Reliability and Stability of Mothers' Reports about their Pregnancies with Twins

Wendy Reich, Richard D. Todd, Cynthia A. Joyner, Rosalind J. Neuman, and Andrew C. Heath

Washington University, School of Medicine, Department of Psychiatry, St Louis, USA

The objective of this study was to determine if mothers' retrospective reports about events in their pregnancies with twins are reliable and stable. Six hundred and twenty-four mothers completed psychiatric interviews about their twins. These interviews also contained questions about the mothers' pregnancies, the perinatal period, and the child's early development. The mothers reported first on one twin and then on the other with interviews spaced from 3 days to 2 weeks apart. Thus mothers reported on the same pregnancy twice. Of these mothers, 47 were re-interviewed 6 to 18 months later by raters blind to the results of the initial interview. The twin design allowed us to compare the short-term reliability of the 624 mothers' reports of the same pregnancy. The re-interview of the 47 mothers enabled us to compare the stability of reports over a longer time period. Agreement between the reports was measured with the kappa statistic. Kappas were good to excellent for the short-term reports of pregnancy for each twin for the 624 mothers. Kappas were equally high for the 47 mothers that were re-interviewed 6 to 18 months later. Mothers show good reliability and stability of reporting about events during pregnancy.

An accurate account of the mother's pregnancy is important to child psychiatry, as there is evidence that many of these prenatal events influence the child's outcome with respect to specific psychiatric disorders. (Milberger et al., 1996; Williams et al., 1998). Thus information about the mothers' behavior, general health, and wellbeing during pregnancy is important for the psychological as well as the physical assessment of the child.

Information about pregnancy is likely to be most accurate if collected when the mother is pregnant (Richardson et al., 1999; Stathis et al., 1999; Williams et al., 1998). Although this approach has produced some important research (Williams et al., 1998), such studies tend to assess particular aspects of the mothers' pregnancy such as cocaine use or smoking to understand possible effects of these behaviors on the child. Furthermore, this prospective approach is not always feasible due to the relatively long wait between the pregnancy and the appearance of symptoms and disorders. For example, researchers interested in particular disorders would be hard put to assess pregnancy variables, then wait to see which children develop the psychopathology of interest. This is particularly true for disorders that have a relatively late onset.

Some studies with a retrospective design have used information from medical records to assess the validity of maternal recall (Githens et al., 1993). However, collection of medical records particularly in large-scale epidemiological studies may be problematic. More importantly much of the critical data such as the mother's substance use or whether or not she ate healthfully might not be available in these records. For these reasons, many studies have to depend on maternal recall in order to retrieve information about pregnancy. Thus the determination of the reliability and stability of maternal recall about their pregnancies is critical

This study builds on and extends the existing literature by using data from a large epidemiological twin study to measure reliability of stability of mothers' reports on their pregnancies.

Materials and Method

The study from which these data were drawn is a large-scale epidemiological study of twins in the state of Missouri (Hudziak et al., 2000). It was designed to investigate genetic factors in attention-deficit/hyperactivity disorder (ADHD). The MOTWIN study assessed subjects from a birth registry of all twins born in Missouri. Children between the ages of 7 and 17 were assessed with the Missouri Assessment of Genetics Interview for Children (MAGIC). This interview is based on the Diagnostic Interview for Children and Adolescents (DICA; Reich, 2000). A test–retest study of questions in the perinatal section of the DICA including questions about early childhood behavior showed good to excellent agreement as measured by the kappa statistic (Reich, unpublished data). The MAGIC is a semi-structured interview designed to assess common psychiatric disorders of children and adolescents. The instrument makes the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*; American Psychiatric Association, 1994) and *International Classifications of Diseases* (*ICD-10*; World Health Organization, 1992) diagnoses. The interview consists of a child version for ages 7 to 12, an adolescent version for ages 13 to 17, as well as a parent interview about children from both age groups. All three interviews ask the same questions with age-appropriate wording.

Like the DICA, the parent version of the MAGIC has a section that asks perinatal questions as well as the early

Address for correspondence: Wendy Reich, Washington University, School of Medicine, Department of Psychiatry, 660 South Euclid, Campus Box 8134, St. Louis Missouri (MO) 63110 USA. Email: wendyr@twins.wustl.edu

development of the child. Included in the perinatal section are two questions about previous pregnancies and 29 questions that refer to the pregnancy itself. (Table 1).

Parents were interviewed separately by different raters to prevent bias. Interviewers were college graduates with either a bachelor's or master's degree. Most had a background in psychology and a number were already familiar with *DSM* nomenclature. The intensive MAGIC training course took one month to complete. The majority of the interviews were conducted in person, although some were conducted by phone.

Parent interviews on both twins were completed for 72.9% of the families contacted. We eliminated all mothers who had been interviewed about both twins on the same day or two days apart (N = 108) as we felt those interviews would be too close in time to represent a true reliability test. We also eliminated fathers' and other caretakers' reports (N = 27). This left us with 624 biological mothers who had reported on the same pregnancy separated by a period of 2 days to approximately 2 weeks (Average time, 9 days). The reliability between the two maternal interviews was examined using the kappa statistic that assesses chance corrected agreement (Bishop et al., 1975; Landis & Koch, 1977). The kappa values are sensitive to small sample sizes and low base rates (skewed distribution of marginal probabilities; Spitznagel et al., 1985). Therefore we interpreted kappa values only when at least 20 mothers reported positive answers.

In addition to the data on the 624 mothers from the main study group, we also obtained follow-up interviews on 47 mothers collected as part of a long-term study of child, adolescent, and parent reporting. The follow-up interval ranged from 6 to 18 months with a mean length of 13 months. This second study allowed us to compare mothers' reports about the same pregnancy over a relatively long period of time (stability). The follow-up study examined the reports of the 47 mothers who reported on Twin 1 (the first-born twin) as compared with their original reports on Twin 1. Similarly, follow-up mothers' reports on Twin 2 (the second-born twin) were compared with their original reports on Twin 2.

Although the follow-up study is not prospective, in that data was not collected at the time of birth, it does measure maternal recall over a longer period of time, thus indicating stability.

Stability, as it is used in this study, measures respondents' recall of events in pregnancy over a relatively long period of time, (at least 6 months) longer than for a traditional reliability (test–retest) study. For this study, the mean length of time for the stability report was 13 months while the mean length of time between reports for the reliability study was 9 days.

The differences between Time 1 and Time 2 in these longitudinal reports were also measured with the kappa statistic. Kappa was calculated only when at least five mothers reported positive symptoms.

In summary, there are three comparison sets of reports on maternal answers to questions about pregnancy. The first is the short-term comparison of answers to questions about the same pregnancy with Twin 1 and Twin 2, asked on two separate occasions from the 624 mothers. The second is the 6 to 18 month follow-up asking about the same pregnancy as part of the follow-up of Twin 1 and the third asks about the same pregnancy as part of the follow-up of Twin 2. Finally, maternal reports were contrasted by the zygosity of their twins. That is, maternal recall was conducted on mothers of monozygotic versus dizgotic twins to see if this made any difference in the quality of the recall.

Results

The results of the comparisons of the reports within the three groups are shown in Table 1. Using the values suggested by Fleiss (1981) to judge the kappas in which 0.75 are excellent, and 0.40–0.74 are fair to good, these kappas show good to excellent agreement between mothers' reports at time 1 and time 2. The original group of 624 mothers interviewed 2 days to 2 weeks apart indicates that mothers report reliably on a variety of questions related to their pregnancy. It is also noteworthy that mothers reliably reported such socially disapproved behaviors as smoking, drinking alcohol, using marijuana, and using street drugs when they were pregnant.

Reliabilities for the 6 to 18 month follow-up were also good to excellent, indicating that mothers can report reliably over longer periods of time (stability of diagnoses). For several questions, agreement could not be determined because the mothers reported too few instances of the particular variable. These are marked with an asterisk in Table 1. There were no significant differences in reporting agreement for the first and second born twin. In part due to the smaller number of respondents in the follow-up phase of the study, some of the confidence intervals are larger than the ones for the 624 mothers. There were no significant differences between the reporting of mothers of monozygotic vs. mothers of dizgotic twins (data not shown).

Discussion

The twin design afforded us a unique opportunity to assess reliability of the pregnancy report because interviewing the mother of the twins about each twin separately meant that each mother was asked about the same pregnancy on two different occasions. The advantage of such a large epidemiologically-derived population is that it examines a representative cross section of the population that would not be available in a study with fewer mothers. To our knowledge, no such study is available in the literature. Further, information pertaining to some of the questions asked retrospectively might not be available in medical records and thus could not be validated in this manner.

Results from these data indicate that mothers are reliable reporters about a variety of pregnancy related events that are of interest to the development of psychopathology. This is particularly important with respect to the use of nicotine, alcohol, illicit substances, and over-the-counter medications as many substances taken during pregnancy have the potential for adversely affecting the fetus. This kind of information may not be retrievable from medical records. Thus data indicating that it can be reported retrospectively is encouraging. Questions such as "mothers

eating healthfully", "mothers wellbeing" avoiding smoking or drinking, or over-the-counter medication, may not be available from a chart. A physician may tell a patient to eat healthfully, but whether or not this advice is followed, is hard to confirm prospectively, or retrospectively from a chart. Agreement for other questions such as "problems at delivery" and "being prescribed medicine by a physician" are likely to be available from medical records, but their reliable reports are encouraging in that they indicates that the reports may be usefully asked retrospectively.

It is worth noting that many questions with the highest kappas tend to be concrete and likely to be remembered. These include questions such as "having a cesarean", "born a breech" and "taking marijuana when pregnant". Most of the questions asked about the pregnancy were written in a concrete style and were asked in an unambiguous manner. This is in accordance with the finding by Mitchell et al. (1986) concerning questions that asked mothers about illegal substances they might have taken when they were pregnant. More data were reported when these substances were referred to by name than when general questions about substance use were asked. It also supports the findings of Wilcox and Horney (1984) indicating that circumstances that make an event more concrete are more likely to be remembered by the mothers.

One limitation to this study has already been alluded to. That is the difficulty judging the accuracy of the information that the parents report. It is hard to imagine that some of the more concrete questions would be endorsed positively twice (particularly in the stability portion of the study) if they were not true. However, we do not know if some mothers are forgetting information and not reporting it on both occasions.

Another limitation is that we do not know if the follow-up information collected on the 47 mothers would be as reliable (stable) over a longer time period than was assessed in this study. This remains to be examined, although it is worth noting that our data presents reliable reports of events that happened at least 7, and in some cases, 18 years ago. It is also possible that a pregnancy resulting in twins might be more memorable to the mothers than would a single birth. The lack of differences between monozygotic and dizgotic twins makes the existence of such a reporting bias less likely. However, to our

Table 1

Comparisons of the Reports Within the Three Groups (Full Set of Mothers, Follow-up Twin 1, Follow-up Twin 2).

Questions (paraphrased)	Full set of mothers (*N* = 624)			Follow-up Twin I (*N* = 47)			Follow-up Twin 2 (*N* = 47)		
Past Pregnancies	Kappa CI			Kappa CI			Kappa CI		
1.Ever had miscarriage	0.81	0.87	0.94	0.79	0.58	0.98	0.84	0.65	0.95
2. How many miscarriages	0.91	0.73	0.83	*			*		
Twin Pregnancies									
1.Happy when found out pregnant	0.79	0.73	0.83	0.78	0.55	1.01	0.75	0.61	0.98
2.How far along / found out / twins	0.86	0.82	0.88	0.87	0.76	0.98	0.83	0.69	0.95
3. Surprised / found out / twins	0.82	0.77	0.87	0.69	0.53	1.01	0.81	0.55	0.93
4. Happy / found out / twins	0.71	0.65	0.76	0.63	0 .39	0.87	0.89	0.75	1.03
5. Nervous about having twins	0.72	0.66	0.77	0.61	0.59	0.78	0.69	0.59	9.08
6. Spotting or light bleeding	0.84	0.79	0.88	0.74	0.05	0.98	0.62	0.42	0.89
7. Heavy bleeding / bed rest	0.50	0.40	0.59	*			*		
8. Nausea / beyond 1st trim.	0.87	0.71	0.95	0.87	0.69	1.04	0.82	0.61	0.91
9. Weight loss over 10 pounds	0.82	0.75	0.89	0.72	0.56	0.79	0.86	0.67	0.94
10. Weight gain over 35 lbs	0.80	0.73	0.88	0.59	0.35	0.97	0.60	0.34	0.74
11.Infections / medical care	0.64	0.54	0.73	0.63	0.58	1.01	0.69	0.46	0.91
12 High blood pressure	0.82	0.76	0.86	0.78	0.55	0.97	0.73	0.49	0.90
13. Water retention	0.75	0.70	0.79	0.52	0.27	0.76	0.70	0.49	0.90
14. Convulsions	0.80	0.71	0.79	*			*		
15. Accidents / medical care	0.57	0.38	0.75	*			*		
16. Emotional problems	0.47	0.33	0.60	*			*		
17. Serious family problems	0.74	0.68	0.79	0.65	0.52	0.88	0.62	0.36	0.87
18. Other illnesses / medical care	0.45	0.31	0.50	*			*		
19. Doctor gave medication	0.79	0.74	0.83	0.81	0.63	0.98	0.79	0.55	0.95
20. Over-the-counter medication	0.71	0.66	0.75	0.64	0.41	0.86	0.70	0.49	0.90
21. Able to eat healthfully	0.62	0.54	0.69	0.61	0.50	0.86	0.62	0.56	0.97
22. Smoked during pregnancy	0.60	0.48	0.76	0.95	0.83	1.05	0.95	0.84	1.05
23. Drank when pregnant	0.66	0.60	0.71	*			*		
24. Took marijuana when pregnant	0.90	0.85	0.95	*			*		
25. Used "street drugs" / pregnant	0.70	0.65	0.76	*			*		
26. Child born prematurely	0.67	0.55	0.79	0.83	0.66	0.98	0.79	0.61	0.96
27. Any problems at delivery	0.67	0.54	0.79	0.62	0.45	0.87	0.68	0.50	0.87
28. Born a breech	0.78	0.63	0.91	*			*		
29. Had a cesarean	0.80	0.65	0.93	1.00	1.00	1.00	1.00	1.00	1.00

Note: CI = confidence intervals

* Kappa not calculated because too few positive reported

knowledge there are no data comparing maternal recall on pregnancy for twin versus singleton pregnancies.

As we have indicated, prospective data on events during pregnancy are more likely to be accurate than those retrieved retrospectively due to the tendency of respondents to forget things that have happened in the past. At issue is whether or not retrospective data are worth collecting. Good to excellent kappas on maternal reports of events during pregnancy indicate that pregnancy information can be reliably retrieved retrospectively. Furthermore, although it may be possible to collect all data about the mother's pregnancy prospectively, validating retrospective data by use of medical records may not be useful with respect to certain information that may not be recorded. Our results are in keeping with the literature cited above, that reports reasonable recall of many perinatal events. However, our large sample size, our methods of calculating agreement, as well as the data from the follow-up study of the 47 mothers (stability of symptoms) make these findings more certain. The relatively large number of questions asked specifically about the pregnancy is reassuring. The next step is a comprehensive study of the validity of retrospective recall of questions pertaining to pregnancy, other perinatal information, and early child development in a large-scale epidemiologically derived population ideally using prospective data and appropriate statistical techniques.

Acknowledgments

This research was supported by NIMH grant MH52813.

References

Bishop, Y. M., Fienberg, S., & Holland, P. (1975). *Discrete multivariate analysis.* Cambridge: MIT Press

Fleiss, J. L. (1981). *Statistical measures for rates and proportions* (2nd ed.). New York: Wiley

Githens, P. B., Glass, C. A., Sloan, F. A., & Entman, S. S. (1993). Maternal recall and medical records: An examination of events during pregnancy childbirth and early infancy. *Birth, 20*(3), 136–141.

Hudziak, J. J., Rudiger, L. P., Neale, M. C., Heath, A. C., & Todd, R. D. (2000). A twin study of inattentive aggressive and anxious/depressed behaviors. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 469–476.

Landis, J., & Koch, G. (1977) The measurement of observer agreement for categorical data. *Biometrics, 33,* 671–679.

Milberger, S., Biederman, J., Faraone, S. V., Chen, L. L., & Jones, J. (1996). Is maternal smoking during pregnancy a risk factor for attention deficit hyperactivity disorder in children? *American Journal of Psychiatry, 153*(9), 1138–1142.

Mitchell, A. A., Cottler, L. B., & Shapiro, S. (1986). Effect of questionnaire design on recall of drug exposure. *American Journal of Epidemiology*, *123*, 670–676.

Reich, W. (2000). Diagnostic interview for children and adolescents (DICA). *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 59–66.

Richardson, G. A., Hamel, S. C., Goldschmidt, L., & Day, N. L. (1999). Growth of infants prenatally exposed to cocaine/crack: Comparisons of a prenatal care and a no prenatal care sample. *Pediatrics, 104*(2), e18.

Spitznagel, E. L., & Helzer, J. E. (1985) A proposed solution to the base rate problem in the kappa statistic. *Archives of Genetic Psychiatry, 42,* 725–728.

Stathis, S. L., O'Callaghan, M., Najman, J. M., Andersen, M. J., & Bor, W. (1999). Maternal cigarette smoking during pregnancy is an independent predictor for symptoms of middle ear disease at five years postdelivery. *Pediatrics, 104*(2), e16.

Wilcox, A. J., & Horney, L. F. (1984). Accuracy of spontaneous abortions recall. *American Journal of Epidemiology, 120,* 727–733.

Williams, G. M., O'Callaghan, M., Najam, J. M., Bor, W., Andersen, A. I. M. S., & Richards, D. U. C. (1998). Maternal cigarette smoking and child psychiatry morbidity: A longitudinal study. *Pediatrics, 102*(11), e11.