# Centralized databases available for describing primary brain tumor incidence, survival, and treatment: Central Brain Tumor Registry of the United States; Surveillance, Epidemiology, and End Results; and National Cancer Data Base[1]

Faith G. Davis, Bridget J. McCarthy,[2] and Mitchel S. Berger

*Department of Biostatistics/Epidemiology, University of Illinois at Chicago, Chicago, IL 60612 (F.G.D, B.J.M.); and Department of Neurosurgery, University of California at San Francisco, San Francisco, CA 94143 (M.S.B.)*

**Characteristics of three databases—the Central Brain Tumor Registry of the United States (CBTRUS) database; the Surveillance, Epidemiology and End Results (SEER) database; and the National Cancer Data Base (NCDB)—containing information on primary brain tumors are discussed. The recently developed population-based CBTRUS database comprises incidence data on all primary brain tumors from 11 collaborating state registries; however, follow-up data are not available. SEER, the population-based gold standard for cancer data, collects incidence and follow-up data on malignant brain tumors only. While not population-based, the NCDB identifies newly diagnosed cases and conducts follow-up on all primary brain tumors from hospitals accredited by the American College of Surgeons. The NCDB is the largest of the three databases and also contains more complete information regarding treatment of these tumors than either the SEER or CBTRUS databases. Additional strengths and limitations of each of these are described, and their judicious use for supporting research, education, and health care planning is encouraged.** *Neuro-Oncology 1, 205–211, 1999 (Posted to Neuro-Oncology [serial online], Doc. 98-22, June 3, 1999. URL <neuro-oncology.mc.duke.edu>)*

Received 2 October 1998, accepted 27 January 1999.

[1]This work was conducted under contract to the Central Brain Tumor Registry of the United States funded by the Pediatric Brain Tumor Foundation of the United States.

[2]Address correspondence and reprint requests to Bridget J. McCarthy, PhD, Department of Biostatistics/Epidemiology, University of Illinois at Chicago, 2121 W. Taylor, M/C 922, Chicago, IL 60612.

[3]Abbreviations used are as follows: CBTRUS, Central Brain Tumor Registry of the United States; SEER, Surveillance, Epidemiology, and End Results; NCDB, National Cancer Data Base; ICDO, International Classification of Disease for Oncology.

[4]Gershman, S.T., Davis, F.G., and McLaughlin, R. (1996) Completeness of reporting for brain and central nervous system neoplasms. Presented at the North American Association of Central Cancer Registries Annual Meeting, Minneapolis, MN, April 17–19.

Brain tumors account for only a small percentage of all cancers (Ries et al., 1998), but the effects of these tumors can be devastating. Rates such as incidence and survival are important measures of the burden of disease in a population, and differing patterns in these rates have provided clues to the etiology of disease (Preston-Martin and Mack, 1996). In addition, variation in the diagnosis, treatment, and care of brain tumors may have a dramatic influence on the prognosis of these patients. Current patterns by tumor subtype of brain tumor incidence, survival, and treatment in the United States have primarily been based on clinical or institutional settings, which are subject to potential referral biases by the systematic inclusion or exclusion of patients with certain characteristics. As such, results may be applicable to similar patient populations but may not be appropriate for describing the disease in the general

population. Data from a few population-based registries of benign and malignant brain tumors, such as the University of Southern California/Los Angeles County Cancer Surveillance Program (Hisserich et al., 1975; Preston-Martin, 1989) and the Greater Delaware Valley Pediatric Tumor Registry (Kramer et al., 1983), have been exceptions to these clinical studies. New technologies for diagnosing and treating brain tumors have been and will continue to be developed, and the impact of better diagnostic procedures on patterns of occurrence, recurrence, and survival in the United States is difficult to assess without extensive population-based data.

Large population-based data sources are needed to provide accurate descriptive statistics, particularly as the number of brain tumors are limited in single clinical settings. Three centralized databases in the United States have the capability of providing statistics for brain and other CNS tumors: CBTRUS,[3] SEER, and NCDB. The objective of this paper is to describe and contrast these databases and their potential uses to develop a better understanding of descriptive statistics that can be generated using these resources.

## Data Sources

CBTRUS is a nonprofit organization that collects population-based incidence data on all primary brain and CNS tumors, benign and malignant, and attempts to utilize data from other resources to more fully describe these tumors. Incidence data are currently available from 11 regions in the U.S. covering the years 1990–1994 (Surawicz et al., 1999). While CBTRUS attempts to provide complete population-based data on all brain and CNS tumors in the U.S., it is limited by individual registry procedures and definitions, such as differences in the types of tumor registries deemed reportable. As an example, one CBTRUS collaborating registry collects benign tumors of the brain and meninges, but does not require reporting of benign tumors of the spinal cord or acoustic nerve (Surawicz et al., 1999). Issues of underascertainment of cases and accuracy of diagnostic classifications are currently being assessed in an effort to develop recommendations for standardizing and improving the quality of brain tumor data received from cancer registries.

SEER is a cancer registry program funded by the National Cancer Institute, and it provides population-based incidence and survival data for all primary malignant cancers. This program began in 1973 and now includes approximately 14% of the U.S. population (Ries et al., 1998). SEER has set the gold standard for cancer data collection internationally and has been used extensively for research, especially through the facilitation of case identification. SEER registries routinely undergo case ascertainment checks and quality control checks through reabstraction studies and through the use of data-editing software to ensure high quality data.

A third centralized database, the NCDB, is supported by the American College of Surgeons and the American Cancer Society and was established to provide hospital-based follow-up data on all primary tumors, benign and malignant, mainly from those institutions accredited by the American College of Surgeons. Few nonaccredited institutions contribute data. Collection of malignant tumors is required for accreditation; however, collection of benign tumors remains voluntary (Standards of the Commission on Cancer, 1998). Although NCDB does not allow incidence rates to be estimated, it does have an extensive description of tumors at diagnosis and includes recurrence and other information on outcomes. While not population-based, this data set is the only large potential source for survival data on benign tumors. This data currently reflects approximately 57% of all cancer patients in the U.S. for the year 1994 (Menck et al., 1997). As this percentage improves, the data will become increasingly valuable and may begin to approximate population-based data. Data quality, although originally not well established, has been improving as greater editing of the data and quality control measures have been instituted (Clive et al., 1995; Smart et al., 1993; Smart et al., 1994).

## Methods

CBTRUS data were compiled from 11 population-based state cancer registries (Surawicz et al., 1999). SEER data were obtained from public use files found on the SEER public use CD-ROM (SEER, 1997). NCDB data were obtained from the American College of Surgeons. A single year common to all data sets (1994) was selected to give the reader a sense of the sample sizes available on an annual basis. Variables from each of the data sets were documented, and distributions of selected demographic and clinical characteristics were compiled using SAS programs (SAS, 1988).

Primary site was divided into three categories defined by the following ICDO (Percy et al., 1990) site codes: intracranial, C70.0 and C70.9–C71.4; skull base, C70.1, C71.5–C71.7, C72.0–C72.5, and C75.1–C75.3; overlapping/not otherwise specified, C71.8–C71.9 and C72.8–C72.9. For the purposes of this report, the SEER site recode of brain and other CNS sites was utilized (SEER, 1997). This recode excluded pituitary and pineal tumors (ICDO site codes C75.1–C75.3) and lymphomas (ICDO histology codes 9590–9970). Benign brain tumors were defined as those with an ICDO fifth digit morphology (behavior) code of 0, while atypical tumors had a code of 1 and malignant tumors had a code of 3. Tumors considered low grade by neurologists (for example, pilocytic astrocytomas) but traditionally coded by the tumor registries in these data sets as malignant were categorized under the malignant category. In the CBTRUS data, the state of Massachusetts did not provide behavior codes for their data ($n$ = 650).

Broad histological groupings of tumors were created using the following ICDO (Percy et al., 1990) histology codes: diffuse astrocytoma, ICDO codes 9410, 9420; anaplastic astrocytoma, ICDO codes 9401, 9411; pilocytic astrocytoma, ICDO code 9421; astrocytoma, not otherwise specified, ICDO code 9400; glioblastoma, ICDO codes 9440–9442; oligodendroglioma, ICDO codes 9450, 9451, 9460; ependymoma, ICDO codes 9391–9394; malignant glioma, ICDO code 9380; neu-

**Table 1.** Characteristics of three tumor registries: CBTRUS, SEER, and NCDB

| Characteristics | CBTRUS | SEER | NCDB |
|---|---|---|---|
| Years | 1990–94 | 1973–94[a] | 1985–94 |
| Regions | Ariz., Colo., Conn., Del., Idaho, Me., Mass., Minn., Mont., N.C., Utah | Conn.; Utah; Iowa; N.M.; Hawaii; Ark.; Atlanta, Ga.; Detroit, Mich.; Los Angeles, San Jose, San Francisco, Calif.; Seattle, Wash. | Mandatory submission of malignant and voluntary submission of benign tumors by hospitals accredited by the American College of Surgeons in the U.S. and Puerto Rico |
| Incidence/follow-up | Incidence | Incidence and follow-up | Follow-up |
| Population/hospital based | Population-based; nonrandom 15% sample of U.S. population | Population-based; nonrandom 14% sample of U.S. population | Hospital-based; estimated to cover approximately 57% of total cases in U.S. |
| Benign/malignant | Both | Malignant | Both[b] |
| Inclusion criteria | Primary brain and CNS tumors selected using the ICDO site codes C70(0:9), C71(0:9), C72(0:9), and C75(1:3) | Primary brain and CNS tumors selected using SEER recode of brain and other CNS (excluding lymphomas, etc., and pituitary and pineal) data from the SEER public use CD-ROM | Primary brain and CNS tumors selected using the ICDO site codes C70(0:9), C71(0:9), C72(0:9), and C75(1:3) |

CBTRUS, Central Brain Tumor Registry of the United States; SEER, Surveillance, Epidemiology and End Results; NCDB, National Cancer Data Base; ICDO, International Classification of Disease for Oncology; Ark., Arkansas; Ariz., Arizona; Calif., California; Colo., Colorado; Conn., Connecticut; Del., Delaware; Ga., Georgia; Me., Maine; Mass., Massachusetts; Mich., Michigan; Minn., Minnesota; Mont., Montana; N.C., North Carolina; N.M., New Mexico; Wash., Washington.

[a]SEER data are now available through 1996.

[b]Benign tumors are reported voluntarily by hospitals.

ronal and mixed, ICDO codes 8680, 8693, 9364, 9490–9491, 9500, 9505–9506; nerve sheath, ICDO codes 9540, 9550, 9560, 9570; meningioma, ICDO codes 9530–9534, 9537–9538; embryonal/primitive/medulloblastoma, ICDO codes 8963, 9470–9473, 9501–9503, 9510, 9443; pituitary adenoma, ICDO codes 8040, 8140, 8146, 8260, 8270–8271, 8280–8281, 8300, 8323, 8333; lymphoma, ICDO codes 9590–9970. These larger categories allowed for general comparisons between the three data sets.

## Results

Characteristics of these three databases are summarized in Table 1. SEER and CBTRUS are population-based data sets that allow estimation of incidence rates, although CBTRUS includes brain and CNS tumors of all behaviors. SEER has a rich 25-year history of data that allows evaluation of trends for both incidence and survival of primary malignant brain tumors from 1973 to the present. These data sets are not mutually exclusive, with some regions—such as Utah and Connecticut—included in all three. Both CBTRUS and NCDB include benign brain and CNS tumors, with the former able to provide incidence rates and the latter able to provide survival rates. Regions which collaborate with CBTRUS systematically attempt to identify all newly diagnosed brain tumor cases, while reporting of benign tumors in the NCDB database is a voluntary decision of the hospitals reporting to this system. While the definition of codes included in rate estimation may vary across standard statistical reports, all three data sets use ICDO codes to describe tumor location, making customized groupings possible.

Most variables available from each of these data resources are summarized in Table 2. Some demographic variables that have a high percentage of informative data are in common to all three data sets: year of birth, race, and gender. Several clinical characteristics with a high percentage of informative data are also in common across these data sets: age at diagnosis, date of diagnosis, primary site, histology, and behavior. While some limited clinical treatment information is available in CBTRUS and SEER, the data from CBTRUS reflect treatment at initial diagnosis and are neither complete nor current. The database with the most extensive clinical and treatment information is NCDB, which contains initial diagnostic information on more recent cases (1990 to 1994) and 5-year follow-up data on cases diagnosed between 1985 and 1989. Use of CBTRUS treatment data is limited to describing patterns of initial treatment, whereas both SEER and NCDB are able to relate treatment characteristics to survival. The extensive clinical data combined with the large number of cases and inclusion of benign tumors make NCDB a unique resource for survival studies.

A comparison of the numbers of cases included in the three databases for the year 1994 is provided in Table 3. NCDB includes the greatest number of cases followed by CBTRUS and then SEER. Data from CBTRUS suggest that tumors coded as benign comprise more than 35% of brain tumors on a population basis, while the lower percentage (22%) of benign brain tumors in NCDB suggests an underreporting of benign tumors based on the voluntary submission of benign data. This proportion is more reasonable if hospitals that did not voluntarily report benign brain tumors to NCDB are excluded from the total, with the proportion of the total number of brain tumors in these hospitals being 28.2% benign and 68.5% malignant. However, the number of glioblastomas identified in one year by NCDB is over four times that of CBTRUS and almost seven times that of SEER. By definition, the numbers

**Table 2.** Comparison of variables available in each of three tumor registries: CBTRUS, SEER, and NCDB

| Variable | CBTRUS[a] | SEER[a] | NCDB[a] |
|---|---|---|---|
| Case/patient ID | A | A | |
| Type of reporting source[b] | A | A | |
| County of residence at diagnosis | E | A | |
| State of residence at diagnosis | C | | A |
| Place of birth | E | C | E |
| Date of birth | A | A (year only) | A |
| Age at diagnosis | A | A | A |
| Race | A | A | A |
| Spanish origin | E | A | C |
| Gender | A | A | A |
| Marital status | E | A | |
| Religion | E | | E |
| Census tract | E | | |
| Income | | B | |
| % Rural[c] | | B | |
| Accession year[d] | | A | |
| Sequence number[e] | E | A | |
| Date of diagnosis | A | A | A (day = C) |
| Primary site | A | A | A |
| Histology | A | A | A |
| Behavior | C[f] | A | A |
| Grade | D | | E |
| Laterality | D | A | A |
| Extent of disease[g] | E | E | E |
| Class[h] | | | A |
| Diagnostic confirmation[i] | A | A | |
| Summary stage[j] | D | E | E |
| Site-specific surgery | E | D | B |
| Reason no cancer-directed surgery | E | A | B |
| Radiation | E | A | B |
| Radiation to CNS | E | A | |
| Radiation sequence with surgery | E | A | C |
| Chemotherapy | E | | B |
| Hormone therapy | E | | A |
| Biological response modifier[k] | E | | B |
| Other treatment | E | | B |
| Residual tumor | E | | D |
| Cancer status[l] | E | | D |
| Recurrence date | | | E |
| Recurrence type | | | E |
| Site of metastases | E | | D |
| Quality of survival | E | | E |
| Vital status | C | A | A |
| Cause of death | | A | E |
| Date of last contact | E | | B |
| Hospital approval category[m] | | | B |
| AHA category[n] | | | A |
| Annual cancer caseload[o] | | | A |

CBTRUS, Central Brain Tumor Registry of the United States; SEER, Surveillance, Epidemiology and End Results; NCDB, National Cancer Data Base.

[a]The percent of records with informative data are A, >95%; B, >90%; C, >80%; D, >70%; E, ≤70%.

[b]Type of reporting source: Indicates source documents used to abstract the cancer being reported rather than the source of the original casefinding.

[c]% Rural: Percent of people in patient's residence area living in rural areas as defined by the census.

[d]Accession year: Year the patient was first seen for the tumor described in this record.

[e]Sequence number: The sequencing of reportable neoplasms in the patient's lifetime, according to the information and rules of the central registry.

[f]Massachusetts did not provide behavior codes for this data. Exclusion of Massachusetts from the data would result in >95% valid codes.

[g]Extent of disease: Site-specific codes for anatomic extent of disease, including tumor size, extension, lymph node involvement, regional nodes positive, and regional nodes examined. Only tumor size and extension are coded for tumors of the brain and CNS.

[h]Class: Nature of the facility's involvement with the case.

[i]Diagnostic confirmation: Best method of confirmation of the cancer being reported.

[j]Summary stage: Stage (in situ, local, regional, distant, etc.) at initial diagnosis or treatment.

[k]Biological response modifier: Indicates treatment with immunotherapy.

[l]Cancer status: Indicates whether evidence of this cancer is present at the date of last contact.

[m]Hospital approval category: Commission on Cancer program approval categories.

[n]AHA category: American Hospital Association control codes.

[o]Annual cancer caseload: Facility's estimate of its annual caseload.

of primarily benign histologies, such as meningioma and nerve sheath tumors, in the SEER data are scarce.

Table 4 shows the wealth of numbers in the NCDB data set for selected tumor characteristics that have accumulated in a relatively short period of time. For example, there are over 1,000 tumors in each of the four grades of astrocytoma classified as not otherwise specified. One must question the accuracy of the appearance of low grades in the glioblastoma category, however, and in tumor categories where grades are not commonly used (for example, nerve sheath tumors). Some informed assumptions and recoding of data may be necessary to increase the utility of this resource, particularly as some concerns about the quality of the data and how it can be interpreted have been raised due to problems with the uniformity of standard references and the use of different coding schemes among software systems and registry groups (Clive et al., 1995).

## Discussion

A key issue with regard to data quality is complete ascertainment of tumors. Two types of underascertainment may be occurring: underreporting (the number of cases not identified) and definitional differences (cases eligible for inclusion). Underascertainment occurs in most registries, but because of definitional issues, it may be particularly acute for brain tumors. A study in Massachusetts documented a rate of underreporting of brain tumor cases of 23.3%, over half of which were benign tumors.[4] Incidence rate patterns may be influenced if underreporting varies by demographic factors, such as race. On the other hand, if all brain tumors are identified, then the overall incidence picture will emerge, even though there may be some misclassification between categories of tumors.

**Table 3.** Distribution of certain variables from each of three tumor registries for the year 1994: CBTRUS, SEER, and NCDB

| Variable | CBTRUS[a] (*n* = 4,801) | SEER[a,b] (*n* = 1,520) | NCDB[a] (*n* = 15,887) |
|---|---|---|---|
| Gender | | | |
| Male | 2,385 (49.7) | 856 (56.3) | 8,145 (51.3) |
| Female | 2,416 (50.3) | 664 (43.7) | 7,742 (48.7) |
| Age at diagnosis | | | |
| 0–44 | 1,543 (32.1) | 574 (37.8) | 5,183 (32.6) |
| 45–64 | 1,508 (31.4) | 408 (26.8) | 5,022 (31.6) |
| 65–74 | 987 (20.6) | 312 (20.5) | 3,354 (21.1) |
| 75–84 | 630 (13.1) | 178 (11.7) | 1,963 (12.4) |
| 85+ | 133 (2.8) | 48 (3.2) | 365 (2.3) |
| Race | | | |
| White | 4,347 (90.5) | 1,336 (87.9) | 14,150 (89.1) |
| Black | 269 (5.6) | 103 (6.8) | 1,193 (7.5) |
| Other/unknown | 185 (3.9) | 81 (5.3) | 544 (3.4) |
| Site | | | |
| Intracranial | 2,764 (57.6) | 884 (58.2) | 9,250 (58.2) |
| Skull base/spinal | 1,211 (25.2) | 227 (14.9) | 3,323 (20.9) |
| Overlapping/NOS | 826 (17.2) | 409 (26.9) | 3,314 (20.9) |
| Behavior[c] | | | |
| Benign | 1,638 (39.5) | 0 | 3,950 (24.9) |
| Atypical | 183 (4.4) | 0 | 468 (2.9) |
| Malignant | 2,330 (56.1) | 1,520 (100) | 11,469 (72.2) |
| Histology[d] | | | |
| Diffuse astrocytoma | 59 (1.2) | 31 (2.0) | 159 (1.0) |
| Anaplastic astrocytoma | 202 (4.2) | 109 (7.2) | 905 (5.7) |
| Pilocytic astrocytoma | 83 (1.7) | 65 (4.3) | 305 (1.9) |
| Astrocytoma, NOS | 338 (7.0) | 183 (12.0) | 1,532 (9.6) |
| Glioblastoma | 1,119 (23.3) | 694 (45.7) | 4,674 (29.4) |
| Oligodendroglioma | 198 (4.1) | 98 (6.4) | 585 (3.7) |
| Ependymoma | 120 (2.5) | 40 (2.6) | 367 (2.3) |
| Malignant glioma | 118 (2.5) | 79 (5.2) | 568 (3.6) |
| Neuronal and mixed | 60 (1.2) | 3 (0.2) | 139 (0.9) |
| Nerve sheath | 322 (6.7) | 5 (0.3) | 483 (3.0) |
| Meningioma | 1,193 (24.8) | 19 (1.3) | 2,803 (17.6) |
| Embryonal | 97 (2.0) | 59 (3.9) | 350 (2.2) |
| Pituitary adenoma | 298 (6.2) | [b] | 718 (4.5) |
| Lymphoma | 224 (4.7) | [b] | 1,049 (6.6) |

See Table 2 for abbreviations; NOS, not otherwise specified.

[a]Number in parentheses is percent of total reported by that agency.

[b]SEER does not include lymphomas (ICDO histology codes 9590–9970) or pituitary or pineal tumors (ICDO site codes C75.1–C75.3) in the brain and other CNS site recode.

[c]In the CBTRUS data set, the state of Massachusetts provided no behavior codes (*n* = 650). Percentages were based on a CBTRUS sample size of 4,151 for this variable.

[d]Not all histologies are presented.

The proportion of benign to malignant tumors registered can be used as a crude indicator of completeness for all primary brain tumors. In the early 1970s, Schoenberg estimated that 30% of brain tumors were benign (Schoenberg et al., 1976). In contrast, at the Mayo Clinic where there was a high autopsy rate, the number of benign tumors equaled the number of malignant tumors (Kurland et al., 1982). In a comparison of CBTRUS data to SEER data, it has been estimated that, with the change in diagnostic procedures which have taken place in the last two decades, the proportion of benign tumors is at least 50% that of malignant tumors, not considering autopsy diagnosis (Davis et al., 1996). Therefore, underascertainment would be more likely to affect incidence rate patterns of nonaggressive tumors such as meningiomas, acoustic neuromas, and pituitary tumors. Alternatively, unknowingly counting the recurrence of benign tumors many years after initial diagnosis may result in an artificially increased incidence rate. Survival rates may also be lower if underreporting of asymptomatic tumors

**Table 4.** Distribution of selected clinical characteristics in patients with selected histologies from the National Cancer Data Base, 1985-94

| Variable | Astrocytoma, NOS[a] (*n* = 15,365) | Glioblastoma[a] (*n* = 30,224) | Oligo-dendroglioma[a] (*n* = 3,074) | Ependymoma[a] (*n* = 2,371) | Malignant glioma[a] (*n* = 4,063) | Nerve sheath[a] (*n* = 3,308) | Meningioma[a] (*n* = 17,850) |
|---|---|---|---|---|---|---|---|
| Primary site | | | | | | | |
| Intracranial | 10,260 (66.8) | 22,661 (75.0) | 2,385 (77.6) | 269 (11.3) | 1,999 (49.2) | 116 (3.5) | 14,532 (81.4) |
| Skull base | 2,056 (13.4) | 606 (2.0) | 166 (5.4) | 1,783 (75.2) | 1,203 (29.6) | 2,951 (89.2) | 1,661 (9.3) |
| Overlapping | 3,049 (19.8) | 6,957 (23.0) | 523 (17.0) | 319 (13.5) | 861 (21.2) | 241 (7.3) | 1,657 (9.3) |
| Grade | | | | | | | |
| I | 1,748 (11.4) | 155 (0.5) | 341 (11.1) | 310 (13.1) | 231 (5.7) | 28 (0.8) | 324 (1.8) |
| II | 3,845 (25.0) | 219 (0.7) | 684 (22.3) | 163 (6.9) | 275 (6.8) | 16 (0.5) | 99 (0.6) |
| III | 4,296 (28.0) | 3,202 (10.6) | 273 (8.9) | 86 (3.6) | 379 (9.3) | 26 (0.8) | 90 (0.5) |
| IV | 2,158 (14.0) | 12,813 (42.4) | 447 (14.5) | 149 (6.3) | 531 (13.1) | 8 (0.2) | 80 (0.4) |
| Other/unknown | 3,318 (21.6) | 13,835 (45.8) | 1,329 (43.2) | 1,663 (70.1) | 2,647 (65.1) | 3,230 (97.6) | 17,257 (96.7) |
| Tumor size | | | | | | | |
| 0–25 mm | 1,260 (8.2) | 1,826 (6.0) | 239 (7.8) | 271 (11.4) | 248 (6.1) | 921 (27.8) | 2,402 (13.4) |
| 25–49 mm | 2,342 (15.2) | 5,481 (18.1) | 472 (15.3) | 320 (13.5) | 548 (13.5) | 572 (17.3) | 3,526 (19.8) |
| 50–100 mm | 1,829 (11.9) | 5,340 (17.7) | 461 (15.0) | 195 (8.2) | 375 (9.2) | 144 (4.4) | 2,394 (13.4) |
| Other/unknown | 9,934 (64.7) | 17,577 (58.2) | 1,902 (61.9) | 1,585 (66.8) | 2,892 (71.2) | 1,671 (50.5) | 9,528 (53.4) |
| Extension[b] | | | | | | | |
| Supratentorial | 2,132 (14.4) | 5,206 (17.3) | 583 (19.1) | 88 (6.6) | 365 (9.6) | 27 (4.0) | 1,372 (8.8) |
| Infratentorial | 309 (2.1) | 215 (0.7) | 25 (0.8) | 30 (2.2) | 188 (4.9) | 33 (4.9) | 87 (0.5) |
| Ventricles | 397 (2.7) | 797 (2.6) | 78 (2.6) | 134 (10.0) | 105 (2.7) | 8 (1.2) | 200 (1.3) |
| Infra/supra | 21 (0.1) | 52 (0.2) | 2 (0.1) | 4 (0.3) | 18 (0.5) | 4 (0.6) | 5 (0.0) |
| Extension | 55 (0.4) | 215 (0.7) | 13 (0.4) | 48 (3.6) | 22 (0.6) | 15 (2.2) | 353 (2.3) |
| Other/unknown | 11,898 (80.3) | 23,644 (78.5) | 2,354 (77.0) | 1,031 (77.2) | 3,119 (81.7) | 580 (87.0) | 13,617 (87.1) |

NOS, not otherwise specified.

[a]Number in parentheses is percent of total reported by the National Cancer Data Base for those years, except where noted in footnote.

[b]Applies only to tumors with ICDO site codes C70.0 and C71.0-C71.9. Percentages are calculated using the total number of cases of each histology with these site codes: *n* = 14,812, astrocytoma, NOS; *n* = 30,129, glioblastoma; *n* = 3,055, oligodendroglioma; *n* = 1,335, ependymoma; *n* = 3,817, malignant glioma; *n* = 667, nerve sheath; *n* = 15,634, meningioma.

occurs more commonly than does underreporting for clinically relevant tumors. Much of the variation in these data sets reflects differences in reporting requirements for brain and CNS tumors, especially nonmalignant tumors.

A number of limitations with respect to the quality of information are also present. Of necessity, these data are less detailed than clinical records are with respect to tumor characteristics and treatment information that are essential for care of the individual. However, for the purposes of surveillance, these data resources are quite robust. Diagnosis of tumor types may vary across institutions and some misclassification by tumor type can be expected. There is also substantial missing or invalid data on some items within each data set—such as place of birth, religion, extent of disease, summary stage, some treatment variables, and recurrence date and type (Table 2)—that limits the generalizability of data pertaining to that item with respect to the population. The American College of Surgeons frequently adds variables to the data items collected by NCDB, with the result that data validity for these new variables may be low. Data validity for long-established variables (such as histology, site, behavior, etc.), however, remains high (Table 2). The usefulness of these variables in describing treatment patterns and recurrence of brain tumors would be greatly improved if more hospital and regional registries accurately collected this information. There also may be some overestimation of incidence rates in the CBTRUS data as a result of personal identifiers not being available to allow for checking of duplicate cases that may be reported from several sources. While personal identifiers are not available in NCDB, a duplicate codes variable is present that allows for the removal of multiple records. This should not be a problem in the SEER data. The lack of personal identifiers in these data sets also limits the ability to use these data resources for studies that may require further information gathering or record linkages. While SEER and CBTRUS cases may be identified through collaboration with the originating central registries, the NCDB database has been constructed in such a way as to prevent any identification of subjects.

These data sets demonstrate the potential value of centralized data collection efforts for compiling descriptive information on incidence, survival, and treatment patterns of rare diseases. However, the potential for underascertainment of cases does limit the interpretation of incidence patterns, and inaccuracies in important clinical characteristics limit all uses of these data. For example, underascertainment of meningiomas may underestimate the true incidence and inappropriately lower the survival rates for these tumors. Efforts to develop editing programs and training programs specific for these tumors for tumor registry and central registry staff may increase the accuracy of this information over time. The addition of

other relevant variables to these data sets, such as occupation, industry, and social class, may allow for further determination of etiologic risk factors associated with primary brain tumors. In the meantime, utilization of these databases can be informative, providing their strengths and limitations are appreciated. While no one data resource provides a complete description of brain and CNS tumors, the judicious use of information from these resources may be informative, and their use for supporting research, education, and health care planning is encouraged.

## Acknowledgments

The authors thank the NCDB Data Request Committee and Herman Menck at the American College of Surgeons for providing the NCDB data set. We thank Jay Hurh and Patti Jukich for their assistance with the data for these tables. The authors would also like to express their appreciation to the CBTRUS collaborators and the many hospital tumor registrars, without whose diligent efforts these data would not be available to the scientific community.

## References

Clive, R.E., Ocwieja, K.M., Karnell, L.H., Hoyler, S.S., Seiffert, J.E., Young, J.L., Henson, D.E., Winchester, D.P., Osteen, R.T., Menck, H.R., and Fremgen, A. (1995) A national quality improvement effort: Cancer registry data. *J. Surg. Oncol.* **58**, 155–161.

Davis, F.G., Malinski, N., Haenszel, W., Chang, J., Flannery, J., Gershman, S., Dibble, R., and Bigner, D.D. (1996) Primary brain tumor incidence rates in four United States regions, 1985–1989: A pilot study. *Neuroepidemiology* **15**, 103–112.

Hisserich, J.C., Martin, S.P., Henderson, B.E. (1975) An areawide cancer reporting network. *Public Health Rep.* **90**, 15–17.

Kramer, S., Meadows, A.T., Jarrett, P., Evans, A.E. (1983) Incidence of childhood cancer: Experience of a decade in a population-based registry. *J. Natl. Cancer Inst.* **70**, 49–55.

Kurland, L.T., Schoenberg, B.S., Annegers, J.F., Okazaki, H., and Molgaad, C.A. (1982) The incidence of primary intracranial neoplasms in Rochester, Minnesota, 1935–1977. *Ann. N. Y. Acad. Sci.* **381**, 6–16.

Menck, H.R., Cunningham, M.P., Jessup, J.M., Eyre, H.J., Winchester, D.P., Scott-Connor, C.E., and Murphy, G.P. (1997) The growth and maturation of the National Cancer Data Base. *Cancer* **80**, 2296–2304.

Michaelis, J. (1995) Towards nation-wide cancer registration in the Federal Republic of Germany. *Ann. Oncol.* 6, 344–346.

Percy, C., VanHolten, V., and Muir, C.M. (Eds.) (1990) *International Classification of Diseases for Oncology.* Second edition. Geneva: World Health Organization.

Preston-Martin, S. (1989) Descriptive epidemiology of primary tumors of the brain, cranial nerves and cranial meninges in Los Angeles County. *Neuroepidemiology* **8**, 283–295.

Preston-Martin, S., and Mack, W.J. (1996) Neoplasms of the nervous system. In: Schottenfeld, D., and Fraumeni, J.F., Jr. (Eds.), *Cancer Epidemiology and Prevention.* Second edition. New York: Oxford University Press. pp. 1231–1281.

Ries, L.A.G., Kosary, C.L., Hankey, B.F., Miller, B.A., Edwards, B.K. (Eds.) (1998) *SEER Cancer Statistics Review, 1973–1995.* Bethesda, Md.: National Cancer Institute.

SAS (1988) *SAS Language and Procedures.* Usage, Version 6. First edition. Cary, NC: SAS Institute.

Schoenberg, B.S., Christine, B.W., and Whisnant, J.P. (1976) The descriptive epidemiology of primary intracranial neoplasms: The Connecticut experience. *Am. J. Epidemiol.* **104**, 499–510.

Smart, C.R., Eberle, C., and Simon, S. (1993) An assessment of National Cancer Data Base quality control. In: Steele, G.D., Winchester, D.P., Menck, H.R., and Murphy, G.P. (Eds.), *National Cancer Data Base: Annual Review of Patient Care, 1993.* Atlanta: American Cancer Society. pp. 103–105.

Smart, C.R., Karnell, L.H., Eberle, C., Zippin, C., Lum, D., Clive, R., Ocwieja, K.Z., the NCDB Data and Quality Management Working Group, and the Quality Control Patient Care Evaluation Study Subcommittee. (1994) In: Steele, G.D., Winchester, D.P., Osteen, R.T., Menck, H.R., and Murphy, G.P. (Eds.), *National Cancer Data Base: Annual Review of Patient Care, 1994.* Atlanta: American Cancer Society. pp. 117–126.

*Standards of the Commission on Cancer, Volume II: Registry Operations and Data Standards.* (1998) Chicago: Commission on Cancer, American College of Surgeons.

Surawicz, T.S., McCarthy, B.J., Kupelian, V., Jukich, P.J., Bruner, J.M., Davis, F.G., and the collaborating registries of the Central Brain Tumor Registry of the United States. (1999) Descriptive epidemiology of primary brain and CNS tumors: Results from the Central Brain Tumor Registry of the United States, 1990–1994. *Neuro-Oncology* [serial online], Doc. 98-13, January 19, 1999. URL neuro-oncology.mc.duke.edu.

SEER (Surveillance, Epidemiology, and End Results) (1997) *SEER 1973–1996 Public-Use CD-ROM.* (1996), National Cancer Institute, DCPC, Surveillance Program, Cancer Statistics Branch. Based on the August 1996 submission.