

csapAIH: uma função para a classificação das condições sensíveis à atenção primária no programa estatístico R*

doi: 10.5123/S1679-49742017000100021

csapAIH: a function to classify ambulatory care sensitive conditions in the statistical software R

Fúlvio Borges Nedel^{1,2}

¹Universidade Federal de Santa Catarina, Departamento de Saúde Pública, Florianópolis-SC, Brasil

²Universitat Autònoma de Barcelona, Facultat de Medicina, Unitat de Bioestadística, Grups de Recerca d'Amèrica i Àfrica Llatines, Barcelona, Espanha

Resumo

Hospitalizações por condições sensíveis à atenção primária (CSAP) são um indicador indireto da efetividade da atenção primária à saúde. Para automatizar a classificação de códigos da Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde – Décima Revisão (CID-10), segundo a lista brasileira de CSAP, e prover funcionalidades para o manejo dos "arquivos reduzidos" da Autorização de Internação Hospitalar (AIH), foi criado um pacote no programa R. O pacote csapAIH contém uma única função homônima, que lê os dados segundo sua natureza (arquivo ou *data frame* com a estrutura da AIH, ou fator com códigos da CID-10) e retorna, segundo as opções definidas, um banco de dados ou vetor contendo a classificação da internação. Este artigo apresenta o pacote e a função csapAIH, seu modo de instalação e uso, além de exemplos de suas funcionalidades que permitem conferir maior agilidade, precisão e validade ao estudo das CSAP no Brasil.

Palavras-chave: Hospitalização; Atenção Primária à Saúde; Pesquisa sobre Serviços de Saúde; Sistemas de Informação em Saúde; Processamento Automatizado de Dados.

Abstract

Hospitalizations due to ambulatory care sensitive conditions (ACSC) are an indirect indicator of primary health care. A package in the R program was developed in order to automatize the classification of codes of the International Statistical Classification of Diseases and Related Health Problems – 10th Revision (ICD-10), according to the Brazilian list of ACSC, and provide functionalities for the management of the "reduced files" of the inpatient hospital authorization (AIH). The csapAIH package contains a homonym function, which reads the data according to its nature (file or data frame with AIH structure, or a factor with ICD-10 codes) and returns, according to defined options, a databank or vector containing the classification for the hospitalization. This article presents the package and the function csapAIH, its installation mode and use, and examples of its functionalities, which may add quickness, precision and validity to the research of ACSC in Brazil.

Keywords: Hospitalization; Primary Health Care; Health Services Research; Health Information Systems; Automatic Data Processing

* O script para a classificação das CSAP segundo a Lista Brasileira de Condições Sensíveis à Atenção Primária foi publicado, originalmente, na tese de Doutorado do autor, defendida em 2009, junto ao Programa de Pós-Graduação em Epidemiologia da Universidade Federal de Pelotas.

Endereço para correspondência:

Fúlvio Borges Nedel – Universidade Federal de Santa Catarina, Centro de Ciências da Saúde, Departamento de Saúde Pública, Campus Universitário Reitor João David Ferreira Lima, Trindade, Florianópolis-SC, Brasil. CEP: 88040-970
E-mail: fulvionedel@gmail.com

Introdução

As hospitalizações por condições sensíveis à atenção primária (CSAP) são um indicador indireto da efetividade da atenção primária à saúde (APS).^{1,2} No Brasil, seu estudo é facilitado pelos seguintes fatores: (i) disponibilização dos "arquivos reduzidos" das bases de dados do Sistema de Informações Hospitalares do Sistema Único de Saúde (BD-SIH/SUS), pelo Departamento de Informática do SUS (Datusus); (ii) publicação da Portaria do Ministério da Saúde (MS/GM nº 221, de 17 de abril de 2008) com a Lista Brasileira de CSAP;^{3,4} e (iii) disponibilização de tabelas de conversão/arquivos .CNV, com suas correspondentes linhas nos arquivos de definição/arquivos .DEF, para a tabulação dessas internações pelo TabWin.⁵

O TabWin, programa criado pelo Datusus para a tabulação das bases de dados de diversos sistemas de informações em saúde existentes no Brasil, tem entre seus atrativos a (i) eficiência na tabulação de grande quantidade de dados, (ii) uma ampla gama de *scripts* pré-definidos para a tabulação das BD-SIH/SUS (entre outros sistemas de informações), a (iii) geração de mapas das unidades territoriais e administrativas de saúde brasileiras, a (iv) capacidade de ler os arquivos comprimidos (extensão .DBC, criada pelo Ministério da Saúde) da Autorização de Internação Hospitalar (AIH) e de (v) criar novos arquivos de bases de dados a partir da junção de arquivos .DBC com seleção de casos e variáveis, além da (vi) integração com o programa R,⁶ permitindo a exportação da tabela gerada para aquele programa.

O R é uma linguagem de computação voltada ao manejo e análise estatística avançada de dados, facilmente extensível mediante funções e pacotes criados pelo próprio usuário, para necessidades específicas. Trata-se de um *software* livre, apropriado a esse tipo de iniciativas, como os 'esquemas de análise': uma série de *scripts* disponibilizados pelo Datusus em seu portal, para a análise de dados em R. Infelizmente, a última atualização da página – bastante antiga – aconteceu em julho de 2005 (<http://www2.datusus.gov.br/DATASUS/tabwin/rx/ListarX.htm>).

Uma tabela, muitas vezes, é inadequada à análise de interesse. Isto é menos frequente nos chamados estudos ecológicos mas, virtualmente, é o caso de todas as pesquisas que têm indivíduos como unidades de análise, a exemplo dos inquéritos hospitalares.⁷ Em tal situação, obriga-se à análise do banco de dados não tabulado, com uma das variáveis indicando se a causa da internação é ou não uma CSAP. Deve-se evitar a seleção causa por causa, para garantir a reprodutibilidade e minimizar a chance de erros na seleção dos casos. Sendo assim, o pesquisador terá de utilizar um programa de manejo ou análise de dados e finalmente, escrever a sintaxe de classificação das causas de internação para criar a variável.

A construção de um *script* para essa classificação é trabalhosa e bastante sujeita a erros, o que impõe alguns cuidados sobre determinados fatores: (i) deve-se conhecer as regras de sintaxe do programa e estar habituado a elas; (ii) os códigos da Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde – Décima Revisão (CID-10) são alfanuméricos, e nem todos têm a mesma extensão (podem ter 3 ou 4 dígitos); e (iii) é grande a quantidade de códigos considerados CSAP, distribuídos entre 19 grupos de causas. Portanto, ferramentas que permitam a automação dessa etapa são oportunas.

Com o propósito de contribuir para o preenchimento dessa lacuna, este trabalho apresenta um pacote do programa R – que inclui uma função homônima – com o objetivo de automatizar a classificação de códigos da CID-10 segundo a Lista Brasileira de Condições Sensíveis à Atenção Primária e, principalmente, facilitar o trabalho com os arquivos descompactados da AIH (rd*.dbf), permitindo, também, operar qualquer vetor do R contendo códigos da CID-10. Este artigo apresenta o pacote e a função csapAIH, seu modo de instalação e uso, além de exemplos de suas funcionalidades.

Desenvolvimento do pacote csapAIH

O R funciona com objetos em um espaço de trabalho. Entre seus objetos mais simples estão os vetores, que combinam diferentes observações sobre uma variável; por exemplo, os códigos da CID-10 registrados como diagnóstico principal em uma série de internações hospitalares. Entre os objetos mais complexos do programa R estão os bancos de dados (*data frames*). Um vetor pode ser anexado como variável a um banco de dados.

Uma função no R é um comando, normalmente dirigido a um objeto, para a execução de uma série de cálculos ou transformações, e retorna um resultado, geralmente possível de ser convertido em um novo objeto. As funções têm seus argumentos, ou opções de execução. Os argumentos podem ser obrigatórios ou não, e ter um padrão de resposta estabelecido (*default*) ou não.

Para a construção da função `csapAIH`, consideraram-se as seguintes premissas:

1. Seu foco principal é o manejo dos arquivos da AIH, e assim a função deve ter por padrão opções voltadas à leitura e manejo dessas bases.
2. Para análises da proporção de CSAP entre as internações, é frequente o usuário interessar-se em excluir as internações que não são devidas a doença ou lesão, como os partos e curetagens por abortamento, e a função deve, portanto, ter uma opção (argumento) que, por padrão, exclua esses registros mas permita o contrário.
3. As AIH de longa permanência (AIH tipo 5: continuidade) são novas AIH geradas para a mesma internação,⁸ e devem ser excluídas para que uma mesma internação não seja considerada mais de uma vez na análise, embora o usuário possa se interessar em manter esses registros; logo, a função deve ter um argumento que, por padrão, exclua esses registros, mas permita o contrário.
4. Os arquivos da AIH têm uma série de variáveis (113, atualmente), muitas de interesse específico e pouco utilizadas na maioria dos estudos; ao trabalhar com esses arquivos, a função deve retornar um banco de dados com um conjunto mínimo – não demasiado extenso – de variáveis sobre o paciente e sua internação.
5. O usuário pode se interessar pela análise de outras bases de dados, ou simplesmente tabular uma série de diagnósticos; logo, a função deve ser capaz de classificar qualquer lista de códigos da CID-10.
6. Ainda que as BD-SIH/SUS registrem a data de nascimento do paciente (campo NASC), o registro da idade em anos completos deve ser feito por um algoritmo combinando os campos IDADE e COD_IDADE; convém que a função automatize esse processo.

A função está organizada em três momentos:

- 1º Leitura dos dados e, para os arquivos da AIH, seleção de registros e criação das variáveis idade (idade civil, segundo o algoritmo recomendado⁸) e faixa etária (detalhada, com 33 categorias; e quinquenal, com 17 categorias); utilizam-se as funções `grepl` do pacote `base`, `read.dbf` do pacote `foreign`⁹ e `read.csv` do pacote `utils`;
- 2º Classificação do código CID-10 do diagnóstico da internação em CSAP, como variável dicotômica, e em seus 19 grupos de causa, como variável politômica; e
- 3º Formatação do resultado retornado, se um banco de dados ou um vetor de classe fator, isto é, uma variável categórica, que pode permanecer como objeto isolado ou ser endereçada a um banco de dados existente no espaço de trabalho do R.

A função tem a seguinte estrutura:

```
csapAIH(x, grupos=TRUE, sihsus=TRUE, x.procobst=TRUE, longa=FALSE, cep=TRUE,
cnes=TRUE, arquivo=TRUE, sep,...)
```

Onde:

x é o alvo da função, sobre o qual serão aplicados os demais argumentos; ele pode ser:

- a) um arquivo em formato `.DBF` ou `.CSV`, com uma base de dados contendo um conjunto mínimo das variáveis das BD-SIH/SUS:
IDENT, DIAG_PRINC, NASC, DT_INTER, DT_SAIDA, COD_IDADE, IDADE, MUNIC_RES, MUNIC_MOV, SEXO, N_AIH, PROC_REA;
- b) um objeto da classe `data.frame`, com o mesmo conjunto mínimo das variáveis das BD-SIH/SUS, presente no espaço de trabalho;
- c) um objeto da classe fator presente no espaço de trabalho – note-se que um fator pode ser tanto um objeto isolado quanto tomado isoladamente de uma base de dados, como em `"base$variavel"`;
grupos: argumento lógico; se `TRUE` (padrão), será criada uma variável com a classificação da internação segundo o grupo de causa CSAP;

sihsus: argumento lógico; `TRUE` (padrão) indica que o alvo da função tem a estrutura mínima da BD-SIH/SUS citada anteriormente;

x.procobst: argumento lógico; TRUE (padrão) exclui as internações pelos procedimentos obstétricos citados na Tabela 1; ignorado se o alvo for um fator, se *sihsus*=FALSE ou se a base de dados incluir códigos da versão anterior da Tabela de Procedimentos do SIH/SUS;

longa: argumento lógico; FALSE (padrão) exclui as internações de longa permanência (AIH tipo 5);

cep: argumento lógico; TRUE (padrão) indica que o alvo da função é uma BD-SIH/SUS e que, além do conjunto mínimo de variáveis, tem o código de endereçamento postal (CEP) do indivíduo; o CEP será incluído no resultado da função; ignorado, se o alvo for um fator ou *sihsus*=FALSE;

cnes: argumento lógico; TRUE (padrão) indica que o alvo da função é uma BD-SIH/SUS e que, além do conjunto mínimo de variáveis, dispõe o número do Cadastro Nacional de Estabelecimentos de Saúde (CNES) do hospital em que ocorreu a internação; o número do CNES será incluído no resultado da função; essa variável foi incluída nos arquivos da AIH em agosto de 2003; para datas anteriores, o argumento deve ser definido como FALSE; ignorado, se o alvo for um fator ou *sihsus*=FALSE;

arquivo: argumento lógico; obrigatório, porém definido automaticamente pela função quando necessário; deve ser definido pelo usuário como FALSE apenas quando *x* contiver, em seu nome, as sequências *dbf* ou *csv* sem que isso seja a extensão de um arquivo; apenas arquivos com a estrutura citada acima podem ser lidos;

sep: permite a leitura de arquivos da AIH em formato CSV; pode ser ";" ou ",".

A exclusão das internações por causas obstétricas é realizada pelos códigos da Tabela de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais (OPM) do SUS. Essa tabela, instituída em 2007,¹⁰ unifica e substitui (desde 2008) as anteriores Tabelas de Procedimentos Ambulatoriais e Hospitalares do SIH/SUS. Assim, o argumento *x.procobst* só é válido para tabela unificada, sendo ignorado se os códigos forem incompatíveis com os listados na Tabela 1.

Para facilitar a distribuição, instalação e carregamento da função no R, e permitir o uso de arquivos de ajuda integrados à sessão de trabalho, foi criado um pacote com o mesmo nome da função: *csapAIH*. Como foi escrito em português, não passa por todos os testes para submissão ao Comprehensive R Archive Network (CRAN),¹¹ o repositório oficial de pacotes do R. Entretanto, em uma sessão do R com conexão à internet, o pacote *csapAIH* pode ser instalado desde seu repositório no portal GitHub (<https://github.com/fulvionedel/csapAIH.git>), por meio do pacote *devtools*, nas plataformas Linux, Mac OS X e Windows. Sua instalação, carregamento e utilização serão demonstrados na sessão seguinte. Maiores informações encontram-se no manual do pacote.¹²

Os arquivos reduzidos da AIH (RD*.dbf) utilizados como exemplos nos arquivos de ajuda e manual do programa, assim como neste artigo, não permitem a identificação do indivíduo e estão disponíveis, para acesso público, no

Tabela 1 – Procedimentos obstétricos cujas internações são excluídas da base de dados com o argumento *x.procobst*=TRUE

Código	Nome do procedimento
03.10.01.001-2	Assistência ao parto sem distócia
03.10.01.002-0	Atendimento ao recém-nascido em sala de parto
03.10.01.003-9	Parto normal
03.10.01.004-7	Parto normal em gestação de alto risco
04.11.01.001-8	Descolamento manual de placenta
04.11.01.002-6	Parto cesariano em gestação de alto risco
04.11.01.003-4	Parto cesariano
04.11.01.004-2	Parto cesariano com laqueadura tubária
04.11.02.001-3	Curetagem pós-abortamento/puerperal
04.11.02.002-1	Embriotomia

Fonte: Tabela de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais (OPM) do Sistema Único de Saúde (SUS) ⁹

portal do Datasus: Transferência de Arquivos >> SIH/SUS.¹³ Nos exemplos apresentados neste relato, o número da AIH foi substituído pelo número de ordem do registro na base de dados.

O *script* para a classificação das CSAP segundo a Lista Brasileira³ foi publicado, originalmente, em tese de Doutorado.¹⁴ O repositório no GitHub é público, o que significa que está aberto a contribuições, sempre bem-vindas.

Instalação do pacote csapAIH

O pacote pode ser instalado a partir de arquivos locais ("csapAIH_0.0.1.zip" ou "csapAIH_0.0.1.tar.gz" – *download* em <https://sourceforge.net/projects/csapaih/>), pelo menu de instalação de pacotes da *front-end*/GUI utilizada (própria do R, RStudio, Rkward...), ou a partir de seu repositório no GitHub. Com o pacote devtools previamente instalado, o comando `devtools::install_github("fulvionedel/csapAIH")` instala o pacote:

```
install.packages("devtools") # instala o pacote devtools
devtools::install_github("fulvionedel/csapAIH") # instala o pacote csapAIH
```

Exemplos de uso

A seguir, são apresentados alguns exemplos de uso da função.

1. Um arquivo reduzido da AIH descompactado (RD*.DBF), armazenado no diretório da sessão de trabalho

Trabalhando com as opções-padrão da função, o usuário deve apenas digitar o nome do arquivo entre aspas (" ") e com a extensão. A função retorna um banco de dados com 16 variáveis (Tabela 2), sem os registros das AIH de longa permanência (repetidas) e das internações por procedimentos obstétricos. O número e a proporção

Tabela 2 – Variáveis da base de dados criada com a aplicação dos argumentos-padrão da função csapAIH sobre um arquivo da AIH ^a

Ordem	Nome	Rótulo	Classe
1	n.aih	Número da AIH ^a	Factor
2	munres	Município de residência	Factor
3	munint	Município de internação	Factor
4	sexo	Sexo	Factor
5	nasc	Data de nascimento	Date
6	idade	Idade	Numeric
7	fxetar	Faixa etária detalhada	Factor
8	fxetar5	Faixa etária quinzenal	Factor
9	csap	CSAP ^b	Factor
10	grupo	Grupo de causa da CSAP ^b	Factor
11	cid	CID-10 ^c	Factor
12	proc.rea	Procedimento realizado	Factor
13	data.inter	Data de internação	Date
14	data.saida	Data de saída	Date
15	cep	Número do CEP ^d	Character
16	cnes	Número do hospital no CNES ^e	Character

a) AIH: autorização de internação hospitalar

b) CSAP: condições sensíveis à atenção primária

c) CID-10: Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde – Décima Revisão

d) CEP: código de endereçamento postal

e) CNES: Cadastro Nacional de Estabelecimentos de Saúde

desses casos são informados em uma mensagem ao final do processo. Como em qualquer ação análoga no R, o resultado deve ser direcionado a um objeto para que permaneça na sessão de trabalho ativa.

O exemplo a seguir, com o arquivo da AIH para o estado de Santa Catarina em janeiro de 2015, apresenta o comando e os primeiros três registros do banco de dados criado:

```
library(csapAIH) # carrega o pacote
# ?csapAIH # mostra a ajuda da função
exemplo1 <- csapAIH("RDSC1501.dbf")
Importados 35.854 registros.
Excluídos 6.216 (17,3%) registros de procedimentos obstétricos.
Excluídos 465 (1,6%) registros de AIH de longa permanência.
Exportados 29.173 registros.
```

Os primeiros três registros do banco de dados criado:

```
head(exemplo1, 3)
n.aih munres munint sexo      nasc idade fxetar fxetar5 csap
1 00000000000001 421660 421660 masc 1985-10-29 29 25-29 25-29 não
2 00000000000002 420540 421660 fem 2015-01-04 0 <1ano 0-4 não
3 00000000000003 420570 421660 fem 2015-01-03 0 <1ano 0-4 não
grupo cid proc.rea data.inter data.saida cep cnes
1 não-CSAP T023 0308010019 2015-01-13 2015-01-16 88122515 2555646
2 não-CSAP P968 0303160039 2015-01-04 2015-01-05 88090840 2555646
3 não-CSAP P240 0303160039 2015-01-03 2015-01-11 88495000 2555646
```

Os arquivos reduzidos da AIH compactados (extensão .DBC), disponibilizados para *download* no portal do Datasus, são organizados por Unidade da Federação (UF) e mês de referência (mês de faturamento) da AIH, com cada arquivo contendo o conjunto de dados para uma UF e mês de referência. Ao expandir esses arquivos para .DBF no TabWin, o usuário tem a opção de unificar a informação referente a um ano – exemplo: todos os meses de 2015 para os arquivos de Santa Catarina – ou a um ano e região – exemplo: todas as internações do Brasil em 2015 – em um único arquivo .DBF. A função pode ser utilizada com esses arquivos da mesma forma, sem mudar em nada seu comportamento.

2. Um arquivo da AIH como data.frame no espaço de trabalho

Se o arquivo da AIH já tiver sido lido e estiver presente no espaço de trabalho como um objeto da classe data.frame, essa situação é identificada pela função, e a sintaxe se modifica porque um objeto no espaço de trabalho é chamado sem estar entre aspas. No exemplo a seguir, o arquivo da AIH, primeiramente, é carregado como objeto no espaço de trabalho (pela função read.dbf, do pacote foreign) e, posteriormente, aplicada a função csapAIH sobre o objeto. Nota-se que a sintaxe de uso da função csapAIH difere do exemplo anterior apenas por não usar aspas:

```
library(foreign)
bd <- read.dbf("RDSC1501.dbf")
exemplo2 <- csapAIH(bd)
Excluídos 6.216 (17,3%) registros de procedimentos obstétricos.
Excluídos 465 (1,6%) registros de AIH de longa permanência.
Exportados 29.173 registros.
```

Os primeiros três registros do banco de dados criado:

```
head(exemplo2, 3)
n.aih munres munint sexo      nasc idade fxetar fxetar5 csap
1 00000000000001 421660 421660 masc 1985-10-29 29 25-29 25-29 não
```

```

2 0000000000002 420540 421660 fem 2015-01-04 0 <1ano 0-4 não
3 0000000000003 420570 421660 fem 2015-01-03 0 <1ano 0-4 não
grupo cid proc.rea data.inter data.saida cep cnes
1 não-CSAP T023 0308010019 2015-01-13 2015-01-16 88122515 2555646
2 não-CSAP P968 0303160039 2015-01-04 2015-01-05 88090840 2555646
3 não-CSAP P240 0303160039 2015-01-03 2015-01-11 88495000 2555646

```

3. Arquivos da AIH sem as variáveis CNES ou CEP

A ausência das variáveis CNES ou CEP no arquivo da AIH resulta em erro quando a função é aplicada com seus argumentos-padrão, logo definidos como FALSE. O exemplo a seguir mostra o erro e a sintaxe adequada ao trabalho com um arquivo da AIH anterior a agosto de 2003, portanto sem a variável CNES (trabalha-se com o arquivo de dados de Santa Catarina referentes a janeiro de 2003):

```

exemplo3 <- csapAIH('RDSC0301.dbf')
Importados 33.596 registros.
Excluídos 0 (0%) registros de procedimentos obstétricos.
Excluídos 698 (2,1%) registros de AIH de longa permanência.
Exportados 32.898 registros.
Error in attr(x, "label") <- value :
tentativa de especificar um atributo em um NULL
names(exemplo3)
Erro: objeto 'exemplo3' não encontrado
exemplo3 <- csapAIH('RDSC0301.dbf', cnes = FALSE)
Importados 33.596 registros.
Excluídos 0 (0%) registros de procedimentos obstétricos.

Excluídos 698 (2,1%) registros de AIH de longa permanência.

```

```

Exportados 32.898 registros.
head(exemplo3, 3)
n.aih munres munint sexo      nasc idade fxetar fxetar5 csap
1 0000000000001 420010 420010 masc 1928-06-24 74 70-74 70-74 sim
2 0000000000002 420253 420010 masc 1938-11-13 63 60-64 60-64 sim
3 0000000000003 421185 420010 masc 2001-08-26 1 1ano 0-4 não
grupo cid proc.rea data.inter data.saida cep
1 g08 J438 76500080 2002-09-30 2002-10-08 89830000
2 g08 J449 76500225 2002-10-17 2002-10-24 89824000
3 não-CSAP J188 76300072 2002-12-10 2002-12-14 89834000

```

Como versões anteriores das BD-SIH/SUS usavam outra codificação para os procedimentos, mesmo com o argumento `x.procobst=TRUE` (padrão da função, portanto não precisa ser anunciado), não foi excluído nenhum procedimento obstétrico.

4. Uma lista de códigos isolada ou como variável em um banco de dados sem a estrutura do SIH/SUS

Se o usuário estiver trabalhando com uma lista de códigos da CID-10 como objeto isolado no espaço de trabalho, deve-se definir o argumento `arquivo=FALSE` e então, a função retorna um banco de dados com as variáveis CID, CSAP e GRUPO. Se o objeto for definido como um fator, é desnecessário o argumento `arquivo=FALSE` porque a função o define internamente, como no exemplo a seguir:

```

cids <- c("A01", "B77", "C03", "D04", "E10", "O23", "P350")
class(cids)

```

```
[1] "character"
csapAIH(factor(cids)) # mesmo que csapAIH (cids, arquivo = FALSE)
csap grupo cid
1 sim g02 A01
2 sim g01 B77
3 não não-CSAP C03
4 não não-CSAP D04
5 sim g13 E10
6 sim g19 O23
7 sim g19 P350
```

Da mesma forma, pode-se trabalhar com uma base de dados sem o formato das BD-SIH/SUS. Aplica-se a função `csapAIH` sobre a variável com os códigos da CID-10 e, se for de interesse do usuário agregar as variáveis criadas ao banco de dados original, isto pode ser feito com a função `cbind`, como no exemplo abaixo, em que se aplica a função sobre uma base de dados de um inquérito hospitalar:⁷

```
head(bage, 3)
gidade gsexo gabep5 gmodelo ghospit gcidint
1 2 masculino D outro/sem USref. HU A040
2 3 masculino C PSF Sta. Casa A041
3 2 feminino D PSF HU A044
exemplo4 <- csapAIH(bage$gcidint, arquivo = FALSE)
head(exemplo4, 3)
csap grupo cid
1 sim g02 A040
2 sim g02 A041
3 sim g02 A044
bage2 <- cbind(bage, exemplo4[1:2])
head(bage2, 3)
gidade gsexo gabep5 gmodelo ghospit gcidint csap grupo
1 2 masculino D outro/sem USref. HU A040 sim g02
2 3 masculino C PSF Sta. Casa A041 sim g02
3 2 feminino D PSF HU A044 sim g02
```

O processo pode ser executado em um único passo:

```
exemplo5 <- cbind(bage, csapAIH(bage$gcidint, arquivo = FALSE)[1:2])
head(exemplo5, 3)
gidade gsexo gabep5 gmodelo ghospit gcidint csap grupo
1 2 masculino D outro/sem USref. HU A040 sim g02
2 3 masculino C PSF Sta. Casa A041 sim g02
3 2 feminino D PSF HU A044 sim g02
```

Finalmente, se for de interesse do usuário criar apenas uma variável dicotômica sobre a classificação CSAP, define-se o argumento `grupos` como `FALSE`:

```
exemplo6 <- csapAIH(bage$gcidint, arquivo = FALSE, grupos = F)
class(exemplo6) ; head(exemplo6)
[1] "factor"
[1] sim sim sim sim sim sim
Levels: sim não
exemplo7 <- cbind(bage, csap=exemplo6)
```


Tabela 3 – Condições sensíveis à atenção primária (CSAP) por grupos de causa – resultado da aplicação da função csapAIH sobre o arquivo RDSP1301.DBF

Grupos de causa	N	% total	% CSAP
g01	680	0,3	2,2
g02	2.487	1,2	8,1
g03	201	0,1	0,7
g04	500	0,3	1,6
g05	469	0,2	1,5
g06	1.276	0,6	4,2
g07	638	0,3	2,1
g08	1.863	0,9	6,1
g09	1.531	0,8	5,0
g10	2.018	1,0	6,6
g11	3.755	1,9	12,3
g12	3.580	1,8	11,7
g13	1.912	1,0	6,2
g14	1.190	0,6	3,9
g15	4.334	2,2	14,1
g16	2.125	1,1	6,9
g17	424	0,2	1,4
g18	891	0,4	2,9
g19	766	0,4	2,5
Total CSAP	30.640	15,4	100,0
Não CSAP	168.723	84,6	–
Total (CSAP e não CSAP)	199.363	100,0	–

The screenshot shows the TabWin application window with the following content:

Arquivo Editar Operações Estatísticas Quadro Gráfico Ajuda

Título: |Movimento de AIH - Arquivos Reduzidos - Brasil - 2008 Subtítulo: |Frequência segundo Sensíveis At.Bas.3

Sensíveis At.Bas.3	Frequência
Total	30.640
1. Doenças preveníveis p/immuniz/condições sensív	680
2. Gastroenterites infecciosas e complicações	2.487
3. Anemia	201
4. Deficiências nutricionais	500
5. Infecções de ouvido, nariz e garganta	469
6. Pneumonias bacterianas	1.276
7. Asma	638
8. Doenças pulmonares	1.863
9. Hipertensão	1.531
10. Angina	2.018
11. Insuficiência cardíaca	3.755
12. Doenças cerebrovasculares	3.580
13. Diabetes melitus	1.912
14. Epilepsias	1.190
15. Infecção no rim e trato urinário	4.334
16. Infecção da pele e tecido subcutâneo	2.125
17. Doença Inflamatória órgãos pélvicos femininos	424
18. Úlcera gastrointestinal	891
19. Doenças relacionadas ao pré-natal e parto	766

Log

```

DEF=Z:\home\fulvio\tabwin\RD2008.DEF
PATH=Z:\home\fulvio\Dropbox\bdshsustia\BR2013\RD*.DBC
Linha=Sensíveis At.Bas.3
Incremento=Frequência
Suprime_Linhas_Zeradas=false
Suprime_Colunas_Zeradas=true
Não_Classificados=0
[Arquivos]
RDSP1301.DBC
Registros_Processados= 199363
Tempo_Decorrido= 0:13
Criado Z:\home\fulvio\Dropbox\tesel\PSF_RS\funcao-aih\CSAP\RDSP
  
```

Mostrar log ao abrir tab

Rodapé: | Arquivo: pantervars2 "CEP&NES" Linhas:19 Chave:0

Figura 1 – Resultado e log da tabulação do arquivo RDSP1301.DBC pelo aplicativo TabWin

```
head(exemplo7, 3)
  gidade   gsexo gabep5      gmodelo ghospit gcidint csap
1      2 masculino   D outro/sem USref.      HU   A040  sim
2      3 masculino   C           PSF Sta. Casa  A041  sim
3      2 feminino   D           PSF           HU   A044  sim
```

O mesmo procedimento pode ser aplicado quando o usuário quiser manter todas as variáveis do arquivo da AIH e apenas acrescentar a classificação CSAP. Por exemplo, o objeto `bd` foi criado anteriormente, com a leitura do arquivo da AIH para Santa Catarina em janeiro de 2015 (`RDSC1501.DBF`). Para acrescentar a classificação CSAP mantendo-se as demais variáveis da base, pode-se executar o seguinte comando:

```
bd <- cbind(bd, csapAIH(bd$DIAG_PRINC, arquivo = FALSE)[1:2])
```

A validade do algoritmo pode ser verificada comparando-se os resultados da função com os que é possível tabular pelo TabWin para um mesmo período. Escolheu-se um mês para São Paulo (arquivo `RDSP1301.dbc/dbf`) por este estado ser o mais populoso e portanto com maior probabilidade de se observar as causas mais raras de internação.

A Tabela 3 mostra os resultados da tabulação das internações por CSAP entre os residentes do estado de São Paulo no mês de competência de janeiro de 2013, realizada pela função `csapAIH` sobre o arquivo `RDSP1301.DBF`. A Figura 1 apresenta os resultados da mesma tabulação, embora realizada pelo TabWin, utilizando-se os arquivos de conversão do Datasus sobre o arquivo comprimido `RDSP1301.DBC`. O número de registros processados, o total de CSAP e a frequência observada em cada grupo são os mesmos:

```
exemplo8 <- csapAIH("RDSP1301.dbf", x.procobst = FALSE, longa = TRUE)
Importados 199.363 registros.
Exportados 199.363 registros.
Tempo decorrido: 22 segundos.
```

Considerações finais

O pacote e a função `csapAIH` apresentam-se como uma ferramenta adicional ao estudo das CSAP no Brasil. Eles foram desenvolvidos para suprir a falta de uma rotina computacional que (i) automatizasse o trabalho de classificação em nível individual da internação e (ii) facilitasse o manejo dos arquivos da AIH.

A função `read.dbf` não consegue ler arquivos `.DBF` com variáveis em que todos os registros são *missings*. Como a função `csapAIH` utiliza `read.dbf` para a leitura desses arquivos, ela sofre a mesma limitação. Nesses casos, antes da leitura no R, tais arquivos devem ser abertos em programa próprio e as variáveis excluídas. Destaca-se como principal limitação do pacote `csapAIH` a impossibilidade de ler os arquivos comprimidos da AIH (extensão `.DBC`), o que, além de evitar a necessidade de prévia expansão dos arquivos, reduz – em muito – o espaço ocupado em disco, já que o arquivo `DBF` resultante é muito maior que seu correspondente `.DBC`. Entretanto, como foi dito anteriormente, o projeto é público e colaboradores são bem-vindos.

A criação de um pacote no R para a classificação das CSAP e outras funcionalidades para o manejo dos arquivos da AIH oferece aos pesquisadores garantia de maior rapidez e validade nesse processo, e às Secretarias Estaduais e Municipais de Saúde, a possibilidade de gerar relatórios automatizados para o monitoramento das CSAP.

A função `csapAIH` automatiza o trabalho com os arquivos da AIH e permite o exercício com outras bases de dados ou uma lista de códigos CID-10 isolada. Sua utilização é simples e, para os usuários não habituados ao R, o resultado pode ser guardado em um arquivo de dados apto a leitura por outros pacotes estatísticos.

Agradecimentos

Agradeço ao Prof. Dr. Antônio Carlos Marasciullo, do Hospital Universitário da Universidade Federal de Santa Catarina (HU-UFSC), a atenta revisão e sugestões sobre a primeira versão do manuscrito.

Referências

1. Ansari Z. The concept and usefulness of ambulatory care sensitive conditions as indicators of quality and access to primary health care. *Aust J Prim Health*. 2007;13(3):91-110.
2. Nedel FB, Facchini LA, Bastos JL, Martín-Mateo M. Conceptual and methodological aspects in the study of hospitalizations for ambulatory care sensitive conditions. *Cienc Saude Coletiva*. 2011;16 supl 1:1145-54.
3. Brasil. Ministério da Saúde. Portaria No 221, de 17 de abril de 2008. *Diário Oficial da República Federativa do Brasil, Brasília (DF)*, 2008 abr 18;Seção 1:70.
4. Alfradique ME, Bonolo PE, Dourado I, Lima-Costa ME, Macinko J, Mendonça CS, et al. Internações por condições sensíveis à atenção primária: a construção da lista brasileira como ferramenta para medir o desempenho do sistema de saúde (Projeto ICSAP – Brasil). *Cad Saude Publica*. 2009 jan-jun;25(6):1337-49.
5. Ministério da Saúde (BR). Datasus. Apresentação TabWin [Internet]. Brasília: Ministério da Saúde; 2010 [citado 2016 jan 06]. Disponível em: <http://datasus.saude.gov.br/apresentacao>.
6. R Foundation for Statistical Computing. R: a language and environment for statistical computing [programa de computador]. Vienna: R Foundation for Statistical Computing; 2009 [citado 2016 jan 06]. Disponível em: <http://www.r-project.org>
7. Nedel FB, Facchini LA, Martín-Mateo M, Vieira LAS, Thumé E. Programa Saúde da Família e condições sensíveis à atenção primária, Bagé (RS). *Rev Saude Publica*. 2008 dez;42(6):1041-52.
8. Ministério da Saúde (BR). Secretaria de Atenção à Saúde. Departamento de Regulação, Avaliação e Controle. Coordenação Geral de Sistemas de Informação. Manual técnico operacional do Sistema de Informação Hospitalar do SUS. Brasília: Ministério da Saúde; 2015.
9. R Foundation for Statistical Computing. Foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ... [Internet]. [Vienna]: R Foundation for Statistical Computing; 2015 [citado 2016 jan 06]. Disponível em: <https://CRAN.R-project.org/package=foreign>
10. Brasil. Ministério da Saúde. Portaria nº 321, de 8 de fevereiro de 2007. Institui a Tabela de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais – OPM do Sistema Único de Saúde – SUS. *Diário Oficial da República Federativa do Brasil, Brasília (DF)*, 2007 fev 09;Seção 1:42.
11. R Foundation for Statistical Computing. Writing R Extensions [Internet]. [Vienna]: R Foundation for Statistical Computing; 2016 [citado 2016 jun 07]. Disponível em: <https://cran.r-project.org/doc/manuals/r-release/R-exts.html>
12. Nedel FB. Package ‘csapAIH’: classificar condições sensíveis à atenção primária [Internet]. [lugar desconhecido]: [editor desconhecido]; 2016 [citado 2016 jun 07]. Disponível em: <http://ufpr.dl.sourceforge.net/project/csapaih/csapAIH-manual.pdf>
13. Ministério da Saúde (BR). Datasus. Transferência de arquivos [Internet]. Brasília: Ministério da Saúde; 2016 [citado 2016 jan 06]. Disponível em: <http://datasus.saude.gov.br/informacoes-de-saude/servicos2/transferencia-de-arquivos>.
14. Nedel FB. Internações hospitalares evitáveis pela atenção primária: estudo do impacto do Programa Saúde da Família sobre as internações por Condições Sensíveis à Atenção Primária no Rio Grande do Sul, Brasil [tese] Pelotas: Universidade Federal de Pelotas; 2009.

Recebido em 21/04/2016
Aprovado em 23/08/2016