What Have We Learned About Uterine Contractions and Preterm Birth?
The HUAM Prediction Study

Jay D. Iams* for the National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network

Measurement of uterine contraction frequency has been employed as a screening test to identify women with increased risk of preterm birth, and as an aid in the early diagnosis of preterm labor. The National Institute of Child Health and Human Development Maternal-Fetal Medicine Units (NICHD MFMU) Network performed a prospective, blinded observational study of uterine contraction frequency to detect and predict preterm labor and birth, respectively. The goal of the study was to assess the sensitivity, specificity, and positive and negative predictive value of various measures of uterine contraction frequency. Data collected from 306 women revealed that contraction frequency was significantly greater in women who would ultimately deliver before rather than after 35 weeks' gestation. However, both sensitivity and positive predictive value of any measure of contraction frequency to predict preterm birth were poor. Contraction frequency did not increase significantly within 1 or 2 weeks of an episode of preterm labor. These results serve to explain the absence of an association between contraction-based surveillance and preterm birth in randomized trials conducted in women at risk of preterm birth.

© 2003 Elsevier Inc. All rights reserved.
History has shown that both camps made points that remain valid today: There are interventions that reduce the morbidity and mortality of pre-maturity, but progress has indeed been limited by our incomplete understanding of the mechanisms that lead to preterm birth.

The use of tocolytics was advocated to prevent preterm birth because the onset of preterm labor was understood to begin with the coalescence of small uterine contractions into larger contractions that then effected cervical change, as observed in studies of primate models. When such use did not produce an obvious decline in the rate of preterm delivery, several explanations were proposed:

1. The medications used were not optimal, either because of inadequate tocolytic effect or unacceptable side effects.
2. The cause of the preterm labor was not understood and thus not appropriately treated with medications directed at stopping uterine muscle contractions.
3. Tocolytic medications were applied too late to have any benefit.

Those who favored Explanation 1 argued for more and better tocolytics drugs; their descendants today advocate prophylactic and/or therapeutic tocolytics medication via either oral or subcutaneous administration. Those who favored Explanation 2, the descendants of Eastman and Taylor, saw vindication in the failure of tocolysis to reduce preterm birth rates, and argue now for therapeutic nihilism until both basic and clinical research reveals an unequivocally effective intervention. Many of this group live, or lived for a time, in Dallas, TX; their views continue to exert a substantial and important influence on clinical practice. Those who favored Explanation 3 saw the failure of tocolysis as an argument to improve the early identification of women with preterm labor so that tocolysis might be more efficacious. This group spawned the March of Dimes Preterm Birth Prevention Trial, a 5-site multicenter trial of a screening and educational program for pregnant women conducted over 5 years. The goal was to identify women at risk for preterm birth through a screening program with the Creasy Score. High-risk women were randomly assigned to a control group that received standard care, or to a study group that received frequent visits and education about how to recognize the earliest signs and symptoms of preterm labor. Once educated to note the symptoms and to self-palpate to detect contractions, the March of Dimes investigators hypothesized that women assigned randomly to the education and self-palpation group would present earlier to the health care system for care, so that tocolytic medication would be more effective. More than 3,000 women were enrolled. The results were negative, ie, there was no difference in the rate of preterm delivery between the study and control groups.

The response of many to these negative results was again based on Explanation No. 3 as the operating hypothesis: Tocolysis as used in the March of Dimes Trial was still too late to be beneficial. Newman et al reported that women taught to self-palpate their contractions did not identify the majority of contractions recorded by a sensitive tocodynamometer. These observations were the basis for the next effort to detect the earliest signs and symptoms of preterm labor in women with commonly recognized risk factors for preterm birth. Technological advances in uterine contraction detection that used a “guard-ring” tocodynamometer, and in electronic transmission of contraction data via the telephone allowed pregnant women to use a contraction monitor at home several times each day, and to transmit their contraction data to a monitoring center where it could be reviewed by specially trained nurses. This approach was based on the belief that 1) uterine activity was higher in women destined for preterm birth, and 2) uterine activity was increased 24 to 48 hours before an episode of preterm labor.

Some reports of pregnancy outcomes in high risk women who used the HUAM device and service were promising, but other studies did not show any difference in the rate of preterm birth when the monitor was used compared to controls. Ultimately, 2 large randomized trials were conducted that both found no difference in the frequency of preterm birth in women who received a contraction monitor compared to those who did not. The study by Dyson et al enrolled more than 2,400 women with risk factors for preterm birth who were randomly assigned to 1 of 3 groups: Daily out-
patient contraction monitoring accompanied by daily nurse contact, daily nurse contact without a contraction monitor, or weekly nurse contact alone. Results are shown in Table 1. There was no benefit to either daily monitoring and contact or daily contact when compared with weekly contact. Women in both intervention groups had more visits and more drug treatment but to no advantage.

The CHUMS2 and Dyson3 trials effectively demonstrated that the contraction surveillance and early treatment approach to prematurity prevention was ineffective. Nevertheless, the reasons for the failure of this approach to reduce the occurrence of preterm birth were not completely explained, and the connection of increased uterine activity to preterm birth risk remained anecdotal strong. As Dyson et al5 were conducting their trial within the Kaiser health care system in northern California, the NICHD MFMU Network was considering a randomized trial of similar design. Ultimately, the Network Steering Committee and NICHD Program Officers decided on a different approach, for both practical reasons (the Dyson trial was enrolling successfully and the cost of a randomized trial within the Network exceeded available funds) and for scientific reasons described in the following paragraphs.

**Rationale for the HUAM Prediction Study**

When the Network Steering Committee reviewed the available literature that related uterine contraction frequency to preterm birth risk in singleton pregnancy, several important unanswered questions were identified:

1. Is the frequency of uterine activity greater in women destined for preterm birth than in women who will deliver at term?
2. Is the frequency of uterine activity greater in women with historical risk factors for preterm birth (eg, a prior preterm birth) greater than in women who have no risk factors?
3. Does uterine activity differ according to gestational age in women destined for preterm birth?
4. Does uterine activity differ according to time of day in women destined for preterm birth?
5. Is the measure of uterine activity an efficient test to: screen for risk of preterm birth; predict preterm birth; or detect preterm labor before it is detected by clinical signs and symptoms?

Although small studies had addressed each of these questions, there were no large blinded observational studies to establish reference values for uterine activity by gestational age, time of day, fetal number, or obstetrical history. Without a firm knowledge of the range of

---

**Table 1. Outcomes of Pregnancy in the Weekly-Contact, Daily-Contact, and Home-monitoring Groups**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>All Women (N = 2422)</th>
<th>Women with Twin Pregnancies (N = 844)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Weekly Contact (N = 798)</td>
<td>Daily Contact (N = 796)</td>
</tr>
<tr>
<td>Preterm birth (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;37 wk</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>&lt;35 wk</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>&lt;32 wk</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Birth weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;1500 g</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>&lt;2500 g</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>No. of unscheduled visits*†</td>
<td>1.2 ± 1.5</td>
<td>1.8 ± 2.0</td>
</tr>
<tr>
<td>Prophylactic to colytic-drug therapy (%) †</td>
<td>12‡</td>
<td>14‡</td>
</tr>
<tr>
<td>Preterm labor &lt;35 wk (%)</td>
<td>23</td>
<td>22</td>
</tr>
</tbody>
</table>

*Plus-minus values are means ± SD. **P < .002 for all comparisons between treatment groups.
†Therapy was given after symptoms appeared but before the criteria for preterm labor were met.
‡P < .01 for the comparisons between the weekly-contact or the daily-contact and the home-monitoring groups for all the women and between the weekly-contact and home-monitoring groups for the women with twin pregnancies.
Reprinted with permission.3
uterine activity in normal pregnancy, it was possible that the uterine activity thresholds used in clinical and research use of HUAM were incorrect. Moore et al found that uterine activity (UA) in normal women studied for 24 hours a day twice weekly from 24 weeks’ through delivery was significantly influenced by the following:

1. Gestational age–UA increases as gestational age advances, especially after 28 weeks.
2. Time of day–UA is greater in the afternoon and evening than in the morning.
3. Rest–UA decreases for 1 to 2 hours after an hour of recumbence.
4. Coitus–UA increases for 1 to 2 hours after coitus.

Similarly, Germain et al studied patterns of uterine activity in 3 groups of women: 1) Low-risk women who delivered at term, 2) Women with a prior preterm birth who delivered at term in the current pregnancy, and 3) Women with a prior preterm birth who delivered prematurely in the current pregnancy. Those in the first 2 groups maintained a normal pattern of uterine activity with fewer contractions in the morning and more in the evening, while high-risk women who delivered preterm had no variation in uterine contractions with time of day.

The Network investigators thus hypothesized that failure to account for these influences on uterine activity could explain the clinical failure of HUAM-based protocols to reduce preterm birth because of an inability to correctly identify women with an increased risk of preterm birth. A protocol was therefore developed to study uterine activity as a screening test for preterm birth risk and as a diagnostic test for early preterm labor. The study protocol was patterned after the Preterm Prediction Study as described by Goldenberg et al in this issue, except that the women enrolled would be chosen to create a population with increased risk of preterm birth, and would in addition to cervical examinations also collect uterine activity data.

The HUAM Prediction Study

Methods

Women with singleton gestations with a prior spontaneous preterm birth between 20 and 36 weeks or second trimester bleeding in the current pregnancy were recruited. A limited number of women with no risk factors were enrolled to allow comparison of contraction frequency in low- and high-risk women. The sample size was chosen to create an 80% power to detect a difference of 0.5 standard deviations [e.g., a difference of one contraction per hour if the standard deviation was no more than 2 contractions per hour]. By enriching the study population with high-risk patients, a sample size of 300 women was appropriate for an endpoint of preterm delivery before 35 weeks’, in which neonatal morbidity is greater. We planned to enroll 50 to 70 low-risk women and 230 to 250 high-risk women.

Eligibility was determined before 22 weeks’ gestation. An ultrasound examination was performed before enrollment to determine gestational age and exclude major fetal anomalies and placenta previa. Women treated before screening with tocolytic medication or a cerclage were excluded. Study visits were scheduled to collect data at 22-24 weeks’ (visit 1), 25-26 weeks’ (visit 2), 27-28 weeks’ (visit 3), 29-30 weeks’ (visit 4), 31-32 weeks’ (visit 5), and ≥33 weeks of gestation (visit 6).

Research nurses trained to use a home contraction monitor (Healthdyne System 37; Matria Inc, Marietta, GA) visited each woman at home to instruct her to use the monitor and transmit data. Uterine activity was recorded for a minimum of 1 hour at least twice daily in 2 sessions at least 2 hours apart, 1 between 0400 and 1559 and the other between 1600 and 0359, on 2 or more days per week from enrollment to 28 weeks. After 28 weeks, 2 additional monitoring sessions per week were required. Data were transmitted immediately after collection to the Data Coordinating Center at The George Washington University Biostatistics Center. Protocol compliance for time, date, and quality of the recordings was followed weekly for each woman. All other study data were transmitted weekly to the center.

The duration of pregnancy and reason(s) for preterm birth were recorded. Providers, investigators, and patients were blinded to results of testing. None of the tests being evaluated were performed outside of the study.

Uterine activity monitor recordings were an-
analyzed according to a standard protocol by four research nurses who were unaware of pregnancy outcome. A contraction was defined as a deflection from baseline with a rounded peak that lasted 40 to 120 seconds. Inverted, “double-peak” and “camel back” contractions were included. “Possible” contractions that were subtle, had a variable baseline, a flat peak, or were accompanied by artifact were not considered contractions. Regular audits of contraction recordings were conducted throughout the study to assure consistent interpretation. Discordant interpretation was not greater than one contraction per hour, and was evenly distributed between increased and decreased contractions.

Analyses of uterine activity were performed to identify relationships with gestational age, time of day [0400-1559 (designated AM) v 1600-0359 (designated PM)], risk status (high v low), and term versus spontaneous preterm delivery <35 weeks’ gestation (after preterm labor or preterm ruptured membranes). Contraction rates per hour were calculated for each patient and each gestational week, and analyzed using a repeated measures random effects model. Because each woman could contribute data only as long as her pregnancy continued, more data were collected for women who remained undelivered. Contraction frequency data from women who delivered after 35 weeks’ was therefore combined for analysis.

Mean contraction frequency increased significantly with gestational age and was increased during p.m. (1600 – 0359) hours regardless of gestational age at delivery (Fig 1). Separate analyses were performed for a.m. and p.m. contraction data. After controlling for gestational age at the time of monitoring, women who delivered before 35 weeks’ had more contractions as measured by both a.m. contraction frequency (P = .09 for 24/7-28/7 weeks’ and P = .03 for 24/7-32/7 weeks’) and p.m. contraction frequency (P < .001 for 24/7-28/7 weeks’ and P = .02 for 24/7-32/7 weeks’).

Results

Among 454 women with singleton pregnancies who met inclusion criteria and consented to enrollment, there were 146 (32.2%) who were not compliant with monitoring, and 2 who delivered within a week of enrollment. The remaining 306 women constitute the study population. Of 274 women at increased risk for preterm birth, 194 (76.4%) had 1 prior preterm birth, 57 (22.4%) had a history of 2 or more preterm births, and 8 (3.1%) had second trimester bleeding. Some had more than 1 risk factor. There were 106 women (34.6%) who delivered before 37 weeks’ gestation, 48 (15.7%) before 35 weeks’, and 18 (5.9%) before 32 weeks’.

The 106 enrolled women recorded 34,908 hours of contraction data, of which 20.8% (7,268 hours) had contractions. Contraction frequency was unrelated to maternal risk status (P = 0.219). Data from low- and high-risk women were therefore combined for analysis.

Mean contraction frequency increased significantly with gestational age and was increased during p.m. (1600 – 0359) hours regardless of gestational age at delivery (Fig 1).

We determined the sensitivity, specificity, and predictive values of mean and maximum a.m. and p.m. contraction frequency at each gestational age interval to predict spontaneous preterm birth before 35 weeks’ gestation. Contraction frequency was dichotomized as <4 v ≥4 per hour when tested as a categorical variable. Receiver operating characteristic curves were constructed for mean and maximum contraction frequency per 2-week gestational age interval.
than did women who delivered after 35 weeks' gestation.

In univariate analyses, maximum contraction frequency was inconsistently related to spontaneous birth before 35 weeks' gestation (Table 2). Because the logistic regression models for maximum a.m. and p.m. contraction frequency had slightly better fit than that for mean contraction frequency, maximum frequency was used in subsequent analyses. Change from baseline contraction frequency was also evaluated but performed less well than either mean or maximum frequency as a screening test for preterm birth. In multivariate analyses, (Table 3) maximum p.m. contraction frequency was significantly related to preterm birth only at 27 to 28 weeks. Sensitivity, specificity, and predictive values for maximum a.m. and p.m. contraction frequency are shown in Table 4. Maximum contraction frequency of 4 or more per hour was a poor screening test for preterm birth. At 22 to 24 weeks', the sensitivity of contraction frequency to predict birth <35 weeks' was below 10%. The sensitivity of maximum p.m. contraction frequency improved at later gestational age intervals, but continued to have low sensitivity and positive predictive value to detect women with increased risk of preterm birth.

We also performed analyses of contraction data to look for a relationship between contraction frequency and an impending episode of preterm labor. Preterm labor was diagnosed as admission to the hospital and treatment for preterm labor in response to contraction frequency accompanied by documented cervical change. There were 108 women with an episode of preterm labor. Mean and maximum contraction frequency for the 1- and 2-week periods before an episode of preterm labor were compared to mean and maximum contraction frequency for the week in which preterm labor was diagnosed to determine a “Delta CPH” or change in contractions per hour. Neither Delta CPH for mean nor maximum contraction frequency was significantly related to a diagnosis of preterm labor (P = .19 and .77, respectively).

The results of the study may be summarized as follows:

1. Is the frequency of uterine activity greater in women destined for preterm birth than in women who will deliver at term? Yes.

### Table 2. Odds Ratio for Spontaneous Delivery at Less Than 35 Weeks, According to the Maximal Daytime and Nighttime Frequency of Contractions

<table>
<thead>
<tr>
<th>Week of Gestation</th>
<th>Odds Ratio (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Women</td>
<td>Daytime (4 a.m.-3:59 p.m.)</td>
</tr>
<tr>
<td>22-24</td>
<td>270 0.9 (0.6-1.3)</td>
</tr>
<tr>
<td>25-26</td>
<td>301 1.2 (1.0-1.5)‡</td>
</tr>
<tr>
<td>27-28</td>
<td>294 1.0 (0.8-1.2)</td>
</tr>
<tr>
<td>29-30</td>
<td>288 1.1 (0.9-1.2)</td>
</tr>
<tr>
<td>31-32</td>
<td>281 1.0 (0.8-1.3)</td>
</tr>
<tr>
<td>&gt;32</td>
<td>266 0.8 (0.6-1.2)</td>
</tr>
</tbody>
</table>

*CI denotes confidence interval.
†P = .02
‡P = .03
§P = .003.

Reprinted with permission.¹

### Table 3. Relation of Contraction Frequency to Occurrence of Spontaneous Birth <35 Weeks

<table>
<thead>
<tr>
<th>Gestational Age</th>
<th>Maximum evening ≥4 contractions/hr</th>
<th>Maximum morning ≥4 contractions/hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weeks</td>
<td>Odds Ratio (95% CI)</td>
<td>Odds Ratio (95% CI)</td>
</tr>
<tr>
<td>22-24</td>
<td>3.0 (0.6-14.6)</td>
<td>3.2 (0.3-33.6)</td>
</tr>
<tr>
<td>27-28</td>
<td>1.0 (0.8-8.7)</td>
<td>0.44.2 (0.1-3.2)</td>
</tr>
<tr>
<td>31-32</td>
<td>1.3 (0.3-5.2)</td>
<td>0.5 (0.1-3.2)</td>
</tr>
<tr>
<td>P value</td>
<td>0.18</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Modified and reprinted with permission.¹

### Table 4. Uterine Contraction Frequency to Predict Spontaneous Birth <35 Weeks

<table>
<thead>
<tr>
<th>Gestational Age</th>
<th>Maximum evening ≥4 contractions/hr</th>
<th>Maximum morning ≥4 contractions/hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weeks</td>
<td>Sensitivity % (95% CI)</td>
<td>Sensitivity % (95% CI)</td>
</tr>
<tr>
<td>22-24</td>
<td>8.6 (8.6-9.4)</td>
<td>28.1 (28.1-28.7)</td>
</tr>
<tr>
<td>27-28</td>
<td>96.4 (96.4-96.7)</td>
<td>88.7 (88.7-89.0)</td>
</tr>
<tr>
<td>31-32</td>
<td>88.3 (88.3-88.7)</td>
<td>91.1 (91.1-91.4)</td>
</tr>
<tr>
<td>P value</td>
<td>0.18</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Modified and reprinted with permission.¹
2. Is the frequency of uterine activity greater in women with historical risk factors for preterm birth greater than in women who have no risk factors? No.

3. Does uterine activity differ according to gestational age in women destined for preterm birth? Yes.

4. Does uterine activity differ according to time of day in women destined for preterm birth? Yes.


**Discussion**

The distinction between association and prediction is a frequent source of confusion in medicine. Our findings highlight the importance of the difference. We found that contraction frequency is significantly increased in women who will deliver before 35 weeks' gestation, but the magnitude of that association was too small to be clinically useful. The results of this study serve to explain in part the failure of HUAM-based screening programs to yield a decrease in the rate of preterm birth. If the assumptions of contraction-based screening are presumed, HUAM must function as a reasonably sensitive screening test in order to apply interventions to reduce preterm birth. Our data indicate that even in a population with an increased risk of preterm delivery, the majority of women destined for preterm birth do not have a contraction frequency that distinguishes them from low-risk women. Similarly, increased contraction frequency in any individual woman is more likely to reflect advancing gestational age or diurnal variation than occult preterm labor. Given these observations, it is not surprising that clinical trials of uterine contraction surveillance of at-risk pregnancies do not show any advantage of such monitoring. The results of other MFMU Network studies summarized in this issue also reveal another reason for the failure of contraction-based prevention programs: their grounding in an incomplete understanding of how preterm labor occurs. There is growing evidence that subacute or even chronic pathophysiologic changes precede an eventual clinical diagnosis of preterm labor or preterm rupture of the membranes. Findings of inflammatory cytokines in second-trimester amniotic fluid, fetal fibronectin expression in cervico-vaginal mucus, cervical shortening as seen on ultrasonography, and increased concentrations of maternal salivary estriol, all detected weeks to months before a preterm birth, provide evidence that spontaneous preterm birth is the result of a long-term process that culminates rather than begins with uterine contractions.

More than two decades of clinical research into the appropriate role of uterine contraction frequency have followed a familiar if not altogether ideal path: An association is noted between an adverse outcome (eg, preterm labor and birth) and a clinical marker – a symptom, sign, or test result (eg, uterine contractions detected by education, self-palpation, and/or tocodynamometry). The association is clinically obvious, and leads to an equally obvious intervention, in this case tocolysis, in the expectation that the adverse endpoint will occur less often if the associated clinical marker is suppressed. Clinical trials are conducted but do not produce the expected result, directing the attention of the research community back to the original association, which when examined in detail proves to be much more complex than previously thought. This has been the path for contraction frequency monitoring, a path accompanied by great controversy in large part because of the presence of private companies as research sponsors and advocates in the process. It is appropriate to remember that the very same path has been and is now being followed for other markers associated with preterm birth such as nutritional and social support, vaginal infections, cervical length, and salivary estriol. What happened with uterine contraction monitoring is typical of clinical research. Explanation No. 2, also known as Dr. Eastman’s dictum, must always be kept in mind: “Only when the factors underlying prematurity are completely understood can any intelligent attempt at prevention be made.” Meanwhile, however, there are patients who need care before “the factors . . . are completely understood.” Our care for them must of necessity be based on the best understanding we have at present tempered by the humility of knowing that we do not really understand, yet.
References


