

Measurement of Substance Use During Pregnancy: Methodologic Issues

Nancy L. Day, Ph.D., Diane K. Wagener, Ph.D., and Paul M. Taylor, M.D.

This paper discusses some methodologic issues in the measurement of substance use, particularly the more complex issue of measuring substance use during pregnancy. Two specific questions are addressed in this presentation:

(1) Whether the questions on use result in accurate assessment of use, and (2) Whether respondents report use accurately. Some of these methodological issues are illustrated with data from our ongoing research on substance use during pregnancy and its effect on outcome.

The assessment of substance use is complicated because there are as many patterns of use as there are users. These patterns, however, can be described by the frequency, quantity, and type of substance. The combination of these three variables defines the pattern of use. Measurement of these parameters will be discussed specifically with reference to marijuana use.

The first descriptor, frequency, is generally considered to be a measure of how much the substance is part of one's life (Room 1977). Clearly, marijuana is a more important part of the life of women who smoke daily compared to women who use marijuana monthly or even less often. This more frequent use has biological implications in terms of exposure, and also has social implications, in that the frequency of substance use correlates very highly with other behaviors, including the use of other substances prior to and during pregnancy (Alpert et al. 1981).

The standard method of collecting data on frequency has been to present the subject with an ordinal ranking of frequencies. There are several problematic issues that need to be considered, however, with this type of assessment. First, from a simple mechanistic viewpoint, the fineness or crudeness of the groupings is an important element in accuracy of assessment. Global categorizations such as daily, often, rarely, or never are so broad as to be meaningless. In addition, qualitative terms such as often and rarely may mean quite different things to different respondents.

At the opposite end of the spectrum, some researchers simply ask how many times the subject has smoked marijuana in a given time period (Greenland et al. 1983). While the specificity of this latter query seems more satisfactory, it too has serious drawbacks. This measure is a rather curious mixture of frequency and quantity and misses the pattern of use.

This kind of metric can be adapted as a time—line follow— back (Sobell et al. 1980; Tennes and Blackard 1980). In this assessment, subjects are asked to recall specific incidents of use for each day over a given time period. Using this technique, it is possible to obtain patterns of use as well as amounts of use because the quantity is measured for each incident of use. However, because this technique asks for recall over a specific period of time, care must be taken to ascertain whether it reflects a usual or aberrant pattern for that subject, as well as whether the subject's recall is accurate. Another variant of this method of data collection is the diary method where respondents are requested to write down all use as it occurs. While this theoretically corrects for some of the memory problems of the time—line technique, it requires a high level of motivation from the subject.

More commonly, researchers have asked subjects to report frequency of use in terms of incidents per day, week, or month. While these data are easy to organize, there are some problems inherent in this method. The issue of grouping is again germane, as the broadness or fineness of the categories can seriously bias estimates. This becomes a problem not so much with the ordinal ranking of subjects, assuming the groupings are not too broad, but with attempts to accurately assess the biological effects of certain doses. The inaccuracies of grouping are also propagated by the need to code data for analysis and to create scales. In most cases, for example, a woman who reports a use level of 1 to 2 times a day will be coded as 1.5 times a day even though she may use only once a day, or she may use twice a day, double the former amount. Due to the need for data reduction, however, these women would be scored as equivalent,

The value of simple frequency assessments is further limited in that the reported frequency will be the usual frequency for that time period. Researchers tend to regard frequency as a constant and seldom allow for variability. This conflicts with reality, since we can document seasonal and situational changes in frequency of substance use. Again, in the assessment of biological, or in the case of pregnancy, teratogenic effects, the “ripples” in the system may have major effects and yet may be missed because of the form of questioning. Practically, even if a respondent volunteers that she used marijuana at a much more frequent rate for a period of time within the time frame considered,

it is very difficult to incorporate this information into the data analysis because of the analytic need for a single, preferably continuous, measure of use.

Quantity, the other major descriptor of patterns of use, is similarly difficult to assess. We ask about the number of joints smoked per incident, for example, and this serves in some respects as a measure of dose, and also as a reflection of whether a subject uses the substance to the point of intoxication, or in more moderate amounts (Room 1977). Thus, it represents a style of use as well as an estimate of the dose received.

Quantity, like frequency, varies. In the measurement of drug use, however, most assessments have been of the usual quantity used, e.g., “How much (how many joints) do you usually smoke?” (National Institute on Drug Abuse 1981). The answers then represent ~ usual pattern, often one that the respondent considers to be more socially appropriate. All use that differs from the usual quantity will be missed in the assessment. Thus, any use greater than the usual, or less than the usual, is not considered.

Another major problem with quantity lies in the estimation of the actual amount consumed. One joint can be of varying sizes and/or varying strengths. In marijuana, unlike alcohol, there is no standardized “proof” of content, and again, unlike alcohol, different methods of consumption yield different degrees of exposure (Nahas 1979; Abel 1980). Additionally, the availability of other substances, such as hashish and sinsemilla, makes the measurement of quantity, or dose, problematic. We also found in our pilot testing that women would report smoking, for example, six or seven joints on an occasion; but, on further questioning, it would turn out that these joints were shared among several friends. Therefore, since marijuana use, unlike the use of many other substances, is a social behavior, it is essential to determine whether the reported quantities were shared or used alone.

Biological assessments are useful measures of the recency of use, and have been used successfully in studies during pregnancy (Greenland et al. 1982). These tests, however, have some of the same drawbacks as the interview data. It is not possible to determine pattern of use from a laboratory test because it only measures exposure at the time the user is tested. Furthermore, there is only a short period of time following marijuana consumption in which laboratory tests can detect or measure levels of cannabinoids in biological fluids of users. This creates a problem for research when we want to cover, for example, the span of a trimester. The researcher must choose between the option of monitoring the women very frequently, i.e., at about 2—week intervals, or less frequently, and having blank spots in the time spanned by

the biological assessment. The first method of frequent assessment could interfere directly with the behavior of the study population. The second method could result in misleading or incomplete data. In addition, laboratory testing of illicit substances will increase both refusal and noncompliance rates, especially in the user groups.

There is a major analytic need for scales that represent the use patterns reported by the subjects. This process is, of course, a necessary step in data reduction. It is highly problematic, however, to take two measures, frequency and quantity, that in themselves are not particularly accurate, and combine them in an algorithm that may just compound the inaccuracy. The most commonly used scale is the usual—quantity/usual—frequency measure.

In our own data, we can illustrate the inaccuracy resulting from asking only about usual quantity and frequency of use. Among the first 290 women in our study who used marijuana, average daily use was calculated to be .7 joints per day, or approximately 5 joints per week. Eliciting maximum quantity increased our estimates of use by 30%, and the addition of questions about use of quantities less than the usual quantity contributed an additional 5.5% to the total reported use. Thus, the usual quantity that our subjects reported represented only 64% of the amount of use reported when we added questions about quantities greater than and less than the usual reported amount. For the first trimester, usual quantity and frequency represented only 57% of the total use.

Among those women who reported using quantities greater than their usual prepregnancy quantity, the mean quantity reported was 3.6 joints. Clearly this represents fairly heavy use and thus greater single dose exposure. On the average, marijuana users who increased their quantity did so about once a week. Quantities less than usual are consumed even more frequently (0.2 per day or about 1.4 times per week). Though occurring more frequently, the relatively small quantities consumed (mean quantity = .9 joints per occasion) do not contribute as much to the overall average daily use.

When measuring the teratogenic effect of drug use during pregnancy, this level of inaccuracy becomes unacceptable because we are misestimating the dose as well as the pattern of delivery of the dose. That is, it is quite possible that a larger exposure at a critical point in the fetal development could have a greater impact than the same dosage spread out over a longer time period. For this reason, it is critical to select measures of quantity and frequency that will allow the data to reflect the variance from the mean as well as the mean experience.

One type of measure that has been used in various forms in fetal alcohol research is the volume variability scale, developed originally by Cahalan et al. (1969). This series of questions ascertains frequency and then asks the subject to indicate for each drinking time whether she drank various numbers of drinks (5 or more, 3—4, 1—2) “nearly every time, more than half of the time, less than half the time, once in a while, or never.” The proportion of drinking incidents spent at each volume level is then calculated to yield a volume—variability score that reflects the variability in quantity.

The Khavari Alcohol Test or KAT scale (Khavari and Farber 1978) is a variant of this technique, though simplified. It asks overall frequency of use and usual quantity. Maximum quantity and frequency of maximum quantity are then ascertained.

These measures also allow one to create different scales that can reflect differing concerns. Calculation of average daily use gives an overall average estimate of use during the time period. With data on variable quantities, it is also possible to look at the frequency with which a woman uses, for example, five or more joints to assess the frequency of use of higher amounts.

There are, of course, a number of other ways to combine the data once they are gathered. But the important point to be made is that each scale highlights one aspect of drug use while suppressing others. A measure of frequent heavy use will reflect those who, for example, use five or more joints when they smoke, but not those who smoke an equivalent amount in more frequent, smaller doses. These different attributes can be compared analytically to explore the relative contributions of the individual scales, as they reflect patterns of use, to the explanation of the variance in the outcome.

Because of the importance of obtaining accurate measures of both marijuana and alcohol use, we conducted an extensive pilot test in our study to develop these assessments. Several variants of traditional questions were pretested using usual quantity and frequency and several scales from alcohol research including the volume variability (VV) scale and the KAT scale discussed above.

Our population is a lower class population of women 18 to 30 years of age who attend the outpatient prenatal clinic at Magee—Womens Hospital in Pittsburgh. sixty percent have graduated from high school, 50% are married, 40% are minorities, predominantly black.

We found that our population initially had problems with the format of the questions. First, they had difficulty because the standard method of assessment asks women to give an overall, total frequency of use, and then to partition that use into the proportion of time she might spend using various quantities. This requires fairly complex mathematical thinking that may not result in accurate measurement under the time and social pressures of the interview. Second, particularly with respect to the marijuana use, it is not clear that women think in terms of frequency first. There has been very little research on how subjects cognitively organize their own assessments of use, but it seemed to be easier for our subjects to think first of quantity and then to describe frequency.

In recognition of this fact, we designed our use assessments to respond to the way our subjects seemed to be organizing their patterns of use. In our study, we first ask questions about quantity, e.g., “When you smoke marijuana, how many joints do you usually smoke?” and then questions about frequency, such as “How often do you smoke this amount?” After this, the respondent is asked “Do you smoke more than this amount?” . . . “How many joints?”

“How often?” . . . and further, “Do you smoke less?” and again, “How many joints?” and “How often?” With few exceptions, our pilot test subjects felt that these questions were easy to answer and described their use accurately. From these questions, it is possible to develop scales that reflect variability in both quantity and frequency, measures of the frequency of heavy use, and measures of average daily use.

We conducted validity tests on these measures comparing our instruments to the VV scale, the KAT scale, and the usual quantity—frequency. There were no significant differences in the rank order of subjects on our measures compared to the other scales. However, the overall average daily use obtained was significantly higher.

Complementary to concern about the validity of the method of assessment is an equally important issue of whether the subjects’ answers are accurate. Two major reasons for inaccuracy are deliberate misrepresentation and errors in recall.

Deliberate misreporting is an important issue when measuring an illicit substance or a substance whose use might be licit but is negatively labeled at particular times, such as alcohol use during pregnancy. We have tried several techniques to increase reporting accuracy. The first was simply to lead into the marijuana questions in a nonthreatening manner. They are imbedded in a section on social supports and drug use by friends.

We have also employed the bogus pipeline technique (Jones and Sigall 1971). This is a method of convincing the subject that you have a laboratory assessment of use when in fact you do not. Interestingly, while the use of the bogus pipeline did significantly increase reporting of other illicit drugs, it did not significantly increase marijuana reporting, leading us to the conclusion that our marijuana assessments were relatively accurate. Further, we discovered during our pilot phase that most women were reluctant to report using any substance currently. However, when asked about use last month or some time in the past, they would respond. As a result, we changed our study design to interview in the fourth month about use during the first trimester, in the seventh month about use during the second trimester, and at delivery we assess the third trimester.

Errors in recall are also important. In the pilot study, we interviewed all clinic women about their use, regardless of month of pregnancy. Women in late pregnancy reported approximately half of the first trimester use that women early in their pregnancy reported. Since there is no reason to believe that these women differed in any way, this represents the problem of accurately recalling actual use during the first trimester, some 6 months earlier. In our study 82% of the women who used marijuana prior to their pregnancy reported a decrease in use during the first trimester. Remembering and accurately reporting a behavior that is changing is difficult, especially many months later.

An accurate assessment of first trimester use is crucial, however, when we are looking for teratogenic effects. The first trimester of pregnancy is the most vulnerable time for the developing fetus. It is the time of major organ and systems development, and the time that major birth defects occur. Also, because of the sequential timing of developmental events during this time period, the same dose of a substance at different time points during organogenesis will have different effects. Thus, it is questionable whether data that average use estimates over the first trimester are sufficiently specific to reflect the relationship between exposure and effect.

A further difficulty with assessment of first trimester use is that though researchers may count conception as the beginning of the pregnancy, it is not clear whether pregnant women really think back to that point when asked to report events during the first trimester.

We have developed a technique to assess the use of marijuana during the first trimester that allows us to circumvent this problem. Early in the interview, women are asked to indicate on a calendar the month and part of the month (beginning, middle, or end) when they: (a) got

pregnant; (b) recognized their pregnancy; and (c) had their pregnancy confirmed by a pregnancy test. Later in the interview, we ascertain marijuana use prior to pregnancy and in the first trimester of pregnancy. Immediately after these questions, the interviewer returns to the calendar and asks the subject whether for the time period between conception and recognition of pregnancy, her marijuana use was more like that which she reported prior to pregnancy or more like the rate she reported for her first trimester. After this, we ask the same set of questions about the time period from recognition to diagnosis. The dates of conception, recognition, and diagnosis and the rates of marijuana use prior to and during pregnancy are used to calculate a month—by—month rate of marijuana use and a weighted estimate of the average daily use for the first trimester.

Sixty—six percent of all marijuana users reported that from conception to recognition of pregnancy, their use was similar to their prepregnancy pattern and 33% reported that from recognition to diagnosis, they were still using marijuana at their prepregnancy rate.

When we looked only at women who had an average daily use of two or more joints prior to pregnancy, 83% said that from conception to recognition their use was similar to their prepregnancy pattern, and 52% reported that from recognition to diagnosis their use was similar to their prepregnancy pattern.

On the average, there were 30.5 days from conception to recognition for the total population, 31.2 days for all marijuana users and 33.2 days for women who smoked two or more joints per day prior to pregnancy. The time span from recognition to diagnosis was also lengthened for these heavier smokers, who had a mean of 22.1 days compared to 20.1 days for all marijuana users and 19.3 days for the total population. Therefore, the heavier users were more likely to continue to use marijuana at their prepregnancy rate, and they used marijuana at this higher rate for a greater part of the first trimester.

The reported and calculated estimates of the average daily use of marijuana are given in table 1. A total of 11.6% of our subjects reported smoking two or more joints per day before their pregnancy, although only 3.4% of the women reported using two or more joints per day during the first trimester of pregnancy, an apparent decrease of two—thirds. However, when we calculated the average daily use as described above, 5.5% of the subjects actually used an average of two or more joints per day during the first trimester. In other words, only about two—thirds of the women who were calculated as using at least two joints per day reported doing so; a sensitivity rate of 63%. The sensitivity rate for those calculated as using one joint

TABLE 1
Reported and Calculated Marijuana
Usage, N=644*

Average Daily Use	Pre-pregnancy Reported Use	Calculated Usage				First Trimester Reported Use
		Month 1	Month 2	Month 3	Average	
≥ 2 per day	11.6	8.1	5.5	3.8	5.5	3.4
1 per day	4.2	3.7	2.8	1.7	3.4	2.0
1-6 per day	15.5	12.0	9.0	7.0	12.5	6.4
Less than weekly	13.7	8.7	11.2	9.7	11.6	7.6
None	55.0	66.9	71.0	77.8	67.0	80.6

*Values given in percent of total.

per day was 60% and for those calculated to have a rate of one to six per week, 51%. This reflects the pattern of misreporting downward. Overall, we found that 56% of the subjects underreported their use when we compared the calculated rate to the actual reported rate. This means that studies that use the more global assessment of first trimester use underestimate the actual amount of exposure in more than half of the cases.

Since the teratogenic effect of any drug can be specific to certain time periods, we calculated rates separately for each month of the first trimester. This allowed us to determine whether the effect is a direct function of extended use over a period of time as long as 3 months, or whether it is the result of exposure over a shorter more specific time period. In table 1 the calculated average daily use is shown for each month of the first trimester. During the first month of pregnancy, 8.1% of the subjects had a calculated use of two or more joints per day. That is, 70% of the women who used marijuana at this rate prior to pregnancy continued at that rate through the first month of pregnancy. However, only 42% of these women reported using that amount for the first trimester, Of those women who we calculated as having used marijuana at the rate of two or more joints per day during the first month but who did not report this rate, 58% actually reported no first trimester use at all.

Thus, the rate that women report for first trimester use is particularly inaccurate during the early part of the

first trimester. In fact, in the absence of estimates of first month use, it is quite clear that the prior—to— pregnancy rate is a better estimator of use during this time period than is the reported first trimester use. This is especially true for estimates of use among heavy users.

The other end of the scale is also problematic in terms of misclassification of subjects. Eighty percent of our subjects reported using no marijuana during the first trimester. However, on closer questioning, only 67% used none during the first trimester. That is, about one out of every five women who reported no marijuana use during the first trimester was actually a user.

We began this paper by pointing out that the measurement of substance use is a complicated issue. We have attempted to highlight some of the areas of difficulty. In particular, it is essential that we work to develop valid instruments to measure substance use and that, particularly for studies of use during pregnancy, these questions be designed to give us information about pattern of use in addition to average use.

An additional serious problem lies in the accuracy of subject's reports about their use, particularly during the first trimester. Our data clearly demonstrate that reports of first trimester use really only represent late first trimester use, the most recent memory. This is not surprising since drug use decreases during the first trimester for most women, and accurate recall is difficult. This means, however, that previous studies using reports of first trimester use underestimated the exposure and seriously underestimated the exposure during the first month of pregnancy.

Abel, E. Prenatal exposure to cannabis: A critical review of effects on growth, development and behavior. *Behav Neural Biol* 29:137—156, 1980.

Alpert, A.; Day, N.; Dooling, E.; Hingson, R.; Oppenheimer, E.; Rosett, H.; Weiner, L.; and Zuckerman,

B. Maternal alcohol consumption and newborn assessment:

Methodology of the Boston City Hospital prospective study. *Neurobehav Toxicol Teratol* 3:195—201, 1981.

Cahalan, D.; Cisin, I.; and Crossley, H. *American Drinking Practices* New Brunswick, NJ: Rutgers Center of Alcohol Studies, Monograph No. 6, 1969.

Greenland, S.; Richwald, G.; and Honda, G. The effects of marijuana use during pregnancy. II. A study in a low— risk home—delivery population. *Drug Alcohol Depend* 11:359—366, 1983.

- Greenland, S.; Staisch, K.; Brown, N.; and Gross, S. The effects of marijuana use during pregnancy. I. A preliminary study. *Am J Obstet Gynecol* 143:488—413, 1982.
- Jones, B., and Sigall, H. The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychol Bull* 76:349—364, 1971.
- Khavari, K., and Farber, P. A profile instrument for the quantification and assessment of alcohol consumption. *J Stud Alcohol* 39:1525—1539, 1978.
- Nahas, G. Current status of marijuana research. *JMIA* 242(25) :2775—2778, 1979.
- National Institute on Drug Abuse. Drug Abuse Instrument Handbook. By Nehemkis, A.; Macan, M.; and Lettieri, D.J., eds. DHHS Pub. No. (ADM) 78—394. Washington, D.C.: Supt. of Docs., U.S. Govt. Print, Off. 1978.
- Room, R. The measurement and distribution of drinking patterns and problems in general populations. In: Edwards, G.; Gross, M.M.; Keller, M.; Moser, J.; and Room, R. eds. *Alcohol-Related Disabilities* Geneva: World Health Organization, Offset Pub. No. 32, 1977. pp. 2-38.
- Sobell, M.B.; Maisto, S.A.; Sobell, L.C.; Cooper, A.M.; Cooper, T.C.; and Sanders, B. Developing a prototype for evaluating alcohol treatment effectiveness. In: Sobell, L.; Sobell, M.; and Ward, B., eds. *Evaluating Alcohol and Drug Abuse Treatment Effectiveness: Recent Advances*. New York: Pergamon Press, 1980.
- Tennes, K., and Blackard, C. Maternal alcohol consumption, birthweight, and minor physical anomalies. *Am J Obstet Gynecol*. 138:774—788, 1988.

This research is supported by two grants from the National Institute on Drug Abuse: “Marijuana, Psychological Factors and Pregnancy Outcome” (DA83209) and “The Effect of Marijuana Use During Pregnancy” (DA88898) and a grant from the National Institute on Alcohol Abuse and Alcoholism: “The Effect of Alcohol Use During Pregnancy” (AA86398).

Authors

Nancy L. Day, Ph.D.
Assistant Professor of Psychiatry and Epidemiology
University of Pittsburgh
Western Psychiatric Institute and Clinic
Pittsburgh, PA

Diane K. Wagener, Ph.D.
Associate Professor of Psychiatry, Biostatistics and
Epidemiology
University of Pittsburgh
Western Psychiatric Institute and Clinic
Pittsburgh, PA

Paul M. Taylor, M.D.
Professor of Pediatrics and Obstetrics and Gynecology
University of Pittsburgh
Magee—Womens Hospital
Pittsburgh, PA

**Click here
to go to
next section**