Carlos Augusto Monteiro[I,II]

Alex Antonio Florindo[I,III]

Rafael Moreira Claro[I,II]

Erly Catarina Moura[I,IV]

Validity of indicators of physical activity and sedentariness obtained by telephone survey

ABSTRACT

OBJECTIVE: To assess the reliability and validity of indicators of physical activity and sedentariness obtained by means of a telephone-based surveillance system.

METHODS: Reliability and validity studies were carried out in two random subsamples (n=110 and n=111, respectively) obtained from the total sample (N=2,024) of adults (≥18 years) studied by the system in the municipality of São Paulo in 2005. Studied indicators included frequency of "sufficiently active during leisure time," "inactive in four domains of physical activity (leisure, work, transportation, and housework)," and "habit of watching television for long periods." Reliability was assessed by comparing results of the original telephone interview with those of another identical interview repeated after seven to 15 days. Validity was assessed by comparing the results of the telephone interview with those of three 24-hour recalls (reference method) carried out in the week following the original interview.

RESULTS: Frequencies obtained for of the three evaluated indicators were either identical or very similar for the first and second telephone interviews. Kappa coefficients ranged from 0.53 to 0.80, indicating good reliability for all indicators. In relation to the reference method, all indicators showed 80% or higher specificity, and sensitivity values were 69.7% for "watching television for long periods," 59.1% for "inactive in four domains," and 50% for "sufficiently active during leisure."

CONCLUSIONS: The indicators of physical activity and sedentariness included in the system seem reliable and sufficiently accurate. If kept operational in coming years, this system may provide Brazil with a useful instrument for evaluating public policies aimed at promoting physical activity and controlling non-transmissible chronic diseases associated with sedentariness.

DESCRIPTORS: Life Style. Diet Surveys. Indicators of Quality of Life. Reproducibility of Results. Validity of Tests. Nutritional Surveillance. Physical Activity.

[I] Núcleo de Pesquisas Epidemiológicas em Nutrição e Saúde. Universidade de São Paulo. São Paulo, SP, Brasil

[II] Departamento de Nutrição. Faculdade de Saúde Pública. Universidade de São Paulo. São Paulo, SP, Brasil

[III] Escola de Artes, Ciências e Humanidade. Universidade de São Paulo. São Paulo, SP, Brasil

[IV] Universidade Federal do Pará. Belém, PA, Brasil

Correspondence:
Carlos Augusto Monteiro
Departamento de Nutrição
Faculdade de Saúde Pública
Universidade de São Paulo
01246-904 São Paulo, SP, Brasil
E-mail: carlosam@usp.br

INTRODUCTION

Global estimates indicate that non-transmissible chronic diseases (NTCDs) determine roughly 60% of all deaths worldwide, and almost half the global burden of disease.[16] In Brazil, it is estimated that NTCDs account for almost two-thirds of all deaths with known cause.[a] The proportion of deaths due to

[a] Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação em Saúde. *Saúde Brasil* 2006: *uma análise da situação de saúde no Brasil.* Brasília: Ministério da Saúde; 2006. 620 p.

NCTDs in Brazilian state capitals increased more than three-fold between the 1930s and '90s.[4] In all regions of the globe, a small group of risk factors determines the great majority of NCTD deaths as well as a substantial fraction of the disease burden related to these diseases. Noteworthy among these factors are unhealthy diets and insufficient physical activity.[16]

In Brazil, the frequency and distribution of risk factors for NTCDs is monitored by means of a telephone-based surveillance system. This system, known as VIGITEL (*Vigilância de Fatores de Risco e Proteção para Doenças Crônicas não Transmissíveis por Inquérito Telefônico* [Surveillance of Risk and Protective Factors for Non-Transmissible Chronic Diseases by Telephone Interview]), has been in operation since 2006 in all 26 Brazilian state capitals as well as in the Federal District.[a] VIGITEL was tested successfully in the city of São Paulo in 2003,[8] and was retested in this same city and in four other capitals in 2005. During the second test in São Paulo, a study of the reliability and validity of the indicators obtained was coupled to the system's normal operation. The present article describes the results pertaining to indicators of physical activity and sedentariness. The reliability and validity of diet-related indicators was described in Monteiro et al.[9]

METHODS

Two systematic subsamples, each with 115 subjects, were extracted from the total sample (N=2,024) of subjects aged 18 years or older surveyed by the VIGITEL system in the city of Sao Paulo, respecting the proportion of men and women in the total sample. VIGITEL sampling procedures have been described in detail elsewhere.[8,a] Five subjects from the first subsample (reliability) and four from the second (validity) either refused to participate in the study or did not complete the required interviews. The reliability study thus included 110 subjects (47 men and 63 women; mean age 45 years; 26.3% with up to 8 years schooling; and 32.7% with 12 or more years schooling), whereas the validity study included 111 subjects (50 men and 61 women; mean age 44 years; 34.2% with up to 8 years schooling; and 27.0% with 12 or more years schooling).

Indicators of physical activity in the VIGITEL system address sufficient physical activity during leisure time, simultaneous inactivity in four domains of physical activity (leisure, work, transportation to work, and household), and the habit of watching television for extended periods of time. Based on the answers provided by the subjects to the questions on physical activity in the VIGITEL questionnaire, the system classifies as "sufficiently active during leisure" subjects who report physical exercise or sport of moderate intensity for at least 30 minutes per day at least five days per week, or sport of vigorous intensity for at least 20 minutes per day at least three days per week. Subjects were classified as "inactive in four domains of physical activity" when they reported 1) not practicing sports or physical exercise at least one day per week; 2) "not walking on a regular basis"; and "not carrying heavy loads on a regular basis" at work (or unemployed for the last three months); 3) not commuting from home to work on foot or by bicycle; and 4) not being responsible for "heavy cleaning" at home. The intensity of exercise or sport reported by the subject is classified *a posteriori* by the system based on a compendium that estimates the energy expenditure associated with different forms of physical activity, attributing moderate intensity to exercise or sports associated with energy expenditures ranging from three to six times that of resting, and vigorous intensity to expenditures equivalent to six or more times that of resting.[1] Finally, the status of "watching television for extended periods" was attributed to subjects who watched television three or more hours per day at least five days per week.

For the reliability study, subjects were contacted by phone seven to 15 days after the original interview by the system, when they were asked to respond again to the block of 12 questions on physical activity. The second interviewer was always different from that of the original interview. The results of these two sequential interviews were compared in terms of frequency of "sufficiently active during leisure," "inactive in four domains of physical activity," and "watching television for extended periods of time," as well of agreement between the individual classification of each subject with respect to these three indicators. In this last case, the degree of agreement between the two interviewers was evaluated using the kappa coefficient, classified as follows: above 0.80 indicates virtually perfect agreement; 0.61 to 0.80, substantial agreement; 0.41 to 0.60, moderate agreement; 0.21 to 0.40, fair agreement; and below 0.21, slight agreement.[3]

For the validity study, subjects responded to three 24-hour recalls addressing physical activity. These surveys consisted of asking subjects to report in detail the type and duration of all physical activity performed in the 24 hours preceding the interview.[10] In the specific case of the present study, if there was no spontaneous report of physical activity in any of the four domains investigated, we directly asked the subject about occasional physical activity in these domains, including type and duration. The same procedure was used for subjects who did not mention "watching television." The 24-hour recalls were administered via telephone in the week following the original interview by the system. Two of the recalls referred to weekdays and the third referred to a Saturday, Sunday, or holiday.

The validity study consisted of comparing the results of the regular VIGITEL telephone interview with those of the three 24-hour recalls (gold-standard). We compared frequencies of "sufficiently active in leisure," "inactive in four domains of physical activity," and "watching television for extended periods," in addition to calculating, for each indicator, the degree of accuracy in classifying the (true) status of each subject as determined by the reference method. We considered as "sufficiently active in leisure" subjects that, in at least two of the three 24-hour recalls, reported performing physical exercise of moderate intensity for 30 minutes or of vigorous intensity for 20 minutes, the classification of exercise intensity being based on the same criteria described above. We considered as "inactive in four domains of physical activity" subjects that, in all three 24-hour recalls, failed to report physical exercise of any type, occupational activities implying walking (at least 30 minutes) or carrying heavy loads, transportation by bicycle or on foot to and from work, and activities related to "heavy cleaning" of the subject's own home. Finally, we classified as positive for "watching television for extended periods" subjects that reported watching television for at least three hours in at least two of three 24-hour recalls.

The degree of accuracy of the telephone interview in determining the true status of each subject was assessed by calculating specificity and sensitivity for each indicator, i.e., the proportion of accurate classifications made by the telephone interview among subjects with "case" or "non-case" status, respectively, according to the reference method.[13]

In addition, in order to assess the validity of indicators obtained by telephone, we compared, based on the three 24-hour recalls, the mean and median daily number of minutes that subjects classified by the telephone interview as "cases" or "non-cases" for each indicator spent on 1) any sport or physical exercise; 2) total physical activity in the four domains studied (leisure, work, transportation to work on foot or by bicycle, and "heavy cleaning" of the home; and 3) watching television. Given the absence of normal distribution in the duration of the activities evaluated, the statistical significance of differences between groups was determined using the non-parametric test for difference between two medians.[7]

The study was approved by the Research Ethics Committee of the Faculdade de Saúde Pública da Universidade de São Paulo.

RESULTS

Table 1 compares the results of the original VIGITEL interviews with those of the repeat interviews. Frequency of "sufficiently active during leisure" was identical for the two series of interviews (24.5%). Frequencies were very similar for "inactive in four domains" (24.6% and 23.6%, respectively), and "watching television for extended periods" (33.6% and 34.6%, respectively). The kappa coefficient indicates substantial agreement for "sufficiently active during leisure" (0.80) and for "inactive in four domains of physical activity" (0.78) and moderate agreement for "watching television for extended periods" (0.53).

Table 2 compares frequencies estimated by the VIGITEL interview with those of the gold-standard. For "inactive in four domains," the difference between values obtained by telephone interview (22.5%) and by the 24-hour recalls (19.8%) was minimal. For the other two indicators, frequencies obtained by telephone interview tended to be slightly overestimated: 26.1% vs. 21.6% for "sufficiently active in leisure" and 35.1% vs. 29.7% for "watching television for extended periods."

The telephone interview showed high specificity (close to or above 80%) for all three indicators. Sensitivity was 69.7% for "watching television for extended periods," 59.1% for "inactive in four domains of physical activity," and 50% for "sufficiently active during leisure."

The mean time per day spent on physical exercise or sports of any nature, estimated based on the 24-hour recalls, was 31.8 minutes for subjects classified by the telephone interview as "sufficiently active during leisure," vs. 8.9 minutes for the remainder of subjects

Table 1. Reproducibility of indicators of physical activity and sedentariness among adults (≥ 18 years) obtained by telephone interview. City of São Paulo, Southeastern Brazil, 2005.

Indicator	Original telephone interview	Repeat telephone interview	Kappa coefficient
Sufficiently active during leisure* (%)	24.5	24.5	0.80
Inactive in four domains of physical activity** (%)	24.6	23.6	0.78
Watch television for extended periods*** (%)	33.6	34.6	0.53

* Subjects who perform physical exercise or sport of moderate intensity for at least 30 minutes per day, five or more days per week, or of vigorous intensity for at least 20 minutes per day, three or more days per week.
** Subjects that: 1) do not perform physical activity or sport at least one day per week; 2) "do not walk frequently" and "do not carry heavy weight frequently" at work; 3) do not commute from home to work on foot or by bicycle; e 4) are not responsible for "heavy cleaning" at home.
*** Watch television for at least three hours per day five or more days per week.

Table 2. Validity of indicators of physical activity and sedentariness obtained by telephone interview in relation to three 24-hour recalls. City of São Paulo, Southeastern Brazil, 2005.

Indicator	Telephone interview	24-hour recalls	Sensitivity of telephone interview	Specificity of telephone interview
Sufficiently active during leisure* (%)	26.1	21.6	50.0	80.5
Inactive in four domains of physical activity** (%)	22.5	19.8	59.1	86.5
Watch television for extended periods*** (%)	35.1	29.7	69.7	79.5

 * Subjects who perform physical exercise or sport of moderate intensity for at least 30 minutes per day, five or more days per week, or of vigorous intensity for at least 20 minutes per day, three or more days per week.
** Subjects that: 1) do not perform physical activity or sport at least one day per week; 2) "do not walk frequently" and "do not carry heavy weight frequently" at work; 3) do not commute from home to work on foot or by bicycle; e 4) are not responsible for "heavy cleaning" at home.
*** Watch television for at least three hours per day five or more days per week.

(median = 20 and zero minutes, respectively; p<0.001). Mean time per day spent on physical activity in the four domains studied (leisure, work, transportation to and from work, and heavy cleaning at home) was 27.5 minutes for subjects classified by the interview as "inactive in four domains" and 139.6 minutes for the remainder (median = zero and 60 min, respectively; p<0.001). Finally, mean time per day watching television was 209.1 minutes for subjects classified by the interview as "watching television for extended periods" and 122.7 minutes for the remaining subjects (median = 203 min and 120 min, respectively; p<0.001).

DISCUSSION

The present study shows that indicators of physical activity and sedentariness obtained through the VIGITEL system show good reproducibility, at both collective (identical or very similar frequencies of the three indicators evaluated in repeated interviews) and individual (kappa coefficients compatible with moderate or substantial agreement in individual classification of exposure) levels. Good reproducibility indicates that interviews are carried out in a standardized fashion, avoiding interpretation or answer induction. It also indicates that subjects are able to understand the questions and have no difficulty answering them, providing answers that do not vary with time. What is expected from a surveillance system such as VIGITEL is that it provide estimates of indicators that, in addition to accurate, are also reproducible. Good reproducibility ensures that temporal variations in these indicators reflect actual variations in behavior among the population, rather than indicator instability.[2,12]

Indicator validity, i.e., in the present scenario, the ability of indicators obtained by VIGITEL to provide results that are similar to those obtained by three 24-hour recalls, was also evaluated from both collective and individual perspectives. On the collective level, the telephone interview revealed a frequency of "inactive in four domains" quite similar to that obtained by the 24-hour recalls (22.5 and 19.8%, respectively). In the case of "sufficiently active during leisure," there was a slight overestimation by the telephone interview, which may indicate subjects' "desire" to be more physically active. The same is not true, however, for "watching television for extended periods," the frequency of which was also slightly higher in the telephone interview than in the recalls.

At the individual level, all three indicators evaluated showed good specificity. Reasonable sensitivity was achieved for "watching television for extended periods" and "inactive in four domains." The low sensitivity found for "sufficiently active during leisure" (50%) may be explained, at least partly, by the fact that the 24-hour recalls investigate only three days, compared to the seven-day reference period.

Still with regards to validity at the collective level, we found that the group of subjects classified as "sufficiently active in leisure" dedicate over threefold more time per day to physical exercise or sport than the remainder of subjects (32 minutes vs. 9 minutes). On the other hand, the group of subjects classified as "watching television for extended periods" spent on average 75% more time watching television than other subjects (209 minutes vs. 120 minutes).

We do not see important limitations in the design used for the reproducibility study, given that the major sources of intra-subject variation and of variation between interviewers were accounted for by repeating the same interview with a different interviewer. Likewise, the kappa coefficient, employed in the analysis of reproducibility of the telephone interview, is the most widely recommended measure for evaluating the reproducibility of instruments used for classifying individuals as exposed or unexposed to a given condition.[13]

Common limitations of validity studies include the use of insufficiently accurate reference methods and samples that are not representative of the population evaluated by the indicator.[13] Regarding the first of these issues, it would have been more appropriate to extend the recall period to seven days, a duration considered

as more adequate for characterizing physical activity patterns,[11] or to employ instruments that directly record physical activity, such as accelerometers.[15] These options, however, were discarded due to the risk of affecting the response rate or even of influencing the subject's usual pattern of physical activity. In any case, issues concerning the precision of the reference method are unlikely to lead to overestimation of the validity of the method under evaluation. Rather, such issues would likely lead to underestimation, which we believe may have been the case, as mentioned, for the sensitivity of the "sufficiently active during leisure" indicator.

As to the representativeness of the current subsample, the probabilistic selection of subjects ensures that our results are applicable to the performance of VIGITEL in the city of São Paulo, but not necessarily in the other cities in which the system is being implemented. In this regard, we believe it will be essential to carry out a similar study in at least one state capital for each of the five Brazilian Regions. In addition to probabilistic

selection, other strengths of the present analysis are calculation of sensitivity and specificity, a recommended procedure given the characteristics of the indicators being strudied,[13] and comparison of daily time dedicated to each of the various activities according to the inclusion or not of subjects in the exposed group for each indicator.

Even though restricted to developed countries, other studies of reproducibility and validity of physical activity indicators obtained by telephone interview have yielded similar results to those of the present study.[5,6,14]

In conclusion, the indicators of physical activity employed by the VIGITEL system appear to be reproducible and sufficiently accurate. The maintenance of the system in years to come will provide the country with a useful instrument for evaluating public policies aimed at promoting physical activity and controlling non-transmissible chronic diseases associated with sedentariness.

REFERENCES

1. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc*. 2000;32(9 suppl): S498-504. doi:10.1097/00005768-200009001-00009

2. Byers T. Nutrition monitoring and surveillance. In: Willet W. Nutritional epidemiology. 2. ed. New York: Oxford University Press; 1998. p. 347-55.

3. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74. doi:10.2307/2529310

4. Malta DC, Cezario AC, Moura L, Morais Neto OL, Silva Jr JB. Construção da vigilância e prevenção das doenças crônicas não trasmissíveis no contexto do sistema único de saúde. *Epidemiol Serv Saude*. 2006;15(3):47-65.

5. Marshall AL, Smith BJ, Bauman AE, Kaur S. Reliability and validity of a brief physical activity assessment for use by family doctors. *Br J Sports Med*. 2005;39(5):294-7. doi:10.1136/bjsm.2004.013771

6. Matthews CE, Ainsworth BE, Hanby C, Pate RR, Addy C, Freedson PS, et al. Development and testing of a short physical activity recall questionnaire. *Med Sci Sports Exerc*. 2005;37(6):986-94.

7. Menezes RX, Azevedo RS. Bioestatística não-paramétrica. In: Massad E, Menezes RX, Silveira PSP, Ortega NRS. Métodos quantitativos em medicina. São Paulo: Manole; 2004. p. inicial-final.

8. Monteiro CA, Moura EC, Jaime PC, Lucca A, Florindo AA, Figueiredo ICR, et al. Monitoramento de fatores de risco para doenças crônicas não transmissíveis

por meio de entrevistas telefônicas. *Rev Saude Publica*. 2005;39(1):47-57. doi:10.1590/S0034-89102005000100007

9. Monteiro CA, Moura EC, Jaime PC, Claro RM. Validade de indicadores do consumo de alimentos e bebidas obtidos por inquérito telefônico. *Rev Saude Publica* 2008; 42(4):582-9.

10. Pereira MA, FitzerGerald SJ, Gregg EW, Joswiak ML, Ryan WJ, Suminski RR, et al. A collection of physical activity questionnaires for health-related research. *Med Sci Sports Exerc*. 1997;29(6 suppl):S1-205.

11. Sallis JE. Seven-day physical activity recall (1985). *Med Sci Sports Exerc*. 1997;29(6 suppl):S89-103.

12. Stein AD, Courval JM, Lederman RI, Shea S. Reproducibility of responses to telephone interviews: demographic predictors of discordance in risk factor status. *Am J Epidemiol*. 1995;141(11):1097-106.

13. Szklo M, Javier Nieto, F. Epidemiology: beyond the basics. Maryland: Aspen; 2004.

14. Timperio A, Salmon JO, Rosenberg M, Bull FC. Do logbooks influence recall of physical activity in validation studies? *Med Sci Sports Exerc*. 2004;36(7):1181-6. doi:10.1249/01. MSS.0000132268.74992.D8

15. Troiano RP. A timely meeting: objective measurement of physical activity. *Med Sci Sports Exerc*. 2006;37(11 supl):S487-9.

16. World Health Organization. Reducing risks, promoting healthy life. Geneva; 2002. (The World Health Report, 2002).