

Original article

Medical record review to recover missing data in a Portuguese birth cohort: agreement with self-reported data collected by questionnaire and inter-rater variability

Elisabete Alves*, Nuno Lunet, Sofia Correia, Vítor Morais, Ana Azevedo, Henrique Barros

Department of Hygiene and Epidemiology, University of Porto Medical School, Porto, Portugal; Institute of Public Health – University of Porto (ISPUP), Porto, Portugal

ARTICLE INFO

Article history:

Received 14 June 2010

Accepted 13 October 2010

Available online 3 April 2011

Keywords:

Epidemiologic methods

Medical records

Interobserver variability

Anthropometry

Cohort studies

ABSTRACT

Objectives: To assess the yield of medical record review to recover missing data originally collected by questionnaire, to analyze the agreement between these two data sources and to determine interobserver variability in clinical record review.

Methods: We analyzed data from a birth cohort of 8,127 women who were consecutively recruited after giving birth from 2005–2006. Recruitment was conducted at all public maternity units of Porto, Portugal. We reviewed the medical records of 3,657 women with missing data in the baseline questionnaire and assessed agreement between these two sources by using information from participants with data from both sources. Interobserver variability was assessed by using 400 randomly selected clinical records.

Results: Data on pregnancy complications and maternal anthropometric parameters were successfully recovered. Agreement between the questionnaire and records in family history data was fair, particularly for cardiovascular disease [$k=0.27$; 95% confidence interval (95%CI): 0.23–0.32]. The highest agreement was observed for personal history of diabetes ($k=0.82$; 95%CI 0.70–0.93), while agreement for hypertension was moderate ($k=0.60$; 95%CI 0.50–0.69). Discrepancies in prepregnancy body mass index classes were observed in 10.3% women. Data were highly consistent between the two reviewers, with the highest agreement found for gestational diabetes ($k=1.00$) and birth weight (99.5% concordance).

Conclusion: Data from the medical records and questionnaire were concordant with regard to pregnancy and well-known risk factors. The low interobserver variability did not threaten the precision of our data.

© 2010 SESPAS. Published by Elsevier España, S.L. All rights reserved.

Revisión de registros médicos para recuperar datos incompletos en una cohorte de nacimiento portuguesa: concordancia con datos recogidos por cuestionario y variabilidad interobservador

RESUMEN

Objetivo: Evaluar el rendimiento de la revisión de registros médicos para completar datos originalmente recogidos por cuestionario, y analizar la concordancia entre ambas fuentes de datos y la variabilidad interobservador en la revisión de registros médicos.

Métodos: Cohorte de nacimiento con 8.127 mujeres reclutadas de forma consecutiva después del parto en todas las maternidades públicas de Porto, Portugal (2005–2006). Se revisaron los registros médicos de 3.657 mujeres con datos incompletos en el cuestionario inicial, y se evaluó la concordancia entre ambas fuentes. La variabilidad interobservador se evaluó en 400 historias clínicas seleccionadas aleatoriamente.

Resultados: La información sobre complicaciones patológicas del embarazo y la antropometría de las madres se recuperó con éxito. La concordancia entre el cuestionario y los registros con respecto a los antecedentes familiares era débil, especialmente para las enfermedades cardiovasculares ($k=0,27$, intervalo de confianza del 95% [IC95%]: 0,23–0,32). La concordancia máxima se observó en los antecedentes personales de diabetes ($k=0,82$, IC95%: 0,70–0,93), mientras que para la hipertensión fue moderada ($k=0,60$, IC95%: 0,50–0,69). Se observaron discrepancias en las categorías de índice de masa corporal antes del embarazo en el 10,3% de las mujeres. Los datos fueron muy concordantes entre los revisores, con el máximo nivel de concordancia para la diabetes gestacional ($k=1,00$), seguida del peso al nacer (99,5% concordantes).

Conclusión: Los registros médicos y la información del cuestionario fueron concordantes para los datos relacionados con el embarazo y los factores de riesgo conocidos. La baja variabilidad interobservador no pone en peligro la precisión de los datos.

© 2010 SESPAS. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Palabras clave:

Métodos epidemiológicos

Registros médicos

Variabilidad interobservador

Antropometría

Estudios de cohortes

* Corresponding author.

E-mail address: ealves@med.up.pt (E. Alves).

Introduction

Clinical records are an important source of information and medical record review is a commonly used data collection method in epidemiological studies. However, data registered in clinical records are not originally collected for research purposes and may not faithfully reflect all the events that happen during a medical consultation^{1,2}. These records are less likely to explicitly include negative self-reports or diagnoses³ and systematically tend to report incomplete information about lifestyles⁴.

Epidemiological studies frequently use personal interviews and self-administered questionnaires as the sole information sources on exposures and outcomes. These methods are regarded as valid tools and provide many advantages for research. Nonetheless, the quality of information obtained through self-report depends substantially on the type of illness^{5,6}, the participants' characteristics^{5,7,8}, the method used to administer the questionnaire⁹ and the questionnaire's design^{10,11}.

Information obtained by self-report and medical record review may not be consistent. Several studies have compared self-reported data collected by questionnaires with medical record abstraction for the assessment of history of cardiovascular diseases (CVD)^{3,7}, pregnancy-related events and birth characteristics^{12–14}, and show that the agreement between the two data sources depends on the data being collected¹⁵. Additionally, any data collection method may be affected by interobserver variability, which might influence medical record reviews when multiple abstractors are involved^{2,16}. In recent years, little attention has been paid to examining interobserver variability in data abstraction from clinical records, even though these methodological aspects influence the quality of the data and are critical for obtaining reliable results^{17,18}.

Given the extensive use of these two data collection methods, the practical implications of their use must be understood, especially when both data sources are complementarily used in the same study. In a Portuguese birth cohort, data on several variables from the baseline evaluation were missing for a large number of participants. We planned to complete these data by abstracting them from medical records at delivery. To the best of our knowledge, no previous Portuguese study has evaluated the yield of missing data recovery by medical record review or agreement between questionnaire and clinical records and between multiple raters in medical record review.

The aim of the present study was to assess the success of missing data recovery by medical record review. Additionally, we evaluated the agreement between data recovered from clinical records and data previously obtained through a structured questionnaire administered to mothers of the cohort, as well as interobserver variability in data abstraction from clinical records.

Methods

This study was based on the birth cohort Geração XXI, to which 8,127 mothers were consecutively invited after delivery. These women gave birth to 8,270 infants. Recruitment was conducted between April 2005 and September 2006 at all public maternity units covering the metropolitan area of Porto, Portugal. All the maternity units (Table 1), except Maternidade Júlio Dinis (MJD), are included in a general hospital, with a variety of medical and surgical specialties, and all correspond to level III maternity units, with differentiated perinatal support. At birth, 91.4% of the invited mothers accepted to participate.

Information on family and personal history of disease and the mothers' anthropometric parameters before pregnancy was

collected during a face-to-face interview conducted by trained interviewers using structured questionnaires within 72 hours of delivery, during the hospital stay. Data on pregnancy complications, gestational age and neonatal characteristics were abstracted from clinical records by the same interviewers.

Missing data recovery by clinical record review and agreement with questionnaire

Of the 8,127 mothers, 3,771 had at least one missing value in personal history of disease, anthropometric parameters, pregnancy complications, blood glucose or oral glucose tolerance test results during pregnancy, or neonatal characteristics at birth (Fig. 1). This was the sample used in the current study.

There were no significant differences between participants with complete and missing data in age [mean (standard deviation, SD): 29.4 (5.5) vs. 29.6 (5.7) years; $p=0.306$], education [median (interquartile range, IQR): 10 (7–12) vs. 10 (7–12) years; $p=0.122$], gravidity (first pregnancy: 48.7% vs. 47.1%; $p=0.156$) and marital status (married/living with a partner: 93.5% vs. 94.3%; $p=0.165$).

Between October 2008 and June 2009, we reviewed the delivery medical records of 3,657 women to recover missing data. The remaining 114 records were not available for review during the period for which we were authorized to consult them, either because they were being used for subsequent patient care or for administrative purposes (Fig. 1).

The clinical records were reviewed by a trained abstractor (EA), who had been an interviewer at the baseline evaluation of Geração XXI, using standardized criteria (Fig. 1). From each medical record, we abstracted data on all the variables of interest, whether or not they were missing in the questionnaire. Since most women had missing values in one or very few variables, we used those who had data available from both the baseline interview and the medical record review to assess agreement between the two data sources in each variable. Therefore, the sample size was not the same for all the variables (Fig. 1).

Interobserver variability

To assess interobserver variability in data abstraction, among the 3,657 records reviewed, data from a randomly selected subsample of 80 clinical records from each hospital were collected independently by two trained abstractors (EA and VM), both of whom interviewed women after delivery, according to the criteria described in Figure 1. We estimated the required sample size to demonstrate, at a 5% significance level, an excellent level of agreement, corresponding to a kappa coefficient of at least 0.85 on past personal and family history of disease, significantly different from a moderate agreement of $k=0.60$ for conditions with a prevalence as low as 2%¹⁹. The expected prevalence of each condition was defined according to the observed prevalence in this sample. We also estimated the required sample size for duplicate review to demonstrate that, for a significance level of 5% and a statistical power of 80%, differences in pre-pregnancy weight, weight at the end of pregnancy, height, gestational age at birth and birth weight were smaller than negligible differences of 1 Kg, 1 Kg, 1 cm, 0.2 weeks and 50 g, respectively, for the mean and standard deviation of each variable observed in this sample and a correlation coefficient between groups not lower than 0.90. To respect all these conditions, a minimum of 400 medical records would have to be independently reviewed by two raters. We did not assess interrater variability by maternity unit.

Table 1
Participants' characteristics at baseline by maternity unit.

	Overall	CHVNG	MJD	HSJ	HSA	HPH
N	3,657	1,269	366	478	304	1,240
Age (years), mean (SD)	29.5 (5.7)	29.6 (5.8)	28.6 (6.0)	29.7 (5.5)	30.4 (5.6)	29.5 (5.5)
Marital status, n (%)						
Married/living with a partner	3430 (94.3)	1,192 (94.6)	330 (90.2)	451 (95.2)	285 (94.1)	1,172 (94.9)
Single/ divorced/widowed	208 (5.7)	68 (5.4)	36 (9.8)	23 (4.9)	18 (5.9)	63 (5.1)
Education (years), median (IQR)	10 (7-12)	9 (6-12)	9 (6-12)	12 (7-16)	11 (7-15)	12 (7-15)
Monthly income (€), n (%)						
< 500	249 (7.2)	103 (8.7)	41 (11.7)	19 (4.3)	35 (12.8)	51 (4.2)
500-1,000	1011 (29.3)	434 (36.7)	122 (34.8)	91 (20.8)	77 (28.1)	287 (23.8)
1,001- 1,501	854 (24.8)	286 (24.2)	82 (23.4)	126 (28.8)	61 (22.3)	299 (24.8)
≥ 1,500	927 (26.9)	275 (23.3)	67 (19.1)	172 (39.3)	67 (24.5)	346 (28.7)
Does not know/Prefers not to answer	410 (11.9)	84 (7.1)	39 (11.1)	30 (6.9)	34 (12.4)	223 (18.5)
Family history of diabetes, n (%)	719 (22.5)	265 (23.5)	77 (23.6)	89 (22.1)	6 (26.2)	222 (20.3)
Family history of CVD, n (%)	514 (16.1)	209 (18.5)	55 (16.9)	59 (14.8)	48 (19.0)	143 (13.1)
Hypertension before pregnancy, n (%)	69 (1.9)	25 (2.0)	4 (1.1)	9 (1.9)	16 (5.3)	15 (1.2)
Diabetes mellitus before pregnancy, n (%)	28 (0.8)	7 (0.6)	1 (0.3)	2 (0.4)	15 (4.9)	3 (0.2)
BMI before pregnancy, n (%)						
<24.9 Kg/m ²	1,213 (71.6)	213 (68.9)	152 (71.7)	246 (71.7)	195 (78.0)	407 (70.3)
25.0-29.9 Kg/m ²	342 (20.2)	65 (21.0)	44 (20.7)	68 (19.8)	40 (16.0)	125 (21.6)
≥ 30 Kg/m ²	138 (8.2)	31 (10.0)	16 (7.6)	29 (8.5)	15 (6.0)	47 (8.1)
Gravidity, n (%)						
1	1,717 (47.1)	567 (44.9)	148 (40.6)	233 (48.9)	149 (49.2)	620 (50.3)
2	1,246 (34.2)	441 (34.9)	136 (37.3)	161 (33.8)	92 (30.4)	416 (33.7)
≥ 3	679 (18.6)	256 (20.3)	81 (22.2)	83 (17.4)	62 (20.5)	197 (16.0)
Multiple pregnancy, n (%)	80 (2.2)	28 (2.2)	7 (1.9)	9 (1.9)	16 (5.3)	20 (1.6)
Gestational hypertensive disorders, n (%) ^a	110 (3.6)	61 (5.5)	13 (4.1)	8 (2.2)	19 (8.8)	9 (0.9)
Gestational diabetes, n (%)	219 (7.3)	128 (11.5)	15 (4.7)	18 (5.1)	26 (12.0)	32 (3.2)
Preterm newborn, n (%) ^b	351 (10.4)	100 (8.4)	38 (11.3)	25 (6.0)	54 (22.1)	134 (11.3)
Low birth weight, n (%) ^c	339 (9.7)	105 (8.5)	42 (12.0)	26 (6.0)	41 (15.4)	125 (10.4)

BMI: body mass index; CHVNG: Centro Hospitalar de Vila Nova de Gaia; CVD: cardiovascular disease; HPH: Unidade Local de Saúde de Matosinhos – Hospital Pedro Hispano; HSA: Centro Hospitalar do Porto – Hospital de Santo António; HSJ: Hospital de São João; IQR: interquartile range; MJD: Centro Hospitalar do Porto – Maternidade de Júlio Dinis; SD: standard deviation.

^a Gestational hypertension and/or pre-eclampsia/eclampsia.

^b Gestational age <37 weeks.

^c Birth weight <2,500 g.

Statistical analysis

For the description of the sample's characteristics, data are presented as counts and proportions for categorical variables, mean and standard deviation for normally distributed continuous variables and median and interquartile range for non-normally distributed continuous variables.

To report the yield of missing data recovery, counts of cases with missing values and counts and proportions for which information was recovered are presented.

To assess the agreement between clinical records and questionnaire information, as well as between observers, for continuous variables, we estimated the mean differences and 95% confidence intervals (95%CI) between the questionnaire and clinical records and between independent reviewers. The proportion of concordant observations was calculated assuming concordance within 1 Kg, 1 cm or 0.1 weeks for weight, height and gestational age, respectively. For categorical variables, we calculated the proportion of concordant observations between the questionnaire and clinical record among all variables registered as “No” and all registered as “Yes” in the questionnaire, considering it as the reference because the aim of this study was to replace missing data on the questionnaire by those abstracted from medical records. We made no assumptions on the comparative validity of each method. In both the agreement between data sources and between reviewers, we calculated the observed proportion of agreement and kappa coefficients and 95%CI. The weighted kappa was used for variables with more than two classes.

Missing data recovery and interrater variability were analyzed for all variables, whether self-reported or abstracted from medical records at the baseline evaluation. The agreement between the questionnaire and medical record was evaluated only for variables that were collected by questionnaire at birth (Fig. 1).

Data analysis was performed using Stata 9.0 (College Station, TX, 2005). Sample size estimation for assessment of interrater variability was performed using WinPEPI²⁰.

Ethics

The study protocol was approved by the Ethics Committee of Hospital de S. João (HSJ). Written informed consent was obtained from all participants at baseline.

Results

Sample characteristics

The mean maternal age at birth was 29.5 years (range: 13 to 45 years). Most women were married or were living with a partner (94.3%) and the median [interquartile range (IQR)] years of education was 10 (7-12). In this sample, 7.2% of the participants had a household monthly income below 500€, while 26.9% had a monthly income above 1,500€. A family history of diabetes and cardiovascular disease (CVD) were reported by 22.5% and 16.1% of the mothers, respectively. The prevalence of

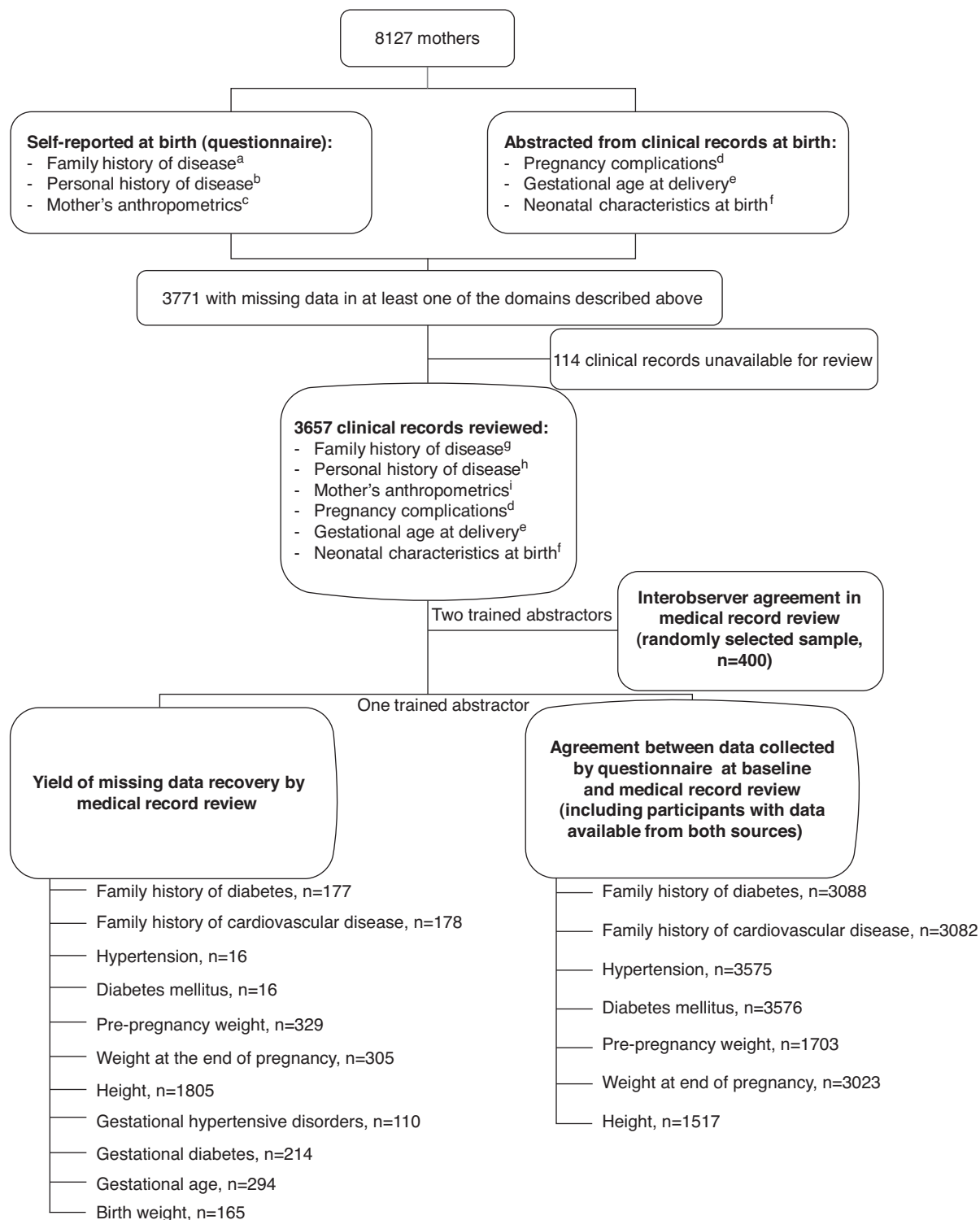


Figure 1. Definition of the samples for missing data recovery, assessment of agreement between data collected by questionnaire and medical record review, and evaluation of interobserver agreement in medical record review.

^a A family history of diabetes mellitus or cardiovascular disease was considered present when participants reported at least one parent or sibling affected by diabetes or by stroke and/or myocardial infarction, respectively.

^b A personal history of pre-pregnancy hypertension and diabetes mellitus was considered present when participants recalled having received a medical diagnosis of these conditions.

^c Usual weight in the 2 years preceding pregnancy and weight immediately before delivery were both obtained through recall information to the nearest 0.1 Kg. Height was measured without shoes by the interviewers to the nearest 0.1 cm. When measurement was not possible, height was reported by the mother as registered in the identity card (35.4% of women with data on height).

^d Gestational hypertension, eclampsia/pre-eclampsia and gestational diabetes were considered only when explicitly recorded as a diagnosis during the current pregnancy.

^e Gestational age was considered as that determined by ultrasound or, when this information was not available, as length of amenorrhea, to the nearest 0.1 weeks.

^f Birth weight, length and head circumference of the newborn were registered to the nearest 1 g and 0.1 cm, respectively.

^g A family history of CVD and diabetes mellitus was classified using the same criteria as those used in the baseline questionnaire. When the affected relative was not clearly identified as a parent or sibling, the family history was considered positive.

hypertension was 1.9%, while the prevalence of diabetes mellitus was 0.8%. Before pregnancy, 20.2% of the women were overweight and 8.2% were obese. This was the first pregnancy for 47.1% of the women; 9.7% delivered a low birth weight newborn, 10.4% had a preterm delivery and 2.2% of the pregnancies were multiple. Gestational hypertensive disorders and gestational diabetes affected 3.6% and 7.3% of the pregnancies, respectively. When compared with women from the other maternity units, those who delivered at Maternidade de Júlio Dinis (MJD) were younger, more likely to live without a partner, had a lower prevalence of pre-pregnancy diagnosis of hypertension and diabetes and had a higher number of previous pregnancies. Mothers selected at the Hospital de Santo António (HSA) were older, more likely to report a family history of diabetes and CVD, to have a multiple pregnancy and to deliver a premature or low birth weight newborn. These women showed the highest prevalence of hypertension and diabetes before and during pregnancy. At MJD and Centro Hospitalar de Vila Nova de Gaia (CHVNG) women had the lowest educational level and 68.1% of the mothers from Hospital de São João (HSJ) had an income higher than 1,000 €/month. Mothers from Hospital Pedro Hispano (HPH) had the lowest prevalence of gestational hypertensive disorders and gestational diabetes (Table 1).

Yield of missing data recovery

The proportion of participants for whom data was recovered from clinical records is presented in Table 2. Overall, the proportion of missing data recovered ranged from zero to 84.1%, depending on the variables considered. Data on previous diagnosis of hypertension and diabetes mellitus were rarely missing and could never be recovered from medical records. Data on family history of diabetes or CVD and on gestational age and birth weight, when missing at baseline, could seldom be recovered from the medical records. This strategy of recovering missing data was most effective in maternal anthropometric parameters and particularly in pregnancy complications. The highest proportion of data for a family history of both diabetes and CVD was recovered in the HPH. The proportion of participants for whom anthropometric parameters was recovered was lowest at HSJ. At all maternity units, data on pregnancy complications could in general be recovered from medical records, while recovery of missing data on gestational age and birth weight was low (Table 2).

One trained reviewer required around 730 hours to perform the clinical record review, including all the procedures for gaining access to the records plus the review itself, corresponding to approximately 90 working days to complete this task and an average of 11.5 minutes per medical record.

Agreement between clinical records and questionnaire information

Table 3 illustrates the agreement between data collected by the questionnaire and those abstracted from the clinical records. Overall, the agreement in family history was fair, particularly for history of CVD ($k=0.27$, 95%CI: 0.23-0.32). At CHVNG the point estimate of the kappa coefficient for family history of disease was

higher than in the remaining hospitals, but this difference was not significant. There was no reference to family history of CVD in any of the clinical record reviewed at HSJ. Agreement in personal history of disease was higher than in family history, particularly for diabetes mellitus ($k=0.82$, 95%CI: 0.70-0.93), with no significant differences among hospitals. In general, the discrepancies found for weight and height indicated higher values in the clinical record than in the questionnaire. Pre-pregnancy weight reported to the interviewer at birth was 0.6 Kg lower than that registered in the clinical record. The difference between the questionnaire and clinical records in women's weight at the end of pregnancy was much smaller (0.3 Kg). Despite the relatively small differences in means, for about half of the women the discrepancy in weight between the questionnaire and clinical record was greater than 1 Kg. For height, the differences between hospitals were quantitatively negligible (Table 3).

The differences observed in weight and height led to changes in the classification by body mass index categories (<25.0, 25.0 to 29.9, ≥ 30 Kg/m²) of 10.3% of women and three out of 871 women were misclassified by two body mass index categories (weighted $k=0.82$, 95%CI: 0.80-0.83). Nonetheless, the variability was randomly distributed, occurring similarly in both directions.

Interobserver variability

The agreement between data collected by the two reviewers was good or very good. The lowest agreement was observed for personal history of diabetes mellitus ($k=0.66$; 95%CI: 0.23-1.00). Data concerning the occurrence of pregnancy complications were highly consistent between the two reviewers, with perfect agreement observed for gestational diabetes ($k=1.00$; 95%CI: 1.00-1.00). The differences observed for women's weight and height were negligible, with at least 95% of the data concordant to the nearest 1 Kg or 1 cm, respectively. The mean difference was not significantly different from zero for the three anthropometric variables assessed. The differences in weight and height between reviewers did not affect the global classification by body mass index categories before pregnancy, with only one out of 151 women changing to an adjacent category (weighted $k=0.99$, 95%CI: 0.98-1.00). Despite the equality of means, gestational age was registered inconsistently by the two reviewers in almost 20% of the records. The highest agreement was observed for birth weight, with 99.5% concordant observations (Table 4).

Discussion

In this study of pregnant Portuguese women, data on pathological complications of pregnancy and maternal anthropometrics were successfully recovered from medical records when missing from the baseline questionnaire, while the yield of the medical record review was low in recovering past family or personal history of disease.

The overall cost of recovering information from clinical records is an important finding of this study. This procedure required one abstractor over an extended period of time. Clinical records at these maternity units were not electronic and the fact that paper records had to be reviewed likely decreased the yield and increased the time involved. The cost-benefit of perform-

^h A personal history of hypertension and diabetes was considered present only when explicitly recorded as a diagnosis and not inferred from blood pressure values, serum glucose or drug use.

ⁱ Pre-pregnancy weight was considered to the nearest 0.1 Kg as self-reported to a health professional, either at early appointments during pregnancy or at birth, while the weight at the end of pregnancy was considered as the weight registered in records at admission for delivery or, when this information was unavailable, at the last medical appointment before birth. Height was considered as registered to the nearest 0.1 cm at admission for delivery or, when this information was unavailable, in the clinical records before birth.

Table 2
Participants with missing information and proportion of recovered data by maternity unit.

	Overall		CHVNG		MJD		HSJ		HSA		HPH	
	Missing data, n	Missing data recovered, n (%)	Missing data, n	Missing data recovered, n (%)	Missing data, n	Missing data recovered, n (%)	Missing data, n	Missing data recovered, n (%)	Missing data, n	Missing data recovered, n (%)	Missing data, n	Missing data recovered, n (%)
<i>Self-reported at birth (questionnaire)</i>												
Family history of diabetes	177	9 (5.1)	59	3 (5.1)	17	1 (5.9)	43	0 (0)	23	0 (0)	35	5 (14.3)
Family history of CVD	178	9 (5.1)	59	3 (5.1)	17	1 (5.9)	44	0 (0)	23	0 (0)	35	5 (14.3)
Hypertension before pregnancy	16	0 (0)	8	0 (0)	1	0 (0)	1	0 (0)	0	0 (0)	6	0 (0)
Diabetes mellitus before pregnancy	16	0 (0)	8	0 (0)	1	0 (0)	1	0 (0)	0	0 (0)	6	0 (0)
Mother's usual prepregnancy weight	329	161 (48.9)	78	42 (53.9)	17	11 (64.7)	17	3 (17.7)	22	12 (54.6)	195	93 (47.7)
Mother's weight at the end of pregnancy	305	55 (18.0)	118	27 (22.9)	40	4 (10.0)	30	2 (6.7)	30	2 (6.7)	87	20 (23.0)
Mother's height	1805	510 (28.3)	951	437 (46.0)	149	24 (16.1)	128	2 (1.6)	44	21 (47.7)	533	26 (4.9)
<i>Abstracted from medical records at birth</i>												
Gestational hypertensive disorders ^a	110	91 (82.7)	61	49 (80.3)	13	11 (84.6)	8	7 (87.5)	19	18 (94.7)	9	6 (66.7)
Gestational diabetes	214	180 (84.1)	127	113 (89.0)	14	9 (64.3)	18	12 (66.7)	23	22 (95.7)	32	24 (75.0)
Gestational age	294	26 (8.8)	85	12 (14.1)	31	5 (16.1)	59	3 (5.1)	60	1 (1.7)	59	5 (8.5)
Birth weight	165	4 (2.4)	27	2 (7.4)	16	0 (0)	45	0 (0)	37	1 (2.7)	40	1 (2.5)

CHVNG: Centro Hospitalar de Vila Nova de Gaia; CVD, cardiovascular disease; HSA: Centro Hospitalar do Porto – Hospital de Santo António; HSJ: Hospital de São João; MJD: Centro Hospitalar do Porto – Maternidade de Júlio Dinis; .

^a Gestational hypertension and/or pre-eclampsia/eclampsia.

ing this process should be considered case-by-case, taking into account the study aims and the expected yield for distinct variables in each setting. However, this is a time-consuming and expensive solution for correction of methodological errors that should not have occurred in the first place, as when using records to retrieve information that was originally also obtained from medical records.

The agreement between self-reported and medical record data was highly variable. Data directly related to pregnancy and well-known risk factors were concordant. Family history of CVD was underreported in the clinical records of the five hospitals. CHVNG showed the highest estimate of agreement for family history of both diabetes and CVD. This difference can be explained by the clinical record format, which varied between maternity units, including a standardized and pre-formatted section for registration of family history of disease at the CHVNG. In general, there was greater consistency between data sources in family history of diabetes than in family history of CVD. Acknowledgement of family history of diabetes as a major risk factor for the development of gestational diabetes²¹ may lead physicians to register these data more systematically and, simultaneously, to increase awareness of these risk factors among pregnant women. A family history of stroke, myocardial infarction and diabetes mellitus can be accurately reported by the participants, when compared with the relatives' self-reports or their death certificates and general practitioners' or hospital notes^{22–24}.

We showed that the agreement between data collected by questionnaire and those abstracted from medical records is highly dependent on the data recording procedures and practices used and the extent to which these are standardized in each hospital. Our overall results are locale-specific, despite reflecting issues that may be present to a greater or lesser extent in each setting; while the overall yield is likely to be similar in distinct settings, the limited external validity of our results needs to be acknowledged. However, we also present the results by institution and interpret them by taking into account our knowledge of the data collection procedures in each hospital, which provides important information that may be generalized with a smaller number of prior assumptions. Recent mothers are likely to provide more accurate information on personal and family history of diseases than the general population. Therefore, the results of this study cannot be generalized to other types of population.

Few studies have previously focused on the agreement between self-reported data and clinical record review for CVD and risk factors in young women before pregnancy^{15,25}. Although the low prevalence and the lack of awareness of hypertension and diabetes diagnosed before pregnancy in young women increase the difficulty of analyzing the agreement^{26,27}, our results are concordant with previous findings. Ramadhani et al²⁵ have reported a substantial agreement between medical records and maternal interviews for non-gestational diabetes data ($k = 0.75$, 95%CI: 0.64–0.86) and similar estimates were found for pre-pregnancy hypertension among pregnant Latin women ($k = 0.68$, 95%CI: 0.46–0.90)¹⁵.

The accuracy of self-reported weight and height, in comparison with objective measurements, has been extensively reported. Overall, weight and body mass index tend to be underreported and height overreported by women^{28,29}, and the same pattern has been observed among pregnant women^{30–32}. In our study, weight and height were systematically higher in the clinical record than in the questionnaire. The longer time frame considered in the questionnaire could translate into underestimation of weight before pregnancy. The questionnaire data on weight at the end of pregnancy was always self-reported when the cohort was assembled, while in some hospitals weight may have been measured by nurses before delivery. Although we have no possibility of knowing if the weight registered in the clinical records was measured

Table 3
Agreement between data collected by questionnaire at baseline and clinical record review by maternity unit.

	Overall	CHVNG	MJD	HSJ	HSA	HPH
Dichotomous variables	Number of "no" or "yes" responses at the questionnaire interview concordant with the medical record / number of mothers responding "no" or "yes" at the questionnaire interview (%)					
<i>Family history of diabetes, n/n (%)^a</i>						
No	2210/2390 (92.5)	814/840 (96.9)	216/233 (92.7)	259/313 (82.7)	156/179 (87.2)	765/825 (92.7)
Yes	515/698 (73.8)	193/256 (75.4)	38/74 (51.4)	75/89 (84.3)	48/64 (75.0)	161/215 (74.9)
Agreement (%)	88.2	91.9	82.7	83.1	84.0	89.0
Kappa (95%CI)	0.66 (0.63-0.70)	0.76 (0.71-0.81)	0.48 (0.37-0.60)	0.58 (0.49-0.67)	0.60 (0.49-0.71)	0.67 (0.61-0.73)
<i>Family history of CVD, n/n (%)^a</i>						
No	2572/2583 (99.6)	886/891 (99.4)	252/253 (99.6)	340/340 (100.0)	198/198 (100.0)	896/901 (99.4)
Yes	95/499 (19.0)	66/204 (32.4)	14/52 (26.9)	0/59 (0.0)	2/45 (4.4)	13/139 (9.4)
Agreement (%)	86.5	86.9	87.2	85.2	82.3	87.4
Kappa (95%CI)	0.27 (0.23-0.32)	0.43 (0.35-0.50)	0.37 (0.23-0.52)	0.00 (0.00-1.00)	0.07 (-0.02-0.16)	0.14 (0.07-0.21)
<i>Hypertension before pregnancy, n/n (%)^a</i>						
No	3474/3507 (99.1)	1,196/1,217 (98.3)	343/345 (99.4)	466/467 (99.8)	282/283 (99.6)	1,187/1,195 (99.3)
Yes	44/68 (64.7)	16/24 (66.7)	2/4 (50.0)	4/9 (44.4)	9/16 (56.3)	13/15 (86.7)
Agreement (%)	98.4	97.7	98.9	98.7	97.3	99.2
Kappa (95%CI)	0.60 (0.50-0.69)	0.51 (0.36-0.67)	0.49 (0.07-0.92)	0.57 (0.25-0.88)	0.68 (0.47-0.89)	0.72 (0.55-0.89)
<i>Diabetes mellitus before pregnancy, n/n (%)^a</i>						
No	3547/3548 (100.0)	1235/1236 (99.9)	348/348 (100.0)	474/474 (100.0)	283/283 (100.0)	1207/1207 (100.0)
Yes	20/28 (71.4)	5/7 (71.4)	0/1 (0.0)	1/2 (50.0)	11/15 (73.3)	3/3 (100.0)
Agreement (%)	99.8	99.8	99.7	99.8	98.7	100
Kappa (95%CI)	0.82 (0.70-0.93)	0.77 (0.51-1.00)	0.00 (0.00-1.00)	0.67 (0.05-1.00)	0.84 (0.69-0.99)	1.00 (1.00-1.00)
Continuous variables						
<i>Mother's usual prepregnancy weight (Kg)</i>						
Questionnaire (mean, SD)	62.0 (11.9)	63.7 (12.6)	59.7 (10.0)	61.3 (11.3)	62.2 (11.3)	62.1 (12.4)
Clinical records (mean, SD)	62.6 (12.5)	64.5 (12.9)	60.1 (10.9)	61.5 (12.1)	64.0 (12.4)	62.7 (12.8)
Mean difference (95%CI)	-0.53 (-0.73 to -0.34)	-0.80 (-1.21 to -0.39)	-0.34 (-0.87 to 0.19)	-0.13 (-0.56 to 0.30)	-1.83 (-3.19 to -0.47)	-0.58 (-0.88 to -0.28)
n concordant/n (%) ^b	753/1703 (44.2)	189/407 (46.4)	91/205 (44.4)	187/431 (43.4)	15/59 (25.4)	271/601 (45.1)
<i>Mother's weight at the end of pregnancy (Kg)</i>						
Questionnaire (mean, SD)	75.4 (11.9)	75.8 (11.7)	74.1 (11.5)	75.2 (11.5)	73.1 (11.3)	76.1 (12.5)
Clinical records (mean, SD)	75.6 (12.2)	76.1 (11.9)	74.4 (11.9)	74.8 (12.2)	73.6 (11.5)	76.5 (12.6)
Mean difference (95%CI)	-0.24 (-0.35 to -0.12)	-0.32 (-0.56 to -0.08)	-0.28 (-0.57 to 0.00)	0.41 (0.01 to 0.81)	-0.52 (-1.00 to -0.04)	-0.35 (-0.48 to -0.22)
n concordant/n (%) ^b	1530/3023 (50.6)	400/936 (42.7)	157/294 (53.4)	241/441 (54.6)	156/261 (59.8)	576/1091 (52.8)
<i>Mother's height (cm)</i>						
Questionnaire (mean, SD)	160.9 (6.3)	159.7 (5.8)	160.3 (6.6)	160.8 (6.3)	161.9 (5.8)	161.1 (6.4)
Clinical records (mean, SD)	161.4 (6.3)	160.7 (6.2)	161.0 (6.5)	161.4 (6.3)	161.8 (5.8)	161.5 (6.4)
Mean difference (95%CI)	-0.49 (-0.64 to -0.35)	-1.07 (-1.64 to -0.49)	-0.76 (-1.14 to -0.38)	-0.62 (-0.93 to -0.32)	0.17 (-0.17 to 0.52)	-0.41 (-0.63 to -0.19)
n concordant/n (%) ^c	688/1517 (45.4)	56/146 (38.4)	99/185 (53.5)	150/341 (44.0)	129/192 (67.2)	254/653 (38.9)

95%CI, 95% confidence interval; CHVNG: Centro Hospitalar de Vila Nova de Gaia; CVD: cardiovascular disease; HPH: Unidade Local de Saúde de Matosinhos - Hospital Pedro Hispano; HSA: Centro Hospitalar do Porto - Hospital de Santo António; HSJ: Hospital de São João; MJD: Centro Hospitalar do Porto - Maternidade de Júlio Dinis; SD: standard deviation.

^a Percentages may not total 100% due to rounding.

^b Within plus or minus 1 Kg.

^c Within plus or minus 1 cm.

Table 4
Interobserver agreement in medical record review.

Dichotomous variables	
<i>Family history of diabetes</i>	
Agreement, n (%)	372 (94.4)
Kappa (95% CI)	0.85 (0.79–0.91)
<i>Family history of CVD</i>	
Agreement, n (%)	372 (99.5)
Kappa (95% CI)	0.89 (0.73–1.00)
<i>Hypertension before pregnancy</i>	
Agreement, n (%)	384 (99.5)
Kappa (95% CI)	0.89 (0.73–1.00)
<i>Diabetes mellitus before pregnancy</i>	
Agreement, n (%)	384 (99.5)
Kappa (95% CI)	0.66 (0.23–1.00)
<i>Gestational hypertensive disorders^a</i>	
Agreement, n (%)	374 (99.2)
Kappa (95% CI)	0.91 (0.80–1.00)
<i>Gestational diabetes</i>	
Agreement, n (%)	388 (100)
Kappa (95% CI)	1.00 (1.00–1.00)
Continuous variables	
<i>Mother's usual prepregnancy weight (Kg)</i>	
Reviewer 1 (mean, SD)	61.8 (11.4)
Reviewer 2 (mean, SD)	61.7 (11.6)
Mean difference (95% CI)	0.048 (–0.16 to 0.26)
n concordant/n (%) ^b	156/165 (94.5)
<i>Mother's weight at the end of pregnancy (Kg)</i>	
Reviewer 1 (mean, SD)	75.5 (12.1)
Reviewer 2 (mean, SD)	75.4 (12.1)
Mean difference (95% CI)	0.043 (–0.05 to 0.14)
n concordant/n (%) ^b	353/371 (95.1)
<i>Mother's height (cm)</i>	
Reviewer 1 (mean, SD)	162.0 (5.7)
Reviewer 2 (mean, SD)	162.1 (5.6)
Mean difference (95% CI)	–0.05 (–0.14 to 0.05)
n concordant/n (%) ^c	302/311 (97.1)
<i>Gestational age (weeks)</i>	
Reviewer 1 (mean, SD)	38.6 (2.2)
Reviewer 2 (mean, SD)	38.6 (2.2)
Mean difference (95% CI)	–0.01 (–0.04 to 0.02)
n concordant/n (%) ^d	324/395 (82.0)
<i>Birth weight (g)</i>	
Reviewer 1 (mean, SD)	3,143 (565)
Reviewer 2 (mean, SD)	3,141 (566)
Mean difference (95% CI)	1.87 (–0.90 to 4.63)
n concordant/n (%) ^e	389/391 (99.5)

95% CI, 95% confidence interval; CVD, cardiovascular disease; SD, standard deviation.

^a Gestational hypertension and/or pre-eclampsia/eclampsia.

^b Within plus or minus 1 Kg.

^c Within plus or minus 1 cm.

^d Within plus or minus 0.1 week.

^e Within plus or minus 1 g.

or self-reported, the higher value obtained by the former seems to support our assumption³². Pregnant women tend to overreport their height³¹. At birth, most participants were measured, which could explain the differences found.

Studies that have examined the accuracy of self-reported height and weight employed to determine body mass index categories have concluded that approximately 80% of women were correctly classified^{29,33}, which is in accordance with our observations.

The interobserver variability was low and did not threaten data precision. Standardized training of abstractors and rigorous quality assurance were proposed as critical criteria to improve the quality and accuracy of clinical record review^{16–18,34,35}. When research involves data collection by distinct observers, the extent to which different observers perceive and record the same information should be evaluated. In the literature, most medical record review studies fail to report on interrater agreement³⁴. Our study showed a low interrater variability, and the discrepancies found in family history of diabetes and CVD between the two reviewers can be attributed to a combination of factors, from the different

structure of the clinical records and the data registered in them to specific problems in accessing information from the charts. When records are not pre-formatted and standardized, professionals are more likely to underreport or to register data in different locations, increasing the difficulty of the data abstraction procedure, which precludes the generalization of these findings to other settings.

This study provides important information for the planning and interpretation of epidemiological studies on pregnant women, but some limitations should be discussed. Despite the large sample size, the evaluation of conditions such as hypertension and diabetes before pregnancy is limited by their very low prevalence. The absence of reference to a medical condition in the clinical record was considered as the absence of the condition itself. We do not believe that this assumption compromises our conclusions as it is unlikely that physicians did not enquire about such conditions or failed to record them when present. When clinical records did not specify which relative had developed the outcomes studied, we considered the family history of disease as positive, which could lead to a higher proportion of family history in clinical records than in the interview. However, given the unusual occurrence of this situation, we do not expect that this factor had a quantitatively significant influence on the results. In this study, there was no gold-standard and differences between data sources could reflect misclassification in the questionnaires, in the clinical records data or both. We were unable to draw conclusions on which method provides higher validity for the data assessed.

In conclusion, data directly related to pregnancy and well-known risk factors can be safely recovered from medical records when missing from self-reported data. The need to implement structured and standardized methods to abstract data from clinical records cannot be overemphasized to enhance data quality, which can subsequently improve the interpretation and generalization of the results obtained.

Author's contribution

Elisabete Alves collaborated in the acquisition, analysis and interpretation of the data and draft the article. Nuno Lunet collaborated in the analysis and interpretation of the data reviewed and revised the article critically for important intellectual content. Sofia Correia helped to supervise the field activities and reviewed the article. Ana Azevedo designed the study, directed its implementation and reviewed and revised the article critically for important intellectual content. Henrique Barros designed and supervised the birth cohort Geração XXI, on which our study is based, and made a substantial contribution to the revisions made after the requests of the reviews. All authors approved the final version.

Conflict of interests

The authors declare that they have no conflict of interests.

References

- Luck J, Peabody JW, Dresselhaus TR, et al. How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. *Am J Med.* 2000;108:642–9.
- Worster A, Haines T. Advanced statistics: understanding medical record review (MRR) studies. *Acad Emerg Med.* 2004;11:187–92.
- St Sauver JL, Hagen PT, Cha SS, et al. Agreement between patient reports of cardiovascular disease and patient medical records. *Mayo Clin Proc.* 2005;80:203–10.
- Fouwels AJ, Bredie SJ, Wollersheim H, et al. A retrospective cohort study on lifestyle habits of cardiovascular patients: how informative are medical records? *BMC Health Serv Res.* 2009;9:59.
- Molenaar EA, Van Ameijden EJ, Grobbee DE, et al. Comparison of routine care self-reported and biometrical data on hypertension and diabetes: results of the Utrecht Health Project. *Eur J Public Health.* 2007;17:199–205.

6. Bowlin SJ, Morrill BD, Nafziger AN, et al. Reliability and changes in validity of self-reported cardiovascular disease risk factors using dual response: the behavioral risk factor survey. *J Clin Epidemiol.* 1996;49:511–7.
7. Okura Y, Urban LH, Mahoney DW, et al. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol.* 2004;57:1096–103.
8. Goldman N, Lin IF, Weinstein M, et al. Evaluating the quality of self-reports of hypertension and diabetes. *J Clin Epidemiol.* 2003;56:148–54.
9. Bergmann MM, Jacobs EJ, Hoffmann K, et al. Agreement of self-reported medical history: comparison of an in-person interview with a self-administered questionnaire. *Eur J Epidemiol.* 2004;19:411–6.
10. Gama H, Correia S, Lunet N. Questionnaire design and the recall of pharmacological treatments: a systematic review. *Pharmacoeconomics and Drug Safety.* 2009;18:175–87.
11. Gama H, Correia S, Lunet N. Effect of questionnaire structure on recall of drug utilization in a population of university students. *BMC Medical Research Methodology.* 2009;9:45.
12. Quigley MA, Hockley C, Davidson LL. Agreement between hospital records and maternal recall of mode of delivery: evidence from 12 391 deliveries in the UK Millennium Cohort Study. *BJOG.* 2007;114:195–200.
13. Rice F, Lewis A, Harold G, et al. Agreement between maternal report and antenatal records for a range of pre and peri-natal factors: the influence of maternal and child characteristics. *Early Hum Dev.* 2007;83:497–504.
14. Troude P, L'Helias LF, Raison-Boulley AM, et al. Perinatal factors reported by mothers: do they agree with medical records? *Eur J Epidemiol.* 2008;23:557–64.
15. Hessel NA, Missett B, Fuentes-Afflick E. Lower agreement on behavioral factors than on medical conditions in self-reported data among pregnant Latina women. *Arch Med Res.* 2004;35:241–5.
16. Eder C, Fullerton J, Benroth R, et al. Pragmatic strategies that enhance the reliability of data abstracted from medical records. *Appl Nurs Res.* 2005;18:50–4.
17. Yawn BP, Wollan P. Interrater reliability: completing the methods description in medical records review studies. *Am J Epidemiol.* 2005;161:974–7.
18. Reisch LM, Fosse JS, Beverly K, et al. Training, quality assurance, and assessment of medical record abstraction in a multisite study. *Am J Epidemiol.* 2003;157:546–51.
19. Altman D. *Practical statistics for medical research.* London: Chapman and Hall; 1991.
20. Abramson JH. WINPEPI (PEPI-for-Windows): computer programs for epidemiologists. *Epidemiol Perspect Innov.* 2004;1:6.
21. Diagnosis and classification of diabetes mellitus. *Diabetes Care.* 2009; 32(Suppl 1):S62–7.
22. Bensen JT, Liese AD, Rushing JT, et al. Accuracy of proband reported family history: the NHLBI Family Heart Study (FHS). *Genet Epidemiol.* 1999;17:141–50.
23. Kee F, Tiret L, Robo JY, et al. Reliability of reported family history of myocardial infarction. *BMJ.* 1993;307:1528–30.
24. Murabito JM, Nam BH, D'Agostino RB, et al. Accuracy of offspring reports of parental cardiovascular disease history: the Framingham Offspring Study. *Ann Intern Med.* 2004;140:434–40.
25. Ramadhani TA, Canfield MA, Waller DK, et al. Medical records vs. interview responses: a comparative analysis of selected variables for linked birth defect cases. *Birth Defects Res A Clin Mol Teratol.* 2004;70:592–6.
26. Macedo ME, Lima MJ, Silva AO, et al. Prevalence, awareness, treatment and control of hypertension in Portugal: the PAP study. *J Hypertens.* 2005;23:1661–6.
27. Scuteri A, Najjar SS, Orru M, et al. Age- and gender-specific awareness, treatment, and control of cardiovascular risk factors and subclinical vascular lesions in a founder population: the SardiNIA Study. *Nutr Metab Cardiovasc Dis.* 2009;19:532–41.
28. Gorber SC, Tremblay M, Moher D, et al. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes Rev.* 2007;8:307–26.
29. Brunner Huber LR. Validity of self-reported height and weight in women of reproductive age. *Matern Child Health J.* 2007;11:137–44.
30. Lederman SA, Paxton A. Maternal reporting of prepregnancy weight and birth outcome: consistency and completeness compared with the clinical record. *Matern Child Health J.* 1998;2:123–6.
31. Oliveira AF, Gadelha AM, Leal Mdo C, et al. Study of validity in self-reported weight and height among pregnant women treated at municipal maternity hospitals in Rio de Janeiro, Brazil. *Cad Saude Publica.* 2004;20 Suppl 1: S92–100.
32. Schieve LA, Perry GS, Cogswell ME, et al. Validity of self-reported pregnancy delivery weight: an analysis of the 1988 National Maternal and Infant Health Survey. NMIHS Collaborative Working Group. *Am J Epidemiol.* 1999;150:947–56.
33. Craig BM, Adams AK. Accuracy of body mass index categories based on self-reported height and weight among women in the United States. *Matern Child Health J.* 2009;13:489–96.
34. Gilbert EH, Lowenstein SR, Koziol-McLain J, et al. Chart reviews in emergency medicine research: where are the methods? *Ann Emerg Med.* 1996;27:305–8.
35. Worster A, Bledsoe RD, Cleve P, et al. Reassessing the methods of medical record review studies in emergency medicine research. *Ann Emerg Med.* 2005;45:448–51.