Clinical trials and tribulations: 17OHPC and preventing recurrent preterm birth

David Young, MD, MSc, FRCSC

In this issue of the Journal, Nelson et al1 report a prospective cohort of 430 consecutive women pregnant with a single fetus who had a history of 1 or more prior singleton spontaneous preterm births (ssPTBs). The women all received intramuscular (IM) 17-alpha hydroxyprogesterone caproate (17OHPC) prophylaxis without a lowering in the preterm birth (PTB) recurrence rate. The investigators were surprised by this in light of the specific recommendation of the Society of Maternal-Fetal Medicine (SMFM)2 for use of 17OHPC for this indication, first in 2012, but just reaffirmed3 in 2017.

The SMFM’s statement was largely based on a randomized controlled trial (RCT) by Meis et al4 in 2003. The Cochrane Library systematic review5 on 17OHPC includes the 3 other RCTs on this subject. The others6-8 all have issues affecting their relevance: conducted in the preultrasound era with subjects who did not have a viable pregnancy6 included; the entry criteria included 2 or more spontaneous abortions6 or prior preterm labor; did not exclude twins6 or cerclages6; the randomization method was undefined.; allocation was not blind7; and were conducted in a time6 or location7,8 compromising generalizability.

The RCT by Meis et al4 had more than twice the sample size (n = 463) of the other 36-8 combined and recruited subjects from 19 Maternal-Fetal Medicine Units Network centers, hence its key role in recommendations of the SMFM2,3,14 and the American College of Obstetrics and Gynecology1 and approval of the Food and Drug Administration (FDA)10 for the marketing of 17OHPC as the first and only agent for the prevention of recurrent PTBs. A detailed evidenced-based medicine critical appraisal11,12 of the RCT by Meis et al4 and the study by Nelson et al1 may be helpful.

Because women with a prior ssPTB13-15 were at high risk for PTB in a subsequent pregnancy, Meis et al4 conducted a multicenter trial to test the effectiveness of 17OHPC compared with placebo in the prevention of recurrent ssPTB in this population. The primary outcome was PTB before 37 weeks (359 days), on which the sample size was calculated. No rationale or insight was provided on this choice of primary outcome, as opposed to a more clinically important one, such as PTB before 35, 32, or 28 weeks, perinatal mortality, or a serious neonatal morbidity measure. Because of their decreased incidence, each of these would have led to an increased, and perhaps unachievable, sample size.11,12

Candidates’ medical records were reviewed to confirm a prior live ssPTB between 20 and 36 weeks 6 days because of spontaneous labor or preterm premature rupture of membranes. Current or planned cerclage was among the appropriate exclusion criteria. Recruited women were currently pregnant between 15 and 20 weeks 3 days confirmed by ultrasound, which also ruled out major anomalies.

Consenting women were given a test IM injection of placebo and were to return in 1 week for randomization. If that return visit did not happen before 20 weeks 6 days, she was not accepted into the trial or randomized. This run-in step allowed potential subjects to better understand the intervention and might predict compliance. The urn method16 of randomization, with stratification by center, was used to create the computerized randomization sequence. There was a 2:1 ratio of assignment to 17OHPC because placebo subjects would receive weekly painful injections with no possibility of direct benefit. Allocation was concealed. Risk of selection bias was thus low5,11.

Study participants returned for weekly IM injections of identically appearing 17OHPC or placebo until 36 weeks’ gestation or delivery. Participants, caregivers, outcome assessors, and data analysts were all blind to group assignment. Prenatal care was otherwise by their usual caregivers. The use of castor oil injection as the placebo has raised some concern17 because of its potential as a uterine muscle stimulant. Castor oil was the vehicle for the 17OHPC injection as well so that all subjects received it if there is any such effect.

From a previous Maternal-Fetal Medicine Network study,18 37% was the anticipated rate of recurrent PTB in the placebo group. Using a type 1 error (2 sided) of 5% and a power of 80%, a sample size of 500 women (334 receiving 17OHPC and 166 receiving placebo) would detect a 33% reduction in PTB from 37% to 25%. No rationale was given for this chosen clinically important difference. This choice meant that a smaller, although arguably substantial reduction in PTB rate of 25%, for example, would not reach statistical significance, with this sample size.19 The smaller the clinically important difference, the larger the sample size needed. The larger the detectable difference sought, the smaller the sample size, but the larger the truly clinically important difference that may be found and not reach statistical significance. In this RCT,4
statistical significance was achieved for the primary outcome, so this did not become an issue, but it is an important decision in trial design.

Analysis was on the intent-to-treat principle (i.e., all subjects were assessed in the group to which they had been randomized without exception). Statistical testing was appropriate, even conservative, because the Wilcoxon rank-sum test for continuous variables makes no assumption regarding normal distribution of data. Although adjustment of the type 1 error was performed in the interim analyses, no consideration was given to the testing of multiple outcomes, which occurred in the analysis. Although specific $P$ values were stated, no declaration was made regarding adjusting the 5% level chosen for attaining statistical significance.

Randomized women were 59.0% non-Hispanic black, 24.4% non-Hispanic white, and 14.9% Hispanic. Subsequent to randomization, race, a known confounding variable, was balanced in the 2 study groups, as were other pretreatment characteristics, with one important exception. There was a greater percentage of women with more than 1 prior ssPTB in the placebo group (41.2% vs. 27.7%, $P = .004$, Fisher exact [2 sided]). Such an imbalance can occur by chance with randomization, unfortunately making the placebo group more at risk for PTB. More prior ssPTBs increases recurrence risk. A known confounder may be addressed in the study design by stratifying randomization on that variable. This RCT was stratified by study center and likewise could have been for such a known risk factor. The downside with increasing the number of stratification factors is that allocation may seem more complex, and some strata may contain small numbers. The number of strata increases by the product of the stratification factors (in this case, 19 centers by 2 prior PTB levels, say). A central 24 hour randomization point by phone or the Internet could facilitate this.

There was no evidence of contamination (placebo group receiving 17OHP or other progesterone) or inequity in cointervention (tocolytic therapy and corticosteroid use). Compliance was high. Side effects were reported, with only injection site swelling being more common with 17OHP. On the basis of a second planned interim analysis, the RCT was stopped early, but just a little, with a sample size of 463 rather than 500. There is growing evidence that early stopping may bias toward benefit of an intervention.

All relevant outcomes were reported. There was no reporting bias. The primary outcome, ssPTB before 37 weeks, was 36.3% with 17OHP and 54.9% with placebo ($P < .001$). Delivery before 35 weeks was also less frequent with 17OHP (20.6% vs. 30.7%, $P = .02$). Similar benefits were seen before 32 weeks ($P = .02$). Outcome data were not available for 4 women, all in the 17OHP group. Curiously, no comment was made on this.

The primary outcome significance test was calculated with a denominator of 306, rather than 310, raising potential for attrition bias. No sensitivity testing was reported. My calculation assuming all 4 missing patients had PTBs, thus 115 PTBs, remains statistically significant ($P = .003$, Fisher exact). Repeating that process for ssPTB less than 35 weeks also changes its level of significance ($P = .039$). When the imbalance in prior ssPTBs was recognized, an adjusted analysis was appropriately performed. The adjusted relative risk of delivery before 37 weeks was 0.70 (95% confidence interval, 0.57–0.85), but the concern and perception linger. The statistical method for the adjustment was not stated.

Despite my comments, the RCT by Meis et al was well designed and conducted. It is much easier to criticize than carry out the ideal clinical trial. The 2 largest concerns are the imbalance in number of women with more than 1 prior ssPTB and the high rate of ssPTB in the placebo group (54.9%). The anticipated rate had been 37%. The high number of women with multiple prior ssPTBs, the early gestation of prior ssPTB, and the need for strong motivation among voluntary participants because of the nature of the intervention were proposed as potential contributors to the high ssPTB recurrence rate in recruits. Nevertheless, these issues continue to be foci of concern among critics.

The investigators were forthright in declaring an initial start of the trial that was halted after 150 subjects had been enrolled because of FDA concerns with the original pharmaceutical company supplying the study drug. No data from these subjects were provided or commented on; however, it has been reported elsewhere that the rate of recurrent ssPTB in the placebo group for this initial start was 36% at the time curtailed, a rate similar to that anticipated in this group and also actually achieved with 17OHP in the published trial. The publication wisely concluded by cautioning that 17OHP may not be effective in women with a lower risk of ssPTB and that results may not be generalizable.

The report by Nelson et al is a prospective cohort with comparison with historical controls, an observational study. A single-center RCT had been considered but believed not feasible because of the cost and recruitment time needed. The study was conducted at Parkland Hospital, which serves the medically indigent women of Dallas County, TX. Parkland had been one of the 19 centers participating in the RCT by Meis et al. One of the authors of this review was a coauthor on the RCT by Meis et al.

A specific high-risk clinic was established for women with the eligibility criteria of singleton pregnancy with prior ssPTB or preterm premature rupture of membranes between 20 0/7 and 35 0/7 weeks’ gestation. When 17OHP was introduced for the prevention of recurrent ssPTB beginning Jan. 1, 2012, all 17OHP patients were followed up. Data collection was conducted under a research protocol with institutional review board approval. Research nurses assessed patients for eligibility and reviewed the obstetric history using a prespecified manual of operation. Women with a prior medically indicated PTB were excluded. Injections began between 16 0/7 and 20 0/7 weeks. The primary outcome chosen was recurrent ssPTB after 35 weeks.

The historical rate of recurrent ssPTB less than 35 weeks at Parkland Hospital was 16.8% prior to 17OHP introduction and was used in sample size calculation. Table 2 in the report
by Nelson et al indicates that this historical cohort consisted of 5787 women from the Parkland population of 1988–2011. For 90% power, a sample size of 413 women was estimated to detect a one-third reduction in recurrent PTB, using a 1-sided alpha of 0.025. A 1-sided test was chosen because a lowering in rate was anticipated. In fact, the sample size value would be the same for a 2-sided alpha of 0.05.

A weakness in the trial design with a historical control group is that the same rigor in data collection and confidence in quality that can be achieved in the prospective cohort cannot be assured. The historical data had been collected before the research protocol was begun. Other background changes in the study population, and health care in general, can occur when comparing participants enrolled after 2012 in the study cohort with historical controls, introducing selection bias. For example, the racial composition of women with a prior ssPTB before 35 weeks at Parkland Hospital in the first half of that period for the historical cohort was as follows: Hispanic, 49%, black, 42% and white, 8%, whereas in the prospective study cohort, it was Hispanic, 80%, black, 17%, and white, 3%. This is a substantial (P < .001, χ²) difference.

There is also an obvious aging in the Parkland population noted, although the published data are not categorized in a way that allows a reviewer to test statistical significance. This changing composition of the population would lower the risk of recurrent ssPTB anticipated in the more recent 17OHPC cohort, a bias favoring 17OHPC.

The intervention, 17OHPC, consisted 250 mg of active drug given IM in a sesame oil vehicle. This avoided the concerns of the castor oil vehicle in the RCT. The drug was compounded by a local pharmacy for $24.99 per dose. Potency and sterility testing confirmed a stable drug dose and freedom from infection. Plasma concentrations of 17OHPC were measured at 24 and 32 weeks in 116 and 101 women, respectively. There was no difference in 17OHPC plasma levels between those who did, or did not, have recurrent ssPTBs, no dose-response gradient. Plasma levels corresponded with concentrations previously reported using a castor oil vehicle.

Between Jan. 1, 2012, and March 31, 2016, 456 consecutive women with a prior ssPTB before 35 weeks received 17OHPC. There could be no blinding of participants or personnel in the conduct of the trial or the assessment of outcomes, hence possible performance and detection biases.

In the 17OHPC prospective cohort, the recurrent ssPTB rate less than 35 weeks at Parkland was actually 24.7%, not less than that of the historical controls (16.8%, P = 1.0, 1 sided) and not less than that in the 17OHPC group in the RCT (20.6%). The rate of recurrent ssPTB less than 35 weeks (16.8%) is much lower in the Parkland historical control group than that in the placebo group (30.7%) of the trial by Meis et al, although Parkland was 1 of the 19 centers contributing subjects to the RCT by Meis et al.

The demographic changes noted in the previous text are a plausible factor in this, so for comparison, a 3:1 matched (on race, body mass index, age, and prior PTB history) control group of 1290 was drawn from the historic subjects. A recurrent ssPTB rate before 35 weeks of 23% was found (P = .45). When analyzed according to specific ssPTB obstetric history, again there was no reduction in ssPTB rates with 17OHPC, although these groups became smaller and with reduced power.

A total of 26 women were excluded from the analysis because of loss to follow-up (n = 21) or delivery before 20 weeks (n = 5) in the 17OHPC cohort. Hence, there was some risk of attrition bias. Arguably, in an intent-to-treat approach, the ssPTBs before 20 weeks in the 17OHPC group would be included as ssPTBs before 35 weeks. Those lost to follow-up, in the extreme, could all have been ssPTBs before 35 weeks or all not. As sensitivity analyses, calculating the resulting ssPTB before 35 weeks rate gives results from 19.7% to 29.0%, none less than the historical control group rate.

An increased incidence of gestational diabetes (13.4%) was diagnosed with 17OHPC treatment, compared with the matched control group (8%, P = .001). This finding supports a recent report.

The Grades of Recommendation, Assessment, Development, and Evaluation approach for assessing quality of evidence and individual study risk of bias would start the RCT by Meis et al with a high rating, subsequently tempered by the unexpectedly very high risk nature of its study population, the imbalance in subjects with more than 1 prior ssPTB favoring the 17OHPC group, and the early stopping decision.

The observational study by Nelson et al with historical controls begins with a low rating but improves with the particularly rigorous manner by which the prospective cohort was followed up. The major limitation of the historical control approach is the lack of confidence in internal validity that is possible with an RCT, in which known and unknown confounders are, to a high probability, balanced by randomization, reducing risk of selection bias.

Why did the report by Nelson et al not confirm the findings of Meis et al? Their study populations’ racial composition was, as outlined in the previous text, quite different (P < .001, χ²). The Parkland prospective cohort’s racial makeup was at lower risk of ssPTB than the control group of Meis et al. The data provided on age distribution and body weight seem to accentuate this baseline risk difference, although, as presented, significance testing is not possible. Interestingly, the frequency of women with more than 1 prior ssPTB was 32.1% (138 of 430) (K. J. Leveno, personal communication), which was less than in the placebo group (41.2%, P = .048, Fisher exact test) of Meis et al but very similar to that of the entire RCT recruits (32.2%, P = 1.0, Fisher exact test).

The effectiveness demonstrated for 17OHPC in the RCT’s high-risk population does not appear to generalize to the Parkland lower recurrent PTB risk population. The large, well-documented prospective cohort is valuable because it presents a real-world scenario, questions the generalizability...
of 17OHPG prophylaxis, and provides additional support for gestational diabetes as a side effect.

A condition of the FDA’s approval of 17OHPG was that a confirmatory, well-designed, multicenter, double-blind RCT was in progress. This international trial, which has pharmaceutical industry funding, has a planned sample size of 1707, with power to assess perinatal mortality and morbidity outcomes, in addition to sPTB less than 35 weeks. There is no indication on the study web site that this RCT has been stratified in any way on the number of prior PTBs. With completion scheduled for mid-2018, the results are eagerly anticipated and germane, even more so now, in light of the report by Nelson et al7 and recent publications, which renew interest in vaginal progesterone, a less costly, less invasive, possibly more effective, and applicable option. The status of the SMFM recommendation on 17OHPG prophylaxis for recurrent sPTB (confirmation, revision to an even more high-risk indication, or complete change), as well as FDA approval, awaits.

REFERENCES


12. Guyatt GH, Sackett DL, Cook DJ; for the Evidenced-Based Medicine Working Group. Users’ guides to the medical literature, II: how to use an article about therapy or prevention. B: what were the results and will they help me in caring for my patients? JAMA 1994;271:59-63.


