

Paweł Daszkiewicz<sup>1</sup>, Zuzanna Daszkiewicz<sup>2</sup>, Piotr Daszkiewicz<sup>3</sup>

## Custom-made database enabling a quick search for patients fulfilling specific criteria. A simple and effective tool for clinical data collection for scientific purposes

Baza danych własnego projektu, umożliwiająca szybkie wyszukiwanie pacjentów spełniających określone kryteria. Proste i skuteczne narzędzie do zbierania materiału klinicznego do celów naukowych

<sup>1</sup> Department of Neurosurgery, The Children's Memorial Health Institute, Warsaw, Poland

<sup>2</sup> Student; Warsaw Medical University, Warsaw, Poland

<sup>3</sup> General Medical Practice "Poradnia Lekarza POZ," Ząbki, Poland

Correspondence: Paweł Daszkiewicz, Department of Neurosurgery, The Children's Memorial Health Institute, Dzieci Polskich 20, 04-730 Warsaw, Poland, tel.: +48 22 815 75 60, e-mail: pmdaszkiewicz@o2.pl

Paweł Daszkiewicz, Klinika Neurochirurgii, Instytut „Pomnik – Centrum Zdrowia Dziecka”, al. Dzieci Polskich 20, 04-730 Warszawa, tel.: +48 22 815 75 60, e-mail: pmdaszkiewicz@o2.pl

### Abstract

**Introduction:** Every evidence-based scientific work requires precise and reliable data. In clinical medicine, this means searching for patients fulfilling specific criteria. Unfortunately, most hospital-operated information systems are designed to process data concerning treatment of specific patients, mainly in its medical and administrative aspects, while a cross-sectional search for patients with a specific pathology at a specific location is difficult if not impossible. Therefore, most of these databases are essentially useless for scientific purposes. **Material and method:** Based on medical records of all patients treated at our facility, we have created a simple database enabling a quick search for patients fulfilling specific criteria. We used Microsoft® Office Access® software installed on a standard PC. Structure and the use of database are described in detail. **Results:** Our database contains basic information about 16,126 patients treated over 38 years and occupies 5.19 MB. To date, 48 queries have been performed, providing reliable scientific material. **Conclusion:** This is an effective and comprehensive tool for the search of patients or items fulfilling specific criteria. To the best of our knowledge, such a tool for the management of clinical data has not been described to-date. After introducing appropriate modifications, a similar database will be very useful in every clinical centre.

**Keywords:** medical records, database, search criteria

### Streszczenie

**Wprowadzenie:** Każda praca naukowa oparta na dowodach wymaga dokładnych i wiarygodnych danych. W medycynie klinicznej oznacza to wyszukiwanie pacjentów spełniających określone kryteria. Niestety, większość szpitalnych systemów informatycznych służy gromadzeniu i przetwarzaniu danych dotyczących leczenia poszczególnych pacjentów, głównie w aspekcie medycznym i administracyjnym, podczas gdy przekrojowe poszukiwanie chorych z określoną patologią w określonej lokalizacji jest trudne albo w ogóle niemożliwe. Dlatego większość tych baz danych jest w zasadzie bezużyteczna dla celów naukowych. **Materiał i metoda:** W oparciu o materiał obejmujący wszystkich chorych leczonych w naszym ośrodku stworzyliśmy prostą bazę danych, umożliwiającą szybkie wyszukiwanie pacjentów spełniających określone kryteria. Zastosowaliśmy oprogramowanie Microsoft® Office Access® zainstalowane na standardowym komputerze osobistym. W pracy opisano szczegółowo strukturę i sposób korzystania z bazy danych. **Wyniki:** Nasza baza danych zawiera podstawowe informacje dotyczące 16 126 pacjentów (stan na styczeń 2019 r.) leczonych na przestrzeni ponad 38 lat i zajmuje 5,19 MB pamięci. Dotychczas przeprowadzono 48 kwerend, uzyskując błyskawicznie wiarygodny materiał kliniczny do badań naukowych.

**Wniosek:** Nasza baza danych okazała się skutecznym, wszechstronnym i niezwykle przydatnym narzędziem do wyszukiwania pacjentów spełniających określone kryteria. Według naszej wiedzy dotychczas nie opisano takiego prostego narzędzia do przetwarzania materiału klinicznego. Po wprowadzeniu odpowiednich modyfikacji podobna baza będzie bardzo przydatna w każdym ośrodku klinicznym.

**Słowa kluczowe:** baza danych, kryteria wyszukiwania, dokumentacja medyczna

## INTRODUCTION

Every evidence-based scientific work must be based on precise data. In clinical medicine, this requires selection of patients fulfilling specific criteria, e.g. a definite pathology at a definite location, a definite genetic mutation, a definite radiologic finding, a definite age, etc. This requires a comprehensive database, enabling search for patients or items fulfilling specific parameters. Unfortunately, most hospital-operated information systems are designed to process data concerning treatment of specific patients, mainly in its medical and administrative aspects, while a cross-sectional search for specific parameters is difficult if not impossible. Therefore, most of these systems are useless for scientific purposes. With this in mind, we devised a simple database comprising 10 parameters, which allows for a quick and easy search through an enormous pool of patients treated over many years.

## AIM OF THE PAPER

The aim of this paper is to present a useful database enabling a quick search for patients fulfilling specific criteria.

## MATERIAL

Records of all patients treated at our facility between January 1980 and December 2018 have been reviewed. To date (January 1, 2019), our database comprises 16,126 patients aged from 1 month to 18 years who received surgical treatment. Due to the profile of our specialty, many of our patients suffer from chronic diseases and thus often require multiple procedures; therefore many of them have been entered more than once (thus enabling an insight into the effectiveness of a particular type of surgery in a particular indication).

## METHODS

Our database uses commercially available Microsoft® Office Access® software installed on a standard personal computer. Every record is composed of 10 fields:

1. Consecutive number of record in the database.
2. and 3. Patient identification fields (name and hospital identification number of the patient, enabling access to their records in the hospital archive).
- 4, 5, and 6. Pathology identification fields (defining nature, location and histological findings – if applicable).

7. Type of surgery identification field (defining the type of surgery performed).

8. Comments field (listing comorbidities, ancillary procedures, encountered problems and complications considered noteworthy).

9. Age at admission.

10. Calendar year of admission.

Fields 4, 5, 6 and 7 have a form of combo-boxes with several options to choose from, greatly facilitating entry of data and preventing typing errors. The order of entry of particular options within each combo-box largely resulted from the initial order of entry of patients when database was being created and is incidental.

In the “location” field, the following options are available: integument (i.e. skin, bone, meninges), brain hemispheres, 3<sup>rd</sup> ventricle, central region, suprasellar area, sella, orbit, posterior cranial fossa, brainstem, vertebral column (i.e. bones and ligaments), vertebral canal (i.e. structures within), peripheral nerve, other.

In the “type of pathology” field, the following options are available: trauma, congenital dysraphic defect, craniosynostosis, neoplasm, abscess, subdural empyema, hydrocephalus, arteriovenous malformation, aneurysm, epilepsy, chronic subdural hematoma, arachnoid cyst, porencephalic cyst, colloidal cyst, carotid-cavernous fistula, arachnitis, post-traumatic liquorrhea, vein of Galen aneurysm, calvarial bone defect, pilonidal sinus, idiopathic hematoma, symptomatic over-drainage syndrome, ischemic stroke, spondylolisthesis, subdural haemorrhage, intervertebral disc herniation, tuberous sclerosis, Apert syndrome, Crouzon syndrome, Dandy-Walker syndrome, syringomyelia, branchial cleft cyst.

In the „histology” field, the following options are available: astrocytoma I (including the most common pilocytic type), astrocytoma II (including the pilomyxoid type), astrocytoma III (i.e. anaplastic type), glioblastoma, gliosarcoma, sarcoma, ependymoma (anaplastic cases were denoted in the “comments” window), medulloblastoma, neuroblastoma, primitive neuroectodermal tumours (PNET), dermoid cyst, epidermoid cyst, malignant tumour not otherwise specified, embryonal sarcoma, optic glioma (for several reasons this was considered a separate entity), von Recklinghausen disease, craniopharyngioma, ganglioglioma, pinealoma, pineoblastoma, adenoma, inflammatory granulation, meningioma (precise definition is provided in the “comments” window), oligodendroglioma, teratoma, hamartoma, schwannoma, mixed glioma, eosinophilic granuloma, other.

In the “type of surgery” field, the following options are available: craniotomy, shunt, burr hole, clipping of aneurysm, infusion test, reconstruction (further details provided in the “comments” window), nerve anastomosis, neurolysis, abscess drainage, reposition of bone fracture, discectomy, aqueduct intubation, brachytherapy, stereotactic biopsy, laminotomy.

All entries have been made over 27 years by the senior author (P.D.). This ensures a uniform interpretation of medical records and consistent choice of options according to uniform criteria. However, prolonged data collection and progress in medical science resulted in the development of novel diagnostic tools and treatment modalities, which were gradually added to the already existing ones, without altering the overall database structure.

Assuming that on the average 2 minutes are needed for one patient, creation of our database required 538 hours to date. This means about 20 working hours annually, which seems reasonable.

## RESULTS

Our database contains 16,126 records (as of January 1, 2019) and occupies 5.19 MB. It enables simple queries, cross queries, and search for duplicates and unmatched data. To date, 48 documented simple queries have been performed, yielding reliable data used in scientific publications plus countless ad-hoc queries needed in current clinical work and administrative reporting.

The system identifies patients fulfilling pre-defined criteria. A more precise definition of search criteria results in a more adequate selection of patients. On the other hand, too precise criteria increase the risk of omitting those who actually should be selected. This is rare, but it does occur. Reliability tests showing the true scale of this phenomenon have not been performed. Therefore, we prefer to use less precise criteria, resulting in a longer, less adequate, but more complete primary list of patients, thus reducing the risk of omitting essential data.

The primary list thus obtained contains names and identification numbers of patients that fulfil pre-defined criteria. In this way, patients’ records from the hospital archive are obtained and final verification of search adequacy is performed. Only confirmed and valid material enters actual scientific analysis.

## DISCUSSION

A dramatic increase of stored data and the necessity to manage an enormous amount of information resulted in an impressive development of branches of informatics dealing with data processing, both in medical science and elsewhere (Carter, 2008; Darling, 1992). This is best reflected by the fact that upon entering “medical documentation databases” Google® search yields 72,800,000 items and 67 databases. Globally, most of these databases actually store publications.

On national level, they store epidemiological data. On hospital level, they store data concerning medical and financial/administrative aspects of treatment of individual patients (Motheral et al., 2003). In practice, obtaining cross-sectional cumulative data on patients fulfilling specific criteria (e.g. patients with a particular type of tumour in a particular location or those in a particular age range) turns out impossible (Hashimoto et al., 2014). This tedious and time-consuming task is usually delegated to younger members of research teams, resulting in “manual” scrolling through enormous pools of data in order to find few items of interest. This is considered a necessary initial step in every serious scientific activity.

Another factor significantly interfering with collection of medical data is duration of storing paper records. According to current Polish regulations (an act issued on November 6, 2008), paper records should be stored for 20 years (30 years in certain cases). After this period, paper records are destroyed and therefore lost for scientific purposes.

Last but not least, the issue of access to hard data in the out-of-hospital setting should be mentioned. The precise number of patients fulfilling a particular criterion may be a robust argument in favour of a given thesis at a conference held abroad. Normally, one must rely on one’s memory or notes prepared beforehand.

Most of existing clinical databases go into great detail. This certainly contributes to reliability, but results in great time- and work-consumption at preparation. Furthermore, such databases occupy a great amount of computer memory, precluding their use on mobile devices (Kathirgamathamby et al., 2011; Müller et al., 2004).

Our database addresses most of these issues. It was devised and created in response to specific needs of a neurosurgery department. It serves to sort the entire population of patients treated at our department since the very beginning to find those fulfilling pre-specified criteria. There is no “expiry date” for data. Our database is very simple, which has both benefits and drawbacks. Recording of patient data takes relatively little time, but subsequent search may yield patients who only in part fulfil desired criteria. Although the primary list is always reviewed and final recruitment is done in a traditional way, owing to our database the working pool is “manageable,” encompassing dozens or hundreds of records instead of thousands. Another advantage is that it occupies relatively little memory space and may be installed on mobile devices, which allows for work in-the-field, certainly keeping in mind the issue of protection of sensitive data. Needless to say, the database is intended for scientific use only. Our database contributed to most of scientific output of our department and proved useful for current administrative reporting, e.g. providing precise number of implanted shunts over consecutive years. It also proved very helpful in the case of patients undergoing combined treatment at other departments.

The structure of our database relies on the fact that a similar pathological process may affect different parts of the

nervous system (e.g. trauma of brain, spinal cord or peripheral nerve) in neurological and neurosurgical diseases, and vice-versa, different pathological processes may affect the same part of the nervous system (e.g. an abscess, vascular malformation or tumour of the central region of brain). Additional categories are histopathological findings and types of surgical procedures performed, which also may be similar in different pathological processes or different in similar processes, depending on a particular clinical situation. For “not otherwise specified” issues (rare, interesting or coexisting problems and complications, additionally performed procedures) the “comments” window was designed. Windows for patient identification, and windows containing data on patient age and calendar year at admission are self-evident and do not require any comment.

For example, a hypothetical 5-year-old child (“John Doe”) admitted in February 2017 and undergoing a posterior fossa craniotomy for cerebellar medulloblastoma will be recorded as: “12345” (consecutive number in database), “Doe” (name), “01234/17” (identification number of hospital record), “posterior fossa” (location), “tumour” (type of pathology), “medulloblastoma” (histological finding), “5” (age at admission) and “17” (year of admission). In this way, by adding or subtracting search parameters, the system will find either a specific patient, or all patients with posterior fossa pathology, all undergoing craniotomy, all medulloblastomas, or patients fulfilling a combination of these criteria. A similar principle was and is still used in large medical databases (Darling, 1992; Kathirgamathamby et al., 2011).

Our intention was to draw attention of the scientific community to the important but largely unsolved problem of finding patients fulfilling specific criteria in the hospital archive. Certainly, every centre is unique in terms of work and patients’ profile, so their database will need other windows and other options. Nevertheless, the general idea appears justified and deserves recommendation.

To the best of our knowledge, such a simple, versatile and effective tool for the management of large pools of data for scientific use has not been described in the literature to date.

## CONCLUSIONS

In our opinion, the novelty (and value) of the paper relies in the following:

1. As opposed to currently available databases used in medical facilities, our database enables rapid and easy search for patients fulfilling specific criteria for scientific purposes. Databases currently available on the market are overly complicated and their practical usefulness

is limited. Of course, the initial creation of our databases needed some time, but the subsequent benefit in terms of saving of time and effort is well worth it.

2. We took advantage of a frequent phenomenon in neurology that the same pathological process may affect several parts of the nervous system (e.g. traumatic injury of brain, spinal cord or peripheral nerve) and vice-versa, the same part of the nervous system may be affected by several different pathological processes (e.g. neoplasm, abscess, injury, congenital defects of cerebral hemispheres). Taking into account other parameters (e.g. patient’s age, treatment modality, calendar year of admission, histopathological findings, etc.), we can construct a query enabling search for patients or items fulfilling prespecified sets of criteria (e.g. children under 3 with hemispheric abscess undergoing craniotomy in 2001).
3. Since our database is relatively simple, it requires relatively little memory space. It may be installed on mobile devices, enabling its use out of ordinary working environment (e.g. at scientific meetings abroad).
4. Our aim is to provide the scientific community with a simple tool that greatly facilitates data mining. Obviously, every centre has its specific features and needs; therefore, a dedicated, custom-made database will have to include other parameters, but the general idea will be the same.

## Conflict of interest

*The authors do not report any financial or personal connections with other persons or organisations which might negatively affect the content of this publication and/or claim authorship rights to this publication.*

## References

- Carter JH: Electronic Health Records. A Guide for Clinicians and Administrators. 2<sup>nd</sup> ed., American College of Physicians, Philadelphia, PA 2008.
- Darling CB: Database technology for medical records. Instr Course Lect 1992; 41: 521–526.
- Hashimoto RE, Brodt ED, Skelly AC et al.: Administrative database studies: goldmine or goose chase? Evid Based Spine Care J 2014; 5: 74–76.
- Kathirgamathamby V, Pitts J, Perry R et al.: An audit of medical documentation on standardised electronic neonatal database (SEND) in a large level-2 neonatal unit. BAPM NNA NNS Posters. Arch Dis Child Fetal Neonatal Ed 2011; 96 Suppl 1.
- Motheral B, Brooks J, Clark MA et al.: A checklist for retrospective database studies – report of the ISPOR Task Force on Retrospective Databases. Value Health 2003; 6: 90–97.
- Müller H, Michoux N, Bandon D et al.: A review of content-based image retrieval systems in medical applications – clinical benefits and future directions. Int J Med Inform 2004; 73: 1–23.