

Semi-automated data collection from electronic health records in a stroke unit in Brazil

Coleta de dados semiautomáticos de registros eletrônicos de saúde na unidade de acidente vascular cerebral no Brasil

Raquel Franco Zambom VALÊNCIO¹, Juli Thomaz de SOUZA², Fernanda Cristina WINCKLER², Gabriel Pinheiro MODOLO³, Natalia Cristina FERREIRA³, Silmeia Garcia Zanati BAZAN², Marcos Christiano LANGE^{2,4}, Carlos Clayton Macedo de FREITAS³, Sergio Alberto Rupp de PAIVA², Rogério Carvalho de OLIVEIRA², Gustavo José LUVIZUTTO⁵, Rodrigo BAZAN³

ABSTRACT

Background: There is a high demand for stroke patient data in the public health systems of middle and low-income countries. **Objective:** To develop a stroke databank for integrating clinical or functional data and benchmarks from stroke patients. **Methods:** This was an observational, cross-sectional, prospective study. A tool was developed to collect all clinical data during hospitalizations due to stroke, using an electronic editor of structured forms that was integrated with electronic medical records. Validation of fields in the electronic editor was programmed using a structured query language (SQL). To store the results from SQL, a virtual table was created and programmed to update daily. To develop an interface between the data and user, the Embarcadero Delphi software and the DevExpress component were used to generate the information displayed on the screen. The data were extracted from the fields of the form and also from cross-referencing of other information from the computerized system, including patients who were admitted to the stroke unit. **Results:** The database was created and integrated with the hospital electronic system, thus allowing daily data collection. Quality indicators (benchmarks) were created in the database for the system to track and perform decision-making in conjunction with healthcare service managers, which resulted in improved processes and patient care after a stroke. An intelligent portal was created, in which the information referring to the patients was accessible. **Conclusions:** Based on semi-automated data collection, it was possible to create a dynamic and optimized Brazilian stroke databank.

Keywords: Stroke; Benchmarking; Artificial Intelligence; Supervised Machine Learning; Emergency Service, Hospital.

RESUMO

Antecedentes: Há alta demanda de dados de pacientes com acidente vascular cerebral (AVC) nos sistemas de saúde de países de baixa e média renda. **Objetivo:** Desenvolver um banco de dados de AVC para integrar dados clínicos ou funcionais e indicadores de qualidade de pacientes com AVC. **Métodos:** Estudo observacional, transversal e prospectivo. Foi desenvolvida uma ferramenta para coletar dados clínicos durante as internações por AVC por meio de um editor eletrônico de formulários estruturados integrado ao prontuário eletrônico. A validação dos campos no editor eletrônico foi programada em linguagem de consulta estruturada (SQL). Para armazenar os resultados da SQL, uma tabela virtual foi criada e programada para atualização diária. Para desenvolver interface entre os dados e o usuário, foram utilizados o *software Embarcadero Delphi* e o componente *DevExpress* para gerar informações apresentadas na tela. Os dados foram extraídos dos campos do formulário e também do cruzamento de outras informações do sistema informatizado, incluindo pacientes internados na unidade de AVC. **Resultados:** O banco de dados foi criado e integrado ao sistema eletrônico do hospital, permitindo coleta diária de dados. Indicadores de qualidade foram criados no banco de dados para que o sistema acompanhasse e realizasse a tomada de decisão com os gestores dos serviços de saúde, resultando em melhoria no processo e no atendimento ao paciente após AVC. Foi criado um portal inteligente, no qual eram registradas as informações referentes aos pacientes. **Conclusões:** Com a coleta de dados semiautomática, foi possível criar um banco de dados de AVC dinâmico e otimizado em unidade de AVC no Brasil.

Palavras-chave: Acidente Vascular Cerebral; Benchmarking; Inteligência Artificial; Aprendizado de Máquina Supervisionado; Serviço Hospitalar de Emergência.






¹Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu, Botucatu SP, Brazil.

²Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu, Departamento de Medicina Interna, Botucatu SP, Brazil.

³Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu, Departamento de Neurologia, Psicologia e Psiquiatria, Botucatu SP, Brazil.

⁴Universidade Federal do Paraná, Complexo Hospital de Clínicas, Curitiba PR, Brazil.

⁵Universidade Federal do Triângulo Mineiro, Departamento de Fisioterapia Aplicada, Uberaba MG, Brazil.

RFZV  <https://orcid.org/0000-0002-8306-3140>; JTS  <https://orcid.org/0000-0003-2227-7505>; FCW  <https://orcid.org/0000-0002-7299-5612>; GPM  <https://orcid.org/0000-0003-1057-5089>; NCF  <https://orcid.org/0000-0003-3476-620X>; SGZB  <https://orcid.org/0000-0002-0607-8189>; MCL  <https://orcid.org/0000-0002-0405-7157>; CCMF  <https://orcid.org/0000-0001-5210-4336>; SARP  <https://orcid.org/0000-0003-4412-1990>; RCO  <https://orcid.org/0000-0002-0586-8226>; GJL  <https://orcid.org/0000-0002-6914-7225>; RB  <https://orcid.org/0000-0003-3872-308X>

Correspondence: Gustavo José Luvizutto; Email: gluvizutto@gmail.com.

Conflict of interest: There is no conflict of interest to declare.

Authors' contributions: GJL, CCMF, RCO, RB: conceptualization, data curation, formal analysis, writing – original draft preparation, project administration, supervision, writing – review & editing; RFZV, JTS, FCW, GPM, NCF: investigation, methodology; RB, SGZB, MCL, SARP: supervision, validation, visualization, roles/writing – original draft, writing – review.

Received on December 22, 2020; Accepted on February 21, 2021.

INTRODUCTION

Stroke is the second leading cause of death and disability worldwide. It has a significant socioeconomic impact on low- and middle-income countries, thereby affecting public health status¹⁻³. In the last 40 years, databank projects have been implemented to provide information on the clinical courses and outcomes of stroke⁴⁻⁸.

Implementation of a databank is important for planning and decision-making by healthcare service managers, in order to improve processes and stroke care outcomes⁹. However, integration of these databanks is limited to large centers and clinical trials, thus restricting access to low and middle-income countries¹⁰. In accordance with the principles of stroke care lines in developing countries, it is necessary for all quality criteria during stroke hospitalization to be registered in electronic medical records, with the aim of establishing new healthcare policies¹¹.

Based on these premises, implementation of a computerized database system containing the main clinical indicators of stroke patients has become a reality in developing countries^{12,13}. Over recent years, the introduction of artificial intelligence has helped in development of databases that are integrated in networks, with the capacity to identify clinical data with greater agility and security^{14,15}.

Therefore, the aim of this study was to develop an interface for accessing stroke patient data and the main benchmarks in acute stroke care, in order to create a semi-automated databank for implementation in low and middle-income countries.

METHODS

Study design, setting and participants

This project with the aim of developing a semi-automated stroke databank was conducted at the stroke unit of Faculdade de Medicina de Botucatu, Botucatu, Brazil. This stroke unit has 10 beds, with an average of 30 patients/month, and belongs to a tertiary-level hospital with 684 beds. The hospital has a hospital information system (HIS) that includes electronic patient records containing clinical and treatment information. All patients who were admitted to the stroke unit were included in this project. The study was approved by the institution's Ethics in Human Research Committee.

Procedures

To organize the information, the tool was designed to include initial data, a clinical summary of the presentation, hospital course (including complications and infections), investigation/complementary examinations (including computed tomography [CT], electrocardiography,

echocardiography, Holter data, carotid and vertebral duplex data, CT angiography, magnetic resonance imaging [MRI] angiography, control CTs or MRIs and transcranial Doppler) and discharge conditions (including etiological diagnosis, medications, discharge with therapy plan, rehabilitation and destination of the patient). The rules for each subgroup were planned so as to validate the completion of certain fields, all of which were mandatory for the service to be completed.

The database interface was developed on the HIS platform, and clinical data were collected based on the following benchmarks for stroke care: 1) prophylaxis for deep venous thrombosis, starting no later than the second day; 2) hospital discharge with antiplatelet therapy for patients with non-cardioembolic stroke; 3) hospital discharge with oral anticoagulation for patients with atrial fibrillation or atrial flutter; 4) use of antiplatelet agents when indicated, starting on the second day of hospitalization; 5) hospital discharge with statins for patients with atherothrombotic stroke; 6) hospital discharge with prophylactic therapy and rehabilitation plan; 7) percentage of patients with acute cerebrovascular disease; 8) length of hospital stay; 9) complications; 10) stroke type-specific ICD-10; 11) hospital mortality; 12) time to CT <25 min; and 13) door-to-needle time ≤60 min.

Execution, monitoring and control phase

To execute the development of the tool, an electronic editor of structured forms was used, which was integrated with the electronic medical record platform used in the institution. Parameterized fields and validation rules were created for essential fields. The rules were created in the field editor. Radio button components were used for information with only one choice and checkbox components were used when the information consisted of multiple choices. In addition, the validation of fields was programmed using a structured language, i.e. in structured query language (SQL)¹⁶.

Mandatory fields were configured in the editor, and each field was parameterized. All patients who were admitted to the stroke unit were included in the database, as this was mandatory at the time of transfer to another ward or at the time of the patient's discharge or death.

Databank storage

To store the results from SQL, a virtual table was created and programmed to update daily at 04:00 h. The interface between the data and user was developed using the program Embarcadero Delphi, and the DevExpress component was used to generate the information displayed on the screen.

Data quality

The stored data were analyzed to check the data quality in accordance with the pre-established rules for creating the tool as mandatory fields with relationships between the fields. A rule was also created to identify patients who were admitted to the stroke unit and then transferred to another

unit, but for whom no discharge summary was filled out at any time during hospitalization. In preparing SQL and business rules for both data analysis and for out-of-base patients, the PL/SQL (Procedural Language for SQL) tool was used via Embarcadero. The data were extracted from the fields of the form and also from cross-referencing of other information from the computerized system, such as patients who were admitted to the stroke unit.

RESULTS

Once extracted, the data were stored in a database, and any type of information could be consulted at any time, thus allowing cross-referencing between the data of the tool and the patients' demographic profile. The indicators that were established according to the area and database were available in almost real time, with daily updates at 04:00 h. The database began registering patients in August 2018 and has included data from approximately 1,000 stroke patients to date.

Both qualitative and quantitative results can be obtained from the detailed data included in the database. The results can be grouped to create filters that can be applied to any column within the database. This provides an overview of the total number of patients for each parameter, thus resulting in development of appropriate public health policies.

In addition to the specific data characteristics of stroke patients, and owing to the integration with the electronic medical record, important clinical data such as hospital course, laboratory test results, imaging tests and surgeries,

as well as the main scales used in the stroke unit, can be displayed on the screen (Figure 1).

The quality indicators (benchmarks) were created in the database for the system to track and perform decision-making in conjunction with healthcare service managers. This resulted in improved processes and patient care after a stroke (Figure 2).

An intelligent portal was created, in which all the information referring to patients' care was available in a single location in a dynamic and objective manner. Development of an interface with a specific database for a hospital area enabled creation of a new concept in relation to the data that were already included in the computerized system, thus allowing new ideas to emerge in order to improve and facilitate the interpretation of large volumes of data (Figure 3).

DISCUSSION

This study reports the development of a tool integrated with electronic medical records that is capable of recording clinical data on patients admitted to a stroke unit and identifying the main indicators of quality of care, using artificial intelligence to improve the quality of care and decision-making process.

Several studies using previous databases have been conducted in low and middle-income countries to identify the quality of healthcare services¹⁷⁻²⁰. Zetola et al. observed the registration of patients in a database with identification, previous clinical history, family history, treatments, previous comorbidities, complementary laboratory tests,

Perguntas	Respostas
NÍVEL DE CONSCIÊNCIA	
NDC QUESTÕES	
NDC ORDENS	
MELHOR OLHAR CONJUGADO	
CAMPOS VISUAIS	
PARÊSIA FACIAL	
MEMBROS SUPERIORES	
MEMBROS INFERIORES	
ATAXIA DE MEMBROS	
SENSIBILIDADE	
MELHOR LINGUAGEM	
DISARTRIA	
EXTINÇÃO E DESATENÇÃO, ANTIGA NEGLIGÊNCIA	

Figure 1. National Institutes of Health Stroke Scale displayed in the electronic medical record.

cardiac tests and neuroimaging tests. Using a computerized system, these data could be extracted dynamically, and their study suggested that a risk factor control program might lead to a reduction in stroke incidence¹⁷.

In 2013, a single database for stroke patients was established in Joinville, Brazil, via a municipal law that required all public and private healthcare establishments to forward monthly information on stroke patients to the Department of Health to create an integrated database. In 2014, the Ministry of Health coordinated a simultaneous study in five cities in Brazil to assess the incidence of, mortality due to and main risk factors for stroke. The study included individuals with stroke and determined the environmental context of stroke, using a national database of stroke patients in Latin America^{21,22}.

The *Rede Brasil AVC* (Brazil Stroke Network), created to improve overall care for stroke patients in Brazil, is a non-governmental organization formed by professionals from various areas with the aim of providing quality care to stroke patients, from prevention to rehabilitation²³. However, population studies to monitor clinical data on stroke patients need to be integrated into networks with shared access and periodic audits. The development of the tool in this study will help facilitate exploration of clinical data and quality indicators in low and middle-income countries.

Studies on the development of systems containing specific data on stroke patients have not been previously

reported. Data collection methods were not reported in previous studies that included databases. Access to information has become increasingly faster and easier through technological advances that enable implementation of structured and standardized electronic databases that use artificial intelligence. Analysis on large datasets of patient characteristics, outcomes from treatments and their costs can help identify the most clinically effective and cost-efficient treatments for a population. Models, analytics, visualization large amounts of data and use of artificial intelligence come together to offer different perspectives on healthcare challenges within the contexts of time and geography. They provide strategic solutions to assist in management, decision-making and future clinical research^{24,25}.



Figure 3. Business intelligence applied to stroke databank.

INDICADORES DE SEGURANÇAS E QUALIDADE - PORTARIA 665 DO MINISTÉRIO DA SAÚDE													
ANO													
MES	1	2	3	4	5	6	7	8	9	10	11	12	13
ANO : 2018													
ANO : 2019													
12	23	23	3	24	7	23	67,86	10,45	7	22	5	4	1
11	21	20	4	26	6	18	86,21	6,81	6	25	3	8	1
10	25	17	1	21	6	20	100	7,97	7	20	2	7	0
09	14	12	3	14	1	15	94,74	9,05	3	16	1	5	1
08	25	22	2	25	4	27	93,55	8,28	6	24	5	11	0
07	35	34	4	37	8	42	93,48	8,75	10	35	5	10	1
06	27	16	4	23	5	20	96,43	6,43	6	23	2	4	0
05	23	20	3	21	1	22	88,46	5,25	3	20	1	3	1
04	16	14	0	18	3	8	100	6,38	4	19	2	5	1
03	18	15	0	18	2	9	100	5,27	2	14	2	7	1
02	30	20	2	28	7	24	93,33	5,78	4	26	3	6	0

1. profilaxia para trombose venosa profunda iniciada até o segundo dia
2. alta hospitalar em uso de antiagregante plaquetário em pacientes com AVC não cardioembólico
3. alta hospitalar em uso de anticoagulação oral para pacientes com Fibrilação Atrial (FA) ou "Flutter", salvo contraindicações
4. uso de antiagregantes plaquetários, quando indicado, iniciado até o segundo dia de internação
5. alta hospitalar em uso de estatina para pacientes com AVC aterotrombótico, salvo contraindicações
6. alta hospitalar com plano de terapia profilática e de reabilitação
7. porcentagem de pacientes com doença cerebrovascular aguda atendidos na Unidade de AVC
8. o tempo de permanência hospitalar do paciente acometido por AVC visando redução do mesmo
9. as seguintes complicações: trombose venosa profunda, úlcera de pressão, pneumonia, infecção do trato urinário
10. CID-10 específico do tipo de AVC à alta hospitalar
11. mortalidade hospitalar por AVC, visando redução da mesma
12. tempo porta-tomografia < 25 minutos
13. tempo porta-agulha < 60 minutos

Figure 2. Safety and quality indicators.

An integrated system significantly integrates separate sectors, and teamwork is recorded. This enables top management to fully analyze healthcare processes. Development of this database for stroke patients will optimize this search time and provide large quantities of relevant information for the area available, which can be used to support future studies^{26,27}.

In conclusion, through development of integration of this tool and electronic medical records, a dynamic and optimized stroke data bank was created. This database will be useful for managing quality indicators, assisting in the planning of actions at the stroke unit and supporting decision-making, and will serve as a basis for future studies and generation of new knowledge.

References

- Johnston SC, Mendis S, Mathers CD. Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling. *Lancet Neurol*. 2009 Apr;8(4):345-54. [https://orcid.org/10.1016/S1474-4422\(09\)70023-7](https://orcid.org/10.1016/S1474-4422(09)70023-7)
- Grysiewicz RA, Thomas K, Pandey DK. Epidemiology of ischemic and hemorrhagic stroke: incidence, prevalence, mortality, and risk factors. *Neurol Clin*. 2008 Nov;26(4):871-95, vii. <https://orcid.org/10.1016/j.ncl.2008.07.003>
- Feigin VL, Lawes CM, Bennett DA, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *Lancet Neurol*. 2009 Apr;8(4):355-69. [https://orcid.org/10.1016/S1474-4422\(09\)70025-0](https://orcid.org/10.1016/S1474-4422(09)70025-0)
- Bronstein K, Murray P, Licata-Gehr E, Banko MA, Kelly-Hayes M, Fast S, et al. The Stroke Data Bank project: implications for nursing research. *J Neurosci Nurs*. 1986 Jun;18(3):132-4. <https://orcid.org/10.1097/01376517-198606000-00005>
- Kunitz SC, Gross CR, Heyman A, Kase CS, Mohr JP, Price TR, et al. The Pilot stroke data bank: definition, design, and data. *Stroke*. Jul-Aug 1984;15(4):740-6. <https://orcid.org/10.1161/01.str.15.4.740>
- Foulkes MA, Wolf PA, Price TR, Mohr JP, Hier DB. The Stroke Data Bank: design, methods, and baseline characteristics. *Stroke*. 1988 May;19(5):547-54. <https://orcid.org/10.1161/01.str.19.5.547>
- Ahmed N, Wahlgren N, Grond M, Hennerici M, Lees KR, Mikulik R, et al. Implementation and outcome of thrombolysis with alteplase 3–4.5 h after an acute stroke: an updated analysis from SITS-ISTR. *Lancet Neurol*. 2010 Sep;9(9):866-74. [https://orcid.org/10.1016/S1474-4422\(10\)70165-4](https://orcid.org/10.1016/S1474-4422(10)70165-4)
- Wahlgren N, Ahmed N, Davalos A, Ford GA, Grond M, Hacke W, et al. Thrombolysis with alteplase for acute ischaemic stroke in the Safe Implementation of Thrombolysis in Stroke-Monitoring Study (SITS-MOST): an observational study. *Lancet*. 2007 Jan;369(9558):275-82. [https://orcid.org/10.1016/S0140-6736\(07\)60149-4](https://orcid.org/10.1016/S0140-6736(07)60149-4)
- Hinchey JA, Shephard T, Tonn ST, Ruthazer R, Selker HP, Kent DM. Benchmarks and determinants of adherence to stroke performance measures. *Stroke*. 2008 May;39(5):1619-20. <https://orcid.org/10.1161/STROKEAHA.107.496570>
- Rossaneis MA, Gabriel CS, Haddad MCFL, Melo MRAC, Bernardes A. Knowledge on health indicators by nurses in hospitalization unities. *Rev Eletr Enf*. 2014 Dec;16(4):769-76. <http://doi.org/10.5216/ree.v16i4.22956>
- Cavalcante RB, Cunha SGS, Bernardes MFVG, Gontijo TL, Guimarães EAA, Oliveira VC. Hospital Information System: use in decision making. *Journal of Health Informatics*. 2012;4:73-9.
- Patrício CM, Maia MM, Machiavelli JL, Navaes AN. O prontuário eletrônico do paciente no sistema de saúde brasileiro: uma realidade para os médicos. *Sci Med*. 2011;21(3):121-31.
- Brazil - Ministério da Saúde. Portaria nº. 665, de 12 de abril de 2012 [accessed on Nov 10, 2020]. Available at: http://bvsms.saude.gov.br/bvs/saudelegis/gm/2012/PRT0665_12_04_2012.html
- Nemati HR, Steiger DM, Iyer LS, Herschel RT. Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decis Support Syst*. 2002 Jun;33(2):14361. [https://doi.org/10.1016/S0167-9236\(01\)00141-5](https://doi.org/10.1016/S0167-9236(01)00141-5)
- Uraikul VU, Chan CW, Tontiwachwuthikul P. Artificial intelligence for monitoring and supervisory control of process systems. *Eng Appl Artif Intell*. 2007 Mar;20(2):115-31. <https://doi.org/10.1016/j.engappai.2006.07.002>
- Project Management Institute. Learn about PMI [accessed on June 06, 2019]. Available at: <https://www.pmi.org/about/learn-about-pmi>.
- Project Management Institute. Guia do Conhecimento em Gerenciamento de Projetos. Guia PMBOK. 6th ed. Newtown Square, PA: Project Management Institute; 2017.
- Araújo CC, Oliveira CA, Cruz C, Daniel D, Souza J, Caserta M. Gerenciamento de Banco de Dados: Análise Comparativa de SGBD'S [accessed on Aug 15, 2019]. Available at: <https://www.devmedia.com.br/gerenciamento-de-banco-de-dados-analise-comparativa-de-sgbd-s/30788>
- Rolim CLRC, Martins M. Quality of care for ischemic stroke in the Brazilian Unified National Health System. *Cad Saúde Pública*. 2011 Nov;27(11):2106-16. <https://doi.org/10.1590/S0102-311X2011001100004>
- Rolim CLRC, Martins M. Computerized tomography utilization for stroke inpatients in the Brazilian Health System. *Rev Bras Epidemiol*. 2012 Mar;15(1):179-87. <https://doi.org/10.1590/S1415-790X2012000100016>
- Araújo JP, Darcis JVV, Tomas ACV, Mello WA. Mortality trend due to cerebrovascular accident in the City of Maringá, Paraná between the years of 2005 to 2015. *Int J Cardiovasc Sci*. 2018 Feb;3(1):56-62. <https://doi.org/10.5935/2359-4802.20170097>
- Zétola VHF, Nôvak EM, Camargo CHF, Júnior HC, Coral P, Muzzio JA et al. Acidente vascular cerebral em pacientes jovens: análise de 164 casos. *Arq Neuro-Psiquiatr*. 2001 Sep;59(3B):740-5. <https://doi.org/10.1590/S0004-282X2001000500017>
- Martins SCO, Sacks C, Hacke W, Brainin M, Figueiredo FA, Pontes-Neto OM, et al. Priorities to reduce the burden of stroke in Latin American countries. *Lancet Neurol*. 2019 Jul;18(7):674-83. [https://doi.org/10.1016/S1474-4422\(19\)30068-7](https://doi.org/10.1016/S1474-4422(19)30068-7)
- Cabral NL, Freire AT, Conforto AB, Dos Santos N, Reis FI, Nagel V, et al. Increase of stroke incidence in young adults in a middle-income country: a 10-year population-based study. *Stroke*. 2017 Nov;48(11):2925-30. <https://doi.org/10.1161/STROKEAHA.117.018531>
- Rede Brasil AVC. Rede Brasil AVC [accessed on Jan 25, 2020]. Available at: <http://www.redebrasilavc.org.br/para-profissionais-de-saude/o-que-e-registro-sits/>
- Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst*. 2014 Feb;2:3. <https://doi.org/10.1186/2047-2501-2-3>
- Wang L, Alexander CA. Stroke care and the role of big data in healthcare and stroke. *Rehabil Sci*. 2016 Nov;1(1):16-24. <https://doi.org/10.11648/j.rs.20160101.13>