Since the use of multivariable survival analysis for pregnancy outcome studies was first introduced into the reproductive epidemiology literature (1), few researchers examining fetal loss have applied these methods, which include Cox proportional hazards and other statistical models. Because logistic regression can yield misleading results, Li et al. (2) are to be commended for electing to make use of these methods. It might appear, therefore, that their study was methodologically strong.

Unfortunately, several key limitations to the study (2) and some of the authors’ decisions limit the credibility of their findings. A synopsis of the concerns follows. First, the response rate was very low. Second, data for early gestational weeks are sparse and may not be representative. Third, most of the cases were interviewed after they miscarried, and the association hinges on the subgroup of these women who both miscarried early and were interviewed after they miscarried. Fourth, the subset analyses are flawed. Fifth, the lack of association with fever undermines the central hypothesis. Sixth, residual confounding from measured risk factors apparently was not available, however.

1. Participation was 39 percent. The authors (2) acknowledge that bias could have occurred, but they do not give much weight to this possibility. With a response rate this low, substantial biases can be present even when the differences between participants and nonparticipants are modest. This bias can occur because the sample could easily be unrepresentative with respect to the joint distribution of miscarriage risk and use of hot tubs or Jacuzzis (any type of whirlpool bath; Jacuzzi Brands, Inc., West Palm Beach, Florida). In table 1, we provide two possible scenarios in which, conditional on outcome, the proportion of participants exposed differs little (i.e., by 3 percent in scenario 1 and by 4–5 percent in scenario 2) from that of nonparticipants. Yet the association is essentially wiped out in the total population. Although these examples use two-by-two tables (applicable in the absence of variable left truncation and confounding from other factors), similar examples could be shown for more complex situations by using survival analysis. The simple point is that small differences in the prevalence of exposure will change the measures of association substantially when the nonparticipants represent a high fraction of the cohort.

A hint about the difference between those who did and did not participate is provided by the reasons given for nonparticipation. “Too busy” was the most common one. Hot tub users who participated were more educated and more likely to exercise and use vitamins than were nonusers who participated; if, in addition, having a miscarriage made an otherwise busy but health-conscious woman even more motivated to be interviewed, bias would be likely. Another clue about the potential for biased participation might have been the difference between gestational age at the time of the positive pregnancy test (this was the definition for entry) for participants versus nonparticipants. This comparative information apparently was not available, however.

2. Li et al. (2) are to be commended for the remarkable early recruitment of pregnant women. Approximately half of the women were recruited 12 days or fewer after missing their first menstrual period. Thirty women were recruited before 28 days had passed from their last menstrual period, that is, on average, before missing their period. The crucial questions are, How representative of the cohort are those who entered the study at an extremely early point in pregnancy? And, how accurate are their gestational dates?

Epidemiologists studying miscarriage are faced with the difficult challenge of trying to ascertain, as early as possible, pregnancy losses. Paradoxically, success in doing so may result in a highly select sample and, if not analyzed carefully, may give rise to biased estimates of effects. The problems of sparse data and selection in the early weeks of pregnancy have been discussed for several decades by authors studying large cohorts (3–7). Given this conundrum, we must be attentive to how the data in the early weeks are handled, when differential entry by underlying risk is possible and, in several studies, has been observed (3,5–7). Moreover, errors in gestational dating may be more prevalent among those...
with extremely early entry times. Li et al. (2) included the early events; that is, their survival analysis was started extremely early in cohort accrual (they report the first risk set as having eight pregnancies). Goldhaber and Fireman (3) cite data from two other cohorts (4, 5) to demonstrate how the choice for when to begin a prenatal life table could lead to unduly influential effects of a small number of early losses on the magnitude of the life table risk for miscarriage. That is, sparse data early on produced excessively high estimates of cumulative risks that would not have resulted if the life table had started later. How much influence the first or first few event(s) had on the measures of association in the hot tub or Jacuzzi analysis and their confidence limits is unclear. It is clear from table 1 of Li et al., however, that hot tub or Jacuzzi users entered the study (had a pregnancy test) earlier than nonusers did.

3. Although the authors (2) describe their study as being prospective, this adjective represents the intent more than the implemented reality. Interviews were conducted “early in pregnancy,” yet more than 60 percent of the cases were interviewed after their miscarriage (hence, after termination of the pregnancy). The authors chose not to exclude these women. Interestingly, the most telling results are actually found not in the Results section or in the tables, but in the Discussion. Here, the authors stratified their cases into four groups based on early and late miscarriages and whether the interview took place before or after the miscarriage occurred. For those who miscarried after 10 weeks, the adjusted hazard ratios were elevated: 2.7 (95 percent confidence interval: 1.5, 4.9) for those interviewed after the miscarriage and 3.2 (95 percent confidence interval: 1.0, 10.3) for those interviewed prior to the event. However, this last stratum included only four exposed cases and therefore should not carry much weight. Statistically, in fact, with a confidence limit ratio of 10, it would have little influence in a truly informative analysis showing the crucial one-way stratification by when the interview took place.

Thus, the only group with a strong and stable association is comprised of women interviewed after the miscarriage. In other words, the prospective portion of the study shows essentially no association at all. How could this occur? Both reporting (“recall”) bias and response (“participation”) bias are possible. We may be looking at events that influence recall. On the other hand, some women might have been “too busy” to be interviewed had they not miscarried, an experience that motivated them to participate. In this latter instance, either the decision to enroll or the ability to schedule an interview within the prescribed time frame for the study could have been influenced by the outcome of the pregnancy. Regardless of the particulars, the best evidence about hot tub or Jacuzzi use and miscarriage comes from those interviewed before the pregnancy terminated.

4. All of the other subset analyses provide little support for the reported association between hot tub use and miscarriage (2). The data were quite sparse (e.g., n’s for exposed cases of 2, 5, 6, and 7) in many strata of temperature, frequency of use,

<table>
<thead>
<tr>
<th>Cases</th>
<th>Noncases</th>
<th>Total no.</th>
<th>Risk ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Proportion</td>
<td>No.</td>
</tr>
<tr>
<td>Participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexposed</td>
<td>144</td>
<td>0.85</td>
<td>809</td>
</tr>
<tr>
<td>Exposed</td>
<td>26</td>
<td>0.15</td>
<td>79</td>
</tr>
<tr>
<td>Nonparticipants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexposed</td>
<td>235</td>
<td>0.88</td>
<td>1,236</td>
</tr>
<tr>
<td>Exposed</td>
<td>32</td>
<td>0.12</td>
<td>168</td>
</tr>
<tr>
<td>Total population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexposed</td>
<td>379</td>
<td>0.87</td>
<td>2,045</td>
</tr>
<tr>
<td>Exposed</td>
<td>58</td>
<td>0.13</td>
<td>247</td>
</tr>
</tbody>
</table>

* Any type of whirlpool bath; Jacuzzi Brands, Inc., West Palm Beach, Florida.
† These scenarios demonstrate how small differences between those persons who self-select to participate and those who do not could have produced the observed data even if no association exists between exposure and outcome in the total population.

TABLE 1. Hypothetical scenarios for exposure of nonparticipants in a study of hot tub or Jacuzzi® use and the risk of miscarriage†

---

Am J Epidemiol 2003;158:938–940
and gestational age at initial use. Findings based on fewer than 10 exposed cases should always be treated with great caution. Furthermore, most women did not know the temperature of their water and were grouped into one category (“missing data”) represented by an indicator variable in the model. This method is known to produce biased results (8).

5. No association was shown with fever. The authors (2) suggest that fever operates differently from hot tubs or Jacuzzis. They also discuss elevation of core body temperature as being the mechanism for the latter. Yet this is exactly what fever does.

6. Women aged 25–29 years were less likely to use hot tubs or Jacuzzis, while those aged 30–34 years were more likely. Maternal age is one of the strongest risk factors for miscarriage: incidence rates rise exponentially after age 30 years (1, 4). A trend in hot tub or Jacuzzi use within categories of age (e.g., if women aged 33–34 years were more likely than those aged 30–31 years to submerge themselves) could produce residual confounding in an analysis in which age is categorized. A linear term in the Cox model for those above age 25 or 30 years or the use of splines could achieve maximal control of such potential confounding.

With respect to methodology, conducting studies of fetal loss is exceedingly difficult, which is perhaps why few risk factors for miscarriage have been identified. Li et al. (2) have led the way with early recruitment and have avoided the easy route of logistic regression, used too frequently for pregnant cohorts with differential left truncation (e.g., Waller et al. (9), Belanger et al. (10)). In this situation, finding the appropriate start time for survival-based methods is a challenge; however, given existing evidence about life tables and the impact of selective entry, inclusion of sparse early risk sets in Cox models appears to be inadvisable. Also notable was the decision not to exclude those women interviewed after they miscarried. Several authors of similarly designed studies elected to exclude such women (9, 10). This issue deserves further discussion and will not be clarified without evidence on the validity of alternative approaches.

With regard to the substantive question, namely, hot tub or Jacuzzi use and miscarriage, the primary evidence was found only among those interviewed after the loss of the pregnancy (2); the lack of association with fever in this and another study (11) is a lingering inconsistency; the low participation rate could easily have produced biased participation; the analysis included potentially influential sparse strata; the subset analyses were also based on sparse and incomplete data and were flawed in how missing information was handled; and residual confounding by maternal age may have occurred. No solid evidence, from this or any other study, links hot tub and Jacuzzi use with miscarriage.

REFERENCES


