

Acurácia das técnicas de relacionamento probabilístico e determinístico: o caso da tuberculose.

Gisele P. Oliveira^I, Ana L.S. Bierrenbach^{II}, Kenneth R. C. Júnior^{III}, Cláudia M. Coeli^{IV}, Rejane S. Pinheiro^{IV}

^I Programa de Pós-Graduação em Saúde Coletiva. Instituto de Estudos em Saúde Coletiva.
Universidade Federal do Rio de Janeiro. Rio de Janeiro, RJ, Brasil

^{II} Instituto de Ensino e Pesquisa. Hospital Sírio Libanês. São Paulo, SP, Brasil

^{III} Instituto de Medicina Social. Universidade do Estado do Rio de Janeiro. Rio de Janeiro, RJ,
Brasil

^{IV} Instituto de Estudos em Saúde Coletiva. Universidade Federal do Rio de Janeiro. Rio de Janeiro,
RJ, Brasil

Os relacionamentos probabilístico e determinístico de registros são técnicas que possibilitam a integração e manejo de bases de dados distintas. Um dos principais problemas que compromete a qualidade da base de dados das notificações de tuberculose (TB) é a presença de registros duplicados. Esse trabalho objetiva analisar a sensibilidade e especificidade das abordagens determinística e probabilística para identificação de registros duplicados de TB, assim como as características dos pares discordantes. Foram analisados todos os registros de TB no período de 2009 a 2011 do estado do Rio de Janeiro. As duas abordagens apresentaram alta concordância quanto à classificação (96%), embora a determinística tenha recuperado 8,8% (n=758) a mais de pares. A presença de valores faltantes para data de nascimento e o baixo percentual da medida de similaridade para o nome do caso foram os principais responsáveis pela não identificação dos registros do mesmo indivíduo pela abordagem probabilística. Para a determinística, a presença de valor faltante foi para o nome da mãe e o menor percentual da medida de similaridade foi para a data de nascimento. As duas técnicas apresentam alta concordância para a classificação como par. Apesar de a técnica determinística ter identificado mais registros duplicados que a probabilística, a segunda recuperou registros não identificados pela primeira. A necessidade e a experiência do usuário devem ser considerados para a escolha da técnica a ser utilizada.

Palavras – Chave: Tuberculose, Sistemas de Informação, Vigilância epidemiológica.

Apoio: Secretaria de Vigilância em Saúde do Ministério da Saúde.