

Debate sobre o artigo de Narvai et al.

Debate on the paper by Narvai et al.

Margareth
Crisóstomo Portela
& Mauricio Teixeira
Leite de Vasconcellos

Escola Nacional de Saúde
Pública Sérgio Arouca,
Fundação Oswaldo Cruz,
Rio de Janeiro, Brasil.
mportela@ensp.fiocruz.br

Instituto de Pesquisa Clínica
Evandro Chagas, Fundação
Oswaldo Cruz, Rio de Janeiro,
Brasil.
mauricio.vasconcelos@ipecc.
fiocruz.br

A redação do texto facilita muito nossa resposta, porque os autores da crítica em questão declararam seu conflito de interesse e porque reconheceram que o artigo de Queiroz et al.¹ foi feito dentro do paradigma da amostragem probabilística e da pesquisa por amostragem, declarando que as “restrições não são infundadas e devem ser cuidadosamente consideradas pelos profissionais de saúde pública, gestores e pesquisadores” (p. 647).

As questões levantadas no artigo de Queiroz et al.¹ foram suscitadas por conta do interesse de utilização dos dados do SB Brasil 2003 e da constatação da ausência das variáveis estruturais do desenho da amostra descrita em sua publicação. Assim, o artigo¹ foi produzido para alertar os pesquisadores da área de que o processo de amostragem do inquérito não havia sido concluído e que os dados disseminados na página da Internet do Ministério da Saúde (<http://www.saude.gov.br>) só permitiam produzir estatísticas da amostra pesquisada. Apesar de isso já ser uma contribuição importante, o artigo prosseguiu propondo uma estratégia para complementar o processo amostral. Hoje sabemos que a estratégia adotada, apesar de trabalhosa, mostrou-se viável uma vez que os pesos amostrais foram calculados e as demais variáveis estruturais do desenho da amostra foram identificadas e inseridas no arquivo de dados, conforme artigo submetido aos *Cadernos de Saúde Pública* em 16 de outubro de 2009.

A questão levantada pelos autores da crítica, sobre quão diferentes as estatísticas da amostra são das estimativas populacionais, contida na discussão sobre o que denominam “validade e representatividade”, depende do desenho da amostra e da variância das variáveis consideradas. Parte dela pode ser alvo de uma análise epistemológica, ou mais precisamente, metodológica, enquanto a outra parte seria observacional, visto depender dos valores das variáveis observadas na pesquisa.

A resposta à parte observacional da pergunta pressupõe a obtenção das estimativas pontuais e de suas variâncias, o que não poderia ser feito sem a conclusão do processo de amostragem.

Como esse processo já foi concluído, tão logo o Ministério da Saúde mostre interesse em incluir essas variáveis adicionais na base de dados que dissemina em seu sítio, os autores dos artigos referidos na crítica – referências 13 a 31 – poderão verificar quão próximas das estimativas populacionais encontram-se as estatísticas amostrais que produziram.

Nunca foi objetivo dos autores do artigo criticado reter a estratégia de finalização do processo de amostragem e as variáveis complementares construídas, gerando uma linha de produção de artigos comparando resultados já publicados com estimativas populacionais obtidas dentro do paradigma da amostragem probabilística. A intenção foi e é, unicamente, contribuir para o aprimoramento científico da área de saúde bucal no Brasil, tendo-se assumido o imenso trabalho de correção e complementação da base de dados, bem como contribuído na chamada pública de 2009 para evitar a reprodução dos mesmos problemas no SB Brasil 2010.

Em relação à parte metodológica, pode-se valer do artigo comemorativo dos 100 anos do nascimento da pesquisa por amostra², em que Leslie Kish advoga que o artigo de Kiar³ corresponde à certidão de nascimento da pesquisa por amostragem, apesar de seus efeitos não terem sido sentidos na primeira metade do período. Segundo Kish, somente na segunda metade do período é que a pesquisa por amostragem decolou em consequência dos avanços acelerados da modelagem em amostragem nas décadas de 1930 e 1940: estratificação, conglomeração, métodos de seleção e múltiplos estágios de seleção.

Em seu artigo de 1934, Jerzy Neyman⁴ estende a alternativa existente (método representativo) com o desenvolvimento de uma base de inferência por meio dos intervalos de confiança. Sua proposta embutia uma nova tese epistemológica sobre indução: o conceito de comportamento indutivo em contrapartida à visão positivista contida no método representativo. Segundo David Teira⁵, da mesma forma que a noção de máxima verossimilhança, trazida para a estatística matemática por Ronald Fisher, o conceito de confiança de Neyman não podia ser simplesmente equacionado como uma probabilidade convencional e requereu uma justificativa epistemológica distinta.

Assim, durante a apresentação de seu método para a Royal Statistical Society, Neyman⁴ (p. 623) afirmou que: “*The term confidence coefficient is not synonymous to the term probability. It means an arbitrarily chosen value of the probability of our being right when applying a certain rule of behavior... The validity of probability statements in the new form of the problem of estimation, which*

has been here so extensively discussed, depends on the permanent use of a system of confidence intervals. This system as a whole (not separate intervals) corresponds to a fixed probability that our predictions are correct".

Em outras palavras, os princípios epistemológicos que regem a pesquisa por amostragem baseiam-se nos intervalos de confiança, construídos com base nos dados observados (estimativa e seu desvio padrão, também chamado de erro padrão) e em probabilidades escolhidas (o nível de significância e seu complemento, o nível de confiança). Dentro dessa visão, o cerne da estatística é a variabilidade e não o valor pontual da estimativa, já que é a variabilidade que determina a precisão de "estar certo", para um nível de significância escolhido.

Em artigo citado na crítica, Reichenheim & Moraes ⁶ (p. 135), ao falarem da validade de domínio, concluem que "*deve ser enfatizado que a capacidade de generalização dos resultados de um estudo não é uma questão de representatividade amostral, e sim, de representatividade inferencial*". Assim, não basta que a amostra seja aleatória e representativa da população, é preciso que ela atenda aos pressupostos inferenciais para que haja a validade de domínio. Reconhecemos a beleza de Guernica, mas jamais usaríamos esta obra única de arte para estimar prevalências de baixas ou fazer qualquer outra quantificação sobre os horrores da guerra que retratou.

É fato que nas amostras autoponderadas (onde o peso amostral é uma constante), algumas estatísticas amostrais constituem-se em estimativas não enviesadas da população. No entanto, totais populacionais só podem ser estimados com emprego dos pesos amostrais, que, por serem constantes, permitem que totais sejam estimados multiplicando proporções amostrais pela dimensão da população. Com exceção desse caso particular, todas as demais amostras não asseguram que estatísticas amostrais sejam estimativas não enviesadas da população, apesar de poderem ser iguais (ou próximas) em caso de variância nula (ou muito pequena). Diante disso, fica fácil entender porque o artigo criticado incluiu a Tabela 1, comparando a dimensão da amostra nos vários estratos com o tamanho da população observado no *Censo Demográfico 2000*: a amostra do SB Brasil 2003 não é autoponderada e, dentro dos princípios que regem a pesquisa por amostragem, requer pesos amostrais. Fica claro, também, que as razões entre as dimensões da população de 2000 e da amostra de 2003 não servem como peso amostral e, portanto, os autores propuseram uma estratégia para cálculo desses pesos. Nesse sentido, o exercício feito na Tabela 1 da crítica não merece

resposta: (1) não existiam os pesos amostrais; e (2) algumas unidades primárias de amostragem (escolas e creches nos municípios de capital, por exemplo) só puderam ser identificadas após a recuperação das informações sobre a seleção da amostra nos municípios, feita entre março de 2006 e dezembro de 2009.

Em relação ao volume de artigos derivados do inquérito, deve-se responder que esta não é uma boa medida da importância de um projeto de pesquisa governamental. Se assim fosse, os censos demográficos não seriam mais realizados, apesar de todos reconhecerem seu uso no planejamento, na divisão tributária entre as esferas de governo e sua condição de levantamento de base de todas as demais pesquisas demográficas (inclusive o SB Brasil 2003, que utilizou os setores censitários, estratificou os municípios e os selecionou com base em estatísticas populacionais derivadas dos censos).

Curioso, também, que em toda a produção citada, a avaliação epidemiológica foi "subordinada" aos paradigmas usuais da estatística. Sem qualquer referência à nova epistemologia que os autores da crítica defendem, nos artigos referidos existe pelo menos um dos seguintes elementos: (1) cálculo de valores *p*; (2) estimativa de "intervalos de confiança" [*sic*], baseados em estatísticas pontuais, razões e variâncias amostrais; (3) testes de hipóteses; e (4) modelos estatísticos.

Além disso, os artigos indicaram, direta ou indiretamente, que a amostra era complexa, mas não abordaram em seus métodos as técnicas usadas para lidar com a fonte de aleatoriedade advinda do desenho de amostra. Houve uma preocupação grande com os cuidados ligados aos erros de medida, sendo desprezada a principal fonte de contribuição para o erro quadrático médio (que no final do dia é o que importa): os erros de amostragem.

Tal fato vem desde o manual da Organização Mundial da Saúde ⁷, referido em vários dos artigos citados, que se dedica quase que exclusivamente aos aspectos relacionados aos erros de medida, definição de traçadores (idades) e padronização da operação da pesquisa em nome da comparabilidade, limitando-se, no tópico de preparação do protocolo de pesquisa, a apenas uma linha sobre amostragem: "*a description of the sampling methods to be used*" ⁷ (p. 10).

Quando o artigo sobre o cálculo dos pesos for publicado, os leitores poderão ver que além dos problemas já discutidos, a base de dados disseminada pelo Ministério da Saúde tem registros duplicados e outros problemas derivados da operação realizada.

Por fim, se um especialista da amostragem estivesse diante do desenho da amostra de uma

pesquisa que incluía objetivos de treinamento e aperfeiçoamento das equipes locais de saúde bucal, como parece ter sido o caso do SB Brasil 2003, ele muito provavelmente aproveitaria esse fato para aumentar a precisão das estimativas, reduzindo o seu nível de conglomeração com o aumento do número de municípios.

1. Queiroz RCS, Portela MC, Vasconcellos MTL. *Pesquisa sobre as Condições de Saúde Bucal da População Brasileira* (SB Brasil 2003): seus dados não produzem estimativas populacionais, mas há possibilidade de correção. *Cad Saúde Pública* 2009; 25:47-58.
2. Kish L. The hundred years' wars of survey sampling. In: *Proceedings of the Conference to Commemorate 100 Years of Sample Survey*. Rome: Centro d'Informazione e Stampa Universitaria; 1996. p. 15-27.
3. Kiar AN. Observations et expériences concernant les dénombremments représentatifs. *Bulletin of the International Statistical Institute* 1895; 9:176-205.
4. Neyman J. On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *J R Stat Soc Ser A* 1934; 97:558-625.
5. Teira D. Milton Friedman, the statistical methodologist. *Hist Polit Econ* 2007; 39:511-27.
6. Reichenheim ME, Moraes CL. Alguns pilares para a apreciação da validade de estudos epidemiológicos. *Rev Bras Epidemiol* 1998; 1:131-48.
7. World Health Organization. *Oral health surveys: basic methods*. 4th Ed. Geneva: World Health Organization; 1997.