

Business Intelligence utilizando tecnologias Web para análise de fatores de risco na ocorrência de doença arterial coronariana

Business Intelligence using Web technology for analysis of risk factors for coronary artery disease

Eliana Rodrigues Espelho Diniz Morais¹
Sílvia Sidnéia da Silva²
Edilson Carlos Caritá¹

Descritores: Business Intelligence; Data Mart; Doença arterial coronariana; OLAP; Data webhouse

RESUMO

Objetivo: O objetivo deste trabalho é a implementação de um ambiente computacional utilizando tecnologias *Web* para analisar dados de saúde, através de ferramentas OLAP (*On-line Analytical Processing*), considerando os fatores de risco para a ocorrência de doença arterial coronariana. **Métodos:** A partir de uma base de dados relacional referente à pesquisa de pacientes com diagnósticos de cardiopatias foi elaborado um modelo multidimensional, gerado um cubo e demonstrado os resultados por meio de uma tabela dinâmica conectada a um servidor OLAP. As consultas foram disponibilizadas utilizando tecnologias *Web*. **Resultados:** Com os resultados obtidos através do cruzamento de informações é possível a identificação de indivíduos doentes ou com predisposição para desenvolver a doença arterial coronariana, e assim, aplicar programas preventivos. **Conclusão:** A implantação de um *Business Intelligence* em uma Instituição de saúde, de qualquer porte, possibilita a busca e interpretação de informações armazenadas para apoio às decisões referentes à vida de um ser humano, não somente de um negócio, como geralmente se pensa.

Descriptors: Business Intelligence; Data Mart; Coronary artery disease, OLAP; Data webhouse

ABSTRACT

Objective: The objective of this paper is to implement a computational environment using web technologies to analyze health-related data through Online Analytical Processing (OLAP) tools on risk factors for coronary artery disease. **Methods:** Based on a relational database of studies on patients diagnosed with heart diseases, a multidimensional model was developed, a cube was generated and the results were presented using a pivot table connected to an OLAP server. Searches were made available using web technologies. **Results:** The results obtained after information crossing allowed to identifying individuals with coronary artery disease or those who have a predisposition to develop it and proposing preventive programs. **Conclusion:** The implementation of Business Intelligence in health care settings of any size allows searching and interpreting stored information to support decisions concerning people's life and not only about business entities as often believed.

Autor Correspondente:
Edilson Carlos Caritá
e-mail: edcarita@uol.com.br

¹ Centro de Ciências Exatas, Naturais e Tecnológicas, Universidade de Ribeirão Preto (UNAERP). Ribeirão Preto (SP), Brasil.

² Curso de Graduação em Enfermagem, Universidade de Ribeirão Preto (UNAERP). Ribeirão Preto (SP), Brasil.

INTRODUÇÃO

As instituições de todas as áreas sabem da importância de suas decisões estratégicas. Na saúde, a utilização de um sistema de apoio à tomada de decisões pode ser de grande auxílio, permitindo a análise de dados através do cruzamento de informações, relacionando problemas de saúde com seus fatores determinantes e identificando os riscos para acometimento de doenças. Os indicadores deverão servir de elementos para a execução de ações de prevenção para melhorar a qualidade de vida dos pacientes.

Uma metodologia em evidência, atualmente, é o *Business Intelligence* (BI), entendido como um conjunto de tecnologias que permite o cruzamento de informações e suporte a análise dos indicadores de desempenho de um negócio. Dessa forma, as ferramentas de apoio a decisão que fazem inferências em um banco de dados histórico são também chamadas de ferramentas de BI⁽¹⁾.

Assim, a implantação de um sistema de apoio à decisão passa a ser um diferencial em uma Instituição de saúde, pois oferece condições para que os níveis gerenciais definam os processos com base em dados consistentes.

Segundo Santos⁽²⁾, sistema de apoio à decisão pode ser definido como um processo organizado e sistemático onde as organizações coletam, analisam e disseminam informações provenientes de fontes internas e externas que são importantes para as atividades do negócio e para os tomadores de decisão. Frente a situação epidemiológica como a de doenças isquêmicas, é de grande valia a implementação desse tipo de sistema.

A doença isquêmica cardíaca, também conhecida como doença arterial coronariana, tem papel de destaque dentre as doenças cardiovasculares, pois a ocorrência destas patologias vem aumentando significativamente com percentuais de 11,8% em 1930 a 45% em 1998, de todas as mortes⁽³⁻⁴⁾. Assim, a doença arterial coronariana, diante da alta incidência e prevalência, no ano de 1990, incluiu-se no grupo das doenças crônico-degenerativas⁽⁵⁾ e vem merecendo investigação visando detecção de suas causas, controle e prevenção. O desenvolvimento do conceito de “fatores de risco” e suas relações com a incidência de coronariopatias decorreram de estudos epidemiológicos prospectivos realizados nos Estados Unidos da América⁽⁶⁾.

No início da década de 60, do século passado, o Estudo de *Framingham* começou a estabelecer relação entre estilo de vida e doença arterial coronariana, demonstrando uma associação consistente entre as características observadas num dado momento, em indivíduos aparentemente saudáveis e a incidência subsequente de coronariopatia. As associações incluem um aumento de concentração de colesterol plasmático,

tabagismo, hipertensão arterial sistêmica, diabetes *mellitus*, obesidade, idade, sexo e a ocorrência de coronariopatia⁽⁶⁾.

Diante destas constatações, evidencia-se a necessidade de estudos para diagnóstico, prevenção e tratamento da doença arterial coronariana utilizando ferramentas computacionais.

O objetivo deste trabalho é a implementação de um ambiente computacional utilizando tecnologias *Web* para analisar dados de saúde, através de ferramentas OLAP (*On-line Analytical Processing*), considerando os fatores de risco para a ocorrência de doença arterial coronariana.

MATERIAIS E MÉTODOS

Para apresentar o conjunto de técnicas e ferramentas que compõe o *Business Intelligence* foi estruturado um *Data Mart*, acessado por ferramentas OLAP, com o objetivo de extrair informações utilizando tecnologias *Web*, referente à pesquisa de pacientes com diagnóstico de cardiopatias.

O desenvolvimento do trabalho se deu partindo de uma base de dados coletada de um estudo realizado no Ambulatório de Cardiologia do Hospital Electro Bonini que pertence à Universidade de Ribeirão Preto – UNAERP. Ela é composta por 103 pacientes do sexo feminino, na faixa etária de 20 a 90 anos, atendidas pelo Sistema Único de Saúde (SUS). As informações coletadas se referem a fatores de risco modificáveis, não modificáveis, valores da pressão arterial aferida na consulta de enfermagem e principais diagnósticos encontrados nos prontuários dos pacientes/usuários⁽⁷⁾.

O modelo relacional foi implementado de acordo com as informações coletadas na pesquisa, que estavam armazenadas em um arquivo do *Access 2000*. Com o uso da ferramenta ETL *Data Transformation Services Import/Export* (SQL Server) os dados de origem foram transformados para garantir a integridade e a limpeza ignorando conteúdos nulos, eliminando duplicidades e corrigindo violação de formato para seguir os padrões definidos no modelo relacional. Assim, a base de dados foi importada e armazenada no Sistema Gerenciador de Banco de Dados (SGBD) *Microsoft SQL Server 2000*.

O modelo entidade relacionamento é demonstrado na Figura 1 e foi criado através da ferramenta *Platinum ERwin*, onde estão relacionadas as tabelas que irão gerar a base de informações considerando o modelo multidimensional.

A estrutura multidimensional, que contém detalhes e dados de agregação em um único modelo lógico e unificado, com suporte para cálculos internos foi projetada, criada e gerenciada por meio da ferramenta *Analysis Services (MSSQL Server)*. Esta ferramenta inclui *wizards* para projeto e implementação dos cubos. O acesso à base de dados que originou o modelo

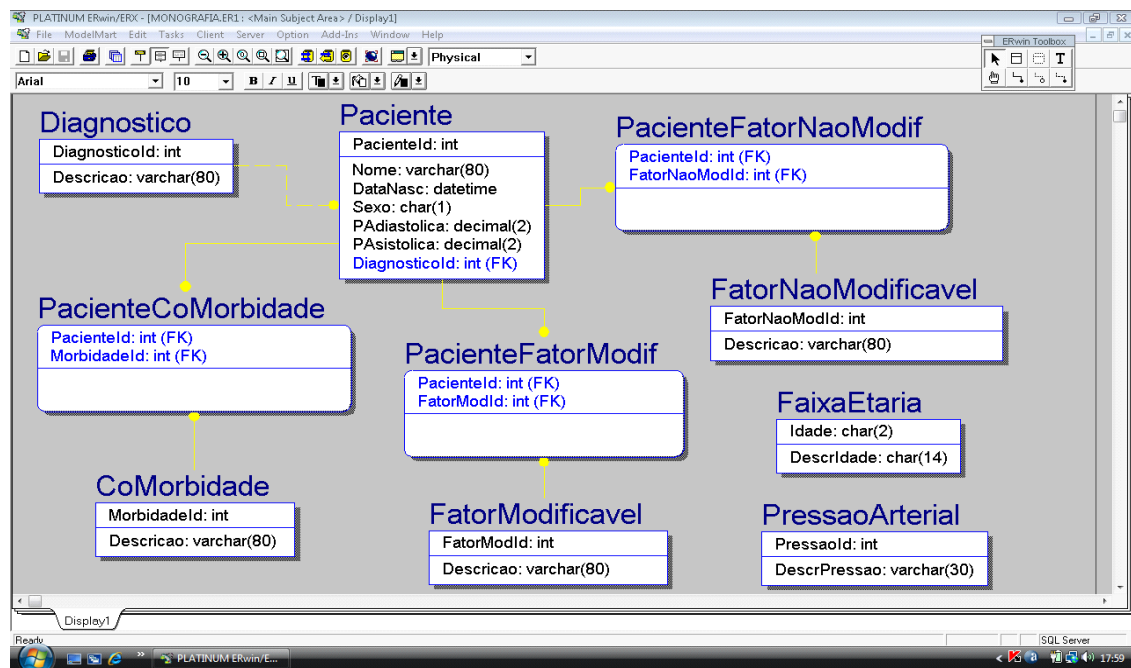


Figura 1 - Modelo Relacional

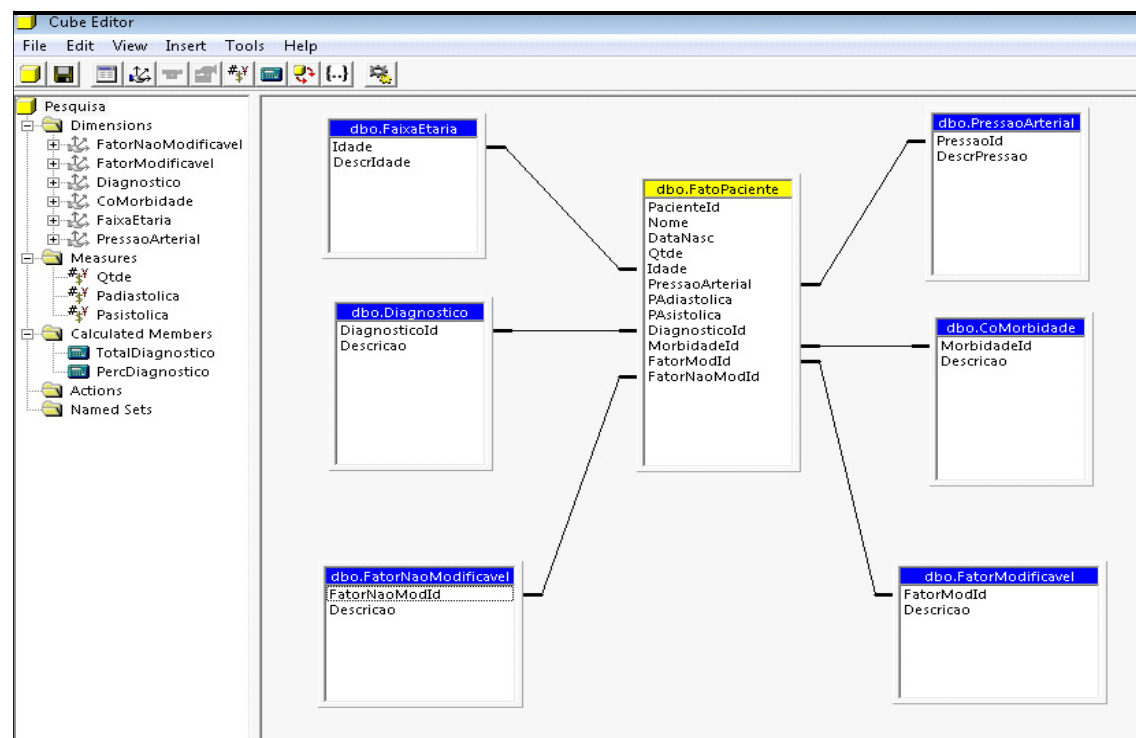


Figura 2 - Modelo Multidimensional

multidimensional foi feito via conexão ODBC (*Open Data Base Connectivity*), disponível para fontes de dados de diversos formatos, utilizando o provedor *Microsoft OLE DB Provider for SQL Server*.

O *Analysis Services* possui um editor para criar o relacionamento das tabelas dimensão com a tabela fato. A tabela fato foi preparada a partir da extração de dados da tabela Paciente do banco de dados relacional e dos atributos que representam as chaves de acesso das tabelas de dimensão.

O esquema de estrutura usado neste projeto foi o modelo estrela que define os elementos de composição do cubo (Figura 2). Este modelo possui seis dimensões que representam os filtros para as consultas e, de acordo com as seleções realizadas, forma-se um conjunto de dados que atendem aos valores escolhidos pelo usuário.

O processamento do cubo baseado em estrutura multidimensional armazena as informações calculando previamente todas as combinações de todos os níveis

de dimensão e agregação, gerando um esquema otimizado para que as consultas realizadas sejam respondidas rapidamente. Foi escolhida a forma de armazenamento MOLAP (OLAP Multidimensional – onde os dados detalhados e agregações são armazenados em formato de cubo do *Analysis Server*). Considerando que os dados armazenados estão em uma única base de dados relacional normalizada não foi necessária a transformação e a limpeza dos dados, pois não havia dados nulos e/ou inválidos.

Ressalta-se que não foram utilizados métodos de agrupamento ou associação, pois as opções de uso de *drill-down* ou *drill-up* ficam a critério do usuário através da interface dinâmica.

Para disponibilizar a visualização dos dados do cubo através de um navegador foi incluído o componente de tabela dinâmica e gráfico no editor de *Websites*, estabelecendo uma conexão com a fonte de dados multidimensional, não requerendo para isso, componentes do *Analysis Services* instalado na máquina do cliente.

A tabela dinâmica, sendo uma ferramenta gráfica, permite que os usuários explorem facilmente as dimensões e as medidas do cubo, obtendo os resultados através do cruzamento de informações. Várias operações são possíveis, como aumentar ou diminuir o nível de detalhe, agrupar as informações, além de alterar linhas e colunas modificando as visões.

O sistema *Web* foi estruturado em páginas HTML (*HyperText Markup Language*) geradas pelo editor de

Websites, para efetuar análises OLAP sobre o cubo multidimensional implementado no *Analysis Services*. O servidor *Web* foi configurado com o *software Apache 2.2.6* e o módulo PHP 5.2.5 (*Hypertext Preprocessor*) que comunica com o Sistema Gerenciador de Banco de Dados.

A atualização do banco de dados multidimensional e o processamento do cubo foram preparados utilizando o DTS (*Data Transformation Services - MSSQL Server 2000*), que fornece um conjunto de ferramentas gráficas e objetos programáveis para a extração, transformação, processamento e consolidação dos dados provenientes da base de dados relacional. Os *packages* criados no DTS foram programados para executar periodicamente estas tarefas.

RESULTADOS E DISCUSSÃO

Na Figura 3 é apresentada uma consulta para demonstrar os principais diagnósticos encontrados entre os pacientes. Observa-se a predominância da hipertensão arterial sistêmica (HAS), representando 56,3% do total de 103 diagnósticos encontrados.

Neste exemplo foi incluída a dimensão faixa etária após a dimensão diagnóstico, possibilitando o cruzamento de informações e assim aumentando o nível de detalhes referente à distribuição de mulheres hipertensas, de acordo com a idade descrita na tabela. Com esta opção de consulta, verifica-se na amostra, que a HAS vem se instalando no sexo feminino,

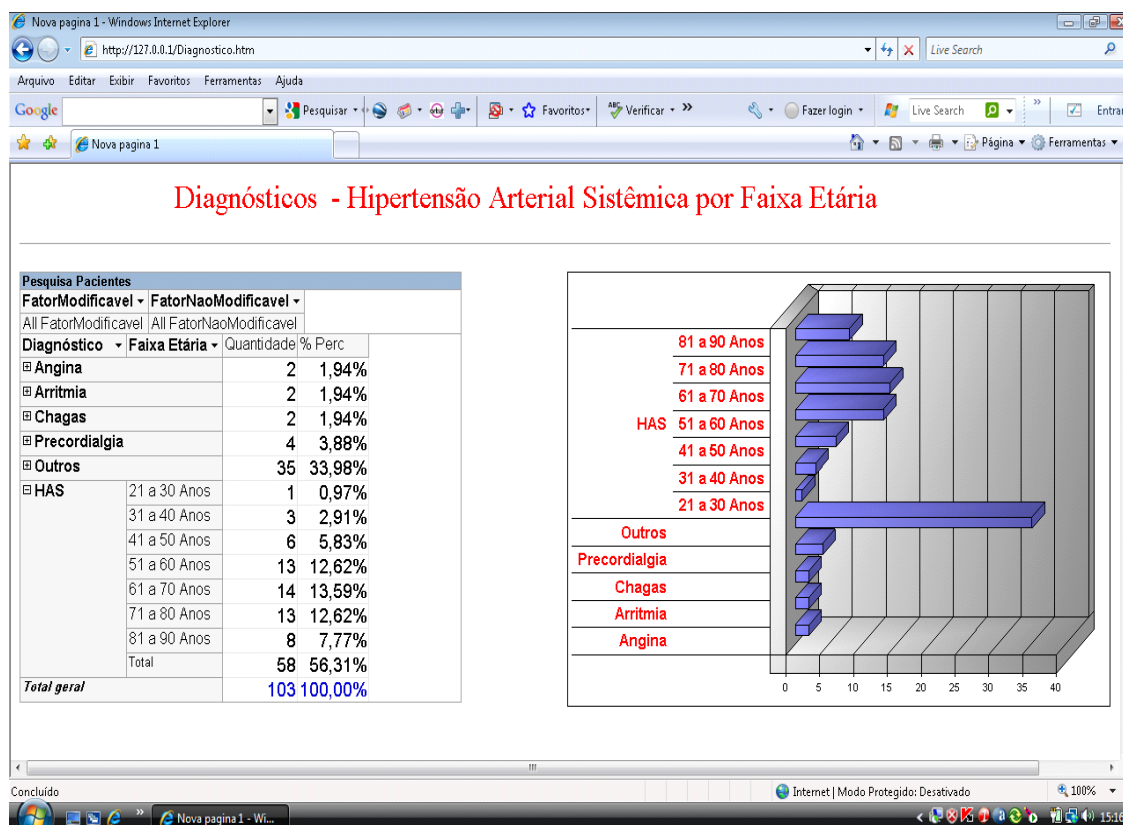


Figura 3 – Tela de Consulta de Diagnósticos por Faixa Etária

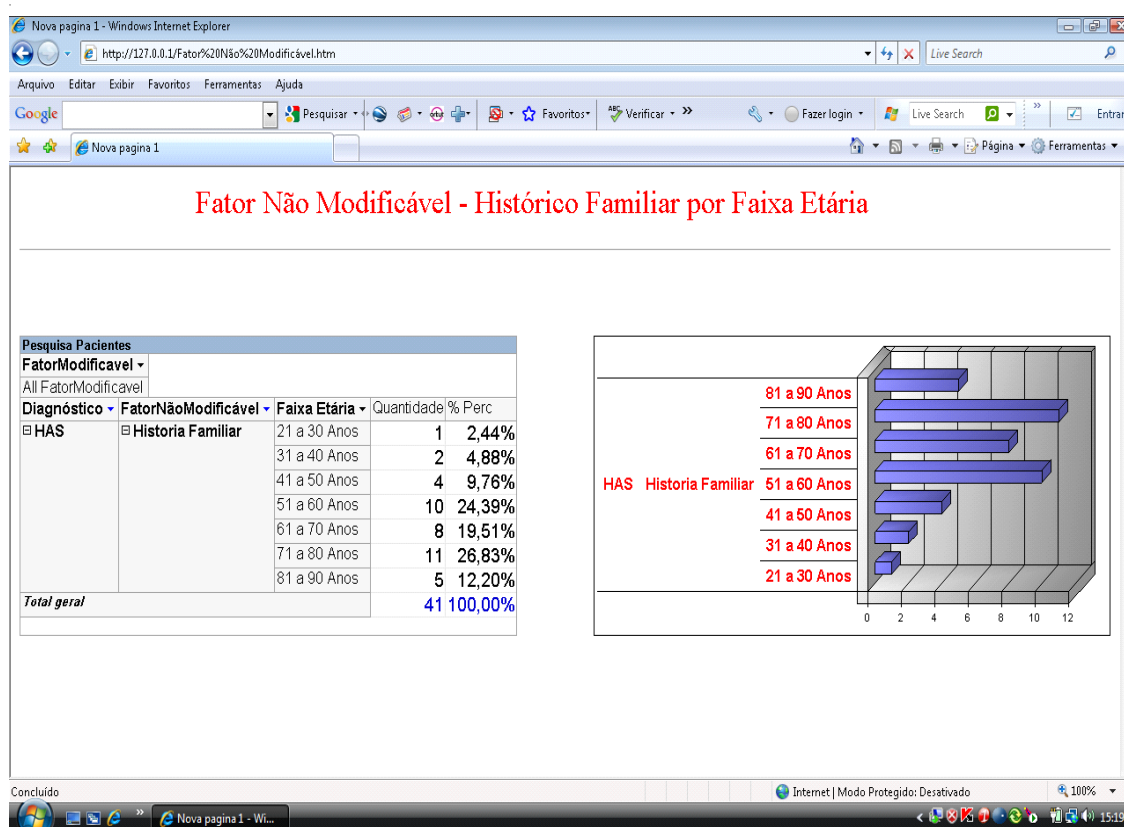


Figura 4 – Tela de Consulta de HAS e História Familiar por Faixa Etária

precocemente, já na faixa etária de 21 a 30 anos. No entanto, o acometimento da patologia, com maior predomínio, existe nas faixas etárias de 51 a 60, de 61 a 70 anos e de 71 a 80 anos, ressaltando que a frequência se mantém elevada até 85 anos, quando ainda existe percentual significativo.

Os estudos demonstram que o risco cardiovascular na idade reprodutiva da mulher é menor⁽⁸⁾, devido ao fator hormonal protetor, porém de acordo com os achados supracitados, constata-se que a mulher contemporânea expõe-se a fatores de risco decorrentes de seu estilo de vida, tornando-se indivíduo-alvo para desenvolvimento precoce de doenças cardiovasculares⁽⁹⁾.

No exemplo mostrado na Figura 4, foi filtrado na dimensão diagnóstico o item HAS e, em seguida, na dimensão fator não modificável o item história familiar e demonstrado, em detalhes, o resultado por faixa etária.

A partir das seleções realizadas forma-se um conjunto de dados que mostra o predomínio da citação da história familiar dentre os fatores de risco para a HAS, em todas as faixas etárias, pelas mulheres do estudo.

A história familiar é considerada fator de risco para aquele indivíduo que possui família onde o pai teve doença cardíaca antes dos 60 anos e a mãe antes dos 65, pois sabe-se que herda genes dos pais⁽¹⁰⁾.

Ainda identificando os fatores de risco, na Figura 5, a análise foi feita combinando a dimensão diagnóstico, filtrando apenas o item HAS com a dimensão fator

modificável e expandindo informações sobre a faixa etária dos pacientes que possuem o tabagismo como fator de risco modificável. Nessa opção de consulta é possível verificar que a indicação desse fator pelas mulheres ocorre, predominantemente, naquelas idades pertencentes à faixa etária de 51 a 60 anos.

Estudos recentes destacam que no período pós-menopausa existe maior risco de o tabagismo levar à ocorrência de doenças cardiovasculares em mulheres⁽¹¹⁾ e, essa informação é trabalhada e discutida como dado epidemiológico importante e prevenível nos programas de educação em saúde com esse gênero, ressaltando-se que este fator apresenta-se associado a 50% dos eventos coronários em mulheres, no gera⁽¹²⁾.

Os valores que permitem classificar os indivíduos adultos acima de 18 anos, de acordo com os níveis de pressão arterial, estão na Tabela 1. Esta classificação utilizada pela Sociedade Brasileira de Cardiologia, Sociedade Brasileira de Hipertensão Arterial e Sociedade Brasileira de Nefrologia é preconizada na V Diretrizes Brasileiras de Hipertensão Arterial⁽¹³⁾.

Analisando os valores de pressão arterial, conforme a classificação descrita na Tabela 1, constata-se que 58,6% das pacientes submetidas à aferição de pressão arterial, se encontram na faixa de hipertensão estágio 1.

De acordo com as informações apresentadas na Figura 6 é possível concluir a necessidade de um domínio mais efetivo das ações que conduzirão a um controle da HAS, pois evidencia-se que essa doença

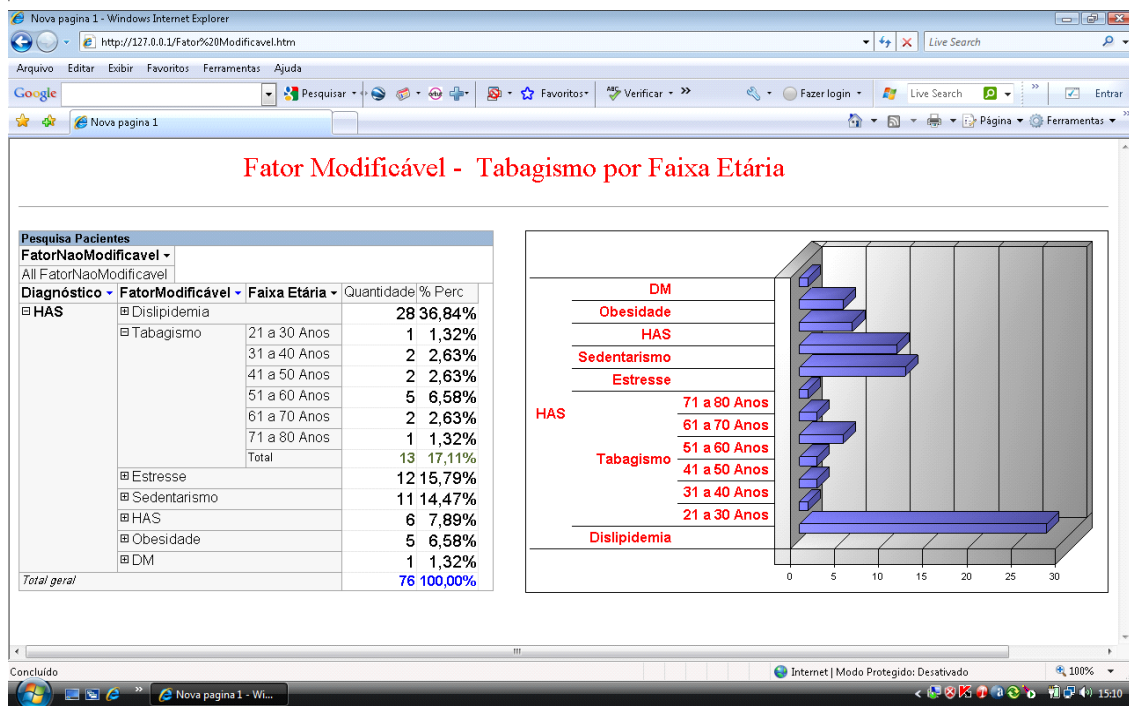


Figura 5 – Tela de Consulta de HAS e Tabagismo por Faixa Etária

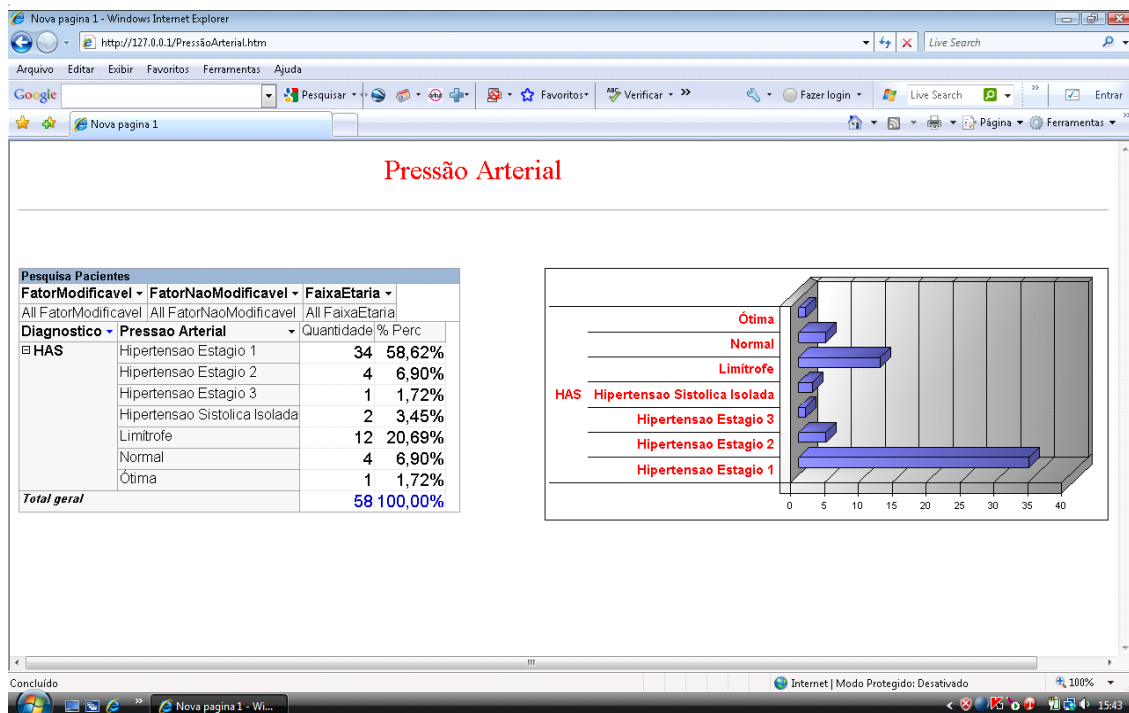


Figura 6 – Tela de Consulta de Pressão Arterial (classificação adotada pela V Diretrizes Brasileiras de Hipertensão Arterial, 2006).

tem elevada incidência e prevalência no mundo⁽¹⁴⁾, bem como nessa amostra.

Com estes tipos de análises é possível relacionar problemas de saúde com seus fatores determinantes, dentro de cada contexto, de acordo com exemplos criados como referência para o entendimento da estrutura preparada para extrair informações.

O resultado alcançado com o uso de tecnologias de BI depende da estratégia que é adotada, pois através

da qualidade da informação que circula, isto é, da qualidade do dado, do sistema de informação e do ambiente computacional existentes, pode-se optar pelo caminho a ser seguido, pautados em decisões consistentes e coerentes.

A ferramenta implementada permite visualizar as informações através de uma grande variedade de opções de consultas e gráficos de forma dinâmica, rápida e flexível.

Tabela 1 – Classificação dos valores de pressão arterial verificados em consultório em indivíduos maiores de 18 anos. V Diretrizes Brasileiras de Hipertensão Arterial, 2006.

Classificação	Pressão sistólica (mmHg)	Pressão diastólica (mmHg)
Otíma	< 120	< 80
Normal	< 130	< 85
Limítrofe	130-139	85-89
Hipertensão Estágio 1	140-159	90-99
Hipertensão Estágio 2	160-179	100-109
Hipertensão Estágio 3	≥ 180	≥ 110
Hipertensão Sistólica Isolada	> 140	< 90

CONCLUSÃO

A implantação da tecnologia de *Business Intelligence* em um serviço de saúde, de qualquer porte, possibilita a busca e interpretação de informações armazenadas para apoio às decisões a respeito da vida de um ser humano, e não somente de um negócio.

Com um sistema de BI os usuários conseguem maior independência e podem realizar as consultas de forma intuitiva e flexível. Independente da origem dos dados ou da forma de extração, transformação e agregação, o mais importante é que a informação deve

“servir” aos usuários finais em uma linguagem que seja compreensível para eles, sem a necessidade de intérpretes.

A análise dos resultados indica os fatores de risco para a doença arterial coronariana e com estas informações é possível a identificação de indivíduos doentes ou com predisposição para desenvolver a doença; aspectos que podem subsidiar programas para a prevenção e controle da patologia.

Conclui-se que os profissionais da saúde podem efetuar consultas e relacionamentos entre os dados de maneira fácil sem a necessidade de um profissional da área de informática.

REFERÊNCIAS

- Colaco M Jr. *Projetando sistemas de apoio a decisão baseados em Data Warehouse*. Rio de Janeiro: Axcel Books do Brasil; 2004.
- Santos RS. *Ambiente para extração de informações através da mineração das bases de dados do Sistema Único de Saúde [Tese]*. São Paulo: Universidade Federal de São Paulo - Escola Paulista de Medicina; 2007.
- Perito MS, Menegali L, Sakae TM. Perfil do atendimento ao paciente internado com síndrome isquêmica aguda sem supradesnível do segmento ST no Hospital Geral Nossa Senhora da Conceição, Tubarão-SC, 2006-2007. *ACM arq. catarin. med.* 2009; 38(2):80-9.
- Timerman RS, Souza MFM, Serrano CV. Aspectos epidemiológicos das doenças cardiovasculares em nosso meio: tendência da mortalidade por doença isquêmica do coração no Brasil de 1979 a 1996. *Rev. Soc. Cardiol. Estado São Paulo.* 2001; 11(4):715-23.
- Pereira JC, Barreto SM, Passos VMA. Perfil de risco cardiovascular e auto-avaliação da saúde no Brasil: estudo de base populacional. *Rev. panam. salud pública = Pan am. j. public health.* 2009 jun; 25(6):491-8.
- Lessa I. *O adulto brasileiro e as doenças da modernidade: epidemiologia das doenças crônicas não transmissíveis*. São Paulo: Hucitec; 1998.
- Braunwald E, Zipes DP, Bonow RO. *O tratado de doenças cardiovasculares*. 7a ed. São Paulo: Elsevier; 2007.
- Castro MENR. *Cardiopatias em mulheres: perfil de clientes atendidas em um hospital escola do interior paulista [Monografia]*. Ribeirão Preto (SP): Universidade de Ribeirão Preto – Curso de Enfermagem; 2008.
- Mínime MH, Gebara OCE, Santos RD. Estratificação de risco de doença coronária em mulheres. *Rev Soc Cardiol Estado de São Paulo.* 2007; 17(4):299-303.
- Castro AP, Neumann AS, Schneider CE, Souza K, Bonaroski LF, Ribas LF, Maluf EMCP. Perfil do risco de doenças cardiovasculares em acadêmicos e médicos do Hospital de Clínicas da Universidade do Paraná. *RBM rev. bras. med.* 2004; 61(6):345-50.
- Konno FY, Santos F, Pinho FM, Massabki PS. Tabagismo na mulher. *Rev Soc Bras Clín Méd.* 2008; 6(1):26-33.
- Vale AAL, Martínez TLR. Prevenção: importância e estratificação de risco. *Atheros.* 2001;12(1):7-12.
- Feldman DN et al. *Doença arterial coronária em mulher: epidemiologia, metodologia, diagnósticos e manejo na síndrome coronária aguda. [Monografia]* São Paulo: Dante Pazzanese; 2008.
- Sociedade Brasileira de Hipertensão; Sociedade Brasileira de Cardiologia; Sociedade Brasileira de Nefrologia. *V Diretrizes Brasileiras de Hipertensão Arterial*. São Paulo; 2006.
- Garcia SMS, Galvão MTG. Aspectos da sexualidade entre portadores de hipertensão arterial. *Rev Nursing.* 2006; 101(9):1045-9.