Multiple imputation to deal with missing objectively-measured physical activity data: findings from two cohorts

Imputação múltipla para lidar com dados faltantes em acelerometria: resultados de duas coortes

AUTHOR'S

Rafaela Costa Martins[1]
Bruna Gonçalves C. da Silva[1]
Cauane Blumenberg[1]
Luiza Isnardi Ricardo[1]
Shana Ginar da Silva[1,2]
João Pedro Ribeiro[3]
Alicia Matijasevich[1,4]
Ana Maria Baptista Menezes[1]
Helen Gonçalves[1]
Fernando César Wehrmeister[1]
Iná dos Santos[1]
Inácio Crochemore-Silva[1,5]
Aluísio J. D. Barros[1]

1 Federal University of Pelotas, Post-graduate Program in Epidemiology, Pelotas, Rio Grande do Sul, Brazil.

2 Federal University of South Frontier, Faculty of Medicine, Passo Fundo, Rio Grande do Sul, Brazil.

3 Federal University of Pelotas, Faculty of Physical Education, Pelotas, Rio Grande do Sul, Brazil.

4 São Paulo University, Faculty of Medicine, Department of Preventive Medicine, São Paulo, São Paulo, Brazil.

5 Federal University of Pelotas, Post-graduate Program in Physical Education, Pelotas, Rio Grande do Sul, Brazil

CORRESPONDING

Rafaela Costa Martins
rafamartins1@gmail.com
Rua Marechal Deodoro, 1160 – 3º piso, Pelotas, Rio Grande do Sul, Brasil.
ZIP CODE: 96020-220.

DOI

10.12820/rbafs.26e0209

ABSTRACT

The objective of this article was to describe patterns of losses of information regarding accelerometer data and to assess the use of multiple imputation to generate physical activity estimates for individuals without accelerometry data. Two birth cohort studies from Pelotas (Brazil) with participants aged 22 and 11-years old assessed objectively measured physical activity differences between complete and imputed cases. Mean values of overall physical activity for complete cases (n_{1993} = 2,985 and n_{2004} = 3,348) and for complete cases plus imputed cases (n_{1993} = 760 and n_{2004} = 79) were described according to predictors. Male individuals, participants with black skin color, and less schooled individuals presented higher averages of overall physical activity than their counterparts. Almost all imputed estimates were comparable to the complete cases, and the highest difference found was 0.7 *mg* for the first quintile of socioeconomic status of the 1993 birth cohort. Multiple imputation is a positive technique to deal with missing data from objectively measured physical activity. It provides a set of relevant variables to be used in order to efficiently predict accelerometer data.

Keywords: Accelerometry; Physical activity; Statistics.

RESUMO

O objetivo desse artigo foi descrever os padrões de perda de informação em dados de acelerometria, além de avaliar o processo de imputação múltipla para estimar o nível de atividade física para indivíduos sem dados de acelerometria. Participantes de duas coortes de nascimentos de Pelotas (Brasil) com 22 e 11 anos participaram do estudo e diferenças entre casos completos e imputados foram avaliadas. A média geral de atividade física para os casos completos (n_{1993} = 2.985 e n_{2004} = 3.348) e para casos completos mais imputados (n_{1993} = 760 e n_{2004} = 79) foi descrita de acordo com os preditores. Indivíduos do sexo masculino, de cor da pele preta e com menor escolaridade apresentaram maiores médias de atividade física geral. Quase todas as estimativas imputadas foram comparáveis com os valores de casos completos, e a maior diferença encontrada foi 0,7 mg para o primeiro quintil de renda na coorte de 1993. Imputação múltipla é uma boa técnica para lidar com dados faltantes de atividade física medida por acelerometria. Essa técnica fornece um gama relevante de variáveis para serem usadas a fim de predizer valores de acelerometria eficientemente.

Palavras-chave: *Acelerometria; Atividade física; Estatística.*

Introduction

Accelerometers are an objective measure of physical activity, quantifying duration, frequency and intensity of physical activity through acceleration signals, movement patterns and their magnitude[1]. The devices currently available are able to store a large amount of data, to consider a wide spectrum of cut-off points for different physical activity intensities, to perform movement pattern recognition and more detailed analyses using raw data[2]. Compared to self-reported methods, accelerometers provide more accurate physiological and mechanical parameters in order to estimate physical activity levels[3]. Due to these advantages, the number of large-scale studies evaluating physical activity by accelerometry has increased significantly[4,5].

However, a few challenges arise when using accelerometers to measure physical activity[6]. For instance, participant's compliance is recognized as a major issue. Large-scale surveys, such as National Health and Nutrition Examination Survey (NHANES), reported that only 25% of the participants provided 7 days of data as proposed by their study protocol[2]. This is mostly attributed to the discomfort or inconvenience of wearing the accelerometer and forgetting to wear the device after taking it off for shower or sleep periods[2,7]. To overcome the compliance challenge, NHANES and International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE) surveys started to adopt strategies such as a 24-hour protocol (not taking the device off overnight) and placing the accelerometer on the wrist, which seems to provide more comfort compared to other sites (e.g. waist, ankle)[8-10].

In the Pelotas (Brazil) Birth Cohorts, the biggest series of accelerometer data collection in low- and middle-income countries so far, objectively measured PA started in all follow-ups in 2010[11,12]. The strategies adopted such as the wrist-attachment and the 24-hour protocol along with a careful logistic of data collection have been provided around 95% of wear time during the period of use requested[11]. However, in spite of strategies to increase compliance in the accelerometer use, losses of follow-up generate missing data that may introduce important bias[13,14].

It is important to note that some participant characteristics might be associated with accelerometer compliance. Lee and colleagues[15] found that older participants (mean age 47.6 years), those with fulltime job, with tertiary education, non-smokers and with high self-reported health wore the accelerometer for significantly more hours than their counterparts. Some strategies could be beneficial to minimize data loss, such as targeted instructions for participants, reminders, diaries and others[16]. On the other hand, even using these methods it is possible that some invalid data is generated, hence it is important to find alternatives to deal with accelerometer missing data and, if needed, interpret the consequences in terms of future associations according to the the distribution of losses. After data collection, multiple imputation is among the several strategies that can be applied to minimize data loss. Multiple imputation uses observed data to help in predicting the outcome, and it is applicable when the data are randomly missing[17]. Regarding accelerometry data, this approach is generally more precise when a greater number of good observed predictors are included, and the results seem to be unbiased when comparing imputed and observed data[18].

Little is known about the distribution of the losses and impact of multiple imputation on the physical activity estimates objectively measured. In this sense, our objective was to describe patterns of losses of information regarding accelerometer data and to demonstrate a real situation of the use of multiple imputation to generate physical activity estimates for individuals without accelerometry data.

Methods

In the calendar years of 1993 and 2004, from January 1st until December 31st, all hospital-born children whose mothers were living in the urban area of Pelotas, a Southern Brazilian city, were eligible to participate in the 1993 and 2004 Pelotas Birth Cohorts. In 1993, out of the 5,265 live births occurring in Pelotas, 5,249 mothers (99.7%) agreed to participate and were interviewed in the hospital soon after delivery composing the original sample of the 1993 birth cohort[19]. In 2004, out of the 4,263 births, 4,231 mothers (99.2%) agreed to take part into the study and composed the original sample of the 2004 birth cohort[20]. Our study focuses on the follow-ups conducted in 2015, when the participants of the 1993 and 2004 Pelotas birth cohorts were 22 and 11 years old, respectively[19,20].

In both cohorts, during the follow-up visits, participants were invited and instructed to wear an Actigraph accelerometer (models wGT3X-BT, wGT3X or the ActiSleep) on the wrist of their non-dominant hand for seven consecutive days. A 24-hour protocol was established, which means that participants were asked

to remain with the device during the full data collection period. We used the ActiLife 6.13.2 software to program the accelerometers. The devices were set in a frequency of 60Hz for data collection. For analysis purpose, data were summarized in five seconds epoch. To be considered a valid case (or a complete case) the participant had to use the device for at least three full days in the 1993 birth cohort, and for four full days in the 2004 birth cohort. Accelerometer data reduction was performed using the GGIR R-package[21], which verifies the sensor calibration error using local gravity as a reference and detects sustained abnormally high values and non-wear periods. Accelerometer data was considered valid when calibration error was lower than 0.02*g* and invalid when detection of sustained abnormally high values were compatible with non-human movement and non-wear detection. At the end, files were considered appropriate for analyses if there was valid data for every 15-min period in at least a 24-h cycle (even when scattered over multiple days). Furthermore, the vector magnitude of activity-related acceleration is calculated using the Euclidian Norm Minus One (ENMO = $\sum \left\| \sqrt{x^2 + y^2 + z^2} - 1g \right|$). Overall physical activity, a total volume of movement, was estimated based on daily means of acceleration in *mg* (gravitational equivalent).

We imputed the missing values of overall physical activity for participants that refused to wear the accelerometer, for those interviewed by telephone that did not wear the accelerometer (e.g. individuals not living in Pelotas), and for those who wore the accelerometer but generated invalid data (less than 3 complete measurement days). We did not impute information for those who did not attend the eligibility criteria for wearing the device (e.g. pregnant women, people with physical or cognitive problems) as any physical activity estimates would be biased. We used multiple imputation (M = 60) with chained equations to impute missing data. The following predictors were used to impute overall physical activity data for the 1993 Pelotas birth cohort: number of people living in the household, schooling (in years of study), monthly family income (in *Brazilian Reais – R\$*), asset index, body mass index (in kg/m²), maternal education (in years of study), alcohol consumption frequency (one or less times per week, two or more times per week), current smoking (yes, no), skin color (black, brown, white, other), and sex (female, male). The same predictors were considered in the 2004 Pelotas birth cohort, except for the participant's frequency of alcohol

consumption and smoking. These predictors were not included because they were not assessed by the 2004 Pelotas birth cohort. In both cohorts, the only predictor with complete information for the whole sample was the sex, the remaining were also imputed.

All variables, except body mass index, skin color and sex, were self-reported by the participants during the 22- and 11-year follow-ups of the 1993 and 2004 birth cohorts, respectively. The body mass index (BMI) was assessed by air-displacement plethysmography method (Bod Pod®), categorized into underweight/normal (<25.0 kg/m²), overweight (25.0-29.9 kg/m²) and obese (≥30.0 kg/m²). Sex and skin color of the participants were observed by the interviewers in previous follow-ups. We calculated the socioeconomic status using a principal component analysis based on self-reported possession of an extensive list of assets by the participants.

Statistical analyses were conducted using Stata 15.1, including multiple imputation (StataCorp. 2017. *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC). Characteristics of participants with valid data and of those with missing data for accelerometry data were described using descriptive statistics. Fischer's exact test was used to compare these samples. Mean values of overall physical activity for complete cases and for complete cases plus imputed cases were described according to the predictors. Statistical significance was assessed using 95% confidence intervals.

Both the 22 years (1993 Cohort) and 11 years (2004 Cohort) follow-up projects were sent and approved by the Federal University of Pelotas Ethics Committee. The protocols numbers for the 22- and 11-year visits were 1.250.366 and 015/15, respectively. The cohort participants, or their caregivers, signed the informed consent term prior to participation in each study.

Results

A total of 3,810 and 3,565 participants were interviewed in the 22-years follow-up of the 1993 Pelotas Birth Cohort and in the 11-years follow-up of the 2004 Pelotas Birth Cohort, respectively (Figure 1). For the 1993 birth cohort, 78.3% had valid accelerometry data, and physical activity estimates were imputed for 760 individuals. In the 2004 birth cohort, 93.9% of the interviewed participants had valid accelerometry data, and the estimates were imputed for only 79 individuals. As for the predictors, the variables with highest number of missing information were maternal education at the 1993 (n = 447) cohort and BMI at the 2004 cohort (n = 76).

Tables 1 and 2 show that skin color, maternal education and BMI are variables that differ between complete cases and losses for accelerometry data in both cohorts. Those with mothers that studied 9 to 11 years

Table 1 – Sociodemographic characteristics of the sample considered complete cases and losses for accelerometry data in Pelotas.

	1993 Birth Cohort			2004 Birth Cohort		
	Complete cases N (%)	Losses N (%)	p-value*	Complete cases N (%)	Losses N (%)	p-value*
Total	2,985 (79.7)	760 (20.3)		3,348 (97.7)	79 (2.3)	
Gender			0.060			0.425
Female	1,549 (78.7)	419 (21.3)		1,626 (97.8)	37 (2.2)	
Male	1,436 (80.8)	341 (19.2)		1,722 (97.6)	42 (2.4)	
Skin color			< 0.001			0.087
White	1,772 (79.4)	461 (20.6)		2,244 (97.3)	62 (2.7)	
Brown	501 (80.7)	120 (19.3)		489 (99.0)	5 (1.0)	
Black	472 (89.4)	56 (10.6)		424 (98.4)	7 (1.6)	
Other	116 (86.6)	18 (13.4)		179 (97.7)	5 (2.3)	
Schooling			0.589			0.663
0-4	82 (78.8)	22 (21.2)		684 (97.6)	17 (2.4)	
5-8	801 (81.2)	186 (18.8)		2,635 (97.8)	57 (2.2)	
9-11	1,217 (79.4)	315 (20.6)		---	---	
12+	885 (78.9)	237 (21.1)		---	---	
Socioeconomic status (quintiles)			0.013			0.380
1st (poorest)	588 (78.7)	159 (21.3)		659 (97.3)	18 (2.7)	
2nd	573 (76.6)	175 (23.4)		666 (98.4)	11 (1.6)	
3rd	604 (80.7)	144 (19.3)		656 (97.9)	14 (2.1)	
4th	592 (79.1)	156 (20.9)		670 (98.1)	13 (1.9)	
5th (richest)	625 (83.6)	123 (16.4)		658 (96.9)	21 (3.1)	
Maternal education			< 0.001			0.007
0-4	607 (83.6)	119 (16.4)		472 (97.7)	11 (2.3)	
5-8	1,098 (81.9)	243 (18.1)		1,151 (98.0)	24 (2.0)	
9-11	647 (79.6)	166 (20.4)		1,104 (98.5)	17 (1.5)	
12+	633 (73.2)	232 (26.8)		621 (95.7)	27 (2.3)	
No. of people living in the household			< 0.001			0.399
0-3	1,665 (77.6)	481 (22.4)		2,185 (97.8)	50 (2.2)	
4+	1,320 (82.5)	279 (17.5)		1,163 (97.6)	29 (2.4)	

Absolute frequencies might not sum up to totals due to missing information; *p-value for Fischer's exact testv

Table 2 – Anthropometric and behavioral characteristics of the sample considered complete cases and losses for accelerometry data in Pelotas.

	1993 Birth Cohort			2004 Birth Cohort		
	Complete cases N (%)	Losses N (%)	p-value *	Complete cases N (%)	Losses N (%)	p-value *
Total	2,985 (79.7)	760 (20.3)		3,348 (97.7)	79 (2.3)	
BMI			< 0.001			0.342
Underweight/Normal	1,682 (83.1)	341 (16.9)		2,127 (97.7)	49 (2.3)	
Overweight	800 (83.5)	158 (16.5)		775 (97.8)	17 (2.2)	
Obesity	503 (65.8)	261 (34.2)		379 (99.0)	4 (1.0)	
Current Smoking			0.032			
Yes	485 (77.0)	145 (23.0)		---	---	
No	2,499 (80.3)	611 (19.7)		---	---	
Alcohol						
One or less times/week	2,433 (80.3)	598 (19.7)	0.013	---	---	
2 or more times/week	382 (75.8)	122 (24.2)		---	---	

Absolute frequencies might not sum up to totals due to missing information; *p-value for Fischer's exact test. BMI = body mass index

were underrepresented in the losses compared to complete cases. Considering only the oldest cohort (1993), white individuals, households with less people, from the second quintile of income, obese individuals, current smokers, and with more alcohol consumption were overrepresented comparing losses and complete cases.

Tables 3 and 4 compare the mean overall physical activity of participants with complete accelerometry data to the complete cases plus imputed cases. Male individuals, participants with black skin color, and less schooled individuals presented higher averages of overall physical activity than their counterparts in both cohorts. After multiple imputation, it is possible to note that almost all imputed estimates were comparable to the complete cases.

We compared mean overall physical activity between complete cases and complete + imputed cases according to some sociodemographic characteristics (Table 3).

The highest difference found in the 1993 birth cohort was 0.7 *mg* for the first quintile of socioeconomic status. For the 2004 birth cohort, the highest difference was 0.2 *mg* for individuals from highly educated mothers. Regarding anthropometric and behavioral characteristics (Table 4), the highest difference of mean overall physical activity was 0.4 *mg* for obese and smoker individuals from the 1993 birth cohort. Similarly, in the 2004 birth cohort the highest difference was also found comparing obese individuals, reaching 0.5 *mg*.

Discussion

This study investigated sociodemographic, anthropometric and behavioral factors associated with participant's noncompliance in obtaining accelerometer data and analyzed the effect of data imputation in physical activity estimates objectively measured using two large samples from the Pelotas (Brazil) birth cohorts. Our

Table 3 – Mean overall physical activity expressed in mg for complete cases and complete cases plus imputed according to sociodemographic characteristics in Pelotas.

	1993 Birth cohort			2004 Birth cohort		
	Complete cases	Imputed cases	Complete cases + imputed	Complete cases	Imputed cases	Complete cases + imputed
	Mean (95%CI)	Mean (95%CI)	Mean (95%CI)	Mean (95%CI)	Mean (95%CI)	Mean (95%CI)
Gender						
Female	35.3 (33.0; 37.6)	35.0 (28.3; 41.7)	35.2 (32.9; 37.6)	60.5 (60.1; 60.8)	59.8 (61.8; 77.1)	60.1 (59.4; 60.7)
Male	41.0 (38.4; 43.6)	40.2 (32.6; 47.7)	40.9 (38.3; 43.4)	69.0 (68.5; 69.5)	69.4 (61.8; 77.1)	69.9 (69.0; 70.8)
Skin color						
White	36.3 (34.4; 38.2)	36.1 (29.9; 42.4)	36.3 (34.3; 38.3)	61.8 (61.5; 62.2)	63.7 (57.2; 70.1)	63.4 (62.7; 64.1)
Brown	40.8 (35.4; 46.1)	40.4 (28.6; 52.3)	40.6 (35.8; 45.4)	72.0 (71.1; 72.8)	70.3 (46.5; 94.2)	68.1 (66.4; 69.7)
Black	42.0 (36.3; 47.7)	40.7 (23.9; 57.5)	41.7 (36.4; 47.1)	73.7 (72.6; 74.8)	72.9 (53.5; 92.3)	70.2 (68.4; 71.9)
Other	37.6 (33.0; 42.1)	36.6 (6.9; 66.2)	37.5 (31.1; 43.8)	68.1 (66.5; 69.6)	63.8 (43.9; 83.7)	66.5 (64.0; 69.0)
Schooling						
0-4	44.9 (35.9; 53.9)	46.4 (17.1; 75.8)	45.2 (35.8; 54.6)	70.9 (69.4; 72.3)	71.0 (59.7; 83.4)	70.8 (69.4; 72.2)
5-8	43.7 (39.3; 48.0)	41.9 (31.1; 52.7)	43.3 (39.2; 47.4)	63.6 (63.0; 64.2)	62.9 (56.4; 69.3)	63.6 (63.0; 64.2)
9-11	38.0 (35.4; 40.5)	36.0 (28.4; 43.7)	37.6 (35.0; 40.1)	---	---	---
12+	32.5 (30.2; 34.8)	34.5 (25.5; 43.6)	32.9 (30.3; 35.6)	---	---	---
Socioeconomic status (quintiles)						
1st (poorest)	39.4 (36.2; 42.5)	36.4 (25.2; 47.6)	38.8 (35.3; 42.2)	70.6 (69.2; 72.1)	69.1 (57.7; 80.5)	70.6 (69.1; 72.0)
2nd	35.7 (33.8; 37.5)	36.7 (25.7; 47.6)	35.9 (33.0; 38.8)	65.7 (64.4; 67.0)	68.4 (54.2; 82.7)	65.7 (64.4; 67.0)
3rd	35.5 (33.6; 37.3)	36.3 (24.3; 48.3)	35.6 (32.9; 38.4)	64.5 (63.2; 65.9)	63.0 (51.1; 74.8)	64.5 (63.2; 65.8)
4th	38.2 (34.7; 41.8)	37.7 (27.1; 48.4)	38.1 (34.5; 41.7)	62.5 (61.2; 63.7)	67.7 (53.8; 81.5)	62.6 (61.3; 63.8)
5th (richest)	41.4 (34.8; 47.9)	40.1 (27.7; 52.4)	41.1 (35.3; 47.0)	62.4 (61.2; 63.6)	59.0 (48.5; 69.5)	62.3 (61.1; 63.4)
Maternal education						
0-4	38.4 (36.0; 40.9)	41.3 (29.4; 53.2)	39.0 (36.0; 42.0)	68.3 (66.7; 69.8)	69.5 (56.1; 82.9)	68.3 (66.7; 69.8)
5-8	39.8 (36.6; 43.0)	37.5 (29.0; 46.0)	39.4 (36.3; 42.4)	67.6 (66.5; 68.6)	68.0 (58.3; 77.7)	67.6 (66.5; 68.6)
9-11	36.7 (32.5; 41.0)	35.5 (26.3; 44.8)	36.7 (32.8; 40.6)	62.5 (61.6; 63.5)	65.9 (53.9; 78.0)	62.6 (61.7; 63.5)
12+	36.0 (32.5; 39.5)	34.4 (21.0; 47.7)	33.6 (29.2; 38.0)	62.8 (61.5; 64.1)	59.4 (50.2; 68.7)	62.7 (61.4; 64.0)
No. of people living in the household						
0-3	37.5 (35.3; 39.7)	36.6 (29.8; 43.4)	37.3 (35.0; 39.6)	63.9 (63.2; 64.6)	63.3 (56.2; 70.4)	63.9 (63.2; 64.6)
4+	38.8 (35.9; 41.6)	38.6 (29.9; 47.2)	38.7 (35.9; 41.5)	67.4 (66.4; 68.4)	67.8 (58.8; 76.7)	67.4 (66.4; 68.5)

Table 4 – Mean overall physical activity expressed in mg for complete cases and complete cases + imputed according to anthropometric and behavioral characteristics in Pelotas.

	1993 Birth cohort			2004 Birth cohort		
	Complete cases	Imputed cases	Complete cases + imputed	Complete cases	Imputed cases	Complete cases + imputed
	Mean (95%CI)	Mean (95%CI)	Mean (95%CI)	Mean (95%CI)	Mean (95%CI)	Mean (95%CI)
BMI						
Underweight/Normal	38.6 (36.4; 40.9)	37.6 (30.9; 44.3)	38.4 (36.1; 40.6)	68.1 (67.7; 68.5)	67.2 (60.7; 73.7)	67.7 (67.0; 68.5)
Overweight	38.1 (35.1; 41.0)	37.2 (27.3; 47.1)	37.8 (34.7; 40.9)	59.9 (59.3; 60.5)	60.9 (49.8; 71.9)	61.3 (60.2; 62.3)
Obesity	36.1 (30.7; 41.5)	36.6 (24.8; 48.4)	36.5 (31.4; 41.6)	57.4 (56.6; 58.2)	56.1 (33.9; 78.3)	58.2 (56.8; 59.7)
Current Smoking						
Yes	40.2 (35.1; 45.2)	38.4 (26.6; 50.2)	39.8 (35.0; 44.5)	---	---	---
No	37.6 (35.8; 39.5)	37.1 (31.4; 42.8)	37.5 (35.7; 39.4)	---	---	---
Alcohol						
One or less times/week	38.0 (36.2; 39.9)	37.0 (31.2; 42.8)	37.6 (35.8; 39.5)	---	---	---
2 or more times/week	40.5 (34.0; 46.9)	38.9 (25.3; 52.5)	39.8 (34.1; 45.5)	---	---	---

BMI = body mass index
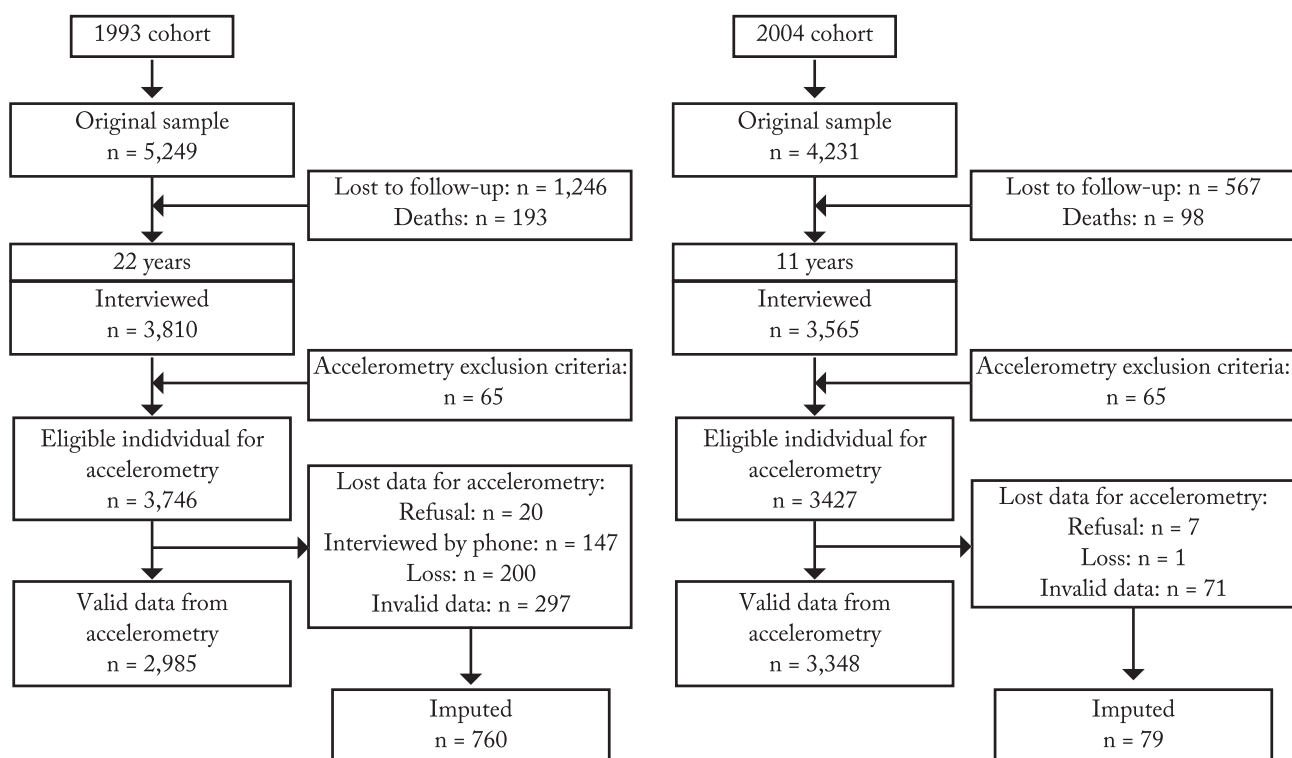
Figure 1 – Flowchart of 1993 and 2004 Pelotas birth cohorts

findings showed that individuals with mothers that studied 9 to 11 years had lower probability to present missing accelerometry data. In the 1993 birth cohort, white individuals, households with less people, from the second quintile of income, obese individuals, current smokers, and with more alcohol consumption were overrepresented comparing losses and complete cases. There were no statistically significant differences in the overall physical activity means between the imputed and complete case estimates according to sociodemographic, anthropometric and behavioral characteristics.

Good science stems from valid measurement and its processing, analysis, and interpretation in an appropriate way. Objective methods to measure physical activity, such as accelerometers, have become a more popular tool over the years, in clinical and research setting, given the high degree of validity to assess patterns of physical activity in free-living conditions[22]. Neverthe-

less, it is challenging to use this device in the fieldwork as the accelerometer is prone to problems with missing data with important implications that can affect data quality, precision, and, consequently, sample size[4,8,23].

The literature recommends including only valid days with enough wearing time[18,24,25]. Non-wear time data is typically deleted prior to estimating overall physical activity. However, these data reduction criteria could lead to a significant decrease in sample size and if summary statistics are computed using the observed data only, these statistics have the potential to be biased and introduce unwanted variance in subsequent estimations[14,22].

In our samples, we did not find differences between the imputed, complete case and imputed + complete case estimates in most of the cases. The high accelerometer compliance (>75% in both cohorts), might explain the similarity between the complete cases and the imputed estimates. However, as missing data is a common problem in studies using accelerometers, assessing the impact of this technique in a situation where there are more opportunities to impute could be insightful. Also, it is important to learn the pattern of the missing data, since its distribution can affect mainly prevalence studies, as well as in association studies where the effects may be diluted or even change directions.

To help increase the data collection efficiency of prospective accelerometry studies, it is very important to investigate the factors associated with participants' compliance. In our study, a higher compliance was observed in participants with non-white skin color, born to mothers with low education, and living in households with more than four people. Additionally, participants with healthy lifestyle (who consumed alcohol one time per week or less and with normal BMI) also had higher compliance. The result that healthier individuals have higher compliance or response rates has been stated elsewhere[15] and highlights potential bias being introduced in many prospective analyses due to differential losses over time. It is challenging to design strategies to increase compliance among groups with high rates of accelerometer missing values. Thus, multiple imputation arises as a good technique to deal with accelerometry missing data during data analysis.

To the best of our knowledge, this is the first study to date to determine the sociodemographic, anthropometric and behavioral characteristics associated with noncompliance and to describe the effect of imputation in estimates for accelerometry data. Our results are inserted in the context of two large birth cohorts from a middle-income country and show that multiple imputation is a good alternative to minimize consequences of missing data and to obtain more valid and comparable data. Multiple imputation has been shown to give unbiased estimates and should be used to replace missing physical activity accelerometry data improving the validity and interpretation of studies. The large amount of variables used as predictors to impute overall physical activity, the high response rate that confirms the good quality of the estimates and the statistical techniques used are strengths of our study. It is known that, in general, the greater the number of predictors and the higher the correlations among variables, the more precise the estimates of missing values will be. Catellier and colleagues[18] mentioned that this approach is generally more precise when a greater number of good observed predictors are included, and the results seem to be unbiased when comparing imputed and observed data.

Although multiple imputation has emerged as a strategy to deal with missing data, some limitations should be noted. We could not formally diagnose that the accelerometry data from both birth cohorts were missing at random. Some sociodemographic and behavioral variables were associated with missing information for physical activity. However, comparing the characteristics of the complete cases to the losses it is possible to note that the missingness can be explained by the known predictors for physical activity. Taking this into consideration, and the low percentage of missingness, we assume that this hypothesis is satisfied in our study. We assessed only overall physical activity in our analysis, caution should be taken when evaluating other types and intensities of physical activity. It is important to note that our results are inserted in a context with low rates of missing values. Studies with higher rates should assess the pattern of missingness in order to guarantee that it occurs at random and check the quality of predictors prior to the multiple imputation.

Even though some differences in the average of overall physical activity between complete and imputed cases, multiple imputation was a good strategy to deal with missing data of objectively measured physical activity. It is important to understand the characteristics of the sample associated with missing values in order to guarantee the high quality of the imputed data. In that sense, multiple imputation methodology is a positive technique to deal with missing data in studies with accelerometers.

Martins et al. Rev Bras Ativ Fís Saúde. 2021;26:e0209

Imputação múltipla em dados de acelerometria

Conflict of interest

The authors declare no conflict of interest.

Author's contributions

Martins RC, Silva BGC, Blumenberg C and Barros AJD participated in the statistical analysis, writing of the article and review of the final version of the article. Ricardo LI, Ribeiro JP and Silva SG participated in the writing of the manuscript and review of the final version of the article. Crochemore-Silva I, Santos I, Wehrmeister FC, Menezes AMB, Matijasevich A and Gonçalves H participated in data collection, coordination of field work, and revision of the final version of the article.

References

1. Chen KY, Bassett DR. The technology of accelerometry-based activity monitors: current and future. Med Sci Sports Exerc. 2005;37(11 Suppl):S490-500.

2. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity re-search. Br J Sports Med. 2014 Jul;48(13):1019–23.

3. Ainsworth B, Cahalin L, Buman M, Ross R. The Current State of Physical Activity Assessment Tools. Prog Cardiovasc Dis. 2015;57(4):387–95.

4. Rowlands AV. Accelerometer assessment of physical activity in children: an update. Pediatr Exerc Sci. 2007;19(3):252–66.

5. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. Med Sci Sports Exerc. 2005;37(11 Suppl):S582-588.

6. Yue Xu S, Nelson S, Kerr J, Godbole S, Patterson R, Merchant G, et al. Statistical approaches to account for missing values in accelerometer data: Applications to modeling physical activity. Stat Methods Med Res. 2018;27(4):1168–86.

7. Belton S, O'Brien W, Wickel EE, Issartel J. Patterns of noncompliance in adolescent field-based accelerometer research. J Phys Act Health. 2013;10(8):1181–5.

8. Fairclough SJ, Noonan R, Rowlands AV, Van Hees V, Knowles Z, Boddy LM. Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers: Med Sci Sports Exerc. 2016;48(2):245–53.

9. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc. 2005;37(11 Suppl):S531-543.

10. Tudor-Locke C, Barreira TV, Schuna Jr JM, Mire EF, Chaput J-P, Fogelholm M, et al. Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the International Study of Childhood Obesi-ty, Lifestyle and the Environment (ISCOLE). Int J Behav Nutr Phys Act. 2015;12(1).

11. da Silva IC, van Hees VT, Ramires VV, Knuth AG, Bielemann RM, Ekelund U, et al. Physical activity levels in three Brazilian birth cohorts as assessed with raw triaxial wrist accelerometry. Int J Epidemiol. 2014;43(6):1959–68.

12. Knuth AG, Assunção MCF, Gonçalves H, Menezes AMB, Santos IS, Barros AJD, et al. Descrição metodoló-gica do uso de acelerometria para mensurar a prática de atividade física nas coortes de nascimentos de Pelotas, Rio Grande do Sul, Brasil, 1993 e 2004. Cad Saúde Pública. 2013;29(3):557–65.

13. Herrmann SD, Barreira TV, Kang M, Ainsworth BE. Impact of accelerometer wear time on physical activity data: a NHANES semisimulation data approach. Br J Sports Med. 2014;48(3):278–82.

14. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Med Sci Sports Exerc. 2012;44:S68–76.

15. Lee PH, Macfarlane DJ, Lam TH. Factors associated with participant compliance in studies using accelerom-eters. Gait Posture. 2013;38(4):912–7.

16. Sirard JR, Slater ME. Compliance with wearing physical activity accelerometers in high school students. J Phys Act Health. 2009;6 Suppl 1:S148-55.

17. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393–b2393.

18. Catellier DJ, Hannan PJ, Murray DM, Addy CL, Conway TL, Yang S, et al. Imputation of missing data when measuring physical activity by accelerometry: Med Sci Sports Exerc. 2005;37(Supplement):S555–62.

19. Gonçalves H, Wehrmeister FC, Assunção MCF, Tovo-Rodrigues L, Oliveira IO de, Murray J, et al. Cohort Profile Update: The 1993 Pelotas (Brazil) Birth Cohort follow-up at 22 years. Int J Epidemiol. 2018;47(5):1389–1390e.

20. Santos IS, Barros AJ, Matijasevich A, Zanini R, Cesar MAC, Camargo-Figuera FA, et al. Cohort Profile Up-date: 2004 Pelotas (Brazil) Birth Cohort Study. Body composition, mental health and genetic assessment at the 6 years follow-up. Int J Epidemiol. 2014;43(5):1437–1437f.

21. van Hees VT, Gorzelniak L, Dean León EC, Eder M, Pias M, Taherian S, et al. Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity. PLoS ONE. 2013;8(4):e61691.

22. Sievänen H, Kujala UM. Accelerometry-Simple, but challenging. Scand J Med Sci Sports. 2017;27(6):574–8.

23. Corder K, Brage S, Ekelund U. Accelerometers and pedometers: methodology and clinical application: Curr Opin Clin Nutr Metab Care. 2007;10(5):597–603.

24. Metzger JS, Catellier DJ, Evenson KR, Treuth MS, Rosamond WD, Siega-Riz AM. Patterns of Objectively Measured Physical Activity in the United States: Med Sci Sports Exerc. 2008;40(4):630–8.

25. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, Mcdowell M. Physical Activity in the United States Measured by Accelerometer: Med Sci Sports Exerc. 2008;40(1):181–8.