

MÉTODOS ESTADÍSTICOS DESARROLLADOS PARA LA ESTIMACIÓN EN LA PREVALENCIA Y PARÁMETROS DE DESEMPEÑO DE TESTS PARA DIAGNÓSTICO CLÍNICO: UNA REVISIÓN DE LITERATURA

José Rafael Tovar Cuevas*

Resumen

Objetivo: hacer una revisión de la bibliografía publicada en revistas especializadas en metodología estadística, acerca de los procedimientos desarrollados para la estimación de la prevalencia y los parámetros de desempeño de pruebas para diagnóstico clínico.

Metodología: se revisaron 52 artículos sobre procedimientos desarrollados para estimar la prevalencia y los parámetros de desempeño de pruebas para diagnóstico clínico, encontrados en revistas especializadas en la publicación de metodologías estadísticas nuevas, para la solución de problemas presentes en el ámbito de la salud.

Resultados: se identificaron tres problemas de interés asociados a la estimación de la prevalencia y a los parámetros de desempeño (sensibilidad y especificidad) en los estudios de evaluación de pruebas para diagnóstico clínico. La presencia de un número de parámetros estimables mayor que la cantidad de datos disponibles para realizar la estimación (falta de identificabilidad), la presencia de individuos con resultado negativo en la prueba (o pruebas) para tamizaje cuyo verdadero estado de salud es desconocido (sesgo de verificación) y la presencia de dependencia entre los resultados de las pruebas en estudios que consideran más de una prueba de tamizaje.

Conclusiones: se encontró una amplia gama de aproximaciones a la solución al problema de presencia de sesgo de verificación, entre las que se incluyen el uso de variables latentes (discretas y continuas) y diferentes formas de reparametrización, además de alternativas para el manejo de la falta de identificabilidad y de independencia entre los resultados de las pruebas. Muchas de las alternativas propuestas se caracterizan por su alto nivel de complejidad y difícil aplicabilidad práctica.

Palabras clave: Prevalencia; Estimación; Sensibilidad; Especificidad; Métodos Estadísticos; Dependencia.

* Escuela de Medicina y Ciencias de la Salud, Universidad del Rosario, Bogotá

STATISTICAL METHODS DEVELOPED IN ORDER TO ESTIMATE THE PREVALENCE AND DEVELOPMENTAL PARAMETERS OF CLINICAL DIAGNOSTIC TESTS: A LITERATE REVIEW

Abstract

Introduction: to do a review of the bibliography published in statistical methodology specialized journals about the procedures developed to estimate the prevalence and performance of clinical diagnostic tests.

Methods: this review includes 52 articles about procedures developed to estimate the prevalence and performance of clinical diagnostic test parameters found in journals specialized in statistical methodology applied to health areas.

Results: three important statistical problems were found: the presence of a number of parameters higher than the pieces of information available to do the estimation procedure, known as lack of the ability to identify, the presence of individuals with negative outcome in the screening tests for comparison, whose actual health conditions are unknown (verification bias) and the dependence between test results.

Conclusions: a wide range of possibilities to address the three problems that characterized the estimation problem was found. To address the verification bias problem many alternatives consider the use of the latent variables and others use different parameters. Many strategies to address the lack of the ability to identify and the dependence between test outcomes were found too. Most of the proposed alternatives are characterized by their complexity and difficult practical applicability.

Keywords: Prevalence, Estimation, Sensibility, Specificity, Statistical Methods, Dependence.

MÉTODOS ESTADÍSTICOS DESENVOLVIDOS PARA ESTIMAR A PREVALÊNCIA E OS PARÂMETROS DE DESEMPENHO DE TESTES PARA DIAGNÓSTICO CLÍNICO; UMA REVISÃO DE LITERATURA

Resumo

Objetivo: realizar uma revisão da bibliografia publicada em revistas especializadas na metodologia estatística sobre os procedimentos desenvolvidos para a estimação da prevalência e os parâmetros de desempenho de testes para diagnóstico clínico.

Materiais e Métodos: revisaram-se 52 artigos sobre procedimentos desenvolvidos para estimar a prevalência e os parâmetros de desempenho de testes para diagnóstico clínico, publicados em revistas especializadas na publicação de metodologias estatísticas novas para a solução de problemas presentes no âmbito da saúde.

Resultados: identificaram-se três problemas de interesse associados à estimação da prevalência e os parâmetros de desempenho (sensibilidade e especificidade) nos estudos de avaliação de testes para diagnóstico clínico. A presença de um número de parâmetros a serem estimados, maior que a quantidade de dados disponíveis para realizar a estimação (falta de identificabilidade), a presença de indivíduos com resultado negativo no teste (ou testes) para triagem cujo verdadeiro estado de saúde é desconhecido (viés de verificação) e a presença de dependência entre os resultados dos testes em estudos que consideram mais de um teste de triagem.

Conclusões: encontrou-se uma ampla gama de aproximações para contornar o problema do viés de verificação, que incluem o uso de variáveis latentes (discretas e contínuas) e diferentes formas de reparametrizações além de alternativas para contornar os problemas da falta de identificabilidade e da falta de independência entre testes. Muitas das propostas encontradas apresentam alta complexidade para a sua execução prática.

Palavras chave: Prevalência; Estimativa; Sensibilidade; Especificidade; Métodos Estatísticos; Dependência.

Introdução

O desenvolvimento de procedimentos clínicos que permitem, com um alto nível de confiança, identificar indivíduos sob efeito de um processo de enfermidade ou infecção é um dos grandes desafios presentes na prática clínica e na saúde pública em geral. Um método diagnóstico pode estar composto por um ou mais procedimentos chamados testes diagnósticos, os quais identificam as alterações sofridas pelo organismo humano quando se encontra doente ou em estado de infecção.

O resultado de um teste diagnóstico é a expressão de uma variável biológica cujo comportamento se altera pela presença de um processo de doença ou infecção e que pode ser ou não diretamente mensurável dentro de uma escala de valores contínua ou discreta. Esse resultado pode ou não ser um reflexo fiel das alterações na saúde da pessoa dependendo de múltiplos fatores associados às características próprias da variável medida e à forma de obtenção das observações, o que tornou necessário o desenvolvimento de indicadores quantitativos que permitam descrever a capacidade de correta classificação de indivíduos presente no teste para diagnóstico clínico.

Materiais e métodos

Realizou-se uma busca em revistas especializadas na publicação de artigos cujo objetivo é a divulgação de novas metodologias para a análise estatística de dados. A busca esteve concentrada em artigos que apresentam novas estratégias de análise para dados obtidos com planejamentos de avaliação de testes para diagnóstico clínico. O diagnóstico

clínico e um tema de interesse tanto para pesquisadores das áreas médicas como estatísticas, nesta revisão se foram selecionadas somente as publicações produto de pesquisas realizadas desde a perspectiva estatística.

Resultados

Estimação da prevalência e parâmetros dos testes para diagnóstico clínico

O problema da estimação dos parâmetros de desempenho de testes diagnósticos tem sido amplamente estudado desde a primeira metade do século passado. É possível identificar três fases no desenvolvimento da pesquisa sobre o tema. Na primeira fase, os pesquisadores concentraram-se no estudo da estimação dos parâmetros de desempenho assumindo planejamentos para coleta de dados que consideravam um teste sob avaliação e um padrão-ouro perfeito ou imperfeito.

Na segunda fase de desenvolvimento estudaram-se planejamentos com dois ou mais testes nos quais devido a restrições clínicas ou éticas, não é possível a aplicação dos testes de triagem no conjunto completo de indivíduos participantes no estudo (viés de verificação) e nos que se assume independência condicional entre os resultados dos testes sob avaliação.

Na última fase, o trabalho concentrou-se na estimação dos parâmetros, assumindo dependência entre os resultados dos testes; situação que pode acontecer ou não em presença de viés de verificação. Em cada uma das fases definidas, tem-se encontrado uma ampla variedade de propostas para a solução do problema de estimação, considerando os dois paradigmas da estatística; o frequentista e o Bayesiano.

Todos os estudos revisados tratam de alguma forma o viés de verificação, assunto que pode ser considerado o problema mais estudado na literatura sobre diagnóstico clínico. O viés de verificação pode se apresentar de diferentes formas dependendo do planejamento estabelecido para o estudo de avaliação de testes. Tem-se viés de verificação quando é preciso avaliar o desempenho de um teste novo e não se tem um padrão-ouro para comparação ou quando só uma parte dos indivíduos avaliados com o teste novo é submetida à verificação por padrão-ouro. Uma das formas mais comuns de viés de verificação é a que acontece quando se tem estudos com dois testes de triagem e só se verificam com padrão-ouro os indivíduos com resultado positivo em algum dos dois testes sob avaliação, de modo que aqueles com ambos os resultados negativos não são verificados.

Estudos que usam métodos frequentistas de estimação

O primeiro estudo encontrado foi realizado por Neyman (1) em 1947. Ele realizou uma avaliação dos procedimentos estatísticos utilizados para estimar os parâmetros de testes quando se tem três níveis de presença do evento de interesse (doença, infecção) na população. Anos depois, Homburguer e colaboradores (2) consideraram a avaliação do desempenho de testes para diagnosticar diferentes tipos de câncer mudando em cada caso o plano de coleta de dados. Nesse mesmo ano, (3) desenvolveu um índice para estimar as proporções de indivíduos bem classificados, quando se tem apenas um teste diagnóstico e populações de doentes e não doentes bem definidas. Mantel (4) estudou a eficiência de diferentes técnicas diagnósticas aplicadas aos casos onde se tem igual número de procedimentos para

cada indivíduo e quando se tem número variável de procedimentos. No ano 1966, (4) aprimoram o índice desenvolvido por (3) e desenvolvem métodos para estimar os parâmetros de interesse num teste novo quando não se tem um padrão-ouro. Os autores usam o conceito de co-positividade como a probabilidade de ter um resultado positivo no teste sob avaliação dado que o teste de referência tem resultado positivo e a co-negatividade como a probabilidade de ter um resultado negativo com o teste novo condicionado num resultado negativo do teste de referência. Rogan and Gladen (5) fazem um estudo sobre as características que deve ter um bom teste diagnóstico e desenvolvem um método estatístico para estimar a prevalência da doença, ajustando a estimativa com as características do teste aplicado numa amostra de indivíduos. Hui e Walter (6) abordam pela primeira vez o problema da estimação dos parâmetros de testes em planejamentos que consideram o uso de dois testes de triagem. Os autores desenvolvem uma metodologia estatística para fazer as estimações dos parâmetros usando máxima verossimilhança e assumindo que os resultados dos testes são independentes condicionalmente ao verdadeiro estado de saúde e propõem o uso de duas ou mais populações com diferentes prevalências da doença, para contornar o problema de falta de identificabilidade no modelo.

O planejamento de (6) é retomado por Thibodeau (7) para desenvolver um método de estimação eliminando a suposição de independência condicional no verdadeiro estado do indivíduo. Numa revisão da literatura com os estudos sobre testes diagnósticos realizados até essa data, Walter e Irwing (8) estudam o problema de se observar um número de parâmetros maior ao número de dados

contidos na tabela cruzada desenvolvida para obter resultados num estudo de avaliação de testes para diagnóstico clínico, o qual, é conhecido como a falta de identificabilidade do modelo estatístico para estimação. Os autores concluem que o problema apenas está presente quando o estudo inclui um ou dois testes para serem avaliados.

Com três testes o número de parâmetros e de graus de liberdade é o mesmo e com quatro ou mais testes os graus de liberdade (número de celas com dados na tabela) superam o número de parâmetros a serem estimados. Nesse artigo os autores tratam de maneira tangencial o problema de testes dependentes e a situação na qual se tem testes com resposta contínua. Hui e Zhou (9) fazem outra revisão sobre o tema, mas dessa vez focando-se nos métodos estatísticos usados para estimar o desempenho de um ou mais testes diagnósticos quando se tem um padrão-ouro perfeito ou imperfeito na presença de viés de verificação.

Schatzkin e colaboradores (10) desenvolveram procedimentos estatísticos para estimar os parâmetros de desempenho em planejamentos com verificação unicamente para aqueles indivíduos com resultado positivo em ambos os testes de triagem. Os autores definem dois índices que permitem comparar a eficiência relativa dos testes avaliados através de quociente de sensibilidades e das taxas de falsos positivos. Cheng e Macaluso (11) aprimoram o estudo de (10) usando planejamentos onde indivíduos com ao menos um dos testes de triagem positivo são verificados pelo padrão-ouro. Sullivan e Alonzo (12) propõem um modelo de regressão logística para estimar as taxas de falsos positivos e falsos negativos quando os dados são

obtidos com verificação somente para indivíduos com pelo menos um resultado positivo nos testes de triagem.

Um método bastante comum para contornar o problema de viés de verificação, ao obter as estimativas dos parâmetros de desempenho dos testes e a prevalência, consiste em assumir que o verdadeiro estado de enfermidade (infecção) é uma variável latente cujo comportamento pode ser ajustado com uma distribuição de probabilidade, de modo que no modelo estatístico usado para a estimação de parâmetros o verdadeiro estado de saúde aparece como um fator aleatório. Outros autores usam esta mesma abordagem mas assumem que a variável latente está relacionada ao fato de ser ou não verificado.

Walter (13) retoma o modelo de (11) e obtém um procedimento que permite estimar todos os parâmetros de interesse (sensibilidades, especificidades e prevalência), usando um modelo que assume a quantidade de sujeitos não verificados como uma variável binomial não observável (latente) e independência entre os testes. As estimativas são obtidas usando métodos de aproximação numérica.

Goetghebeur e colaboradores (14) propõem um modelo com classes latentes quando não existe um padrão-ouro e tem-se muitos testes pouco confiáveis que poderiam ser usados como tal. Van der Merwe e Maritz (15) consideram o modelo de (13) e acrescentam um parâmetro de correlação entre os testes para diagnóstico. Garret e colaboradores (16) utilizam a técnica de classes latentes e comparam seus resultados com aqueles obtidos a través de algoritmos comumente utilizados em psiquiatria, como o cluster

análise e a análise fatorial. Bertnatsky e colaboradores (17) propõem um modelo assumindo o verdadeiro estado de saúde como variável latente para ser usado com três ou mais testes diagnósticos. Outros autores como Begg (18), Baker (19), Qu e colaboradores (20), Yang e Becker (21), Zhou (22) e Albert e colaboradores (23,24), também utilizaram essa técnica para estabelecer o modelo estatístico e obter as estimativas dos parâmetros de desempenho tomando o viés de verificação como uma variável latente do modelo.

Em vista do grande uso de modelos com classes latentes na abordagem estatística do diagnóstico clínico, Rindskopf e Rindskopf (25) e Rindskopf (26) fazem uma revisão crítica sobre o uso desses métodos, e avaliam suas deficiências como procedimentos estatísticos. De outro lado, Cronin e colaboradores (27) apresentam um estudo de simulação que coloca em evidência os erros dos estudos com quantidades pequenas de falsos positivos que usam classes latentes na correção do viés de verificação.

Estudos que usam métodos Bayesianos de estimação

Muitos autores têm contornado os problemas encontrados na estimação dos parâmetros de testes usando métodos baseados na inferência Bayesiana. Do mesmo modo, que sob o enfoque clássico, os problemas estudados são basicamente os mesmos (viés de verificação, a falta de identificabilidade e a dependência), mas as estratégias usadas para obter as estimativas são outras, o que traz novas opções para os analistas de dados. Um dos primeiros estudos sob o enfoque Bayesiano para contornar os problemas relacionados ao diagnóstico clínico,

encontrado nesta revisão, é o artigo de Fryback (28), que faz um estudo sobre o erro de assumir independência condicional entre fatores clínicos em situações onde é necessário avaliar um número estabelecido de aspectos clínicos que identificam a presença da enfermidade. Sox (29), num artigo direcionado a clínicos e especialistas da área da saúde, faz uma breve revisão sobre o uso do teorema de Bayes no momento de fazer diagnóstico clínico incluindo diferentes testes diagnósticos. O autor enfatiza na interpretação das probabilidades a priori e a posteriori e as relações que podem ser estabelecidas entre as mesmas.

Gastwirth e colaboradores (30) desenvolvem um método Bayesiano para ser utilizado em situações onde se tem prevalências próximas de zero. O foco principal do artigo é estimar a probabilidade preditiva de um indivíduo ser positivo para um evento de interesse quando a probabilidade de dito evento é pequena demais e ilustram o método proposto com dados de doadores de sangue. Johnson e Gastwirth (31) desenvolvem uma metodologia Bayesiana para aproximar distribuições preditivas utilizando grandes amostras e comparam seus resultados com os valores exatos das distribuições. Joseph e colaboradores (32) consideram o planejamento diagnóstico com viés de verificação e propõem um método para obter as estimativas dos parâmetros de desempenho dos testes, nos casos onde se tem um ou dois testes de triagem e não existe padrão-ouro. Eles desenvolvem o método usando variáveis latentes e métodos de Monte Carlo assumindo independência entre os testes. O autor propõe também um método para elicitar distribuições Beta(a,b) a priori e uma forma de contornar o problema de falta de identificabilidade

sob a perspectiva Bayesiana. No mesmo ano, Joseph e Gyorkos (33) apresentam um método para estimar as razões de verossimilhança entre características dos testes diagnósticos (sensibilidade e especificidade) usando as distribuições simuladas num ambiente MCMC (Monte Carlo e Cadeias de Markov).

Enoe e colaboradores (33) fizeram uma revisão dos métodos estudados para estimar os parâmetros de desempenho com ênfase na forma como as metodologias Bayesianas aprimoraram e contornaram as limitações dos métodos frequentistas usados até esse momento. Esses autores também propõem um procedimento para obter distribuições Beta (a,b) a priori para os parâmetros de desempenho dos testes. Dendukuri e Joseph (34) retomam sobre o trabalho de (32), mas dessa vez concentrando-se no estudo da dependência entre os testes, para o qual usam modelos de efeitos fixos e de efeitos aleatórios comparando os resultados.

Utilizando a aproximação de (32), Martinez e colaboradores (35) estabelecem um método Bayesiano de estimação para a prevalência e os parâmetros de desempenho dos testes em planejamentos que incluem a informação de uma covariável para todos os indivíduos participantes no estudo. No seguinte ano, os mesmos autores (36) aprimoram seu método de estimação considerando os estudos que tem viés de verificação, informação de covariáveis e dependência entre testes. Achcar e colaboradores (37, 38) estudam dois cenários para propor métodos de estimação de prevalência e parâmetros de desempenho. No primeiro, tem-se um teste diagnóstico em avaliação, informação de covariáveis e não existe padrão-ouro implicando que o estado de saúde é considerado variável latente. No

segundo, assume-se uma situação na qual uma parte dos indivíduos é verificada e a outra não. Três anos depois, em 2008, Martínez e colaboradores (39), retomam este estudo e acrescentam as estimativas dos valores preditivos e das quantidades de indivíduos com diagnóstico positivo para uma enfermidade usando dados de câncer de cervix. Um ano depois (2009), os mesmos autores (40) acrescentam ao trabalho anterior uma covariável e aplicam a mesma metodologia para desenvolver estimativas de três testes diagnósticos em ausência de um padrão-ouro.

Da mesma forma que na perspectiva clássica, alguns autores estudaram o viés de verificação incluindo no modelo de estimação uma variável binária latente (o indivíduo foi ou não verificado). Martinez e colaboradores (41) usam métodos Bayesianos de estimação incluindo esta aproximação em estudos que tem só um teste diagnóstico, uma parte dos indivíduos é verificada pelo padrão-ouro e a outra não. Buzoianu e Kadane (42) também modelam o viés de verificação como se fosse uma variável latente aplicando modelos de regressão logística com ligação logito quando se tem um teste diagnóstico. Aragon e colaboradores (43) generalizam o estudo de (42) para situações onde se tem dois testes diagnósticos e comparam os resultados obtidos supondo independência com aqueles obtidos assumindo dependência condicional.

Estudos que abordam o problema de dependência entre testes diagnósticos

O problema da dependência entre os testes de triagem e seu efeito sobre os parâmetros de desempenho dos testes é o menos estudado e mais recentemente identificado. Embora esse problema

tenha sido abordado de forma tangencial em muitos dos trabalhos descritos nas duas seções anteriores, considera-se é importante entrar um pouco mais em detalhe na revisão dos estudos que concentraram seu interesse na abordagem do mesmo.

Vacek (44) usa o planejamento proposto por (6) assumindo duas populações e dois testes diagnósticos em ausência de padrão-ouro e realiza um estudo analítico sobre o comportamento da associação entre testes com resposta binária usando como parâmetro a covariância. A autora faz uma análise da forma como a presença de associação entre testes afeta as estimativas da sensibilidade, a especificidade e a prevalência quando se usa o método de máxima verossimilhança. Num único estudo encontrado na sua classe, Brenner (45) aborda a situação onde se pode ter apenas um ou vários testes diagnósticos de resposta em escala contínua dentro de planejamentos com forma paralela ou em série. O autor avalia o comportamento dos parâmetros dos testes quando se mudam os valores dos pontos de corte e os valores das covariâncias entre testes depois de serem dicotomizados.

Torrance-Rynard e Walter (46) propõem modelos de classes latentes para o viés e o real estado de saúde. Os autores calculam as estimativas de máxima verossimilhança para a dependência e os parâmetros dos testes e comparam com os resultados obtidos assumindo independência entre testes. Qu e Hadgu (47) modelam múltiplos testes usando modelos lineares generalizados (ligação logito) e obtêm os estimadores de máxima verossimilhança dos parâmetros e da matriz de variâncias-covariâncias usando a quadratura Gauss-Hermite para aproximar distribuições normais com dados discretos.

Georgiadis e colaboradores (48) propõem uma reparametrização um a um dos parâmetros de desempenho em termos da correlação entre os testes e, utilizando uma abordagem Bayesiana, obtêm as estimativas para a reparametrização e depois estimam os parâmetros de interesse. Sullivan e Janes (49) desenvolvem um método que permite obter de forma analítica as estimativas dos parâmetros em situações onde não existe padrão-ouro e comparam os resultados obtidos usando a sua aproximação com os obtidos por meio de modelos com classes latentes.

Entre todos os artigos revisados, o único que trata diretamente o problema de estimação da dependência é o realizado por Bohning e Patilea (50). Eles desenvolvem dois índices para medir a associação entre testes diagnósticos a partir da tabela de probabilidades observadas em situações com verificação somente em indivíduos com ao menos um resultado positivo. Os autores estimam as quantidades de indivíduos realmente doentes e não doentes entre o grupo de pessoas com ambos os testes de triagem negativos de modo que, com os dados completados é possível estimar os parâmetros de interesse.

Discussão

Quando se desenvolve um novo teste diagnóstico para alguma doença ou infecção, é preciso avaliar a eficiência do mesmo realizando estudos de campo que consideram planejamentos para a coleta de dados com grupos (amostras) de sujeitos que apresentam características clínicas associadas à presença do fenômeno (sintomas) e grupos de sujeitos que não as apresentam. Os dados obtidos são utilizados no desenvolvimento de um modelo estatístico para estimação,

no qual os parâmetros são os índices quantitativos que identificam a capacidade de classificação do teste. Estudar as características desses parâmetros e o comportamento das suas estimativas tem sido alguns dos temas de interesse dentro da pesquisa nas áreas médica e estatística. Neste momento, conta-se com um importante número de estudos sobre o tema o que tem permitido identificar problemas tanto estatísticos quanto metodológicos na realização do procedimento de estimação. No campo estatístico as pesquisas tem se concentrado no estudo de três grandes problemas: o viés de verificação, a falta de identificabilidade no modelo estatístico e a dependência entre resultados dos testes.

A falta de identificabilidade é um problema que consiste no fato de haver um número de parâmetros para estimar maior do que o número de observações disponíveis para realizar a estimação. Esse problema foi amplamente abordado por (8) numa revisão sobre os procedimentos de estimação de parâmetros de testes., concluindo-se que a falta de identificabilidade só está presente quando se analisam dados obtidos com planejamentos que incluem um ou dois testes sob avaliação e um padrão-ouro (ou um teste de referência) para verificação. Com três ou mais testes de triagem não é possível ter falta de identificabilidade. Inicialmente, a estratégia utilizada por vários autores como (44), para obter as estimativas usando máxima verossimilhança, foi colocar restrições num subconjunto de parâmetros e assumi-los como conhecidos. Na perspectiva Bayesiana, o problema tem sido contornado assumindo-se distribuições a priori informativas sobre um subconjunto ou o conjunto de parâmetros, alternativa

que tem como limitante o fato de se obter estimativas cuja amplitude de valores está limitada ao posto da distribuição a priori utilizada.

A presença de viés de verificação é o problema mais estudado, tanto na área estatística quanto na área médica, dada a importância que tem seus efeitos sobre as estimativas dos parâmetros, que pode ter fortes implicações na saúde pública. Em revisões realizadas com estudos publicados em revistas de pesquisa médica, Gupta e Roehrborn (51) expressam que, entre 1302 artigos publicados em grandes revistas médicas como, o *England Journal of Medicine*, somente 46% dos estudos utilizaram alguma estratégia para contornar o viés de verificação. Mais recentemente, (27) encontraram que um grande número das revisões sistemáticas sobre avaliação de testes para diagnóstico tem se concentrado na proporção de estudos nos quais o modelo estatístico não considera a correção do viés de verificação. Esses autores citam alguns exemplos, como uma revisão entre estudos pediátricos realizados entre 1987 e 1989, na que se encontrou que, o 40% (15/42) dos estudos apresentava resultados com viés de verificação e outra revisão com 112 estudos realizados entre 1978 e 1983, dos quais só 51 (46%) apresentaram ajustes aos dados para eliminar o viés de verificação. De acordo com esses autores, conclui-se que, nos estudos realizados recentemente, a correção do viés aparece com maior frequência. Na revisão aqui apresentada, foi encontrada uma ampla gama de aproximações para contornar o problema do viés de verificação, que incluem o uso de variáveis latentes (discretas e contínuas) e diferentes formas de reparametrizações além de alternativas para contornar os problemas

da falta de identificabilidade e da falta de independência entre testes.

A partir da revisão realizada, pode-se concluir que, a maioria dos estudos tem desconhecido que muitos métodos de diagnóstico clínico incluem a medição de traços biológicos cuja resposta é expressa em escala contínua e que, devido ao fato de serem medidos no mesmo indivíduo, esses traços necessariamente apresentam algum tipo de dependência que pode ou não ser explicada como um fenômeno de comportamento linear ou de concordância. Muitas propostas apresentadas para contornar o problema de dependência partem do pressuposto de que os testes tem resposta binária, sem considerar que as observações tomam essa apresentação depois de serem dicotomizadas por meio de um ponto de corte estabelecido a partir de critérios clínicos. Dentro das propostas apresentadas para estudar a dependência, geralmente são utilizadas reparametrizações complexas e modelos com variáveis latentes que exigem o uso de métodos iterativos computacionalmente complexos para obter os valores das estimativas. Os autores que consideraram a continuidade nas variáveis dos testes, assumem também que as mesmas ajustam-se à distribuição normal bivariada de probabilidades, pressuposto que em muitas situações de diagnóstico clínico pode não se cumprir.

A respeito da dependência entre os resultados dos testes, este é um assunto que tem sido abordado mais recentemente. A

maioria dos autores assume estrutura binária com dependência suscetível de ser modelada usando correlação de Pearson. Somente num dos artigos revisados considerou-se o uso de testes diagnósticos que respondem a traços contínuos. É importante ressaltar que em planejamentos nos quais é preciso medir traços contínuos com posterior dicotomização, a estrutura do modelo estatístico para estimação deve considerar a presença da dependência entre as variáveis medidas como um parâmetro a mais para ser estimado e cujo efeito deve ser isolado para se obter os verdadeiros valores dos parâmetros de interesse (prevalência, sensibilidades e especificidades). Quando as estimativas são obtidas utilizando-se os resultados binários, perde-se uma importante quantidade da informação contida nos dados contínuos e a estrutura de dependência entre variáveis não aparece de modo evidente no modelo proposto. Essa situação não acontece quando a distribuição de probabilidade dos traços biológicos medidos é uma Normal bivariada e a relação entre os mesmos é considerada linear. Nesses casos, a covariância das variáveis binárias obtidas depois da dicotomização tem uma relação funcional com o coeficiente de correlação de Pearson das variáveis contínuas originalmente medidas. Por outro lado, pode acontecer que as análises iniciais dos dados não ofereçam suficiente evidência para assumir as hipóteses de normalidade e dependência linear, de modo que a proposta de análise referida não é válida, sendo necessário utilizar outras formas de abordar o problema.

REFERENCIAS

1. Neyman J. Outline of statistical treatment of the problem of diagnosis. *Pub. Health Reports*. 1947, 62:1449-1456.
2. Homburguer F., Pfeiffer P., Page O., Rizzone G., Benotti J. Evaluation of diagnostic tests for cancer. *Cancer*. 1950 jan, 3:15-25.
3. Youden W.J. Index for rating diagnostic tests. *Cancer* 1950, 3:32-35.
4. Mantel N. Evaluation of a class of diagnostic tests. *Biometrics* 1951. 7:240-246.
5. Gart J. Buck A. Comparison of a screening test and reference test in epidemiologic studies. II. A probabilistic model for the comparison of diagnostic tests. *American Journal of Epidemiology* 1966, 88:593-602.
6. Rogan W. Gladen, B. Estimating prevalence from the results of a screening test. *American Journal of Epidemiology*, 1978, 107(1):71-76.
7. Hui S. Walter S. Estimating the error rates of diagnostic tests. *Biometrics* 1980, 36:167-171.
8. Thibodeau, L.A. Evaluating diagnostic tests. *Biometrics* 1981, 37:801-804.
9. Walter S. Iwig L.M. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *Journal of Clinical Epidemiology* 1988, 41:923-937.
10. Schatzkin A., Connor R., Taylor P., Bunnag B. Comparing new and old screening tests when a reference procedure cannot be performed on all screenees. *Am. J. Epid.* 1987. 125(4): 672-678.
11. Cheng H., Macaluso M. Comparison of the accuracy of two tests with a confirmatory procedure limited to positive results. *Epidemiology* 1997. 8:104-106.
12. Sullivan, M., Todd A. Comparing disease screening tests when true disease status is asserted only for screen positives. *Biostatistics* 2001. 2:249-260.
13. Walter S. Estimation of test sensitivity and specificity when disease confirmation is limited to positive results. *Epidemiology* 1999. 10:67-72.
14. Goethebeur E., Linev J., Boelaert M., Van Der Stuyft P. Diagnostic test analyses in search of their gold standard: latent class analyses with random effects. *Statistical Methods in Medical Research* 2000. 9:231-248.
15. Van Der Merwe L., Maritz, S. Estimating the conditional false-positive rate for semi-latent data. *Epidemiology* 2002. 13:424-430.
16. Garret E., Eaton W., Zeger S. Methods for evaluating the performance of diagnostic tests in the absence of a gold standard: a latent class model approach. *Statistics in Medicine* 2002. 21:1289-1307.
17. Bertnasky S., Joseph L., Belisle P., Boivin J.F., Raghu R., Moore A., Clarke A. Bayesian modelling of imperfect ascertainment methods in cancer studies. *Statistics in Medicine* 2005. 24:2365-2379.
18. Begg C., Greenes R. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* 1983. 39:207-215.
19. Baker S. Evaluating multiple diagnostic tests with partial verification. *Biometrics* 1995. 51:330-337.
20. Qu Y., Ming T., Kutner M. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 1996. 52:797-810.
21. Yang I., Becker, M. latent variable modeling of diagnostic accuracy. *Biometrics* 1997. 53:948-958.
22. Zhou, X.. Correcting for verification bias in studies of a diagnostic test's accuracy. *Statistical Methods in Medical Research* 1998. 7:337-353.
23. Albert P., Lori. A. Cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. *Biometrics* 2004. 60:427-435.

24. Albert P., McShane L., Shih J. and The U.S. National Cancer Institute Bladder Tumor Marker Network. Latent class modeling approaches for assessing diagnostic error without a gold standard: with application to p53 immunohistochemical assays on bladder tumors. *Biometrics* 2001. 57:610-619.
25. Rindskopf, D., Rindskopf W. The value of latent class analysis in medical diagnosis. *Biometrics in Medicine* 1986. 5:21-27.
26. Rindskopf D. The use of latent class analysis in medical diagnosis. Papers presented at the Annual Meeting of the American Statistical Association 2002. 2912-2916.
27. Cronin A., Vickers A. Statistical methods to correct for verification bias in diagnostic studies are inadequate when there are few false negatives: a simulation study. *BMC Medical Research Methodology* 2008. 8:75, 1-9.
28. Fryback D. Baye's theorem and conditional nonindependence of data in medical diagnosis. *Computers and Biomedical Research* 1978. 11:423-434.
29. Sox H. Probability theory in the use of diagnostic tests: An introduction to critical study of the literature. *Annals of Internal Medicine* 1986. 104:60-66.
30. Gastwirth J., Johnson W., Reneau D. Bayesian analysis of screening data: Application to AIDS in blood donors. *The Canadian Journal of Statistics*. 1991. 19(2):135-150.
31. Johnson W., Gatswirth J. Bayesian inference for medical screening tests: Approximations useful for the analysis of Acquired Immune Deficiency Syndrome. *Journal of Royal Statistical Society, Series B* 1991. 53(2):427-439.
32. Joseph L., Gyorkos T., Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* 1995. 141:263-272.
33. Joseph L., Gyorkos T. Inferences for likelihood ratios in the absence of a "Gold Standard". *Medical Decision Making* 1995. 16:412-417.
34. Enøe, C., Georgiadis M., Johnson W. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Preventive Veterinary Medicine* 2000. 45:61-81.
35. Dendukuri N., Joseph L. Bayesian approach to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001. 57:158-167.
36. Martinez E., Achcar J., Louzada F. Análise Bayesiana do desempenho de dois testes diagnósticos quando indivíduos com resultados negativos em ambos os testes não são verificados por um padrão-ouro. *Revista Brasileira de Biometria* 2004. 22(3):21-32.
37. Martinez E., Achcar J., Louzada F. Bayesian estimation of diagnostic tests accuracy for semi-latent data with covariates. *Journal of Biopharmaceutical Statistics* 2005. 15:809-821.
38. Achcar J., Martinez Edson., Louzada F. Binary data in the presence of covariates and misclassifications: A Bayesian approach. *Brazilian Journal of Probability and Statistics* 2005. 19:65-84.
39. Martinez E., Louzada F., Derchain S.F., Achcar J., Gontijo R., Ssarian L., Syrjänen K. Bayesian estimation of performance measures of cervical cancer screening tests in the presence of covariates and absence of a gold standard. *Cancer Informatics* 2008. 6:33-46.
40. Martinez E., Louzada F., Achcar J., Syrjänen K., Derchain S.F., Gontijo R., Sarian L. Bayesian estimation of performance measures of screening tests in the presence of covariates and absence of a gold standard. *Brazilian Journal of Probability and Statistics* 2009. 23:68-81.
41. Martinez E., Achcar J., Louzada F. Estimators of sensitivity and specificity in the presence of verification bias: A Bayesian approach. *Computational Statistics and Data Analysis* 2006. 51:601-611.
42. Buzoianu M., Kadane, J. Adjusting for verification bias in diagnostic test evaluation: A Bayesian approach. *Statistics in Medicine* 2008. 27:2453-2473.
43. Aragon D., Martinez E., Achcar J. Bayesian estimation for performance measures of two diagnostic tests in the presence of verification bias. *Journal of Biopharmaceutical Statistics* 2010. 20:821-834.

44. Vacek P. The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* 1985. 41:959-968.
45. Brenner H. How independent are multiple “independent” diagnostic classifications?. *Statistics in Medicine* 1996. 15:1377-1386.
46. Torrance-Rynard V., Walter, S. Effects of dependent errors in the assessment of diagnostic tests performance. *Statistics in Medicine* 1997. 16:2157-2175.
47. Qu, Y., Hadgu A. A model for evaluating sensitivity and specificity for correlated diagnostic test in efficacy studies with an imperfect reference test. *Journal of the American Statistical Association* 1998. 93:920-928.
48. Georgiadis M., Johnson W., Gardner I. Singh R. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2003. 52:63-76.
49. Sullivan M., Janes H. Insights into latent class analysis of diagnostic test performance. *Biostatistics* 2007. 8(2):474-484.
50. Böhning, D., Patilea V. A capture-recapture approach for screening using two diagnostic tests with availability of disease status for the test positives only. *Journal of American Statistical Association* 2008. 103(481):212-221.
51. Gupta A., Roehrborn, C. Verification and incorporation biases in studies assessing screening tests: prostate-specific antigen as an example. *Urology* 2004. 64(1):106-116.