Nos. 05-380 & 05-1382

AUG 10 2006

IN THE

# Supreme Court of the United States

ALBERTO R. GONZALES, ATTORNEY GENERAL,
*Petitioner*,

—v.—

LEROY CARHART, ET AL.,
*Respondents*.

ALBERTO R. GONZALES, ATTORNEY GENERAL,
*Petitioner*,

—v.—

PLANNED PARENTHOOD FEDERATION OF AMERICA, INC., ET AL.,
*Respondents*.

ON WRITS OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE EIGHTH CIRCUIT AND THE UNITED STATES COURT OF APPEALS FOR THE NINTH CIRCUIT

## BRIEF OF *AMICI CURIAE* STATISTICIANS PROFESSOR GEORGE W. COBB, PROFESSOR MARY W. GRAY, PROFESSOR NORMAN HENDERSON, PROFESSOR JOHN J. MCARDLE, PROFESSOR JAMES TRUSSELL, AND PROFESSOR JEFFREY A. WITMER IN SUPPORT OF RESPONDENTS

MOLLY S. BOAST
*Counsel of Record for the Amici*
CHRISTIAN R. EVERDELL
AARON BLOOM
JOHN S. CRAIG
DEBEVOISE & PLIMPTON LLP
919 Third Avenue
New York, New York 10022
(212) 909-6000

*Counsel for Amici Curiae*

BLANK PAGE

## TABLE OF CONTENTS

Page

Page

# TABLE OF AUTHORITIES

Page

## FEDERAL CASES

**MISCELLANEOUS**

BLANK PAGE

## INTEREST OF *AMICI CURIAE*[1]

*Amici curiae* are professors at prominent universities and colleges, as well as distinguished statisticians in their fields. Even with the breadth of their professional interests, *amici* are united in their belief that statistical analysis used in judicial decision-making must rest on sound scientific principles. *Amici* are aware that the interpretation of scientific data often plays a role in courts' adjudication of facts and law. Absent professional guidance, courts may misapply statistical principles and render decisions based, at least in part, on invalid or unsupportable assumptions.

*Amici* believe that portions of the record below reflect a misapplication of sound statistical methodology. Misleading and incorrect statements by some of the government's expert witnesses were relied upon by two of the district courts, and have been perpetuated in briefs submitted to this Court by Petitioner and supporting *amici*. *Amici* here believe that correcting those misleading and incorrect assertions, and providing an accurate interpretation of the data upon which those assertions purport to rely, will contribute to the Court's understanding and help it reach a well-reasoned, scientifically-principled decision in these highly important appeals.

---

[1] Pursuant to Supreme Court Rule 37, *amici curiae* state that this brief has not been authored in whole or in part by counsel for a party in this case, and no entity other than the *amici* or their counsel made a monetary contribution to the preparation or submission of this brief. Letters of consent to the filing of this brief have been lodged with the Clerk.

*Amici curiae* include the following:

**George W. Cobb, Ph.D.**
Robert L. Rooke Professor of Mathematics and Statistics
Mount Holyoke College

**Mary W. Gray, Ph.D.**
Professor, Department of Mathematics and Statistics
American University

**Norman Henderson, Ph.D.**
Professor (Emeritus), Department of Psychology
Oberlin College

**John J. McArdle, Ph.D.**
Professor, Department of Psychology
Director, Quantitative Area
University of Southern California

**James Trussell, Ph.D.**
Professor, Economics and Public Affairs
Director, Office of Population Research
Princeton University

**Jeffrey A. Witmer, Ph.D.**
Professor, Department of Mathematics
Oberlin College

## INTRODUCTION AND SUMMARY OF ARGUMENT

The relative safety of intact dilation and extraction ("intact D&X") compared to dilation and evacuation by disarticulation ("D&E") is a critical question in the scientific debate surrounding the Partial-Birth Abortion Ban Act of 2003, which all parties agree prohibits, at a minimum, intact D&X. At the center of this debate is a peer-reviewed study by Dr. Stephen T. Chasen and colleagues, which compared the complication rates of intact D&X and D&E in a group of 383 women who received these procedures at The New York Weill Cornell Medical Center. The study authors found the comparative complication rates of the two procedures to be similar, suggesting that women undergoing intact D&X face no increased risk compared to those undergoing D&E. Stephen T. Chasen *et al.*, *Dilation and evacuation at ≥ 20 weeks: comparison of operative techniques*, 190 Am. J. Obstet. Gynecol. 1180 (2004) (the "Chasen study") (attached as Appendix A).[2]

The Chasen study was the subject of extensive expert witness testimony in the trial court records in Nebraska, New York, and California, where several of the government's experts incorrectly asserted, with varying degrees of certitude, that portions of the data show that intact D&X may be riskier than D&E. *Amici* believe those statements were misleading because the numerical differences upon which they were based are not "significant" as measured by

[2] *Amici* use the term "intact D&X," as opposed to "intact D&E" or simply "D&X," to reflect the term Dr. Chasen used in his published study. *See* Chasen, *Dilation and evacuation, supra,* at 1180-83.

accepted methods of statistical analysis. Those misstatements were incorrectly relied upon by some of the lower courts and now reappear in briefs submitted to this Court. *Amici* believe it is critical to ensure that the Court has a scientifically accurate understanding of the Chasen study and the conclusions about the relative risk of intact D&X and D&E that can, and cannot, validly be drawn from its data.

At the root of the government experts' misstatements is the meaning of the statistical concept of probability value (or "p-value"), which measures the statistical significance of observed differences in outcomes among members of a group. *Amici* therefore seek to do the following: (1) provide the Court with an accurate description of the concept of p-value; (2) explain why the relevant p-values in the Chasen study are not statistically significant and therefore do not support the assertions of the government's experts that intact D&X may be more risky than D&E; and (3) highlight additional statistical evidence from a follow-up study published by Dr. Chasen that further undermines those experts' assertions.

## SUMMARY OF THE CHASEN STUDY

The objective of the Chasen study, published in the peer-reviewed *American Journal of Obstetrics and Gynecology* in 2004, was "to evaluate the relative safety of dilation and evacuation and intact D&X in patients undergoing surgical abortion late in the second trimester." Chasen, *Dilation and evacuation, supra,* at 1181. Prior to the publication of the Chasen study, no data had been formally published regarding the complication rates of intact D&X. Dr. Chasen sought to collect data on the complications that resulted from intact D&X and D&E in order to assess whether intact D&X posed a greater health risk than D&E.

Dr. Chasen's study sample included a group of 383 women who underwent either intact D&X or D&E at twenty weeks' gestation or greater at The New York Weill Cornell Medical Center between June 1996 and June 2003. Dr. Chasen retrospectively identified cases for inclusion in the study sample by reviewing operative reports to determine the extraction technique used. Of the 383 women included in the study, 120 (31.3%) received intact D&X and 263 (68.7%) received D&E. Dr. Chasen used this sample to assess whether there was any difference between the two surgical variations in the number or severity of operative and post-operative complications. "Complications" was defined in the study as "any situation requiring unplanned intervention," including "unplanned hospital admission, repair of any genital tract lacerations, return to the operating room for additional procedures, and blood transfusion." *Id.* at 1181.

Dr. Chasen also identified a smaller sample of 62 women from among the larger group of 383 who later became pregnant and sought prenatal care at The New York

Weill Cornell Medical Center. Of those 62 women, 17 had originally undergone intact D&X and 45 had originally undergone D&E. Dr. Chasen used this sample to assess whether there was any difference between the two surgical variations in subsequent adverse pregnancy outcomes. Specifically, Dr. Chasen examined the incidence of spontaneous preterm birth, which occurs when a woman unexpectedly delivers prematurely in the third trimester before the fetus has been carried to term.

Dr. Chasen found that operative and post-operative complications "occurred with similar frequency" in the two patient groups—5.0% (6 out of 120) of the intact D&X group compared to 4.9% (13 out of 263) of the D&E group. *Id.* at 1182. With regard to subsequent pregnancy outcomes, Dr. Chasen found that 11.8% (2 out of 17) of the intact D&X group and 4.4% (2 out of 45) of the D&E group experienced spontaneous preterm birth in a later pregnancy. The percentage difference in the incidence of subsequent preterm birth, however, was not statistically significant when measured by a p-value calculation, the established method of assessing the statistical significance of different outcomes. Specifically, the p-value calculation of 0.30 was far greater than 0.05, the threshold at or below which an observed difference is deemed statistically significant. Dr. Chasen ultimately concluded, based on these data, that "intact D&X appears to have similar complication rates as dilation and evacuation" and that "[t]he observed complication rates and subsequent obstetric outcomes appear comparable between the 2 techniques." *Id.* at 1183.

## ARGUMENT

### I. Testimony That Intact D&X Patients In The Chasen Study Showed An Increased Risk of Subsequent Preterm Birth Is Inaccurate And Should Be Disregarded.

In his testimony about the Chasen study before Judge Casey in the Southern District of New York, one of the government's expert witnesses, Dr. Steven L. Clark, ascribed a false significance to the numerical difference in the incidence of subsequent preterm birth in the intact D&X group versus the D&E group. Dr. Clark did this in two ways. First, he asserted that patients who received intact D&X showed a "threefold increased risk of premature birth," N.Y. Tr. 2386:4-2386:13, even though the numerical difference between the two groups was not statistically significant. One cannot validly conclude from these data that patients undergoing intact D&X have an increased risk of subsequent preterm birth compared to patients undergoing D&E. Second, Dr. Clark compounded his error by asserting that the 0.30 p-value in the Chasen study indicated a 70% chance that there was in fact a true difference between intact D&X and D&E in the likelihood of causing subsequent preterm births. N.Y. Tr. 2425:15-2426:18. Such an assertion fundamentally misstates the conclusions that can be drawn from a p-value calculation. A p-value helps one evaluate whether an observed difference is statistically significant; it does not indicate the probability that there is a true difference between two groups.

Dr. Clark's misstatements about the Chasen study data, p-value calculations, and statistical significance persist in the record before this Court. For example, although Judge Casey

struck down the Partial-Birth Abortion Ban Act, he mistakenly endorsed Dr. Clark's testimony in his opinion. *See Nat'l Abortion Fed'n ("NAF") v. Ashcroft*, 330 F. Supp. 2d 436, 476-77 (S.D.N.Y. 2004). The Solicitor General, in his brief to this Court in *Gonzales v. Carhart*, reiterates Dr. Clark's misleading assertion that the Chasen study "showed a nearly threefold increase in the risk of premature birth" in women who had undergone intact D&X, without mentioning the lack of statistical significance of those data. Petr.'s Br. (*Carhart*) 38 n.12. Similarly, the *amicus* brief filed by the American Association of Pro-Life Obstetricians and Gynecologists (AAPLOG) *et al.* cites Dr. Clark's testimony about the supposed "threefold increased risk of premature birth" as evidence of the "possible dangers of D&X." Br. of *Amici Curiae* AAPLOG *et al.* 19-20.

Because these assertions disregard well-settled statistical principles for interpreting data, or worse, fundamentally misstate those principles, *amici* seek to provide the Court with an accurate explanation of the meaning of a p-value calculation and to explain from a statistical standpoint why the Chasen study data do not support the conclusion that intact D&X causes more complications than D&E in subsequent pregnancies.

### A. It Is Widely Accepted In The Scientific Community That An Observed Difference Between Two Groups Is Statistically Significant Only When The P-value Equals 0.05 Or Less.

The Chasen study was designed to evaluate the relative safety of intact D&X and D&E by examining data from a sample of patients who received one or the other of these two surgical variations. In any study that seeks to draw a conclusion about the population as a whole from a smaller

sample, it is necessary to determine whether any differences observed in the sample are "statistically significant." If the differences are not "statistically significant," it is more likely that they reflect variation due to chance. Only when observed differences have been determined to be "statistically significant" may these differences validly be attributed to the general population. Statisticians perform this analysis by calculating a probability value (or "p-value"), which accounts for factors such as a small sample size and a small number of events that might skew the observed data and suggest a false difference.

To use a straightforward example, suppose we wish to compare how often a certain penny and a certain nickel turn up "heads." We begin by forming what is called a "null hypothesis"—a hypothesis that there is no difference between the frequency with which the penny will turn up heads and the frequency with which the nickel will turn up heads. We then test the null hypothesis by flipping each coin repeatedly and counting the number of times that each coin turns up heads. Suppose we flip each coin 5 times and find that the penny turns up heads in 4 of the 5 flips (or 80% of the time) and the nickel turns up heads in 2 of the 5 flips (or 40% of the time). Even though the observed frequencies clearly differ numerically, it would be a mistake to reject the null hypothesis based on these limited data and conclude that the penny is twice as likely as the nickel to turn up heads. The number of flips is too small for us to question seriously the null hypothesis. It is entirely plausible that we would observe 4 out of 5 heads when tossing a "fair" penny and 2 out of 5 heads when tossing a "fair" nickel simply by chance. In contrast, we would feel more confident about rejecting the null hypothesis, and concluding that the penny and the nickel are in fact different, if we flipped the coins 5,000 times each

and found that the penny turned up heads 4,000 times while the nickel turned up heads 2,000 times.

Statisticians formally compare observed data with the null hypothesis by calculating a p-value; ultimately, it is the p-value that determines whether the null hypothesis will be rejected. Statisticians calculate a p-value by first assuming that the null hypothesis is true—that there is actually no difference in the frequency with which the penny will turn up heads and the frequency with which the nickel will turn up heads. Given this assumption, a p-value then measures the probability that we would see a difference in the sample similar to or greater than the difference we actually observed. To return to the coin flip example, a p-value measures the likelihood that the difference we observed—that is, the difference in the number of heads that turned up for the penny versus the nickel—or an even greater difference, would in fact occur, assuming there was no *true* difference between them. In doing so, a p-value accounts for the distortion caused by the small number of flips for each coin.

It is generally accepted within the scientific community that an observed difference between two outcomes is significant only if the p-value is less than or equal to 0.05.[3] A p-value at or below the 0.05 threshold indicates that if there is no true difference in the population, the likelihood of observing the difference we observed in our sample is small—only 5%. If the probability of observing such a

[3] A p-value of 0.05 is the broadest cut-off for statistical significance recognized by the scientific community. Designation of more restrictive p-value thresholds of 0.01 or even 0.001 as indicators of stronger statistical significance is also generally accepted.

difference is that low or lower, it is very unlikely that the observed difference can be explained by chance. At that point, statisticians feel comfortable rejecting the null hypothesis and concluding that there is a "statistically significant" difference between the two outcomes. A p-value greater than 0.05, on the other hand, indicates sufficient possibility that the observed difference was due to chance that it would be inappropriate to reject the null hypothesis and conclude that the difference was significant. This statistical convention is well-settled and uncontroversial. *See* David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, *in Fed. Judicial Ctr., Reference Manual on Scientific Evidence* 83, 121-125 (2d ed. 2000). Indeed, the government's experts, including Dr. Clark, did not contest this point in their testimony. *See* N.Y. Tr. 2425:15-2425:17 (Clark); N.Y. Tr. 2197:8-2197:10 (Sprang); Neb. Tr. 1721:18-1722:8 (Lockwood).

In the context of the Chasen study, p-value calculations help to assess whether any differences between intact D&X and D&E observed in the study sample were likely due to chance or were significant enough for one validly to conclude that a similar difference would likely occur in the general population of all women undergoing intact D&X and D&E. Only a p-value of 0.05 or less would support the latter conclusion.

### B. The Chasen Study Found No Statistically Significant Difference In The Incidence Of Subsequent Preterm Birth Between The Intact D&X Group And The D&E Group Based On A P-value Of 0.30.

Because the Chasen study data pertaining to differences in subsequent preterm birth rates between the intact D&X

and D&E groups is not statistically significant, it is inaccurate to conclude, as Dr. Clark did, that intact D&X showed a "threefold increased risk" in subsequent preterm birth compared to D&E. N.Y. Tr. 2386:10. Dr. Clark focused solely on the *numerical* difference in the percentage of patients who experienced subsequent preterm births in the intact D&X group compared to the D&E group. He did not take into account the effect of the small number of events and the relatively small sample size that, in this case, renders the numerical difference statistically inconsequential. In sum, Dr. Clark's assessment of the data disregards the importance of p-value.

To be sure, the raw numbers show that the patients who received intact D&X experienced a higher rate of subsequent preterm births than those who received D&E. Out of a total of four preterm births, two occurred in each patient group. This translates to an event rate of 11.8% (2 out of 17 patients) in the intact D&X group versus 4.4% (2 out of 45 patients) in the D&E group, a roughly threefold difference. The numerical comparison of these event rates, however, is only the beginning of the analysis. To draw any valid scientific conclusion from these data about the relative risk of preterm birth from undergoing intact D&X versus D&E, it is necessary to evaluate how likely it is that the observed difference between the two groups is an accurate representation of what would occur in the broader population of women undergoing these procedures. In other words, one must determine whether the threefold numerical difference observed in the Chasen study would likely replicate itself in a larger sample, or whether, as in the coin flip example, the observed difference is easily explained by chance.

Dr. Chasen appropriately considered this question in his study by calculating a p-value to determine whether the observed difference in preterm births was significant. The resulting p-value of 0.30 falls well above the recognized threshold for statistical significance of 0.05 or less, indicating that there is enough of a possibility that the difference was caused by chance that it would be invalid to conclude that there is increased risk for the intact D&X group. Dr. Chasen correctly determined that the difference was insignificant and thus concluded that the "subsequent obstetric outcomes appear comparable between the 2 techniques." Chasen, *Dilation and evacuation, supra*, at 1183.

Dr. Charles Lockwood, the Chairman of the Department of Obstetrics, Gynecology and Reproductive Sciences at Yale University and one of the government's expert witnesses, agreed that these data lacked significance and could not validly support a conclusion that intact D&X subjected women to increased risk in later pregnancies. With regard to the difference in subsequent preterm births, Dr. Lockwood testified:

> It's not statistically significant, so I wouldn't . . . put an enormous amount of weight on it. . . . [T]here were additional risk factors in the D&X group for subsequent premature delivery. . . . [F]ormally as a clinician-researcher, I wouldn't draw any conclusions from it.

Neb. Tr. 1721:23-1722:4; *see also* Cal. Tr. 971:10-971:13 (testimony of government's expert witness, Dr. Watson A. Bowes, Jr., that the Chasen study "certainly does not suggest that D&E where the fetus is removed intact is any less safe than where the [fetus] is disarticulated").

The conclusions of Dr. Chasen and Dr. Lockwood that the data do not indicate an increased risk of subsequent preterm birth in the intact D&X group are based on an accurate understanding of p-value and are consistent with accepted methods of statistical analysis. These conclusions were also endorsed by the district courts in Nebraska and California. *See Carhart v. Ashcroft*, 331 F. Supp. 2d 805, 1022-23 & n.150 (D. Neb. 2004) (crediting the testimony of Dr. Lockwood over that of Dr. Clark about the risk of subsequent preterm birth); *Planned Parenthood Fed'n of Am. v. Ashcroft*, 320 F. Supp. 2d 957, 1001 (N.D. Cal. 2004) (finding that the Chasen study does not support a conclusion that intact D&X may lead to an increased risk of subsequent preterm birth).

*Amici* urge this Court to disregard the testimony of Dr. Clark about the relative risk of preterm birth in subsequent pregnancies for intact D&X and D&E, and similarly to reject any arguments by the Solicitor General or supporting *amici* based on that testimony.

### C. Dr. Clark's Statements About The Meaning Of A P-value Of 0.30 Are Incorrect And Should Be Disregarded.

In addition to drawing a statistically invalid conclusion about the Chasen study data, Dr. Clark made two inaccurate statements about the meaning of p-value in his testimony that are not in accord with accepted principles in the scientific community.

First, Dr. Clark implied that p-values of 0.30 and 0.05 were substantively similar. In his testimony about the difference in the incidence of spontaneous preterm birth in the two patient groups, Dr. Clark conceded that the result

would be statistically significant only if the p-value were less than 0.05, but then erroneously suggested that the 0.30 p-value in the Chasen study was roughly equivalent to a p-value of 0.05:

> Well, I would say that if the P value is less than .05 then there is absolutely no question. Usually we don't accept something as absolute scientific truth unless the chances are less than five percent. *In this particular case it's just stretching it a little bit to say 30 percent chance*, but you're roughly correct.

N.Y. Tr. 2426:2-2426:8 (emphasis added). Contrary to Dr. Clark's statement, a p-value of 0.30 does not even approach the level of statistical significance indicated by a p-value of 0.05. A p-value of 0.30 means that, if intact D&X and D&E truly posed the same risk of subsequent preterm birth and we conducted the Chasen study multiple times, we would expect to observe the same (or greater) difference in subsequent preterm births that we observed in the Chasen study 30% of the time. Given that we would observe the same (or greater) difference relatively frequently—almost one third of the time—it would be inappropriate to reject our null hypothesis that intact D&X and D&E are similar. Only when the likelihood of observing that difference is small—5% percent or less—can we validly conclude that the difference we observed is significant. Dr. Clark's assertion that a p-value of 0.30 can be "stretched" to equate to a p-value of 0.05 is simply incorrect.

Second, Dr. Clark erroneously testified that 1 minus 0.30, or 0.70 (70%), could measure whether there was an actual difference between intact D&X and D&E in the risk of subsequent preterm birth:

> [T]here is a 30 percent chance this occurred by chance and a 70 percent chance that it in fact is a true, meaningful, increased risk . . . .

N.Y. Tr. 2429:20-2429:22; *see also id.* at 2426:12-2426:13 ("[T]here is a 70 percent chance that in fact this is a real difference."). This misstatement was subsequently adopted in the district court's opinion in *NAF*. *See NAF*, 330 F. Supp. 2d at 476-77.

Dr. Clark's statements fundamentally misconstrue the meaning of p-value. A p-value is not the probability that there is no real difference between two groups, and one minus a p-value is not the probability that there is an actual difference between two groups. P-values are calculated *assuming that there is no real difference between the two groups*. A p-value can only tell us the likelihood that an observed difference between two groups would be repeated in subsequent studies, assuming the null hypothesis were true. To use the Chasen study as an example, a p-value of 0.30 means that, if the Chasen study were rerun multiple times, and there were in fact no true difference in subsequent obstetric outcomes between intact D&X and D&E, then one would expect to see differences as large or larger than the one observed in the Chasen study in 30% of the subsequent trials and differences smaller than the one observed in the Chasen study in 70% of the subsequent trials. N.Y. Tr. 2683:5-2683:19 (Howell). The 0.30 p-value does not measure the percentage likelihood that there is a "real difference" between intact D&X and D&E that caused the difference in subsequent preterm birth rates.

Dr. Clark's testimony that the Chasen study indicates a 70% chance of increased risk of subsequent premature birth

following intact D&X is therefore incorrect and should be rejected.

### D. Dr. Chasen's Follow-up Study Suggests That The Risk Of Subsequent Preterm Birth Was Correlated To The Patient's Medical Indication For Abortion, Not To The Procedure She Received.

Dr. Chasen's follow-up study, published in 2005 in the *American Journal of Obstetrics and Gynecology*, provides further evidence that the higher percentage of subsequent preterm births observed in the intact D&X group was not caused by any heightened risk of the procedure itself, but instead was related to the medical condition of the woman seeking the initial abortion. Stephen T. Chasen *et al.*, *Obstetric outcomes after surgical abortion at ≥ 20 weeks' gestation*, 193 Am. J. Obstet. Gynecol. 1161 (2005) (the "Chasen follow-up study") (attached as Appendix B).

The Chasen follow-up study examined data from the same patient sample of 383 women in the original Chasen study, but had a different objective. The Chasen follow-up study sought to "identify risk factors for spontaneous preterm birth" and, as a secondary goal, to assess whether intact D&X or D&E was "associated with higher rates of spontaneous preterm birth in future pregnancies." *Id.* at 1162. With over a year of additional follow-up data available since the publication of the original study, the number of subsequent pregnancies among the 383 women increased from 62 to 120.[4] Of these 120 subsequent pregnancies, 102 reached the

---

[4] *Amici* in support of Petitioner imply that the addition of 58 pregnancies from the time of the original Chasen study to the time

third trimester and could be evaluated for subsequent preterm birth rates. Within that sample, spontaneous preterm births occurred in only 7 of the pregnancies, 3 in the intact D&X group and 4 in the D&E group.

The Chasen follow-up study identified two factors that appeared to contribute to the risk of subsequent preterm birth in the study population, neither of which related to the abortion procedure the woman had previously undergone: (1) whether the patient had received an abortion due to preterm premature rupture of membranes (PPROM) or spontaneous cervical dilation, medical conditions already known to carry a heightened risk for spontaneous preterm birth in subsequent pregnancies, and (2) whether the patient's subsequent pregnancy involved multiple fetuses, such as twins or triplets, which typically deliver prematurely. *Id.* at 1162-63.

Dr. Chasen found that 27.3% (3 out of 11) of the women in the study population who underwent abortion due to PPROM or spontaneous cervical dilation experienced subsequent preterm births, compared to 4.4% (4 out of 91) of

---

of the follow-up study is suspect. *See* Br. of *Amici Curiae* AAPLOG *et al.* 25 n.91 (the Chasen follow-up study used the dataset from the original study, "but somehow [found] twice as many prenatal patients"). This contention has no basis. The original Chasen study contained data on 62 subsequent pregnancies in 62 women; in other words, one subsequent pregnancy per woman. The follow-up study, on the other hand, contained data from 120 pregnancies in 89 women; in other words, it included women who had multiple subsequent pregnancies. Hence, while the number of *subsequent pregnancies* in the sample nearly doubled, the number of *prenatal patients* did not.

the women who underwent abortion for other reasons. That difference was statistically significant with a p-value of 0.03. Dr. Chasen also found that women who received an abortion due to PPROM or spontaneous cervical dilation made up a significantly greater percentage of the patients who had received intact D&X compared to those who received D&E (25% versus 4.3%, p-value = .001), and that the rates of subsequent preterm birth among patients who had undergone the earlier abortion for other medical reasons "were nearly identical" among those who had undergone intact D&X and those who had undergone D&E (4.2% versus 4.5%). *Id.* at 1163.

From these data, Dr. Chasen concluded that "[t]he association between surgical abortion and subsequent spontaneous preterm birth in these women is related to the indication for abortion, rather than the abortion procedure itself" and that "[s]urgical abortion with either variation of D&E late in the second trimester should not be considered a risk factor for subsequent spontaneous preterm birth." *Id.* Although Dr. Chasen conceded that the sample size of his follow-up study was too small to rule out the possibility that either intact D&X or D&E may increase the risk of subsequent preterm birth, the study provides concrete data supporting an alternative explanation for the numerical difference in the incidence of preterm birth between the intact D&X and D&E groups.

*Amici* believe the results of the follow-up study support Dr. Chasen's explanation for the difference in preterm birth rates observed between the intact D&X and D&E groups. The follow-up study's statistical evidence is certainly more scientifically compelling than the speculative assertion offered by Dr. Clark in his testimony that the numerical

difference may have resulted from the greater cervical dilation that occurs with intact D&X. N.Y. Tr. 2386:12. The district court in *Carhart*, which considered both explanations in connection with the original Chasen study, concurred. *See Carhart*, 331 F. Supp. 2d at 963-64, 1022-23; *see also* Cal. Tr. 971:10-971:13 (Bowes) (agreeing that the Chasen study does not suggest that intact D&X is "any less safe" than D&E).

Accordingly, this Court should not credit Dr. Clark's speculative causation hypothesis or the arguments based on that speculation put forth by the Solicitor General and supporting *amici*. *See* Petr.'s Br. (*Carhart*) 37-38 n.12; Br. of *Amici Curiae* AAPLOG *et al.* 19-20.

## II. Testimony That The Chasen Study Data Showed "Trends" Toward Increased Risk In The Intact D&X Group Is Unfounded And Contradicted By Additional Data From Dr. Chasen's Follow-Up Study.

Unlike Dr. Clark, the government's other expert witnesses did not assert that the numerical increase in the incidence of subsequent preterm birth in the intact D&X group suggested an actual difference between intact D&X and D&E. Instead, Dr. M. LeRoy Sprang and other government experts made a similarly invalid claim by testifying that the numerical increase, while not statistically significant, represented a "trend" that was cause for concern. Neb. Tr. 1158:8-1159:10 (Sprang); N.Y. Tr. 2122:8-2122:21 (Sprang); Cal. Tr. 1106:9-1107:15 (Sprang); N.Y. Tr. 2547:4-2547:8 (Cook). Dr. Sprang also mistakenly identified another purported "trend" in the data on cervical lacerations and all genital tract lacerations, which showed a roughly threefold and twofold numerical increase, respectively, in the

intact D&X group. Neb. Tr. 1159:15-1160:2; N.Y. Tr. 2122:22-2123:16, 2124:22-2125:8; Cal. Tr. 1103:21-1104:6. According to Dr. Sprang, both of these "trends" "would support that D&X has more risk than D&E." Neb. Tr. 1160:8-1160:11.

Dr. Sprang's assertions, like those of Dr. Clark, are not grounded in accepted principles of statistical analysis and should be disregarded, as should the portions of Judge Straub's dissent in *NAF* and the government's supporting *amicus* briefs that credit Dr. Sprang's testimony. *See Nat'l Abortion Fed'n v. Gonzales*, 437 F.3d 278, 309 (2d Cir. 2006) (Straub, J., dissenting); Br. of *Amici Curiae* Congressman Ron Paul *et al.* 9-10; Br. of *Amici Curiae* AAPLOG *et al.* 20 n.73.

### A. The P-values In The Chasen Study Do Not Indicate Even Borderline Statistical Significance.

As explained above, statisticians require a p-value of 0.05 or less before concluding that an observed difference is statistically significant. As p-value gets closer to 0.05, it becomes more difficult to attribute the observed difference to chance. Where the p-value falls just shy of that mark, statisticians may assign a borderline or marginal significance to the observed difference. Generally, only a p-value falling between 0.05 and 0.10 qualifies as marginally significant. Thus, if the p-value approaches 0.05, but does not meet or fall below that threshold, statisticians will not reject the null hypothesis by concluding that the observed difference is statistically significant, but they might determine that the null hypothesis should be tested further. Outside this p-value range, there is still enough likelihood that the difference was caused by chance that it would be invalid to conclude that the

observed difference is even marginally significant, or, to use Dr. Sprang's terminology, that there is a "trend."[5]

In the Chasen study, the observed difference in the incidence of spontaneous preterm birth between the intact D&X and D&E groups was 11.8% versus 4.4%, with a p-value of 0.30. That measure falls well outside the range of marginal significance. Dr. Chasen did not calculate p-values for the operative and post-operative complications he measured, which included cervical lacerations and all genital tract lacerations (cervical, perineal, and labial), but he did report the number of those complications that occurred in each treatment group. Dr. Chasen found that 3 patients in the intact D&X group and 2 patients in the D&E group, or 2.5% versus 0.8%, experienced cervical lacerations during the procedure. Similarly, 4 patients in both the intact D&X and D&E groups, or 3.3% versus 1.5%, experienced genital tract lacerations. Chasen, *Dilation and evacuation, supra,* at 1182. The p-values for these differences are 0.18 and 0.26, respectively. Both fall well outside the 0.05 to 0.10 range and therefore do not support a finding of even marginal statistical significance.

Dr. Sprang's own testimony identified the very reason why it is inappropriate to draw any conclusions from these

---

[5] As a technical matter, the use of the term "trend" in this context is inappropriate. The statistical term "trend" generally refers to changes that occur over time; for example, an increase in the unemployment rate each year for 10 years accurately can be termed a trend. In contrast, the Chasen study contains no indication that the rate of complications from D&E or intact D&X increases or decreases over time. A true test of trend in results requires a different statistical test than was performed in this study.

numbers: "[W]hen you look at [the data] because the numbers were so small it's hard to draw meaningful conclusions." N.Y. Tr. 2121:16-2121:17. It is precisely because so few of these events occurred in the study, a factor that the p-value calculations of 0.30, 0.18 and 0.26 take into account, that it is invalid to conclude that these differences are statistically significant or even marginally significant. That is, there are not enough data in the study on spontaneous preterm births, cervical lacerations, or genital tract lacerations to conclude for any of these events that a "trend" exists in the intact D&X group. Indeed, the government's own experts could not agree on which of these differences indicated a "trend" and which did not. *Compare* Neb. Tr. 1158:8-1160:11; N.Y. Tr. 2122:8-2125:8 (Sprang) (finding "trends" in all three events), *with* N.Y. Tr. 2419:15-2419:24, 2416:4-2416:13 (Clark) (stating that intact D&X posed "significant hazards to future pregnancies," but that intact D&X and D&E showed "no difference in any short-term outcome"), *and* Neb. Tr. 1719:16-1722:8 (Lockwood) (testifying that he "wouldn't draw any conclusions" from the difference in spontaneous preterm births and that "it would be extraordinarily unlikely that these two procedures have markedly different occurrences in the rate of . . . short-term complications.").[6]

---

[6] Furthermore, all three of the serious complications in the Chasen study that required admission of the patient to the surgical intensive care unit—one amniotic fluid embolus, one pulmonary embolus, and one uterine perforation—occurred in the D&E group. *See* Chasen, *Dilation and evacuation, supra*, at 1182. By Dr. Sprang's logic, these data would indicate a "trend" against D&E for all serious complications, despite the obvious paucity of data.

*Amici* do not mean to imply that there would be no benefit in further examining the complications or subsequent obstetric outcomes of intact D&X and D&E. The data from the Chasen study taken alone, however, simply do not suggest—even marginally—that there is an increased risk of operative or post-operative complications, or subsequent preterm birth, among women who receive intact D&X, compared to those who receive D&E. *Amici* therefore urge this Court to disregard the unfounded testimony of Dr. Sprang and others on the statistical meaning of the Chasen study data.

### B. The Numerical Difference In The Incidence Of Subsequent Preterm Birth Between The Intact D&X And D&E Groups Decreased In Dr. Chasen's Follow-up Study.

Dr. Chasen's follow-up study offers additional evidence that the numerical differences observed in the original Chasen study are not indicative of true differences in the subsequent obstetric outcomes of intact D&X and D&E, but are simply the consequence of small numbers. As noted above, the follow-up study reported data from an additional 58 subsequent pregnancies from within the original study population of 383 women. Chasen, *Obstetric outcomes, supra*, at 1162. These additional data nearly doubled the number of subsequent pregnancies that could be used to measure the incidence of spontaneous preterm birth, up from 62 in the original Chasen study to 120 in the follow-up study. Dr. Chasen found that, with the increased sample size, the difference in the rate of subsequent preterm births between the intact D&X and D&E groups narrowed considerably from 11.8% (2 out of 17) versus 4.4% (2 out of 45) in the original study to 9.4% (3 out of 32) versus 5.7% (4 out of 70)

in the follow-up study, down from a roughly threefold numerical difference to a difference of just over one and one half. *Id.* at 1163. Similarly, the p-value for this difference more than doubled from 0.30 in the original study to 0.68 in the follow-up study, departing even further from the 0.05 threshold for statistical significance. *Id.*

The Chasen follow-up study illustrates perfectly why it is invalid to conclude that there is an actual difference between two outcomes when the p-value does not equal or fall below 0.05. The data from the follow-up study suggests that the difference in spontaneous preterm births observed in the original Chasen study was the result of chance fluctuations due to small numbers. Adding more subsequent pregnancies to the sample, similar to adding more coin flips to our example, caused the observed difference to decrease, the p-value to increase, and the so-called "trend" to evaporate. As a result of the Chasen follow-up study, there is now even less reason to credit the claims of Dr. Sprang.[7]

---

[7] *Amici* disagree with the brief of the AAPLOG *amici*, which argues that "because the D&X data were excluded when D&X was performed in the presence of [PPROM] and/or advanced cervical dilation," no valid conclusions about preterm birth can be drawn from the Chasen follow-up study. Br. of *Amici Curiae* AAPLOG *et al.* 25 n.91. The 9.4% and 5.7% figures reported in the study represent the rates of *all* preterm births in intact D&X and D&E groups, respectively, *including* patients in those groups who received an abortion due to PPROM and/or advanced cervical dilation.

## III. Courts Require A P-value Of 0.05 Or Less, Or Its Statistical Equivalent, Before Concluding That An Observed Difference Is Statistically Significant.

Courts analyzing statistical evidence have widely adopted the principle that a p-value of 0.05 or less, or a roughly equivalent measure of at least 2–3 standard deviations between an expected outcome and an observed outcome, is the appropriate gauge for determining whether observed differences among groups are statistically significant, as opposed to the result of mere chance. Accordingly, courts have required a p-value calculation of 0.05 or less, or a minimum of 2–3 standard deviations, before crediting a claim of actual difference between two groups.[8]

For example, in analyzing claims of racial discrimination in the grand jury selection process in *Castaneda v. Partida*, 430 U.S. 482 (1977), this Court relied upon the principle that statistical significance is achieved where a difference of greater than two or three standard deviations is observed between the expected outcome and the

---

[8] Standard deviations can be used to measure the degree to which an observed result differs from the expected result, assuming the null hypothesis were true. Generally speaking, a p-value of 0.05 is equivalent to 1.96 standard deviations. For simplicity's sake, it is commonly said that a p-value of 0.05 is analogous to 2 standard deviations and a p-value of 0.01 is analogous to 3 standard deviations. As such, an observed outcome of greater than two or three standard deviations from the expected outcome is generally required for a finding of statistical significance.

observed outcome. *See id.* at 497 n.17.[9] In *Castaneda*, a difference of twelve standard deviations between the racial composition of a randomly selected grand jury pool and the pool at issue was held sufficient to establish the defendant's *prima facie* case of discrimination. *See id.* at 496; *see also Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 309 n.14, 311 n.17 (1977) (analyzing racial composition of workforce and repeating *Castaneda* principle that a difference of greater than 2–3 standard deviations would cast doubt on hypothesis that workforce was hired randomly without regard to race).

The courts of appeal take a similar approach in analyzing the meaning of statistical evidence in a variety of contexts. *See, e.g.*, *Smith v. Xerox Corp.*, 196 F.3d 358, 366-67 (2d Cir. 1999) (requiring p-value of 0.05 or less, equivalent to 2 standard deviations, to establish statistical significance of data supporting disparate impact claims); *Benson v. Tocco, Inc.*, 113 F.3d 1203, 1209 (11th Cir. 1997) (observing that standard deviation of 3.04 between rate of termination for employees under age 40 and those over age 40, and standard deviation of 2.66 between rate of

---

[9] *Amici* note that whereas the *Castaneda* court discussed the significance of the 2–3 standard deviation threshold in the context of a binomial distribution of data, the Chasen data are based on hypergeometric distributions. Although equating standard deviations to p-values for data based on hypergeometric distributions is more complicated than for a binomial or a normal distribution, particularly when the sample size is relatively small, the principle that an outcome of 2–3 standard deviations is roughly equivalent to a p-value of 0.05 or less remains the same, subject to minor fluctuations depending on the distribution and symmetry of the data.

termination for employees under age 50 and those over age 50, would support an inference of discriminatory termination); *Ford v. Seabold*, 841 F.2d 677, 688-89 & n.12 (6th Cir. 1988) (rejecting statistical evidence relating to composition of grand jury pool because "probability [that number of nonwhites in jury pools would deviate from the population] did not reach the 0.05 level, the level which most statisticians . . . deem statistically significant" and observing that "[t]he two or three standard deviation benchmark applied in *Castaneda* is essentially equivalent to probability values of 0.05 and 0.01 respectively"); *Eldredge v. Carpenters 46 N. Cal. Counties Joint Apprenticeship & Training Comm.*, 833 F.2d 1334, 1340 n.8 (9th Cir. 1987) (holding that level of significance of 0.045 in percentage difference of men and women admitted to apprenticeship program gave rise to inference of discrimination); *Segar v. Smith*, 738 F.2d 1249, 1282-83 (D.C. Cir. 1984) (observing that "social scientists usually accept a study that achieves statistical significance at the .05 level").

Accordingly, this Court should not hesitate to conclude that the p-value calculations from both Chasen studies of 0.68, 0.30, 0.26, and 0.18—all well above the 0.05 threshold—fail to demonstrate a statistically significant, or even a marginally significant, difference in the rate of spontaneous preterm births or cervical/genital tract lacerations between the intact D&X and D&E groups. Any suggestion that the numerical differences between the two groups signifies something more meaningful than mere chance finds no support in fundamental principles of statistical analysis, and is at odds with courts' established method of interpreting statistical evidence.

## CONCLUSION

The Chasen study and its follow-up are critical pieces of scientific evidence concerning the safety of intact D&X. If the Court chooses to consider the Chasen data in its constitutional analysis of the Partial-Birth Abortion Ban Act of 2003, *amici* believe those data should be evaluated objectively and in accordance with widely accepted principles of statistical analysis. Straightforward p-value calculations indicate that there is no evidence in the Chasen data upon which to base a conclusion that intact D&X is more dangerous than D&E. The Court should not be misguided by the unfounded and speculative assertions to the contrary of Dr. Clark, Dr. Sprang, and other government experts. For the foregoing reasons, *amici* urge this Court to disregard those misstatements, along with the arguments of Petitioner and supporting *amici* that perpetuate them, and instead to evaluate the results of the Chasen data using principles of interpretation that form the standard relied upon by the scientific community.

Respectfully submitted.

MOLLY S. BOAST,
*Counsel of Record for the Amici*
CHRISTIAN R. EVERDELL
AARON BLOOM
JOHN S. CRAIG
DEBEVOISE & PLIMPTON LLP
919 Third Avenue
New York, NY 10022
(212) 909-6000

*Counsel for Amici Curiae*

August 10, 2006

BLANK PAGE

## APPENDIX A

American Journal of Obstetrics and Gynecology
(2004) 190, 1180–3

**American Journal of Obstetrics and Gynecology**

www.elsevier.com/locate/ajog

**EDITORS' CHOICE**

# Dilation and evacuation at ≥20 weeks: Comparison of operative techniques

**Stephen T. Chasen, MD,* Robin B. Kalish, MD, Meruka Gupta, Jane E. Kaufman, MD, William K. Rashbaum, MD, Frank A. Chervenak, MD**

*Department of Obstetrics and Gynecology, Weill Medical College of Cornell University, New York, NY*

Received for publication October 16, 2003; revised December 19, 2003; accepted December 22, 2003

**KEY WORDS**

Abortion
Dilation and evacuation

**Objective:** The objective of this study is to compare the relative safety of 2 techniques for surgical abortion late in the second trimester.

* Reprint requests: Stephen T. Chasen, MD, 525 E 68th St, Room J130, New York, NY 10021.

*E-mail:* stchasen@med.cornell.edu

 doi:10.1016/j.ajog.2003.12.034

**Study design:** Retrospective review of patients who underwent surgical abortion at ≥20 weeks' gestation at our hospital from June 1996 through June 2003. Records were reviewed to determine whether the technique used was dilation and evacuation or intact dilation and extraction. Subsequent pregnancies at our hospital were identified, and obstetric outcomes were recorded. Categorical data were compared with Fisher exact test and $\chi^2$ analysis. Continuous data were compared with Mann-Whitney *U* test.

**Results:** Three hundred eighty-three patients met inclusion criteria. Intact dilation and extraction was performed in 120 cases, and dilation and evacuation was used in 263. Intact dilation and extraction was associated with higher parity, later gestational age, and more preoperative cervical dilation. There was no difference in procedure time or estimated blood loss in the 2 groups. Complications occurred in 19 cases (5.0%), and occurred with similar frequency in the 2 groups. We identified 62 subsequent pregnancies. There were no second-trimester miscarriages. Spontaneous preterm birth occurred in 2 of 17 (11.8%) pregnancies in the intact dilation and extraction group, compared with 2 of 45 (4.4%) in the dilation and evacuation group ($P = .30$).

**Conclusion:** Outcomes appear similar between patients undergoing dilation and evacuation and intact dilation and extraction after 20 weeks' gestation. Subsequent obstetric outcomes are similar between the 2 groups. The technique for surgical abortion should be determined by the physician on the basis of intraoperative factors.



---

In the United States, induced abortion late in the second trimester is uncommon. In 1999, 9643 abortions

were performed at ≥21 weeks' gestation, representing only 1.5% of total abortions reported.[1] Dilation and evacuation is the most common method used for second trimester abortion[1] and is considered the safest abortion technique in the second trimester.[2-5]

Dilation and evacuation involves preoperative dilation of the cervix with osmotic dilators, such as laminaria, which are placed in the cervix for 1 or more days before operative evacuation. All published reports of dilation and evacuation have described the use of grasping forceps to remove the fetus and placenta. With the use of forceps, fetal parts are grasped, and the fetus and placenta are disarticulated as they are removed. Low complication rates for this procedure have been reported.[2-5]

A variant of dilation and evacuation can be performed when sufficient cervical dilation is present. In these cases, the fetus is delivered via breech extraction. If the fetus is not in the breech presentation, internal podalic version may be performed. In most cases, after delivery of the body, the fetal head will become lodged in the cervix, and cranial decompression with suction must be performed to complete delivery. This procedure has been referred to as intact dilation and extraction, or D&X. There are no published data regarding the frequency or complication rate of this procedure. Despite this, some have stated that this procedure poses serious maternal risks and is less safe than dilation and evacuation.[6,7]

The objective of this study is to evaluate the relative safety of dilation and evacuation and intact D&X in patients undergoing surgical abortion late in the second trimester.

## Materials and methods

This study was a review of patients who underwent surgical abortion at ≥20 weeks' gestation at the New York

Weill Cornell Medical Center from June 1996 to June 2003. All procedures included in this study were performed by 1 of 2 physicians (S.T.C. and W.K.R) who are skilled in both techniques, dilation and evacuation and intact D&X. Institutional Review Board approval was obtained for this study.

Gestational age was confirmed by ultrasound in all cases. All cases ≥24 weeks' gestation were performed because of fetal demise. Dilation of the cervix was achieved with insertion of laminaria. Laminaria were placed preoperatively on 2 consecutive days, unless the cervix was sufficiently dilated at the initial examination. Women who went into labor after laminaria placement and delivered without dilation and evacuation or intact D&X were not included in this study.

The surgical technique was determined when the patient was examined under anesthesia, and was based on cervical dilation and fetal position. Both physicians performing these procedures used similar criteria in determining the optimal surgical technique. All cases were considered complete when, after evacuation of the fetus and placenta, hemostasis was apparent after sharp and suction curettage. Blood loss was estimated by the surgeon. Procedures were performed in an ambulatory setting, unless patients had been hospitalized before the procedure for medical or obstetric conditions. The type of anesthesia used for the procedure, general, regional, or intravenous sedation, was at the discretion of the anesthesiologist. Intraoperative ultrasound guidance was not routinely used.

Operative reports were reviewed to determine the technique used in each case. If the fetus was delivered intact in the breech presentation to the level of the umbilicus or higher, the procedure was considered an intact D&X, whether the entire fetus was removed intact

or decompression of the head was required. In some cases, the presenting fetal head was well applied to the cervix and was initially decompressed with suction, followed by intact delivery of the fetus. These cases were also considered intact D&Xs, as disarticulation with forceps was not required. All other cases, which were performed with multiple insertions of forceps, were categorized as dilation and evacuation.

Complications included any situation requiring unplanned intervention. These included unplanned hospital admission, repair of any genital tract lacerations, return to the operating room for additional procedures, and blood transfusion. Subsequent pregnancies at our hospital were identified from a review of the medical records. We did not obtain follow-up on patients routinely, so subsequent pregnancies with prenatal care obtained elsewhere were not included in this analysis. Thirty-eight patients in this series, 26 of whom underwent dilation and evacuation and 12 of whom underwent intact D&X, were included in a prior publication describing obstetric outcome after second trimester abortion.[8]

Demographic characteristics and outcomes were compared on basis of operative technique. Categorical variables were analyzed with the use of 2-tailed Fisher exact test and $\chi^2$ analysis where appropriate. Mann-Whitney *U* test was used to analyze continuous variables, which were not normally distributed. Analysis was performed with SPSS 11.0 (Chicago, Ill). A *P* value <.05 was considered significant.

## Results

Three hundred eighty-three patients met inclusion criteria. Intact D&X was performed in 120 cases (31.3%), and dilation and evacuation was used in 263 (68.7%).

Demographic characteristics of the 2 groups are listed in Table I. Intact D&X was associated with later gestational age, higher parity, and younger maternal age.

Indications for surgical abortion are listed in Table II. Preterm cervical dilation and/or preterm premature rupture of membranes (PPROM) were more common in those who underwent intact D&X. Abnormal fetal karyotype was more common in those who underwent dilation and evacuation.

Intraoperative variables are listed in Table III. Overall, laminaria were used in 96.1% of cases. Laminaria were less likely to be used in patients undergoing intact D&X. All 15 patients in whom laminaria were not used had preterm cervical dilation. Preoperative cervical dilation was greater in those who underwent intact D&X.

**Table I** Demographic characteristics of patients undergoing dilation and evacuation

| | D & E (n = 263) | D & X (n = 120) | *P* value |
|---|---|---|---|
| Median maternal age (range) | 34 y (16-45) | 32 y (12-43) | .01* |
| Median parity (range) | 0 (0-7) | 1 (0-5) | .04* |
| Median gestational age at D & E (range) | 21 wks (20-27) | 23 wks (20-25) | <.001* |
| Prior vaginal delivery | 81 (30.8%) | 50 (41.7%) | .05† |
| Prior cesarean delivery | 49 (18.6%) | 26 (21.7%) | .49† |
| Multifetal pregnancy | 13 (4.9%) | 8 (6.7%) | .48† |

D & E, Dilation and evacuation.

* Mann Whitney *U* test;

† Fisher exact test.

**Table II** Indication for dilation and evacuation*

| Indication for D & E | D & E (n = 263) | D & X (n = 120) | *P* value† |
|---|---|---|---|
| Abnormal fetal karyotype | 112 (42.6%) | 33 (27.5%) | .005 |
| Structural fetal abnormality | 96 (36.5%) | 47 (39.2%) | .65 |
| Intrauterine fetal demise | 27 (10.3%) | 14 (11.7%) | .72 |
| Premature cervical dilation/PPROM | 14 (5.3%) | 20 (16.7%) | .001 |
| Other | 22 (8.4%) | 13 (10.8%) | .45 |

* Some patients had more than 1 indication;
† Fisher exact test.

No differences were noted in procedure time or estimated blood loss between the 2 groups.

In the entire group of 383 patients, complications occurred in 19 cases (5.0%). Complications occurred with similar frequency in the dilation and evacuation and intact D&X groups (4.9% vs 5.0%; $P > .999$). The 6 complications occurring in the intact D&X group included 4 genital tract lacerations (3 cervical mucosal lacerations and 1 perineal laceration) noted and successfully repaired at the time of surgery; 1 case of excessive bleeding requiring the return to the operating room on the day of surgery for curettage; and 1 patient readmitted because of heavy bleeding on postoperative day 13 who underwent curettage for retained placenta.

Complications in patients undergoing dilation and evacuation included 4 genital tract lacerations (2 cervical lacerations and 2 labial lacerations) noted and successfully repaired at the time of surgery; 2 cases of

excessive bleeding requiring the return to the operating room on the day of surgery (for curettage in 1 case, and repair of cervical lacerations in 1 case); 2 postoperative admissions for intravenous antibiotics because of endometritis; 1 admission for observation after excessive bleeding; and 1 admission because of severe nausea from general anesthesia.

**Table III** Intraoperative variables

| | D & E (n = 263) | D & X (n = 120) | *P* value |
|---|---|---|---|
| Laminaria | 259 (98.5%) | 109 (90.8%) | .001* |
| Median preoperative cervical dilation (range) | 3 cm (0-6) | 5 cm (2-10) | .001† |
| Anesthesia | | | |
| General | 92 (35.05%) | 53 (44.2%) | .22‡ |
| Regional | 9 (3.4%) | 3 (2.5%) | |
| Local/conscious sedation | 162 (61.6%) | 64 (53.3%) | |
| Median procedure time (range) | 22 min (6-60) | 22 min (6-45) | .54† |
| Median estimated blood loss (range) | 100 mL (40-1500) | 100 mL (20-1200) | .38† |

* Fisher exact test;
† Mann-Whitney *U* test;
‡ Chi-square analysis.

Three patients, all of whom underwent dilation and evacuation, had complications requiring admission to the surgical intensive care unit. One patient, with a fetal

demise, had an amniotic fluid embolus with disseminated intravascular coagulation requiring transfusion of blood and clotting factors. One patient was diagnosed with sepsis and pulmonary embolus. One patient had a uterine perforation at the site of a cesarean delivery scar and required exploratory laparotomy and blood transfusion.

Forty-five women (17.1%) in the dilation and evacuation group and 17 women (15.0%) in the intact D&X group had a subsequent pregnancy and received care at our medical center. There were no second-trimester spontaneous abortions in either group. Spontaneous preterm birth occurred in 2 of 17 (11.8%) pregnancies in the intact D&X group compared with 2 of 45 (4.4%) in the dilation and evacuation group ($P = .30$). Both spontaneous preterm births in the intact D&X group occurred in women at high risk for prematurity: one woman, who underwent intact D&X caused by PPROM at 23 weeks' gestation, subsequently delivered at 32 weeks, and the other underwent intact D&X at 23 weeks' gestation because of cervical incompetence with advanced cervical dilation, and subsequently delivered at 35 weeks.

**Comment**

On the basis of this series of 383 patients, surgical abortion late in the second trimester is a safe procedure. There were low rates of complications in procedures performed with either technique. As described in a prior publication,[8] we did not see a high rate of spontaneous preterm birth in subsequent pregnancies in those who received obstetric care at our medical center.

Because this study was retrospective, we cannot state that one technique is superior to the other. Procedures performed with intact D&X occurred at later gestational ages and with more preoperative cervical dilation.

Although more advanced gestation might be expected to increase complication rates, more preoperative cervical dilation might be expected to facilitate these procedures.

Differences between maternal age, gestational age, and indication for surgical abortion were noted between the 2 groups. The difference in gestational age may be due to changes in the cervix that occur as pregnancy progresses. With advancing gestation, the cervix may be more likely to respond to laminaria placement with greater dilation. Intact D&X requires more dilation, as the intact fetus cannot pass through a minimally dilated cervix. In addition, more women in the intact D&X group underwent abortion caused by premature cervical dilation or PPROM. The incidence of these conditions increases with advancing gestation, and the spontaneous cervical dilation occurring in these cases probably facilitated intact D&X.

The difference in maternal age is likely because of the higher rate of fetal aneuploidy in the dilation and evacuation group. Amniocentesis is typically performed at 16 to 18 weeks, and advanced maternal age is the most common indication. Most cases of fetal aneuploidy are identified by 20 weeks. This could explain why those in the dilation and evacuation group were older, and underwent abortion at earlier gestational ages.

Our approach of performing intact D&X when possible is intended to minimize the use of forceps in extracting the fetus. We believe that the use of forceps to grasp the fetus can cause inadvertent trauma to the uterine wall. At these gestational ages, evacuation of a fetus can require multiple insertions of forceps, and intact D&X avoids this. Though we believe our low complication rate validates our approach, we acknowledge that the retrospective nature of this study precludes us from

concluding with certainty that intact D&X prevented adverse outcomes.

Another important limitation is the relatively small number of patients receiving prenatal care at our hospital in subsequent pregnancies. Most patients returned to their referring obstetrician for future obstetric care, and we were unable to assess subsequent pregnancy outcomes of patients delivered at other institutions. We believe that significant bias as a result would be unlikely, though we cannot be certain that outcomes in pregnancies followed at our hospital are representative of outcomes in all subsequent pregnancies. Though we are reassured by the low number of complications in subsequent pregnancies in both groups, we acknowledge our lack of power to conclude that subsequent pregnancy outcomes are not different.

Some have stated that intact dilation and extraction poses serious risks to the health of a woman beyond the risks associated with dilation and evacuation. Such putative risks include higher rates of cervical incompetence, uterine rupture, abruption, amniotic fluid embolism, and trauma to the uterus.[6,7] We are not aware of any published data supporting these statements. In our patients, the overall rate of complications was comparable between those undergoing dilation and evacuation and intact D&X. No patient undergoing intact D&X experienced uterine rupture, amniotic fluid embolism, or required blood transfusion. Because our approach is to perform intact D&X when possible on the basis of cervical dilation and fetal position, it is unlikely that intact D&X could have been performed in these patients undergoing dilation and evacuation who experienced severe complications.

In conclusion, our data affirm that abortion after 20 weeks' gestation with intact D&X appears to have sim-

ilar complication rates as dilation and evacuation when performed by experienced physicians. The observed complication rates and subsequent obstetric outcomes appear comparable between the 2 techniques. In accordance with the American College of Obstetricians and Gynecologists statement of policy,[9] our data supports that the most appropriate technique for surgical evacuation of pregnancy after 20 weeks' gestation should be based on intraoperative factors. Attempts to regulate intact D&X on the basis of concern for maternal well-being cannot be supported by available evidence.

## References


1. Elam-Evans LD, Strauss LT, Herndon J, Parker WY, Whitehead S, Berg CJ. Abortion surveillance—United States, 1999. MMWR Morb Mortal Wkly Rep 2002;51:1-28.
2. Grimes DA, Schutz KF, Cates W, Tyler CW. Mid-trimester abortion by dilation and evacuation: a safe and practical alternative. N Engl J Med 1977;296:1141-5.
3. Grimes DA, Hulka JF, McCutchen ME. Midtrimester abortion by dilation and evacuation versus intra-amniotic instillation of prostaglandin F2: a randomized clinical trial. Am J Obstet Gynecol 1980;137:785-90.
4. Schneider D, Halperin R, Langer R, Caspi E, Bukovsky I. Abortion at 18-22 weeks by laminaria dilation and evacuation. Obstet Gynecol 1996;88:412-4.
5. Autry AM, Hayes EC, Jacobson GF, Kirby RS. A comparison of medical induction and dilation and evacuation for second-trimester abortion. Am J Obstet Gynecol 2002;187:393-7.
6. Partial Birth Abortion Act of 2002, HR 760, US Congress (2003).
7. Sprang ML, Neerhof MG. Rationale for banning abortions late in pregnancy. JAMA 1998;280:744-7.
8. Kalish RB, Chasen ST, Rosenzweig LB, Rashbaum WK, Chervenak FA. Impact of midtrimester dilation and evacuation on

subsequent pregnancy outcome. Am J Obstet Gynecol 2002;187: 882-5.

9. American College of Obstetricians and Gynecologists. Statement on so-called "partial birth abortion" laws. Washington, DC: The College; 2002.

## APPENDIX B

American Journal of Obstetrics and Gynecology (2005) 193, 1161-4

**American Journal of Obstetrics & Gynecology**

www.ajog.org

# Obstetric outcomes after surgical abortion at ≥20 weeks' gestation

**Stephen T. Chasen,[a,*] Robin B. Kalish,[a] Meruka Gupta,[b] Jane Kaufman,[b] Frank A. Chervenak[a]**

*Division of Maternal-Fetal Medicine,[a] Department of Obstetrics and Gynecology,[b] Weill Medical College of Cornell University, New York, NY*

Received for publication March 1, 2005; revised May 9, 2005; accepted May 25, 2005

**KEY WORDS**

Dilation and evacuation
Spontaneous preterm birth

---

Presented at the Twenty-Fifth Annual meeting of the Society for Maternal Fetal Medicine, February 7-12, 2005, Reno, Nev.

* Reprint requests: Stephen T. Chasen, MD, Weill Medical College of Cornell University, 525 East 68th Street, J-130, New York, NY 10021.

*E-mail*: stchasen@med.cornell.edu

 doi:10.1016/j.ajog.2005.05.078

**Objective:** The purpose of this study was to describe obstetric outcomes after surgical abortion at ≥20 weeks, and to identify risk factors for subsequent spontaneous preterm birth.

**Study design:** Patients who had surgical abortion at ≥20 weeks' gestation from 1996 to 2003 and received subsequent prenatal care at The New York Weill Cornell Medical Center were identified. Indication for abortion, operative technique, and subsequent pregnancy outcomes were reviewed. Student *t* test, Fisher exact test, and Mann-Whitney *U* were used where appropriate.

**Results:** One hundred and twenty pregnancies in 89 women were identified. Thirteen (10.8%) ended with early miscarriage, and 5 were electively terminated. Of the remaining 102 pregnancies, 7 ended with spontaneous preterm birth. Those who experienced preterm birth were more likely to have undergone abortion due to cervical dilation and/or preterm premature rupture of membranes (PPROM) (27.3% vs 4.4%; $P = .03$). Those with a multifetal pregnancy in the subsequent pregnancy were more likely to have preterm birth (75.0% vs 4.3%; $P < .001$). In patients who underwent dilation and evacuation (D&E) for reasons other than cervical dilation and/or PPROM, rates of spontaneous preterm birth were identical between those who had intact dilation and extraction (D&X) and D&E using forceps (4.2% vs 4.5%; $P = 1.0$).

**Conclusion:** In those who have undergone D&E at ≥20 weeks, only a history of midtrimester cervical dilation and/or PPROM or a current multifetal pregnancy were associated with spontaneous preterm birth.

Dilation and evacuation (D&E) is considered the safest method of midtrimester abortion.[1,2] Although D&E has been categorized as a risk factor for future spontaneous preterm birth, published studies have not documented such an association.[3,4] Relatively few second trimester abortions are performed beyond 20 weeks' gestation; in the United States, only 1.4% of abortions were performed beyond 20 weeks in 2001.[5] Because these procedures typically require more cervical dilatation, it is plausible that D&E late in the second trimester could be a risk factor for future preterm birth.

Most D&Es involve disarticulation of the fetus with forceps. A variation of D&E, "intact dilation and extraction," or "intact D&X," has been described.[6] Intact D&X involves extraction of the largely intact fetus through the dilated cervix, and generally requires more cervical dilation. Complication rates for intact D&X are similar to those for D&E using forceps, although intact D&X is usually done later in pregnancy.[7] It has been suggested, in the absence of supportive data, that intact D&X in particular may increase the risk of subsequent preterm birth.[8,9]

Our objective was to describe the rate of and identify risk factors for spontaneous preterm birth in women who have undergone D&E at ≥20 weeks' gestation. A secondary objective was to determine whether operative technique is associated with higher rates of spontaneous preterm birth in future pregnancies.

## Material and methods

Patients who underwent dilation and evacuation at our hospital at ≥20 weeks' gestation from 1996 to 2003, and who received subsequent prenatal care at The New York Weill Cornell Medical Center were identified by searching computerized hospital records using the appropriate

procedure codes. Gestational age had been confirmed by ultrasound in all cases. Medical records were reviewed to collect data about the terminated pregnancies, including the technique of surgical abortion, and the outcomes of subsequent pregnancies. This study was approved by the Institutional Review Board of our institution.

All patients underwent serial laminaria insertion before surgical abortion, unless spontaneous cervical dilation had already occurred. Laminaria were placed preoperatively on 2 consecutive days, and operative evacuation was performed on the third day. At each insertion, as many laminaria as could be accommodated by the cervix were placed. Patients received doxycycline the days of laminaria insertion and the day of operative evacuation for antibiotic prophylaxis. The operative technique was categorized as "D&E" if the procedure required disarticulation of the fetus with forceps. The operative technique was categorized as "intact D&X" if the fetus was delivered largely intact without disarticulation with forceps.[7] Spontaneous preterm birth was defined as delivery at <37 weeks' gestation resulting from preterm labor or preterm premature rupture of membranes (PPROM).

Fisher exact test was used to compare categorical outcomes. Student *t* test and Mann-Whitney *U* were used to compare continuous data. Statistical analysis was performed using Statistical Package for the Social Sciences (SPSS Release 11.0, SPSS, Inc, Chicago, Ill) software.

## Results

There were 120 pregnancies in 89 women included in the study. These pregnancies were identified from a group of 383 patients who underwent surgical abortion at ≥20 weeks' gestation from 1996 to 2003. The outcomes of 62 of these pregnancies were briefly summarized in a pre-

vious publication that focused on operative techniques of surgical abortion late in the second trimester.[7]

The median maternal age at the time of surgical abortion was 33 years (interquartile range 29-37 years), and the median interval between abortion and the estimated date of confinement in a subsequent pregnancy was 18 months (interquartile range 14-33 months). The median gestational age at abortion was 21 weeks (interquartile range 20-23 weeks).

First trimester spontaneous abortion occurred in 13 (10.8%) pregnancies. Five pregnancies (4.2%) were electively terminated. There were no cases of second trimester pregnancy loss. The remaining 102 pregnancies occurred in 80 women. There were 93 deliveries at term. One woman with lupus nephritis underwent surgical abortion at 24 weeks' gestation because of fetal demise. Her 2 subsequent pregnancies were delivered at 31 weeks and 28 weeks because of HELLP syndrome. Spontaneous preterm birth occurred in the remaining 7 pregnancies (6.9%), which are described in Table I.

There were no significant differences in maternal age or gestational age at time of abortion with regard to the likelihood of subsequent spontaneous preterm birth. The association between indication for surgical abortion and subsequent spontaneous preterm birth can be seen in Table II. Of those undergoing abortion due to PPROM or spontaneous cervical dilation, 27.3% experienced spontaneous preterm birth in a subsequent pregnancy, compared with 4.4% of those undergoing abortion for other indications ($P = .03$; OR 8.2 [1.5-43.0]).

Those with a multifetal pregnancy after surgical abortion had a 75% incidence of spontaneous preterm birth, compared with 4.3% of those with a singleton pregnancy ($P < .001$; OR 70.5 [5.9-837.2]). In 87 singleton pregnancies in which surgical abortion was done for reasons

**Table I** Pregnancies with spontaneous preterm birth after surgical abortion at ≥20 weeks

| Case | Indication for surgical abortion | Gestational age at abortion | Operative technique | Gestational age at next delivery | Number of fetuses |
|---|---|---|---|---|---|
| 1 | Aneuploidy | 20 weeks | D&E | 30 weeks | Singleton |
| 2 | Aneuploidy | 21 weeks | D&E | 31 weeks | Twin |
| 3 | PPROM | 23 weeks | Intact D&X | 32 weeks | Singleton |
| 4 | PPROM/IUFD | 24 weeks | Intact D&X | 35 weeks | Singleton |
| 5 | PPROM | 21 weeks | D&E | 36 weeks | Singleton |
| 6 | Aneuploidy | 20 weeks | D&E | 36 weeks | Twin |
| 7 | Aneuploidy | 23 weeks | Intact D&X | 36 weeks | Triplet |

other than PPROM or cervical dilation, spontaneous preterm birth occurred in only 1 (1.1%).

The operative technique used was D&E in 70 cases and intact D&X in 32. There was no significant difference in the rate of spontaneous preterm birth between those who had undergone D&E versus intact D&X (5.7% vs. 9.4%; $P = .68$). PPROM and/or spontaneous cervical dilation were more common indications in those who had undergone intact D&X (25% vs 4.3%; $P = .001$). In the 91 patients who underwent abortion for other indications, the rates of subsequent spontaneous preterm birth were nearly identical between those who had undergone D&E and intact D&X (4.5% vs 4.2%; $P = 1.0$).

## Comment

D&E is the most common procedure used to terminate a pregnancy in the second trimester in the United States. Lower complication rates have been described for D&E compared with medical induction.[1,2] D&E may also be preferable to some patients because it is typically done on an outpatient basis, and women do not have to endure labor. When Grimes et al attempted to perform a randomized clinical trial to compare D&E with medical induction using misoprostol and mifepristone, most women did not consent to randomization and opted to undergo D&E instead.[10] While most women may regard D&E as preferable to induction, many might reconsider if D&E were found to be a risk factor for subsequent preterm birth.

Based on our data, only those who have undergone surgical abortion late in the second trimester because of PPROM and/or premature cervical dilation are at high risk for spontaneous preterm birth in subsequent pregnancies. The high rate of subsequent spontaneous preterm birth seen in our patients with PPROM or cer-

vical dilation before surgical abortion (27.3%) is comparable to published data describing the rate of spontaneous preterm birth after midtrimester loss.[11] The association between surgical abortion and subsequent spontaneous preterm birth in these women is related to the indication for abortion, rather than the abortion procedure itself. The other significant association between multifetal pregnancy and preterm birth is obviously not surprising.

**Table II** Indication for abortion and subsequent spontaneous preterm birth

| Indication for surgical abortion | No spontaneous preterm birth (n = 95) | Spontaneous preterm birth (n = 7) | *P* value |
|---|---|---|---|
| Fetal abnormality | 69 (72.6%) | 4 (57.1%) | 0.40 |
| Intrauterine fetal demise | 12 (12.6%) | 0 (0%) | 1.0 |
| PPROM or spontaneous cervical dilation | 8 (8.4%) | 3 (42.9%) | 0.03 |
| Other | 6 (6.3%) | 0 (0%) | 1.0 |

Fisher exact test used for statistical comparison.

The rate of spontaneous preterm birth in singleton pregnancies after surgical abortion for indications other than PPROM and/or cervical dilation was only 1.1%. No patient who underwent surgical abortion using intact D&X for indications other than PPROM or cervical dilation delivered prematurely, with the exception of a triplet pregnancy delivered at 36 weeks. Our data do not support the notion that achieving advanced cervical dilation with laminaria is a risk factor for preterm birth. In fact, we have found that in women undergoing D&E

from 14 to 24 weeks, the risk of subsequent preterm birth was inversely correlated with the degree of cervical dilation achieved before surgical abortion.[3]

Because of the relatively small number of patients studied, we cannot exclude the possibility that surgical abortion late in the second trimester may increase the risk of spontaneous preterm birth. In our hospital, the rate of spontaneous preterm birth in singleton pregnancies is approximately 5%. To have an 80% likelihood of detecting a 50% increase in the rate of spontaneous preterm birth to 7.5% at a significance level of 0.05, we would need 539 patients. In this study with 102 pregnancies, our power to detect this magnitude of an increase in spontaneous preterm birth was only 34%.

In conclusion, surgical abortion at ≥20 weeks' gestation is not associated with a high risk for spontaneous preterm birth in future pregnancies. Surgical abortion with either variation of D&E late in the second trimester should not be considered a risk factor for subsequent spontaneous preterm birth.

## References

1. Stubblefield PG, Carr-Ellis S, Borgatta L. Methods for induced abortion. Obstet Gynecol 2004;104:174-85.

2. Autry AM, Hayes EC, Jacobson GF, Kirby RS. A comparison of medical induction and dilation and evacuation for second-trimester abortion. Am J Obstet Gynecol 2002;187:393-7.

3. Kalish RB, Chasen ST, Rosenzweig LB, Rashbaum WK, Chervenak FA. Impact of midtrimester dilation and evacuation on subsequent pregnancy outcome. Am J Obstet Gynecol 2002; 187:882-5.

4. Schneider D, Halperin R, Langer R, Caspi E, Bukovsky I. Abortion at 18-22 weeks by laminaria dilation and evacuation. Obstet Gynecol 1996;88:412-4.

5. Strauss LT, Herndon J, Chang J, Parker WY, Levy DA, Bowens SB, et al. CDC. Abortion surveillance—United States, 2001. MMWR Surveill Summ 2004;53:1-32.

6. Haskell WM, Easterling TR, Lichtenberg ES. Surgical abortion after the first trimester. In: Paul M, Lichtenberg ES, Borgatta L, Grimes DA, Stubblefield PG, editors. A clinician's guide to medical and surgical abortion. New York (NY): Churchill Livingston; 1999. p. 123-54.

7. Chasen ST, Kalish RB, Gupta M, Kaufman JE, Rashbaum WK, Chervenak FA. Dilation and evacuation at > or =20 weeks: comparison of operative techniques. Am J Obstet Gynecol 2004;190:1180-3.

8. Partial Birth Abortion Act of 2002, HR 760, US Congress (2003).

9. Sprang ML, Neerhof MG. Rationale for banning abortions late in pregnancy. JAMA 1998;280:744-7.

10. Grimes DA, Smith MS, Witham AD. Mifepristone and misoprostol versus dilation and evacuation for midtrimester abortion: a pilot randomised controlled trial. BJOG 2004; 111:148-53.

11. Goldenberg RL, Mayberry SK, Copper RL, Dubard MB, Hauth JC. Pregnancy outcome following a second-trimester loss. Obstet Gynecol 1993;81:444-6.