Pathol. Oncol. Res. (2016) 22:129–133
DOI 10.1007/s12253-015-9982-z

ORIGINAL ARTICLE

The Establishment and Utility of a Free Online Database of Primary Bone Tumors

Hairong Xu[1] · **Robert P. Seifert**[2] · **Xiaohui Niu**[1] · **Yuan Li**[1] · **Marilyn M. Bui**[2,3]

Received: 18 January 2015 / Accepted: 1 September 2015 / Published online: 16 September 2015

Keywords Online database · Bone tumors · Establishment · Epidemiology · Tool

Introduction

Primary bone tumors are rare. It is estimated that 3010 individuals (1680 male and 1330 female) were diagnosed with malignancies of bone and joints in 2013 in the United States. [1] The yearly incidence is about 0.9 per 100,000. In comparison, the 2013 incidence of prostate and breast cancer was estimated to be 238,590 and 234,580 respectively. [1] The rarity of this entity together with the non-specific clinical symptoms and complex imaging findings, make this a challenging diagnosis for clinicians. [2–4] However, early and accurate diagnosis is the key to appropriate treatment and better clinical outcome.

Primary bone tumors follow a somewhat rigid age-specific distribution. [5, 6] For instance, between the ages of 10–20 years, there is an incidence peak for osteochondroma and osteosarcoma. [7] Beyond age 40 years, there is a steady increase in chondrosarcoma, myeloma and lymphoma. [7] The use of epidemiological data, including the patient's age, sex and tumor location, has been the standard 1st approach to establishing a differential diagnosis. This information is augmented by radiological and histological studies to render an accurate diagnosis. However, in many instances, comprehensive epidemiology data can indicate a definite diagnosis before obtaining diagnostic tissue via an invasive procedure. [8, 9] Although epidemiological information on bone and soft tissue tumors can be gathered from literature searches and published sources, an online source would provide greater convenience. In addition, a patient database could provide a real-time and up-to-date source unlike literature and textbook searches. [10–13] To the best of our knowledge, based on Chinese and English literature searches, no such database has so far been described.

A database is defined as a data structure composed of organized and interrelated data. They are designed to facilitate the collection of data and support the acquisition of relevant information. [14] Here, we describe our recently built, free, online database of primary bone tumors, including its construction and the main practical utility. This may serve as a reference for users interested in constructing a free online database. It will also be useful for sarcoma physicians and researchers alike.

Materials and Methods

Data Resource and Extraction

Ji Shui Tan Hospital (JST) in Beijing, China, a healthcare facility with 1000 beds, specializes in musculoskeletal diseases. The Orthopaedic Oncology Surgery Department was established in the 1970s with its mission to diagnose and treat primary bone tumors. From 1973 to 2012, nearly 10,000 patients with primary bone tumors were treated at JST. Data from 9200 patients, including demographic features (such as

✉ Xiaohui Niu
niuxiaohui@263.net

✉ Marilyn M. Bui
Marilyn.Bui@moffitt.org

[1] Department of Orthopedic Oncology Surgery, Beijing Ji Shui Tan Hospital, Peking University, Beijing, China

[2] Department of Pathology and Cell Biology, University of South Florida College of Medicine, Tampa, FL, USA

[3] Anatomic Pathology Department, H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, Tampa, FL 33612, USA

age, sex, location, diagnosis), treatment protocols (such as chemotherapy regimens, surgical managements, surgical margin), and prognosis (such as recurrence, systemic relapse, state of being, functional outcome) have been collected and stored in one FoxPro (Microsoft, Inc) database, which is the clinical database at JST. All of the diagnoses of primary bone tumors were histologically confirmed by two senior sarcoma pathologists, who were trained in Canada or the United States. The diagnostic criteria published in the WHO Classification of Bone and Soft Tissue Tumor (2002) was followed. [15].

To demonstrate the epidemiologic characteristics of primary bone tumors, the first step in our database design was deciding what parameters to choose. After a review of the literature, age, sex, location, and diagnosis were believed to be the most significant factors contributing to the epidemiology characteristics. [2, 5, 6, 8, 16–19] Therefore, an Excel spreadsheet encompassing the above information was extracted from the FoxPro database. To protect the patients' personal health information, data transferred to the spreadsheet was de-identified.

Database and Website Design

WordPress (WordPress Foundation) a free, open-source blogging tool and full content management system based on PHP and MySQL [20, 21], was used for web site design. WordPress includes many free themes, thus saving costs needed for web page design. For the primary bone tumor online database, the CyperChimps "Responsive Theme" was chosen as the template (Fig. 1). Indeed, thousands of free themes, plugins and widgets provided by the WordPress community (http://wordpress.org) allow users to create as many functional modules as needed.

Although the WordPress web design program is free, a commercial web hosting service is still recommended to ensure the quality of services. For our database, Bluehost (http://www.bluehost.com) was chosen to facilitate the web hosting management and domain naming.

Design of Search Engine and Database Connection

Through the web management backend system, the de-identified spreadsheet data was easily imported into the website's MySQL database. Age and sex were given continuous variable and dichotomous variable attributions, respectively. Both tumor location and diagnosis were attributed with string variables. Locations included femur, tibia, humerus, etc., in long bones and scapular, pelvis, clavicle, etc. in flat bones. Locations for long bones were further subcategorized into proximal (prox), shaft (shaft), and distal (dist).

After a detailed survey among 20 musculoskeletal tumor surgeons, 4 search modules were suggested, including diagnosis, age, location, and combination of age, gender, location, and diagnosis (named complex search). One PHP-based search engine script was programed to take the user's requests, connect to the database, and then display the results. The database was updated with new cases via the backend management system, as before.

To monitor user statistics, we utilized a free website tracker system provided by Google Analytics (http://www.google.com/analytics/). This tracking system helped the website administrator analyze the information regarding the database users, potentially providing ideas for further improvements to the website.

Results

We established a free online database of primary bone tumors (http://www.sarcoma-jst.org). This database is easily accessed on Apple or Windows computers or any type of tablet or

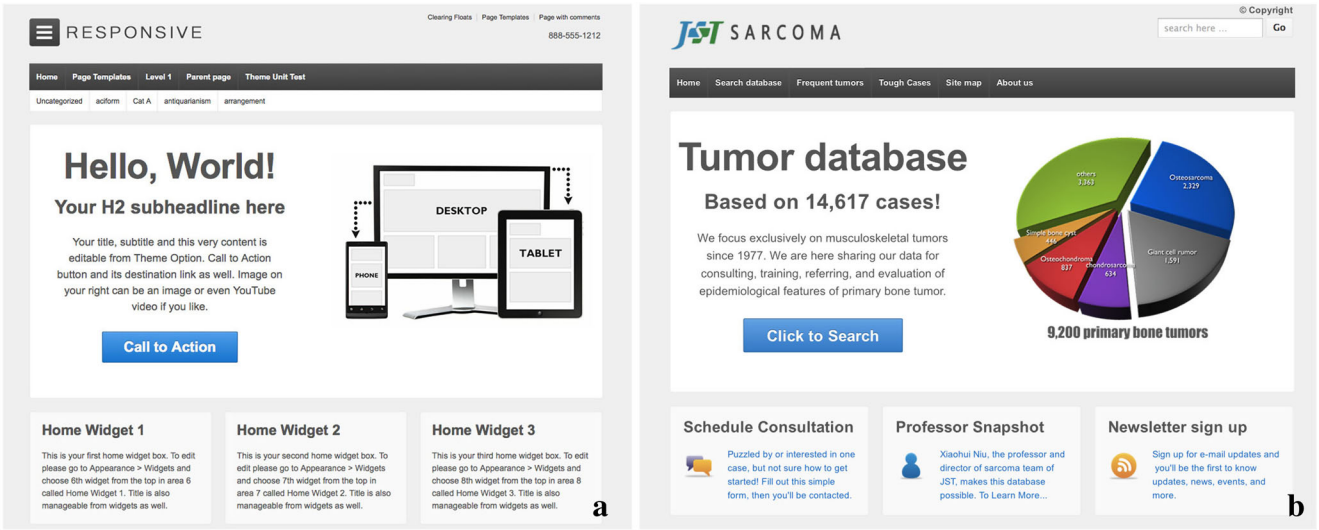

Fig. 1. **a** Original "Responsive" website design template. **b** Index page of our database website, after relevant template text and figures were replaced

modern mobile phone. The total cost of establishing this online database was estimated to be $500, with the maintenance fee estimated to be around $60 annually.

Search by Diagnosis

Users can search by diagnosis by typing the correct name of the tumor into the text entry field using standard terminology as defined by the WHO Classification of Bone and Soft Tissue Tumor (2002). [15] To help users select the correct terminology, a complete companion list including the names of all primary tumors is provided on the website. For example, if a user inputs "osteosarcoma" into the search field, the database will exhibit how many cases have been identified followed by figures showing age and gender distributions, as well as a table of location distribution (Fig. 2 and Table 1). One advantage of this search strategy is that the users can retrieve cases of osteosarcoma subtypes such as conventional osteosarcoma by inputting convention osteosarcoma into the text-entry field.

Search by Age

For age-related searches, the database was designed to output cases ±5 years of the input age. For example, if the user types 23 into the text-entry field, the database will show all patients between ages 18 and 28 years. In addition to showing how many cases have been identified, the database will also display figures of age and gender distributions, as well as tables of location and diagnosis distributions.

Search by Location

To search by location, the database has a pull-down menu designed for choosing one certain location. That is, when a long bone such as femur or tibia is chosen, an additional pop-up menu will allow the user to further select distal, proximal,

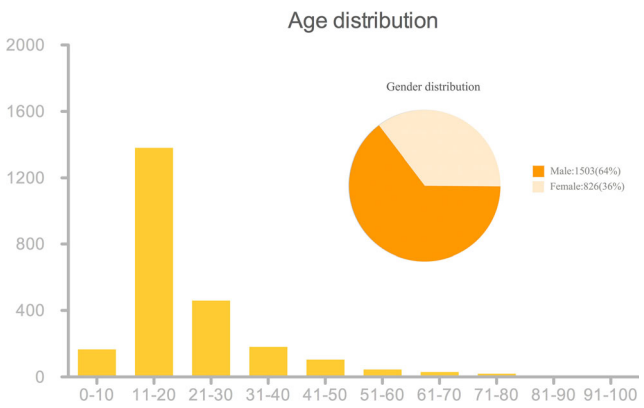

Fig. 2 Age and gender distribution graph, in response to real search of "osteosarcoma" in our online database. As shown here, osteosarcoma appears to be mostly occur in the second decade of life and shows a slightly male predilection

Table 1 Location distribution shown in the form of a table in response to an online search. Percent distributions are given in decreasing order

Location distribution			
Location		N of cases	Percentage
femur	dist	1049	45.04 %
	prox	130	5.58 %
	shaft	62	2.66 %
tibia	prox	555	23.83 %
	dist	56	2.4 %
	shaft	6	0.26 %
humerus	prox	206	8.84 %

or shaft. For example, by selecting femur and distal, the database will show 2207 cases, followed again by figures showing age and gender distributions, as well as one table showing diagnosis distribution.

Complex Searches

The "complex search" was designed to help users analyze epidemiologic characteristics of a select group of cases defined by age, gender, location, and diagnosis. For example, as shown in Fig. 3, by selecting "humerus, distal" and "female" as the first condition and "humerus, proximal" and "female" as the second condition, leaving the entry fields of age and diagnosis blank, the system will identify female patients with tumors at humeral ends, with demographic data again shown through figures and tables. Through this search module, users can retrieve almost any specified group by defining age, gender, location, and diagnosis.

Utilization Examples

A 53-year-old female patient visited the JST clinic, complaining of moderate pain in her right groin area. X-ray, CT, and MRI scans revealed a 2 × 2 cm cartilaginous lytic lesion with a sclerotic border in her right acetabulum. An invasive core needle biopsy, revealed an enchrondroma, a benign cartilage tumor. Before any definitive surgery was done, we conducted an epidemiological search using this newly established online database. By using a search-by-location strategy, we identified 724 cases in the pelvis. The epidemiology results showed the following diagnosis distribution: 192 cases (26.5 %) with primary chondrosarcoma, 62 cases (8.56 %) with osteochondroma, 18 cases (2.49 %) with chondroblastoma, and only 1 case (0.14 %) with enchondroma. Because the epidemiology results indicated that enchondroma was extremely rare in pelvis, and a limited biopsy may underestimate the tumor grade, an open biopsy was performed for further investigation. This new biopsy demonstrated primary chondrosarcoma, grade 2 of 3. This diagnosis warranted a local

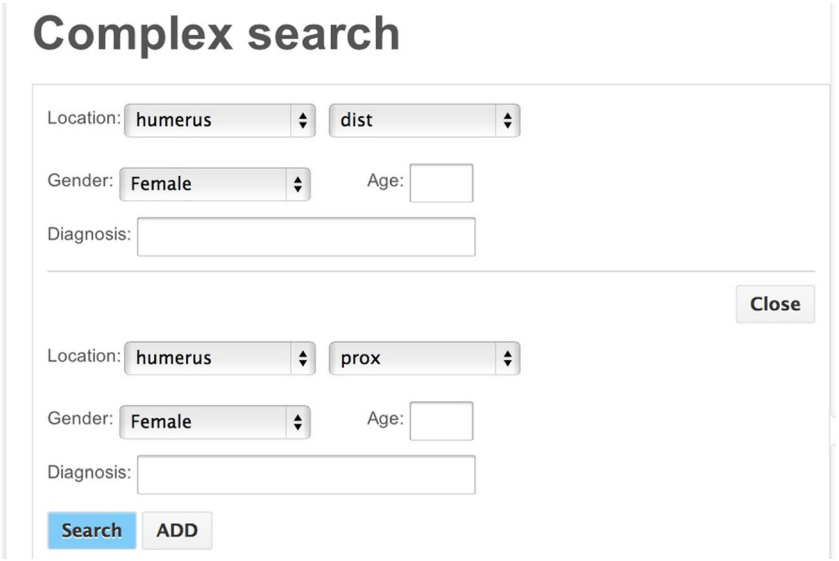

Fig. 3 The "complex search" module was designed to allow searches with various and unlimited combinations of age, gender, location, and diagnosis

resection for definitive surgery rather than curettage, which is the treatment of choice for an enchondroma. Core biopsy has its limitations, such as limited sampling, and may not include the most high-grade area. In addition, there is some histologic overlap between the features of enchondroma and a grade 2 chondrosarcoma. Thus, the core biopsy underestimated the definitive diagnosis and grade of the tumor. In this case, the online epidemiological study was a useful adjunct to facilitate the correct diagnosis and treatment.

In addition to clinical application, the database can be used for research purposes. We conducted a head-to-head comparison between the established database and a Mayo Clinic record search in order to identify any potential epidemiology differences between Chinese and American populations [22].

Discussion

In recent years, online databases have become fundamental in both research and clinical practice because they provide an effective and efficient way to collect, store, and selectively retrieve information. [23–25] In the design of a tumor database, it is very important to define certain variables based on the intended objective. More variables in the database are not necessarily better. Irrelevant and/or redundant variables should be removed in order to highlight what is most important. For our database, we chose age, sex, tumor location, and histological diagnosis based on the goal of demonstrating epidemiology characteristics of primary bone tumors. We found that additional details of treatment schedule and prognosis will not be of further benefit for our purposes.

Given the large set of data in the original FoxPro database, the manual translation of these records in both initial establishment continued expansion of the online database was never a viable option. Fortunately, Excel-formatted data could be easily imported into the MySQL database. Excel-formatted data is now easily created from most modern database programs (for example, Sybase, DB2, Access) by using an export wizard. [26–28] Although the MySQL database system stores the information in an arbitrarily, it allows for sorted results to be reported out by defining the required parameters. [29] Because the MySQL database can be smoothly manipulated and queried by using PHP, a free and open resource software engine, PHP-based scripts enable the database to be searchable by defining specific terms. An important reason for choosing the free and open source web design program WordPress was its innovative and user-friendly backend management interface. [30] This software provided an easy and efficient platform for website maintenance by non-IT professionals.

Although searches can be conducted on Apple or Windows computer, the use of tablets and mobile phones offer the benefit of portability. For use on portable devices, Responsive Web Design (RWD), a series of standards developed in 2010, [31, 32] permit a website to display clearly on every device, including desktop computers, tablets, and mobile phones. Therefore, we chose the free "Responsive Theme" (CyberChimps, http://cyberchimps.com) as the website template for our database, thus enabling the user to search the database online even when not sitting at a desk (that is, in clinic or in classroom via a mobile phone).

At the time of this writing, 9200 cases of primary bone tumors have been included into our online database. The database will be updated quarterly. We believe that the methods described in this research are generalizable and this report can serve as a resource for users who are interested in constructing similar online tumor databases. In addition, our established database (http://www.sarcoma-jst.org) can be a useful tool for sarcoma physicians taking care of patients, as well as

researchers interested in epidemiology studies of primary bone tumors.

Acknowledgments We thank Ms. Rasa Hamilton at Moffitt Cancer Center for her assistance in medical editing, Dr. Qing Zhang, Dr. Lin Hao, Dr. Yi Ding and other colleagues from Orthopaedica Oncology Surgery Department of JST hospital for their contribution of the establishment of this online database, Dr. Xiaoyuan Huang and Dr. Yi Ding from Pathology Department of JST hospital for their contribution in the pathological diagnosis, and Ms. Brenda Barrow for assisting in manuscript submission.

References

1. Siegel R, Naishadham D, Jemal A (2013) Cancer statistics, 2013. CA Cancer J Clin 63:11–30
2. Franchi A (2012) Epidemiology and classification of bone tumors. Clin Cases Miner Bone Metab 9:92–95
3. Limb D, Dreghorn C, Murphy JK, Mannion R (1994) Primary lymphoma of bone. Int Orthop 18:180–183
4. Widhe B, Widhe T (2000) Initial Symptoms and Clinical Features in Osteosarcoma and Ewing Sarcoma*. J Bone Joint Surg Am 82:667
5. Campanacci M (1999) Bone and soft tissue tumors: clinical features, imaging, pathology and treatment. Springer Verlag, Wien
6. Unni KK, Inwards CY (2010) *Dahlin's bone tumors: general aspects and data on 10,165 cases*. Lippincott Williams & Wilkins, Philadelphia
7. Anfinsen KP, Devesa SS, Bray F, et al. (2011) Age-period-cohort analysis of primary bone cancer incidence rates in the United States (1976–2005). Cancer Epidemiol Biomark Prev 20:1770–1777
8. Wu JS and Hochman M. *Bone Tumors: A Practical Guide to Imaging*. Springer: New York, 2012, p.414.
9. Davies AM, Sundaram M, James SJ (2009) Imaging of bone tumors and tumor-like lesions: techniques and applications. Springer, New York, p 716
10. Higgs BW, Elashoff M, Richman S, Barci B (2006) An online database for brain disease research. *BMC Genomics* 7:70
11. Rainville I, Garber JE (2008) Familial cancer database online. Lancet Oncol 9:925–926
12. Khan AM, Zaidi SM, Zaffar Bharati MS (2009) Use of on-line databases by faculty members and research scholars of jawaharlal nehru university (JNU) and jamia millia islamia (JMI), new Delhi (India): a survey. *The International Information &amp*. Libr Rev 41:71–78
13. Castaldi PJ, Cho MH, Cohn M, et al. (2010) The COPD genetic association compendium: a comprehensive online database of COPD genetic associations. Hum Mol Genet 19:526–534
14. Neufeld ML, Cornog M (1986) Database history: from dinosaurs to compact discs. J Am Soc Inf Sci 37:183–190
15. Fletcher CD, Unni KK, Mertens F (2002) *Pathology & genetics: tumours of soft tissue and bone*. World Health Organization, Geneva
16. Baena-Ocampo LC (2009) Ramirez-perez E, linares-gonzalez LM and delgado-chavez R. Epidemiology of bone tumors in Mexico city: retrospective clinicopathologic study of 566 patients at a referral institution. Ann Diagn Pathol 13:16–21
17. Hogendoorn PCW, Athanasou N, Bielack S, et al. (2010) Bone sarcomas: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 21:v204–vv13
18. Kaatsch P (2010) Epidemiology of childhood cancer. Cancer Treat Rev 36:277–285
19. Iarc. Who Classification of Tumours of Soft Tissue and BoneIARC WHO Classification of Tumours SeriesVolume 5 of World Health Organization Classification of Tumours. 4 ed.: World Health Organization: Geneva, 2013, p.427
20. Chen X, Choi JH (2010) Designing online collaborative location-aware platform for history learning. J Educ Technol Dev Exch 3: 13–26
21. Williams B, Damstra D, Stern H (2012) *Professional WordPress: Design and Development*. John Wiley & Sons, Hoboken
22. Niu XH, Xu HR, Inwards CY, Li Y, Zhang Q, Ding Y, Letson GD, Bui MM (2015) Primary bone tumors: epidemiological comparison of 9,200 patients treated at Beijing Ji Shui Tan Hospital, Beijing, China with 10, 165 patients at Mayo Clinic, Rochester, MN, USA. Arch Pathol Lab Med 139(9):1149–1155
23. Junghans TB, Sevin IF, Ionin B, Seifried H (2004) Cancer information resources: digital and online sources. Toxicology 198:177–193
24. Lawrence DW, Laflamme L (2009) Using online databases to find journal articles on injury prevention and safety promotion topics: how do SafetyLit subscribers use other databases? Saf Sci 47:1–8
25. Blamey RW, Hornmark-Stenstam B, Ball G, et al. (2010) ONCOPOOL - a european database for 16,944 cases of breast cancer. Eur J Cancer 46:56–71
26. Saracco CM, Roth MA, Wolfson DC (2004) Enabling distributed enterprise integration with WebSphere and DB2 information integrator. IBM Syst J 43:255–269
27. Taylor B, Adurty N, Bradley S, Taylor CK (2010) The official new features guide to sybase ASE 15. Wordware Publishing, Inc., USA
28. Baranyi J, Tamplin ML (2004) ComBase: a common database on microbial responses to food environments. J Food Prot® 67:1967–1971
29. Zawodny JD and Balling DJ. High performance MySQL: optimization, backups, replication, load balancing & more o'reilly: sebastopol; 2008
30. Jones KM, Farrington P-A (2011) Using WordPress as a library content management system. American Library Association, Chicago
31. Marcotte E (2011) Responsive web design, Editions Eyrolles
32. Frain B (2012) Responsive web design with HTML5 and CSS3. Packt Publishing, UK