

NOTA TÉCNICA

Proposta de Integração de Dados do Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH-SUS) para Pesquisa

A Proposal to Integrate Data from the Hospitals Information System of the Unified Health System (SIH-SUS) for Research Use

Mônica R. Campos

Universidade Estadual do Rio de Janeiro

Mônica Martins

Fundação Oswaldo Cruz

José de C. Noronha

Universidade Estadual do Rio de Janeiro

Claudia Travassos

Fundação Oswaldo Cruz

Resumo

Nesta nota técnica apresenta-se um método para a criação de banco de dados gerado dos arquivos do Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH-SUS), colocados à disposição em CD-ROM pelo Ministério da Saúde. O banco de dados resultante deste processo contém dados que estão dispersos em vários arquivos com estrutura e objetivos diversos. Neste exemplo específico utilizaram-se os seguintes arquivos mensais: Movimento de AIH (MA); Movimento de Prestadores (MT) e Procedimentos Autorizados (PA).

Palavras-Chave

Informações Hospitalares; Integração de Bancos de Dados; Sistema de Informações Hospitalares (SIH-SUS).

Summary

Presented in this technical report is a method to develop a database originating from files of the Ministry of Health Hospital Admission Information System (SIH-SUS - Sistema de Informações Hospitalares do Sistema Único de Saúde), which are available to the public on CD-ROM. This database contains data from different files, each with a different structure and objective. In this example, three files were utilized on a monthly basis: Hospital Admission Authorizations (AIH -MA); information of health services covenant to SUS (MT) and Authorized Procedures (PA).

Key Words

Inpatient Information System; Data Linkage; SIH-SUS.

Apresentação

Informações geradas a partir de bancos de dados desenhados para fins administrativos - como o Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH-SUS) - têm sido empregadas para avaliar e acompanhar a qualidade de sistemas e serviços de saúde, apesar de apresentarem limitações.^{1,2} Destacam-se como principais vantagens dos bancos de dados administrativos o grande volume de casos registrados e o reduzido tempo entre a ocorrência do evento e seu registro no sistema. Considera-se, assim, que a utilização de tais bancos pode representar um importante avanço nas ações de avaliação e melhoria da qualidade dos serviços de saúde.

Esta nota técnica tem como objetivo divulgar o método de integração de dados utilizado para a criação de um banco de dados extraído do SIH-SUS, que pode ser aplicado em estudos que incorporam variáveis dos pacientes, do processo de cuidado e dos profissionais que prestaram serviços.

Esta nota técnica tem como objetivo divulgar o método de integração de dados utilizado para a criação de um banco de dados extraído do SIH-SUS, que pode ser aplicado em estudos que incorporam variáveis dos pacientes, do processo de cuidado e dos profissionais que prestaram serviços. Estes dados estão disponíveis neste sistema em diferentes arquivos com estruturas diversas. Desta forma, sempre que se fizer necessário utilizá-los conjuntamente, procedimentos específicos de integração de dados serão requeridos. Tendo em vista o grande volume de dados disponíveis no SIH-SUS e a complexidade do processo de integração de dados, buscou-se desenvolver uma estratégia voltada para aumentar a sua eficiência e diminuir as chances de erro na criação do novo banco.

Apresenta-se um exemplo de integração de dados utilizado em uma pesquisa (Morbidade e desempenho nos hospitais do Sistema Único de Saúde: estudo de tendências temporais. Coordenadora - Claudia Travassos; Financiamento - PAPES/FIOCRUZ; Projeto nº113) que analisa a adequação do cuidado prestado aos pacientes submetidos à cirurgia coronariana. Para avaliar em suas múltiplas facetas estas

internações, faz-se necessária a utilização de informações sobre os pacientes, a equipe médica, os hospitais e os procedimentos realizados durante a internação, com vistas a construção do modelo explicativo do risco de morrer dada a cirurgia.

O SIH-SUS produz um volume expressivo de informações, tendo como fonte primária de dados o instrumento de Autorização de Internação Hospitalar (AIH), que é o documento utilizado para o reembolso dos serviços prestados sob regime de internação nos hospitais com vínculo com o SUS. Fornece informações para o gerenciamento do Sistema, contendo informações detalhadas sobre o paciente, os prestadores e o cuidado prestado. Os dados do SIH-SUS são colocados à disposição em CD-ROM, pelo DATASUS, com periodicidade mensal para cada um dos municípios do país, desde 1993. Para melhor compreensão do instrumento AIH pode-se dividi-lo em seis blocos: a) identificação do hospital; b) identificação do paciente; c) identificação do responsável pelo paciente; d) caracterização da internação; e) procedimentos especiais; e f) serviços profissionais. Estes dados estão distribuídos por diversos arquivos com objetivos e estruturas diferenciados.

Dada a estrutura dos diversos arquivos disponíveis nos CD-ROM mensais, a criação do banco de dados para a pesquisa implicou a integração de três arquivos abaixo destacados:

- Arquivos Movimento de AIH (tipo **MA**), cujos registros correspondem a cada AIH enviada pelos hospitais, agregadas por município e por mês.
- Arquivos Movimento de Prestadores (tipo **MT**), cujos registros correspondem a cada ato médico realizado em cada internação, por prestador direto de serviço, agregados pelo país e por mês. Os registros contêm a variável Número da AIH que identifica cada internação. Vale destacar a

existência no CD-ROM mensal do arquivo tipo “TA” - ocorrência de atos médicos na AIH - que possui estrutura similar ao arquivo “MT”. Segundo comunicação pessoal da equipe do DATASUS, o arquivo “TA” inclui as internações dos hospitais federais com verba própria (natureza=31) que são pagos pelo sistema que se utiliza da AIH para reembolso. Essas internações que ocorrem apenas no Estado do Rio de Janeiro não estão contabilizadas nos arquivos “MT”.

- Arquivos Procedimentos Autorizados na AIH (tipo PA), cujos registros correspondem a cada Procedimento Especial realizado em cada internação, agregados por município e por mês.

Processo de elaboração da base de dados

Na criação da base de dados da pesquisa utilizaram-se os arquivos dos CD-ROM mensais de 1996, selecionando-se apenas os registros de interesse para a pesquisa, ou seja, aqueles cujo Procedimento Realizado tinha o código “32011016” - cirurgia coronariana com extra-corpórea. (Procedimento Realizado: variável do formulário AIH utilizada como unidade de pagamento pelo mecanismo de reembolso aos hospitais adotado pelo SUS. Estão classificados e codificados na Tabela de Procedimentos do SUS, disponível nos CD-ROM). Dado o caráter agudo de tal procedimento, a problemática de dupla contagem de registros por inclusão indevida de registros com AIH do tipo 5 (Crônico e Fora de Possibilidade Terapêutica) não se verifica.

Nos três tipos de arquivos apresentados, tanto as nomenclaturas das variáveis quanto os formatos dos arquivos são similares, seguindo a padronização dos arquivos no CD-ROM, diferindo apenas quanto à localização nele. Ou seja, todos têm o mesmo nome para cada variável, formato tipo DBF e estão compactados em DBC. (DBC - Extensão

de arquivos em formato compactado disseminados pelo DATASUS). Vale lembrar que a estrutura dos diretórios e dos arquivos no CD-ROM de movimentação mensal é por Unidade da Federação e secundariamente por tipo de arquivo (MA, MT e PA), com cada arquivo desagregado por município.

Para elaboração de uma base de dados nacional e anual, faz-se necessária a integração dos arquivos municipais em cada base de dados mensal e, posteriormente, a integração destas em uma única base anual. Este procedimento tem que ser realizado para cada tipo de arquivo (MA, MT e PA). Para a realização desses processos de integração utilizou-se o programa executável APPENDA, disponível no CD-ROM, que executa tanto a descompactação do formato DBC para formato DBF, quanto a integração dos arquivos de municípios e, posteriormente, dos mensais.

A fim de otimizar o processo de descompactação e integração dos arquivos para compor-se um arquivo único com dados de Brasil para o ano de 1996 percorreram-se os seguintes passos:

1. Criação de bancos de dados mensais para o país para cada tipo de arquivo:

- criação de um diretório para cada tipo de arquivo – MA, MT e PA – e consecutivamente a criação de um subdiretório para cada mês, uma vez que os arquivos em cada CD-ROM mensal possuem o mesmo nome;
- transferência de cada arquivo do CD-ROM mensal (por exemplo do diretório d:\RJ\MA\MA*.DBC) para o disco rígido (neste caso, c:\MA\JANEIRO\MA*.DBC);
- execução do programa APPENDA, no disco rígido, convertendo-se, por exemplo, arquivos do tipo MA de DBC para DBF e agregando-se os municípios com dados referentes ao mesmo mês.

É importante ressaltar que, uma vez

criadas as bases mensais em DBC para cada tipo de arquivo – MA, MT e PA –, onde os dados de cada município se encontram dispostos nesses subdiretórios mensais, utiliza-se, como mencionado acima, o programa APPENDA para agregar os municípios. Porém, ao expandir-se de DBC para DBF, é recomendável, para otimização de tempo e de espaço em disco rígido, expandir-se somente a base de dados de interesse, ou seja, recomenda-se a utilização do processo de filtragem para a seleção apenas dos registros necessários ao estudo. Os filtros utilizados nesta pesquisa são descritos logo a seguir por tipo de arquivo.

- Arquivos tipo **MA**: a variável PROC_REA que é o procedimento realizado, foi utilizada como filtro para seleção das unidades de observação. Logo, somente os registros cujo PROC_REA = “32011016” – procedimento de cirurgia coronariana com extra-corpórea – foram incluídos na base de dados.
- Arquivos do tipo **MT**: a variável ATO_PROF que, neste caso, assume o código do Procedimento Realizado, foi utilizada como filtro para seleção de unidades de observação, ou seja, utilizaram-se somente registros cujo ATO_PROF = “32011016” – procedimento de cirurgia coronariana com extra-corpórea.
- Arquivos do tipo **PA**: neste tipo de arquivo não existe uma variável a ser utilizada como filtro para as unidades de observação de interesse do estudo. Por isso, foi necessário agregar-se os registros sem filtro, o que não foi dificultado devido ao reduzido número de variáveis neste tipo de arquivo. A filtragem das unidades de observação ocorreu *a posteriori*, como será descrito em detalhes mais à frente.

É importante destacar, no entanto, que os processos de descompactação e integração de arquivos são realizados para

cada mês e para cada tipo de arquivo, o que corresponde, neste exemplo, a 36 vezes a execução dos processos acima. Assim, para reduzir as chances de erro na estrutura dos arquivos gerados em cada subdiretório e diminuir o tempo de trabalho, não se recomenda a seleção das variáveis de interesse nessa etapa do processo. Dessa forma, evita-se agregar arquivos com estruturas de colunas (o conjunto de variáveis) diferentes devido a erros na seleção de variáveis em cada um dos momentos de criação de cada um dos arquivos mensais. É mais indicado que a seleção das variáveis de interesse seja feita apenas uma única vez, isto é, na etapa de integração dos arquivos mensais.

2. Criação de banco de dados anual para o país para cada tipo de arquivo:

- A partir dos arquivos mensais, gerados na etapa anterior, para cada tipo de arquivos – MA, MT e PA – procede-se à integração deles em arquivos anuais, criando-se, por exemplo, o arquivo MA_96.DBF com a utilização do programa APPENDA.
- Simultaneamente, procede-se à seleção das variáveis de interesse no momento da integração dos arquivos mensais.

Processo de vetorização do banco de dados tipo “MT”

Nesta etapa do processo os arquivos tipo “MA”, “MT” e “PA” já se encontram com seus dados mensais e municipais agregados para o país, formando três bases de dados anuais. Para a criação de uma base de dados única, constituída a partir da integração destes três novos arquivos, é necessária a padronização dos registros, já que estes diferem entre si, como citado anteriormente. Tal divergência ocorre porque uma única AIH pode se repetir em várias linhas de um mesmo arquivo: tanto nos casos em que tenham sido autorizados mais de um Procedimento Especial (no máximo cinco) numa mesma internação (arquivos “PA”), ou, ainda, sejam referidos, para efeito de pagamento, vários integrantes

da equipe responsável pela internação (no máximo sete – arquivos “MT”). Assim, tal padronização, no caso específico desta pesquisa, buscou uniformizar os registros (linhas do arquivo) de forma a que representem uma única AIH, isto é, uma única internação.

Ao padronizar os registros dos arquivos “MT” e “PA”, pela variável número da AIH, faz-se necessária a execução de um procedimento específico denominado vetorização. Isto é, a criação de novos registros correspondendo, agora, a uma internação (uma única AIH) que contém em suas colunas (variáveis) as informações que anteriormente se encontravam distribuídas em vários registros nos bancos de dados originais. Dito de outra forma, a vetorização corresponde à transformação de registros (linhas) de um banco de dados em variáveis (colunas) de outro banco. Por exemplo, no caso do arquivo “MT”, a vetorização implicou transformar os registros definidos pela variável TIPO_ATO com categorias de 1 a 7 em um novo conjunto de sete variáveis no banco de dados no qual cada registro corresponde a um único número da AIH (internação). Cabe observar que a variável TIPO_ATO, tem vinte e uma categorias, mas, neste exemplo, trabalhou-se apenas com as sete categorias iniciais referentes à composição da equipe cirúrgica. As outras categorias não continham informação nos CD_ROM mensais para o ano de 1996 nas internações no procedimento Cirurgia Coronariana com Extra-Corpórea.

Após a agregação realizada no passo 2 do item “elaboração da base de dados”, descrito anteriormente, gerou-se um banco de dados tipo “MT” no qual foram computados 68.003 registros nas 11 variáveis. Tais registros representam o total de participantes das diversas equipes responsáveis pelas 13.680 internações ocorridas no Brasil, em 1996, no procedimento cirurgia coronariana com extra-corpórea. Vale notar que tal quantitativo expressa o fato de que cada equipe nestas cirurgias é geralmente,

constituída, em média, por cinco integrantes: um primeiro cirurgião, três auxiliares cirúrgicos e um anestesista.

Etapas do processo de vetorização do arquivo tipo “MT”:

- Definição da variável de referência para “vetorização” que no caso específico desta pesquisa, é a variável “TIPO_ATO”.
- Partição do arquivo original “MT_96.dbf” (contendo registros não-vetorizados por AIH) em sete arquivos, cada um contendo apenas registros correspondentes a uma única categoria da variável de referência. Para a realização desta etapa utilizou-se o programa SPSS 8.0, gerando os arquivos “MT01.sav”, “MT02.sav”, “MT03.sav”, “MT04.sav”, “MT05.sav”, “MT06.sav” e “MT07.sav”.
(*SELECT CASES / IF “TIPO_ATO” = “x”*)
- Execução do processo de integração dos arquivos gerados na etapa anterior, tendo a variável “Número de AIH” como chave de ligação entre os arquivos. Tal etapa deve ser executada arquivo por arquivo, consecutivamente. Nesta etapa, procede-se à transformação dos registros (linhas) dos bancos de dados (“MT0?.sav”) em variáveis (colunas) de outro banco (“MT_FIM.sav”). Ver ao final o programa SPSS - Partição e Merge para Vetorização por Número da AIH.
(*MERGE/ADD VARIABLES*)
- Atribuição de novo nome às variáveis contidas em cada um dos arquivos parciais. Por exemplo, para o arquivo MT01.sav, renomeia-se a variável “TIPO_ATO” para “TIP_ATO1”. Este procedimento é realizado concomitantemente ao procedimento anterior de integração dos arquivos.

Com a realização do processo de vetorização, foi possível, então, apurar-se que, do total de integrantes nas

diversas equipes (68003), 65% é de profissionais do próprio hospital, enquanto que os 35% restantes são profissionais contratados, em que este último percentual atinge seu máximo para o primeiro cirurgião, cerca de 43% de contratados. Observa-se, ainda, a distribuição e composição dos integrantes da equipe cirúrgica por tipo de contrato no conjunto das internações de interesse (Tabela 1).

(“count”) e a variável Número da AIH. Sua origem foi o banco de dados tipo “MA” com agregados em nível nacional para 1996, contendo todas as AIH’s cujo procedimento fosse “32011016” (arquivo ma_96.dbf).

- definição da variável de referência, do arquivo “PA”, para realização do processo de vetorização que no caso específico deste arquivo é a variável “SEQUENCIAL”, que indica a

Tabela 1 - Composição da equipe cirúrgica por tipo de contrato em internações de cirurgia coronariana com extra-corpórea. Brasil, 1996. (Arquivo “MT_FIM.sav”)

Composição da Equipe Cirúrgica	Tipo de Contrato Profissional				Total
	Profissional do Hospital		Profissional Contratado		
1º Cirurgião	7.803	57,4	5.802	42,6	13.605
1º Auxiliar Cirúrgico	8.081	60,2	5.351	39,8	13.432
2º Auxiliar Cirúrgico	8.333	65,1	4.472	34,9	12.805
3º Auxiliar Cirúrgico	8.020	68,3	3.717	31,7	11.737
Demais Auxiliares Cirúrgicos	2.819	99,9	2	0,1	2.821
Anestesista	9.145	67,6	4.386	32,4	13.531
Consulta Clínica	72	100,0	0	0,0	72
Total de profissionais nas equipes cirúrgicas	44.273	65,1	23.730	34,9	68.003

Fonte: SIH-SUS.

Processo de vetorização do banco de dados tipo “PA”

O banco de dados do tipo “PA” gerado nas etapas anteriores, possui as seguintes variáveis: Número da AIH (“N_AIH”), SEQUENCIAL e o Procedimento Especial (“PRCD_AUT”) e contém 1.732.148 registros que correspondem ao total de procedimentos especiais pagos no Brasil no ano de 1996, em todas as internações. Como já mencionado, não existe no arquivo “PA” uma variável que possa ser utilizada como filtro dos registros de interesse, exigindo a realização de um processo destinado a selecionar apenas as AIH objeto.

Etapas do processo de vetorização do arquivo tipo “PA”:

- Seleção das internações de interesse a partir da criação de um arquivo denominado “N_AIH.sav”, contendo uma variável de controle

posição relativa do Procedimento Especial nos cinco campos disponíveis para esta variável no formulário da AIH.

- fracionamento do arquivo original “PA_96.dbf” (contendo registros não-vetorizados por AIH e ainda não selecionados pelas internações de interesse) em cinco arquivos, cada um contendo apenas registros correspondentes a uma única categoria da variável de referência. Para a realização desta etapa utilizou-se o programa SPSS 8.0, gerando os arquivos “PA01.sav”, “PA02.sav”, “PA03.sav”, “PA04.sav”, “PA05.sav”. Assim, cria-se um arquivo para cada “ordem” de procedimento especial indicado pela variável “SEQUENCIAL”.

(*SELECT CASES / IF “SEQUENCIAL” = “x”*)

- execução do processo de integração dos arquivos gerados nas etapas anteriores, isto é, integração do arquivo "N_AIH.sav" com cada um dos novos arquivos "PA0?.sav", tendo a variável "Número da AIH" como chave de ligação entre os arquivos.

(MERGE/ADD VARIABLES)

- Na criação dos arquivos "PA0?.sav", imediatamente após o processo de integração, procede-se à seleção das internações de interesse em cada arquivo, a partir da filtragem dos registros de interesse com base na variável "Count", oriunda do arquivo "N_AIH.sav".

("SELECT CASES" / IF "count" = "1")

- Integração dos arquivos "PA0?.sav", contendo apenas as internações de interesse. Tal etapa deve ser executada arquivo por arquivo consecutivamente, onde procede-se à transformação dos registros (linhas) dos bancos de dados ("PA0?.sav") em variáveis (colunas) de outro banco ("PA_FIM.sav").

(MERGE/ADD VARIABLES)

- Atribuição de novo nome às variáveis contidas em cada um dos arquivos parciais. Por exemplo, para o arquivo PA01.sav, renomeiam-se as variáveis "SEQUENCIAL" para "SEQ1" e "PRCD_AUT" para "PRC_AUT1". Este procedimento é realizado concomitantemente ao procedimento anterior de integração dos arquivos.

O banco de dados resultante do processo acima, de integração de dados dos arquivos tipo **MA**, **MT** e **PA**, permite, por exemplo, a realização de análises sobre variações nas taxas de mortalidade hospitalar segundo características do paciente, da equipe cirúrgica e do processo de cuidado. Com relação ao processo de cuidado, a análise pode

incorporar o uso de procedimentos diagnósticos e terapêuticos específicos (procedimento especiais) como por arteriografia, albumina e nutrição parenteral. Por exemplo, apurou-se que, em 52% das 13.680 internações no procedimento cirurgia coronariana ocorreu pelo menos um primeiro Procedimento Especial (Tabela 2).

Considerou-se este processo eficiente na criação de uma base de dados originária de arquivos diversos, com registros variados e com grande volume de casos.

Tabela 2 - Distribuição de frequência da ocorrência de procedimentos especiais em internações de cirurgia coronariana com extra-corpórea. Brasil, 1996. (Arquivo "PA_FIM.sav")

Número de procedimentos	Frequência de internações	Percentual
0	2.184	16,0
1	7.130	52,1
2	3.085	22,5
3	984	7,2
4	233	1,7
5	64	0,5
Total	13.680	100,0

Fonte: SIH-SUS.

Exemplo:

Programas SPSS

Criação dos bancos de dados

MT01 a MT07

GET

FILE='C:\PAPES\MT\MT_96.sav'.

EXECUTE.

FILTER OFF.

USE ALL.

SELECT IF(tipo_ato = "01").

EXECUTE.

SAVE OUTFILE = 'C:\PAPES\MT\MT01.sav'

/COMPRESSED.

Comentário1: O programa foi repetido para cada categoria da variável "tipo_ato" = 01 até 07, gerando-se desta forma os arquivos mt01.sav até o MT07.sav

Programa para integração de arquivos

```

GET
FILE='C:\PAPES\MT\MT07.sav'.
EXECUTE .
SORT CASES BY
N_AIH (A).
/RENAME tipo_ato = tip_ato7
cgc_cpf=cgc_cpf7 id_cgc_c=id_cgc_7
qtd_ato=qtd_ato7 sp_ponto=sp_pont7
tipo=tipo7 val_ato=val_ato7
MATCH FILES /TABLE=*
/FILE='C:\PAPES\MT\MT06.sav'
/RENAME (d_r = d0 ) tipo_ato=tip_ato6
cgc_cpf=cgc_cpf6 id_cgc_c=id_cgc_6
qtd_ato=qtd_ato6 sp_ponto=sp_pont6
tipo=tipo6 val_ato=val_ato6
/BY N_AIH
/DROP= d0.
EXECUTE.
SORT CASES BY
N_AIH (A). SAVE OUTFILE =
'C:\PAPES\MT\MT_FIM.sav'
/COMPRESSED.

```

Comentário 2: O programa acima também foi repetido para cada um dos arquivos (MT07.sav à MT01.sav), agregando-se a cada passo um novo conjunto de sete variáveis associadas a cada categoria da variável “TIPO_ATO”, ou seja, a cada integrante da equipe responsável pela mesma cirurgia. Tal processo tem como resultado um arquivo único (“MT_FIM.SAV”) que contém internações por linha e conjuntos de variáveis (7) de cada integrante da equipe cirúrgica por coluna.

Referências bibliográficas

1. Iezzoni LI. Assessing quality using administrative data. *Annals of Internal Medicine* 1997; 8(part 2) : 666-674.
2. Wray NP, Ashton CM, Kuykendall DH, Hollingsworth JC. Using administrative databases to evaluate the quality of medical care: a conceptual framework. *Social Science and Medicine* 1995; 12 : 1707-1715.