Human Reproduction, Vol.28, No.3 pp. 819–827, 2013
Advanced Access publication on November 21, 2012 doi:10.1093/humrep/des405

human reproduction

ORIGINAL ARTICLE *Reproductive epidemiology*

Validity of self-reported data on pregnancies for childhood cancer survivors: a comparison with data from a nationwide population-based registry

A. Overbeek[1,2,*], M.H. van den Berg[1], C.W.P.M. Hukkelhoven[3], L.C. Kremer[4], M.M. van den Heuvel-Eibrink[5], W.J.E. Tissing[6], J.J. Loonen[7], A.B. Versluys[8], D. Bresters[9], G.J.L. Kaspers[1], C.B. Lambalk[2], F.E. van Leeuwen[10], and E. van Dulmen-den Broeder[1] on behalf of the DCOG LATER/VEVO Study Group

[1]Division of Pediatric Oncology/Hematology, Department of Pediatrics, VU University Medical Center, PO Box 7057, Amsterdam 1000 MB, The Netherlands [2]Department of Obstetrics and Gynecology, VU University Medical Center, PO Box 7057, Amsterdam 1000 MB, The Netherlands [3]The Netherlands Perinatal Registry, Mercatorlaan 1200, Utrecht 3528 BL, the Netherlands [4]Department of Pediatric Oncology, Emma Children's Hospital/Academic Medical Center, PO Box 22660, Amsterdam 1100 DD, The Netherlands [5]Department of Pediatric Oncology, Sophia Children's Hospital/Erasmus MC University Medical Center, PO Box 2060, Rotterdam 3000 CB, The Netherlands [6]Department of Pediatric Oncology, Beatrix Children's Hospital/University Medical Center Groningen, PO Box 30001, Groningen 9700 RB, The Netherlands [7]Department of Pediatric Oncology, Radboud University Nijmegen Medical Center, PO Box 9101, Nijmegen 6500 HB, The Netherlands [8]Department of Pediatric Oncology, Wilhelmina's Children's Hospital, University Medical Center Utrecht, PO Box 85090, Utrecht 3508 AB, The Netherlands [9]Leiden University Medical Center, Willem-Alexander Children's Hospital, PO Box 9600, Leiden 2300 RC, The Netherlands [10]Department of Epidemiology, Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam 1066 CX, The Netherlands

*Correspondence address. Tel: +31-20-444-1409; Fax: +31-20-444-5122; E-mail: a.overbeek@vumc.nl

Submitted on August 15, 2012; resubmitted on October 9, 2012; accepted on October 11, 2012

STUDY QUESTION: To what degree do records registered in the Netherlands Perinatal Registry (PRN) agree with self-report in a study questionnaire on pregnancy outcomes in childhood cancer survivors (CCSs)?

SUMMARY ANSWER: This study suggests that self-reported pregnancy outcomes of CCSs agree well with registry data and that outcomes reported by CCSs agree better with registry data than do those of controls.

WHAT IS KNOWN ALREADY: Many studies have shown that childhood cancer treatment may affect fertility outcomes in female CCSs; however, these conclusions were often based on questionnaire data, and it remains unclear whether self-report agrees well with more objective sources of information.

STUDY DESIGN, SIZE, DURATION: In an nationwide cohort study on fertility (inclusion period January 2008 and April 2011, trial number: NTR2922), 1420 CCSs and 354 sibling controls were invited to complete a questionnaire regarding socio-demographic characteristics and reproductive history. In total, 879 CCSs (62%) and 287 controls (81%) returned the questionnaire.

PARTICIPANTS/MATERIALS, SETTING, METHODS: The current validation study compared the agreement between pregnancy outcomes as registered in the PRN and self-reported outcomes in the study questionnaire. A total of 589 pregnancies were reported in CCSs, and 300 pregnancies in sibling controls, of which 524 could be linked to the PRN.

MAIN RESULTS AND THE ROLE OF CHANCE: A high intra-class correlation coefficient (ICC) was found for birthweight (BW) (0.94 and 0.87 for CCSs and controls, respectively). The self-reported BWs tended to be higher than reported in the PRN. For gestational age (GA), the ICC was high for CCSs (0.88), but moderate for controls (0.49). CCSs overestimated GA more often than controls. The Kappa values for method of conception and for method of delivery were moderate to good. Multilevel analyses on the mean difference with regard to BW and GA showed no differences associated with time since pregnancy or educational level.

LIMITATIONS, REASONS FOR CAUTION: Not all pregnancies reported could be linked to the registry data. In addition, the completeness of the PRN could not be assessed precisely, because there is no information on the number of missing records. Finally, for some outcomes there were high proportions of missing values in the PRN registry.

WIDER IMPLICATIONS OF THE FINDINGS: Our study suggests that questionnaires are a reliable method of data collection, and that for most variables, self-report agrees well with registry data.

STUDY FUNDING/COMPETING INTEREST: This work was supported by the Dutch Cancer Society (grant no. VU 2006-3622) and by Foundation Children Cancer Free. None of the authors report a conflict of interest.

TRIAL REGISTRATION NUMBER: NTR2922 http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=2922.

Key words: validity of self-reported data / pregnancy outcome / long-term survivors / childhood cancer

Introduction

Advances in childhood cancer treatment over the past decades have significantly improved survival, resulting in a rapidly growing number of childhood cancer survivors (CCSs). However, in females there is evidence that both chemo- and radiotherapy may adversely affect reproductive function (Bath, 2002; Green *et al.*, 2002; Reulen *et al.*, 2009; Signorello, 2010). In studies conducted so far, primary outcomes considered in the field of female reproductive function of CCSs after anti-cancer treatment have included ovarian function, premature menopause, uterus function, actual fertility, pregnancy outcomes or a combination of these outcomes. The majority of studies have been large cohort studies in which the outcomes of interest were obtained through interviews or mailed questionnaires. However, the validity of such self-reported data may be limited, and may potentially lead to biased results.

Several studies have assessed the accuracy and validity of self-reported data regarding pregnancy outcomes in healthy women. It appears that correlations between medical charts and maternal self-report are high for birthweight (BW), gestational age (GA) and method of delivery (Olson *et al.*, 1997; Rice *et al.*, 2007). Antenatal and perinatal complications and time to pregnancy, however, are reported less accurately (Coolman *et al.*, 2010; Cooney *et al.*, 2009; Olson *et al.*, 1997; Rice *et al.*, 2007).

We performed a validation study to assess the accuracy of pregnancy outcomes reported by CCSs and sibling controls in a mailed questionnaire.

Materials and Methods

Study population

The current study is part of a Dutch nationwide study on reproductive function, ovarian reserve, premature menopause and pregnancy outcomes in CCSs, the so-called DCOG LATER-VEVO study. The study population, procedures and data collection methods have been described previously in detail (Overbeek *et al.*, 2012). In short, CCSs eligible for the nationwide study were selected from a cohort of patients treated for childhood cancer at one of the seven Dutch pediatric oncology and stem cell transplant centers between 1963 and 2002. The collaborative group for long-term effects after childhood cancer (LATER) has designed and implemented an electronic database, in each center, which includes patient and treatment details of all patients treated for cancer before the age of 18 years. The inclusion criteria for the DCOG LATER-VEVO study and the current study were identical and were defined as having been treated for a malignancy or central nervous system tumor before the age of 18, having survived for at least 5 years after diagnosis, being alive and being at least 18 years at study entry. The DCOG LATER-VEVO study was limited to female survivors. Women were excluded if they were not able to speak or read Dutch and if they had severe mental sequelae. All CCSs were asked for permission to invite their sister(s) for participation in the control group of the DCOG LATER-VEVO study. Eligible sisters were never diagnosed with cancer, had to be able to read and speak Dutch and had to be 18 years or older. The DCOG LATER-VEVO study consisted of three parts: a questionnaire, the provision of a blood sample and a transvaginal ultrasound of the reproductive organs, with the two latter parts requiring a hospital visit. Eligible CCSs and sibling controls could decide either to refuse or to participate and, in case of participation, whether to take part in one, two or all three parts of the study. Data collection for the current study took place between January 2008 and April 2011. Women were eligible for the present study if they reported in the questionnaire ever having been pregnant. The exclusion criteria were pregnancy terminated before 24 weeks and date of birth of the offspring before 1 January 1985 or after 31 December 2009.

Data collection

For the purpose of the current study, only the questionnaire data were taken into account. The questionnaire is an adaptation of a well-tested questionnaire used by the Department of Epidemiology of the Netherlands Cancer Institute in a Dutch cohort study on long-term effects of ovarian stimulation for IVF (de Boer *et al.*, 2003; van Leeuwen *et al.*, 2011). The questionnaire addresses (amongst others) the following issues: socio-demographic characteristics, wish to have children and reproductive history, and for each pregnancy, data on maternal age, method of conception, duration of pregnancy, complications, pregnancy outcomes and details of the offspring.

Medical information regarding pregnancy outcomes was derived from the Netherlands Perinatal Registry (PRN), a nation-wide population-based registry, in which data of three medical registries (midwives, obstetricians and pediatricians/neonatologists) are combined. After delivery, standardized digital forms are entered in the nationwide database. These items are recorded by the caregiver, who is provided a standard manual with additional information on the definitions. The data are sent annually to the national registry office, where a number of range and consistency checks are conducted. False records are sent back to the caregiver with a request to correct them. Data in this registry are available from 1985 onwards. As from 1999, the PRN has included ~95% of all ~180 000 deliveries at >16 completed weeks of gestation in the Netherlands (The Netherlands Perinatal Registry/Stichting Perinatale Registratie Nederland, 2011). The missing 5% is due to the fact that some midwives

and general practitioners involved in obstetric care do not provide data for the PRN registry or because the records sent from the midwife practices are not received properly by the PRN. Before 1999, the PRN registry was less complete; however, exact proportions of completeness are not known as these records were not linked to Statistics Netherlands. In the PRN registry, the following variables are recorded: method of conception (natural, hormonal stimulation, IUI, controlled ovarian hyperstimulation, IVF and other), method of delivery, BW (in grams), GA (in weeks) and highest diastolic blood pressure (in mmHg). In the DCOG LATER-VEVO questionnaire, women reported on method of conception (natural, hormonal stimulation, IUI, IVF/ICSI), complications during pregnancy (hypertension and growth retardation of the child), method of delivery (spontaneous, assisted delivery, induction of labor, Caesarean section), date of birth of the baby, GA, sex and BW. All participants gave written informed consent for data abstraction from medical records. The records of self-reported pregnancies from the questionnaire were linked to the PRN by using both the mother's date of birth and the child's date of birth as linkage keys. If linkage led to multiple hits, the sex of the baby and the postal code of the mother were used in an attempt to correctly link self-reported data with PRN data. If it was not possible to link self-reported data with a unique corresponding record in the PRN, the data were not included in the current study.

Statistical analysis

The data were checked for a normal distribution. Data of continuous variables are presented as the mean [standard deviation (SD)] if normally distributed or as the median and interquartile range (IQR) if not normally distributed. In case CCSs and controls were compared, an independent Student's *t*-test was used when data were normally distributed. Mann–Whitney *U*-test was used to compare CCSs and controls when data were not normally distributed. We calculated intra-class correlation coefficients (ICCs) and confidence intervals for the continuous variables, BW and GA. ICCs were calculated using a two-way random effects model. BW and GA were categorized [BW: very low birthweight (<1500 g), low birthweight (1500–2500 g), normal birthweight (2500–4000 g) and high birthweight (>4000 g); GA: preterm (<37 weeks), term (37–42 weeks), and post-term (>42 weeks)]. For BW, GA and method of delivery, the PRN was considered the gold standard. Sensitivity, as well as specificity, was calculated. Sensitivity was defined as the proportion of those with the condition (as defined by the PRN) who are correctly classified by the questionnaire. Specificity was the proportion of those without the condition (as defined by the PRN), who are correctly classified by the questionnaire. For method of conception and pregnancy complications, the PRN could not serve as the gold standard, since these variables are not registered consistently. Therefore, for these variables, only reliability measures were calculated. In order to assess the agreement between self-reported data and PRN data for categorical variables, the proportion of overall agreement and Cohen Kappa statistics were calculated. The proportion of overall agreement, which is the proportion of cases for which PRN and self-report agree, is a crude descriptive measure that is informative, useful and easy to interpret, but does not distinguish between agreement on positive ratings and agreement on negative ratings and does not take into account chance. Therefore we also calculated Kappa values. According to Landis and Koch, variables with values of Kappa >0.75 can be considered as an excellent agreement, values of 0.40–0.75 as a moderate agreement, and values below 0.40 as a poor agreement (Landis and Koch, 1977). In case a woman had reported more than one pregnancy, only the first reported pregnancy was used for the calculation of validity and reliability outcomes in order to avoid dependency of observations. To determine which variables were independently associated with overall agreement, multilevel analysis was performed, allowing the correction for the clustering of pregnancies for one woman. Variables included in this analysis were time between delivery and questionnaire, maternal age and educational level. Analyses were performed using SPSS software (version 15.0, SPSS, Inc., Chicago, IL) and MLWin (version 2.24, Centre for Multilevel Modelling, University of Bristol).

Results

During the inclusion period of the current study, 1420 CCSs and 354 sibling controls were invited for the DCOG LATER-VEVO study. In total, 879 CCSs (62%) and 287 controls (81%) returned the DCOG LATER-VEVO questionnaire. In 289 CCSs, 589 pregnancies were reported and 300 pregnancies were reported in 123 controls. There were 160 pregnancies in the survivor group and 67 pregnancies in the control group that had to be excluded because the date of birth was before 1 January 1985 or after 31 December 2009 or unknown and/or the pregnancy was terminated before 24 weeks of GA. This resulted in 429 pregnancies of CCSs and 233 pregnancies of controls included in the study. Linkage to the PRN database yielded 488 unique hits. In 37 cases linkage led to multiple hits in the PRN. All but one could subsequently be identified by the sex of the baby and/or by the postal code of the mother. No matching records were found in the PRN database for 72 pregnancies in the survivor group and 65 pregnancies in the control group. Finally, 357 pregnancies reported by 218 CCSs and 167 pregnancies reported by 105 controls were included in the data-analyses (Fig. 1). There were 78 deliveries (28%) of CCSs and 49 pregnancies (29%) of controls which took place before 1999. Of those that could not be linked, 22 (35%) and 17 (39%) pregnancies of CCSs and controls, respectively, took place before 1999. Table I presents the basic characteristics of the CCSs and the controls for the included group as well as of those who could not be linked to the PRN records. The mean ages at completion of the questionnaire (SD) were 34.5 (6.3) and 36.7 (7.0) years for CCSs and controls, respectively, for those who could be included in the validation study ($P = 0.03$) and 36.8 (7.0) and 38.1 (6.5) years for CCSs and controls who could not be linked to PRN records, respectively. The mean maternal age at delivery (SD) was 29.2 (4.3) years for CCSs, whereas controls were 28.8 (4.0) years ($P = 0.47$), comparable to the age of those who could not be linked, namely 28.4 (5.1) and 28.3 (4.1) for CCSs and controls, respectively. The median durations (IQR) from child birth to questionnaire were 4.7 (10.7) and 5.5 (14.5) for CCSs and controls, respectively ($P = 0.05$), whereas in the non-linked group these were 7 (10) and 7 (9.8) for CCSs and controls, respectively.

The ICC between self-report and registry regarding BW was high [0.94 (95% confidence interval: 0.91, 0.96) and 0.87 (95% confidence interval: 0.83, 0.90) for CCSs and controls, respectively]. For GA, the ICC was also high for CCSs (0.88, 95% confidence interval: 0.85, 0.91), but moderate for controls (0.49, 95% confidence interval: 0.32, 0.62).

In Tables II and III, the results of the comparison between self-reported data and PRN data regarding various categorical pregnancy outcomes are presented. For the categories of low and normal BW, the Kappa values in the survivor group were high (0.87 and 0.79, respectively); however, the corresponding values were moderate in the control group (0.42 and 0.61, respectively). The high BW category

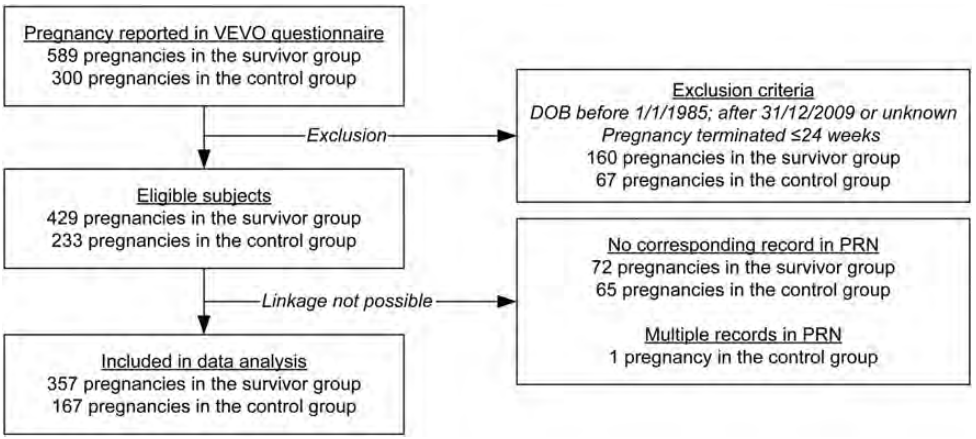

Figure 1 Description of the recruitment of eligible women from the DCOG LATER-VEVO study population. DOB, date of birth.

Table 1 Characteristics of participants of the DCOG LATER-VEVO Study eligible for the validation study, The Netherlands, 2008–2011.

Characteristic	Included in the validation study				Record not linked with PRN			
	CCSs (n = 218)		Controls (n = 105)		CCSs (n = 63)		Controls (n = 44)	
	n	%	*n*	%	*n*	%	*n*	%
Age (years) at questionnaire								
<30	58	26.6	19	18.1	15	23.8	3	6.8
30–35	66	30.3	29	27.6	13	20.6	14	31.8
35–40	54	24.8	30	28.6	18	28.6	16	36.4
>40	40	18.3	27	25.7	17	27.0	11	25.0
Maternal age (years) at first born								
<25	44	20.2	25	23.8	14	22.2	11	25.0
25–30	107	49.1	39	37.1	29	46.0	16	36.4
30–35	59	27.1	3	31.4	16	25.3	15	34.1
>35	8	3.7	8	7.6	4	6.3	2	4.5
Duration (years) from child birth to questionnaire								
<2	50	22.9	20	19	17	27.0	5	11.4
2–5	69	31.7	22	21	12	19.0	11	25
5–8	38	17.4	23	21.9	8	12.7	9	20.5
>8	61	28.0	39	37.1	26	41.3	19	43.1
Maternal educational level[a]								
Lowest	5	2.3	0	0	0	0	0	0
Low	38	17.4	15	14.3	7	11.1	5	11.4
Medium	103	47.2	46	43.8	31	49.2	19	43.1
High	72	33.0	44	41.9	25	39.7	20	45.5

[a]Educational level: lowest (special education), low (primary school), medium (secondary school) and high (college or university).

was scored with a moderate agreement (0.71 and 0.62 in CCSs and controls, respectively). BWs reported in the questionnaire tended to be higher than those reported in the PRN. Of the 23 discrepancies in the survivor group, BW was overestimated by self-report in 19 cases (83%). In the control group, self-report overestimated the BW category in 13 of the 16 cases (81%). For this category, there were no missing values. In both groups, categories of GA were reported with a moderate agreement (preterm delivery: 0.79 and 0.72; term delivery: 0.75 and 0.65, for CCSs and controls, respectively). When evaluating discrepancies between self-report and registry data, GA in

Table II Comparison of the self-reported data from the DCOG LATER-VEVO questionnaire with data from the Netherlands Perinatal Registry regarding various pregnancy outcomes in childhood cancer survivors.

	Reported in Netherlands Perinatal Registry					Agreement	Kappa value	95% CI
Birthweight								
Reported in questionnaire	Very low birthweight (<1500 g)	Low birthweight (1500–2500 g)	Normal birthweight (2500–4000 g)	High birthweight (>4000 g)	Missing			
Very low birthweight (<1500g)	3	0	0	0	0	98	0.59	0.23–0.95
Low birthweight (1500–2500 g)	0	19	4	0	0	98	0.87	0.76–0.98
Normal birthweight (2500–4000 g)	0	1	145	0	0	91	0.79	0.70–0.88
High birthweight (>4000 g)	4	0	14	28	0	92	0.71	0.59–0.83
Gestational age								
Reported in questionnaire	Preterm (<37 weeks)	Term (37–42 weeks)	Post-term (>42 weeks)	Missing				
Preterm (<37 weeks)	40	6	0	0		93	0.79	0.69–0.89
Term (37–42 weeks)	10	159	0	0		91	0.75	0.65–0.86
Post-term (>42 weeks)	0	3	0	0		99	n/a	n/a
Method of conception								
Reported in questionnaire	Natural	Hormonal	IUI	IVF/ICSI	Missing			
Natural	151	2	0	0	43	97	0.82	0.67–0.97
Hormonal	1	2	0	0	2	98	0.56	0.12–1.01
IUI	0	0	3	1	3	99	0.85	0.57–1.14
IVF/ICSI	2	0	0	7	1	98	0.81	0.61–1.02
Method of delivery								
Reported in questionnaire	Spontaneous	Vacuum/forcipal extraction	Caesarean section	Missing				
Spontaneous	96	8	1	21		88	0.76	0.67–0.85
Vacuum/forcipal extraction	2	41	0	1		95	0.86	0.77–0.94
Caesarean section	12	0	36	0		93	0.81	0.71–0.91
Pregnancy complications								
Reported in questionnaire	Hypertension	No hypertension	Missing					
Hypertension	28	17	3			90	0.59	0.45–0.73
No hypertension	10	136	24			90	0.59	0.45–0.73

the CCSs group was often overestimated in the self-reported questionnaire [13 of 19 cases (68%)]. However, within the control group GA was overestimated as frequently as underestimated (4 versus 4 cases).

The Kappa value for method of conception varied largely per method (0.82 and 0.71 for natural conception, 0.56 and 0.66 for controlled ovarian hyperstimulation, 0.81 and 0.85 for IVF/ICSI, for CCSs and controls, respectively). When examining the discrepancies between self-report and PRN with regard to method of conception, no specific direction of misclassification could be discerned in the survivor group or in the control group. Moreover, 49 of the 218 cases (22%) in the survivor group were missing in the PRN registry, 43 of which were self-reported as natural conceptions. For this category in the control group, 33 of 105 cases (31%) were missing in the PRN, of which 32 cases were self-reported natural conception.

Table III Comparison of the self-reported data from the DCOG LATER-VEVO questionnaire with data from the Netherlands Perinatal Registry regarding various pregnancy outcomes in controls.

	Reported in Netherlands Perinatal Registry					Agreement	Kappa value	95% CI
Birthweight								
Reported in questionnaire	Very low birthweight (<1500 g)	Low birthweight (1500–2500 g)	Normal birthweight (2500–4000 g)	High birthweight (>4000 g)	Missing			
Very low birthweight (<1500 g)	1	0	0	0	0	100	1.00	1.00–1.00
Low birthweight (1500–2500 g)	0	2	3	0	0	95	0.42	0.004–0.84
Normal birthweight (2500–4000 g)	0	1	73	0	0	86	0.61	0.44–0.78
High birthweight (>4000 g)	0	1	11	13	0	89	0.62	0.44–0.81
Gestational age								
Reported in questionnaire	Preterm (<37 weeks)	Term (37–42 weeks)	Post-term (>42 weeks)	Missing				
Preterm (<37 weeks)	9	4	0	0		94	0.72	0.50–0.93
Term (37–42 weeks)	2	87	0	1		92	0.65	0.43–0.87
Post-term (>42 weeks)	0	2	0	0		98	n/a	n/a
Method of conception								
Reported in questionnaire	Natural	Hormonal	IUI	IVF/ICSI	Missing			
Natural	65	1	0	0	32	96	0.71	0.39–1.02
Hormonal	0	1	0	0	1	99	0.66	0.04–1.28
IUI	1	0	0	0	0	99	n/a	n/a
IVF/ICSI	1	0	0	3	0	99	0.85	0.56–1.14
Method of delivery								
Reported in questionnaire	Spontaneous	Vacuum/forcipal extraction	Caesarean section	Missing				
Spontaneous	50	0	0	25		94	0.86	0.75–0.98
Vacuum/forcipal extraction	1	14	0	0		98	0.92	0.81–1.03
Caesarean section	4	1	10	0		94	0.77	0.57–0.96
Pregnancy complications								
Reported in questionnaire	Hypertension	No hypertension	Missing					
Hypertension	8	6	1			90	0.61	0.37–0.85
No hypertension	2	64	24			90	0.61	0.37–0.85

For the method of delivery, the Kappa values were 0.76 and 0.86 for spontaneous delivery, 0.86 and 0.92 for vacuum or forcipal extraction and 0.81 and 0.77 for Caesarean section in CCSs and controls, respectively. CCSs reported the method of delivery as spontaneous in nine cases, whereas PRN stated additional techniques were used to enable the delivery (vacuum or forcipal extraction or even Caesarean section). Conversely, in 14 cases, the survivor reported an assisted delivery, while PRN records showed a spontaneous delivery. Of the six discrepancies between PRN and self-report regarding the method of delivery in the control group, all were due to controls reporting an assisted delivery, while PRN mentioned a spontaneous delivery. With regard to the method of delivery, 22 cases in the survivor group (10%) and 25 cases in the control group (24%) were missing in the PRN data, 21 (95%) and 25 (100%) of which, respectively, were self-reported spontaneous deliveries. The Kappa values for pregnancy-induced hypertension were 0.59 for CCSs and 0.61 for controls. For this variable there were 27 discrepancies between PRN and self-report in the survivor group and 8 in the control group, the direction of which was non-specific. In the survivor group, 27 records were missing in the PRN, 3 of which were reported as hypertension by self-report. In the control group, 25 records were missing in the PRN registry, one of which was reported as hypertension in the questionnaire.

In Table IV, validity measures are presented for BW, GA and method of delivery. Sensitivity ranged from 42.9 to 100% in CCSs, and from 33.3 to 100% in controls. Specificity was good, ranging

Table IV Measures of validity of birthweight, gestational age and method of delivery.

	CCSs			Controls		
	Positive predictive value (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Sensitivity (%)	Specificity (%)
Birthweight						
Very low birthweight (<1500 g)	100	42.9	100	100	100	100
Low birthweight (1500–2500 g)	82.6	95.0	98.0	40.0	33.3	97.0
Normal birthweight (2500–4000 g)	99.3	89.0	98.2	98.7	83.9	94.4
High birthweight (>4000 g)	60.9	100	90.5	52.0	100	87.0
Gestational age						
Preterm (<37 weeks)	87.0	80	96.4	69.2	81.8	95.7
Term (37–42 weeks)	94.1	94.6	80.0	97.8	93.6	81.8
Post-term (>42 weeks)	n/a	n/a	98.0	n/a	n/a	98.1
Method of delivery						
Spontaneous	91.4	87.2	89.5	100	90.9	100
Vacuum/forcipal extraction	95.4	83.7	98.6	93.3	93.3	98.5
Caesarean section	75.0	97.3	92.5	66.7	100	92.7

Data from the Netherlands Perinatal Registry were considered gold standard.

from 89.5 to 100% in CCSs and from 81.8 to 100% in controls. In addition, the influence of educational level, maternal age and the time between delivery and questionnaire on the difference in BW and GA between self-reported and registered data was assessed by multilevel analysis. None of these factors were associated with a less accurate recall of pregnancy outcomes. More specifically, there were no large differences between self-reported and registered data regarding BW and GA with longer time since birth, lower educational level or higher maternal age.

Discussion

In this study, the agreement between self-reported pregnancy outcomes and registry data in female CCSs and controls was examined. Our results show that the validity of pregnancy outcomes reported by CCSs is good for GA and BW. However, sibling controls reported GA with only a moderate agreement. CCSs as well as controls tended to overestimate BW and CCSs more often overestimated GA. Hypertension in pregnancy and controlled ovarian hyperstimulation were reported with moderate-to-good agreement, whereas the method of conception and delivery was reported with a good agreement. The sensitivity and specificity were high for CCSs and controls for BW, GA and method of delivery. A higher maternal age at child birth, a longer time since pregnancy and a lower educational level were not associated with lower agreement in either group. Overall, it seemed that some self-reported pregnancy outcomes of CCSs (specifically birthweight and gestational age) agreed better with the registry data than those reported by controls, indicating a potential source of differential misclassification. This might be due to the increased awareness of late effects and a higher frequency of medical follow-up among CCSs.

To our knowledge, no studies have been conducted to assess the validity of self-reported data on pregnancy outcomes in CCSs. However, with respect to the self-report of late effects in general, it has been found that CCSs report a wide range of late effects in significantly greater numbers than recorded in medical notes (Schwartz *et al.*, 2010; Taylor *et al.*, 2010). Furthermore, CCSs appear to show a biased response style, indicating a systematic tendency to deny difficulties on QOL measures (O'Leary *et al.*, 2007). Olson *et al.* conducted a study in which mothers of CCSs were interviewed on pregnancy and delivery information. The authors concluded that the validity and reliability of maternally reported pregnancy and delivery information may vary with the nature of the factor of interest (i.e. BW, pregnancy complications etc.), but is affected little by time from birth or case–control status (Olson *et al.*, 1997). In concordance with our study, high correlations for BW and GA were found. However, self-report by mothers of CCSs on pregnancy-induced hypertension and method of delivery had low validity and reliability scores, whereas in our study, there seem to be moderate-to-high agreement scores. The study described by Olson *et al.* took place in the 80s, whereas our study was conducted >20 years later. The differences in the validity of self-reported data on pregnancy complications and medical interventions may be due to the different time frame in which both studies were conducted. Possibly, the level of communication between the patient and her physician has changed in this period, leading to a better recall of pregnancy complications and medical interventions in our study. Finally, the way in which the method of delivery and medical conditions during pregnancy were questioned may have differed between the study of Olson *et al.* and our study, which may also have caused differences in accuracy. Rice *et al.* (2007) evaluated the agreement between maternal report and medical records in women who gave birth to a child following IVF. Correlations between self-report and medical records for BW were comparable to our results. The Kappa value for delivery via Caesarean section, however, was 1.00, in contrast to 0.81 in our study. In the study of Rice *et al.*, a distinction was made between Caesarean

sections in general and emergency Caesarean sections. The Kappa value for the emergency Caesarean sections in the study of Rice *et al.* was 0.78. Moreover, the authors pointed out that their results may overestimate recall rates of pregnancy outcomes in the general population, because women who were pregnant after IVF treatment may better recall pregnancy-related outcomes (Rice *et al.*, 2007). In accordance with the study of Rice *et al.*, Tomeo *et al.* and Sou *et al.* described a validity of 100% for the report of Caesarean sections in a group of women recruited from the general population, whereas induction of labor and forcipal or vacuum extraction were reported less sensitively (sensitivity and specificity of 93 and 86% for induction of labor and 26 and 74% for assisted delivery) (Tomeo *et al.*, 1999; Sou *et al.*, 2006).

When interpreting our results, the limitations of this study should be considered.

First, we were not able to link 17% (72/429) of the pregnancies in the survivor group and 28% (65/233) in the control group. In other recent studies, higher linkage rates have been reached (CWPM Hukkelhoven, Perinatal Registry Utrecht, personal communication, 2012). In these studies, in which recent pregnancies are evaluated, not only date of birth of the child, but also BW, GA and postal code could be used as linkage keys. Due to the objective of our study, i.e. investigating whether self-reported pregnancy outcomes such as BW and GA agree with those in the registry, it was not possible to use these outcomes as linkage keys. Postal code could often not be used, as it was likely that participants who reported a pregnancy from many years ago may have changed address since they were pregnant. It could also be that due to input errors of birth dates, either in the registry or in the self-reported data, some pregnancies could not be linked.

Secondly, the completeness of the PRN could not be assessed precisely, because there is no information on the number of missing records. Records in the PRN database agree with records of Statistics Netherlands for 95% and slightly less in earlier years (The Netherlands Perinatal Registry/Stichting Perinatale Registratie Nederland, 2011). However, the database of Statistics Netherlands can also not be considered complete, as it contains no data on births among women who stay illegally in the Netherlands. In this study, pregnancy records were available from 1985 onwards, and it may very well be possible that many of the older records in the PRN could not be linked due to incompleteness of the registry or due to recall bias of the participants. Indeed, the proportion of records in which date of birth was reported before 1999 was larger for non-linked records than for linked records.

Thirdly, for BW, GA and method of delivery, the PRN database was considered the gold standard. However, 12% of the records of the CCSs and 24% of those of the controls were missing in the PRN with regard to the method of delivery. Agreement on Caesarean section was only 0.77 in controls and 0.81 in CCSs. In other studies, recall of Caesarean section often agrees for the full 100% (Tomeo *et al.*, 1999; Sou *et al.*, 2006; Rice *et al.*, 2007). It is unlikely that women forget (to report) such an operation. As Kappa values are < 0.8, one could therefore cautiously question the quality of the PRN registry. Unfortunately, input errors in the PRN cannot be excluded, nor can they be quantified, as, for this study, there was no alternative source of pregnancy records available.

Finally, for many cases, the categories of the outcomes 'method of delivery' and 'method of conception' had missing values in the PRN. This can be explained by the fact that providing information on method of conception was only mandatory for gynecologists; for midwives this was optional. This may lead to a differential misclassification, as pregnancies under gynecologic supervision are often more complicated than those supervised by midwives. In this way, it may seem that an alternative method of conception, other than spontaneous, influences pregnancy outcomes and complications.

In conclusion, our results indicate that for the most important outcomes regarding fertility and pregnancy, self-report in CCSs is consistent with registry parameters. The use of questionnaires in CCSs to assess pregnancy outcomes therefore seems justified. However, since we observed differences in accuracy between CCSs and controls, differential misclassification should be considered when interpreting the data. In addition, one should realize that especially BW and GA are more often overestimated in self-reported questionnaires.

Acknowledgements

The authors gratefully acknowledge the assistance of Priyanta Malhoe and Elise Wendel in the data abstraction.

Authors' roles

C.B.L., F.E.L. and E.D.B. designed the study and E.D.B. and M.H.B. were responsible for the grant acquisition. A.O. and M.H.B. acquired, analysed and interpreted all data, and drafted the manuscript. C.W.P.M.H. helped conduct the data analysis of the registry data. G.J.L.K., C.W.P.M.H., C.B.L., F.E.L. and E.D.B. helped revise the manuscript. L.C.K., M.M.H.-E., W.J.E.T., J.J.L., A.B.V. and D.B. provided the study subjects and approved the final version of the manuscript.

Funding

This work was supported by the Dutch Cancer Society (grant no. VU 2006-3622) and by Foundation Children Cancer Free.

Conflict of interest

None declared.

References

Bath LEW. Late effects of the treatment of childhood cancer on the female reproductive system and the potential for fertility preservation. *BJOG* 2002;**109**:107–114.

Coolman M, de Groot CJ, Jaddoe VW, Hofman A, Raat H, Steegers EA. Medical record validation of maternally reported history of preeclampsia. *J Clin Epidemiol* 2010;**63**:932–937.

Cooney MA, Buck Louis GM, Sundaram R, McGuiness BM, Lynch CD. Validity of self-reported time to pregnancy. *Epidemiology* 2009; **20**:56–59.

de Boer EJ, den Tonkelaar I, te Velde ER, Burger CW, van Leeuwen FE. Increased risk of early menopausal transition and natural menopause after poor response at first IVF treatment. *Hum Reprod* 2003; **18**:1544–1552.

Green DM, Whitton JA, Stovall M, Mertens AC, Donaldson SS, Ruymann FB, Pendergrass TW, Robison LL. Pregnancy outcome of

female survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. *Am J Obstet Gynecol* 2002;**187**:1070–1080.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–174.

O'Leary TE, Diller L, Recklitis CJ. The effects of response bias on self-reported quality of life among childhood cancer survivors. *Qual Life Res* 2007;**16**:1211–1220.

Olson JE, Shu XO, Ross JA, Pendergrass T, Robison LL. Medical record validation of maternally reported birth characteristics and pregnancy-related events: a report from the Children's Cancer Group. *Am J Epidemiol* 1997;**145**:58–67.

Overbeek A, van den Berg MH, Kremer LC, van den Heuvel-Eibrink MM, Tissing WJ, Loonen JJ, Versluys B, Bresters D, Kaspers GJ, Lambalk CB *et al.*. A nationwide study on reproductive function, ovarian reserve, and risk of premature menopause in female survivors of childhood cancer: design and methodological challenges. *BMC Cancer* 2012;**12**:363.

Reulen RC, Zeegers MP, Wallace WH, Frobisher C, Taylor AJ, Lancashire ER, Winter DL, Hawkins MM. Pregnancy outcomes among adult survivors of childhood cancer in the British Childhood Cancer Survivor Study. *Cancer Epidemiol Biomarkers Prev* 2009;**18**:2239–2247.

Rice F, Lewis A, Harold G, van den Bree M, Boivin J, Hay DF, Owen MJ, Thapar A. 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.

Schwartz LA, Mao JJ, Derosa BW, Ginsberg JP, Hobbie WL, Carlson CA, Mougianis ID, Ogle SK, Kazak AE. Self-reported health problems of young adults in clinical settings: survivors of childhood cancer and healthy controls. *J Am Board Fam Med* 2010;**23**:306–314.

Signorello LBM. Stillbirth and neonatal death in relation to radiation exposure before conception: a retrospective cohort study. *Lancet* 2010;**376**:624–630.

Sou SC, Chen WJ, Hsieh WS, Jeng SF. Severe obstetric complications and birth characteristics in preterm or term delivery were accurately recalled by mothers. *J Clin Epidemiol* 2006;**59**:429–435.

Taylor N, Absolom K, Michel G, Urquhart T, Gerrard M, Jenkins A, Lee V, Vora A, Eiser C. Comparison of self-reported late effects with medical records among survivors of childhood cancer. *Eur J Cancer* 2010; **46**:1069–1078.

The Netherlands Perinatal Registry/Stichting Perinatale Registratie Nederland (2011). Perinatal Care in the Netherlands 2008. Utrecht: Stichting Perinatale Registratie Nederland, 2011.

Tomeo CA, Rich-Edwards JW, Michels KB, Berkey CS, Hunter DJ, Frazier AL, Willett WC, Buka SL. Reproducibility and validity of maternal recall of pregnancy-related events. *Epidemiology* 1999; **10**:774–777.

van Leeuwen FE, Klip H, Mooij TM, van de Swaluw AM, Lambalk CB, Kortman M, Laven JS, Jansen CA, Helmerhorst FM, Cohlen BJ *et al.* Risk of borderline and invasive ovarian tumours after ovarian stimulation for in vitro fertilization in a large Dutch cohort. *Hum Reprod* 2011;**26**:3456–3465.

Appendix

Members of the DCOG LATER-VEVO study group include the following: Beerendonk CCM (Radboud University Nijmegen Medical Center), van den Berg MH (VU University Medical Center, Amsterdam), Bökkerink JP (Radboud University Nijmegen Medical Center), van den Bos C (Emma Children's Hospital/Academic Medical Center, Amsterdam), Bresters D (Willem-Alexander Children's Hospital, Leiden University Medical Center), van Dorp W (Sophia Children's Hospital/Erasmus MC University Medical Center, Rotterdam), van Dulmen-den Broeder E (VU University Medical Center, Amsterdam), van Engelen MP (Wilhelmina's Children's Hospital, University Medical Center Utrecht), van den Heuvel-Eibrink MM (Sophia Children's Hospital/Erasmus MC University Medical Center, Rotterdam), Huizinga GA (Beatrix Children's Hospital/University Medical Center Groningen), Jaspers MWM (Academic Medical Center-University of Amsterdam), Kaspers GJL (VU University Medical Center, Amsterdam), Kremer LC (Emma Children's Hospital/Academic Medical Center, Amsterdam), Lambalk CB (VU University Medical Center, Amsterdam), Laven JS (Sophia Children's Hospital/Erasmus MC University Medical Center, Rotterdam), van Leeuwen FE (Netherlands Cancer Institute, Amsterdam), Loonen JJ (Radboud University Nijmegen Medical Center), Louwerens M (Leiden University Medical Center), Overbeek A (VU University Medical Center, Amsterdam), van der Pal HJ (Emma Children's Hospital/Academic Medical Center, Amsterdam), Ronckers CM (Dutch Childhood Oncology Group—Longterm Effects after Childhood Cancer (LATER), Simons AHM (University Medical Center Groningen), Tissing WJE (Beatrix Children's Hospital/University Medical Center Groningen), Tonch N (Academic Medical Center, Amsterdam), Verkerk ECM (Dutch Childhood Oncology Group—Longterm Effects after Childhood Cancer (LATER), and Versluys AB (Wilhelmina's Children's Hospital, University Medical Center Utrecht).