C. cooperative clinical trials. *Chem Oncol* 4:290-294, 1980
8. Tygstrup N: Achieving an adequate sample size: The multicenter trial. In: *The Randomized Clinical Trial and Therapeutic Decisions*, Tygstrup N, Lachin JM, Juhl E, Eds. New York: Marcel Dekker, 1982, pp 213-218
9. Weiss DG, Williford WO, Collins JF, Bingham SF: Planning multicenter clinical trials: A biostatistician's perspective. *Controlled Clin Trials* 4:53-64, 1983
10. Fleiss JL: Multicentre clinical trials: Bradford Hill's contributions and some subsequent developments. *Stat Med* 1:353-359, 1982
11. Lagakos SW, Pocock SJ: Randomization and stratification in cancer clinical trials: An international perspective. In: *Cancer Clinical Trials: Methods and Practice*, Buyse ME, Staquet MJ, Sylvester RJ, Eds. Oxford: Oxford University Press, 1984, pp 276-286
12. Pocock SJ, Lagakos SW: Practical experience of randomization in cancer trials: An international survey. *Br J Cancer* 46:368-375, 1982

13. Grizzle JE: A note on stratifying versus complete random assignment in clinical trials. *Controlled Clin Trials* 3:365-368, 1982
14. McHugh R, Matts J: Post-stratification in the randomized clinical trial. *Biometrics* 39:217-225, 1983
15. Meier P: Statistical analysis of clinical trials. In: *Clinical Trials: Issues and Approaches*, Shapiro SH, Louis TA, Eds. New York: Marcel Dekker, 1983, pp 155-189
16. Sylvester RJ, Pinedo HM, DePauw M, et al: Quality of institutional participation in multicenter clinical trials. *N Engl J Med* 305:852-855, 1981
17. Chakravorti SR, Grizzle JE: Analysis of data from multiclinic experiments. *Biometrics* 31:325-338, 1975
18. Mielke P, McHugh RB: Two-way analysis of variance for the mixed model with disproportionate sub-class frequencies. *Biometrics* 21:308-323, 1965
19. Machin D, Staquet MJ, Sylvester RJ: Advantages and defects of single-center and multicenter clinical trials. In: *Controversies in Cancer: Design of Trials and Treatment*, Tagnon HJ, Staquet MJ, Eds. New York: Masson, 1979, pp 7-15
20. George SL: Practical problems in the design, conduct and analysis of cooperative clinical trials. *Proc IX Int Biomet Conf* 1:217-224, 1976
21. Cox DR: *Planning of Experiments*. New York: Wiley, 1958
22. Greenberg BG: Conduct of cooperative field and clinical trials. *Am Stat* 13(3):13-17, 28, 1959
23. Armitage P: *Statistical Methods in Medical Research*. New York: Wiley, 1971
24. Fleiss JL: *Design and Analysis of Clinical Experiments*. New York: Wiley, 1986
25. Food and Drug Administration. Draft guideline for the format and content of the statistical section of an application. Docket no. 85D-0246, 1985
26. Yates F: The analysis of multiple classifications with unequal numbers in the different classes. *J Am Stat Assoc* 29:51-66, 1934
27. Overall JE: General linear model analysis of variance. In: *Coordinating Clinical Trials in Psychopharmacology: Planning, Documentation, and Analysis*, Levine J, Ed. Washington, D.C.: U.S. Government Printing Office, 1979