Using Clinical Classification Trees to Identify Individuals At Risk of STDs During Pregnancy

Trace S. Kershaw [assistant professor] and **Jessica Lewis [research associate]**
Department of Epidemiology and Public Health, and the Center for Interdisciplinary Research on AIDS, Yale University, New Haven, CT.

Claire Westdahl [assistant professor and director of nurse-midwifery] and **Yun F. Wang [assistant professor]**
Emory University School of Medicine, Atlanta.

Sharon Schindler Rising [executive director]
Centering Pregnancy and Parenting Association, Chesire, CT.

Zohar Massey [research associate] and **Jeannette Ickovics [professor]**
Department of Epidemiology and Public Health, and the Center for Interdisciplinary Research on AIDS, Yale University, New Haven, CT.

Abstract

CONTEXT—Few studies have used classification tree analysis to produce empirically driven decision tools that identify subgroups of women at risk of STDs during pregnancy. Such tools can guide care, treatment and prevention efforts in clinical settings.

METHODS—A sample of 647 women aged 14−25 attending two urban obstetrics and gynecology clinics in 2001−2004 were surveyed in their second and third trimesters. Baseline predictors at the individual, dyad, and family and community levels were used to develop a classification tree that differentiated subgroups of women by STD incidence at 35 weeks' gestation. Logistic regression analyses were conducted to assess whether the classification tree groups or commonly used risk factors better predicted STD incidence.

RESULTS—Nineteen percent of women had an incident STD during pregnancy. Classification tree analysis identified three subgroups with a high STD incidence (33−61%), one with a moderate incidence (16%) and three with a low incidence (6−11%). Women in subgroups with high STD incidence included those not living with the partner with whom they conceived and those who had a moderate or a high level of depression, a history of STDs and a low level of social support. A logistic regression model using groups defined by the classification tree analysis had better predictive ability than one using common demographic and sexual risk predictors.

CONCLUSION—This classification tree identified risk factors not captured by traditional risk screenings, and could be used to guide STD treatment, care and prevention within the prenatal care setting.

Young pregnant women are a group at particularly high risk of STDs.[1,2] In a systematic literature review of sexual risk and STDs among pregnant teenagers and those who have given birth, STD incidence was 19−39% during pregnancy, and 14−39% between six and 10 months postpartum. In addition, adolescent mothers were twice as likely as their nulliparous peers to acquire an STD during a 12-month period. Furthermore, 78−88% of pregnant women engaged

Author contact: Trace.Kershaw@yale.edu.

in sex without condoms during pregnancy, and pregnant women were only about one-fifth as likely as nonpregnant women to use condoms.[3]

Current guidelines from the Centers for Disease Control and Prevention (CDC) recommend STD screening for all high-risk women-usually defined as those younger than 25 and those with multiple or new partners.[4] Frequently, screening for all high-risk pregnant women occurs only during initial prenatal care visits, so an infection that is acquired later in pregnancy may go undetected. Some clinics retest women who are deemed high-risk (e.g., those testing STD-positive at their initial visit) or exhibit symptoms; however, testing procedures differ from clinic to clinic, and screening behavior can vary by provider.[5] In addition, because an STD diagnosis and treatment are not enough to have long-term effects on risky behavior,[6] an emphasis on prevention is necessary to achieve long-term risk reduction.

Prevention programs integrated with prenatal care can capitalize on women's motivation to have a healthy pregnancy and child. HIV and other STD prevention programs have been successfully integrated in psychiatric care, drug treatment and palliative medicine settings.[7–9] This model for prevention could be extended to include prenatal care.

For interventions to be most effective in clinical settings, they should be easy to implement and tailored to meet the needs and risks of particular patients. Thus, it may be beneficial to identify subgroups of women at varying risk of STDs during pregnancy, and implement treatment, care and prevention programs that best meet the needs of those individuals. To successfully do this, providers need to go beyond the typical characteristics obtained during prenatal care assessments. A 2002 Institute of Medicine report and more recent research syntheses suggest that ecological approaches to health, and to STD prevention, are particularly useful and needed.[3,10,11] Ecological systems theory emphasizes that factors from many levels (e.g., individual, dyad, family and community) can influence health.[3,10–12] The few studies on young people's sexual activity that have included factors from many levels have shown that dyad, family and community factors independently contribute to risk behavior above and beyond individual factors.[13–15]

In addition to using primarily individual-level predictors, most studies that have tried to estimate STD risk have used standard analytic techniques (e.g., logistic regression). However, these methods are limited because they assume primarily linear relationships between the predictors and outcomes that may not be realistic in applied settings. Furthermore, results from these techniques (e.g., odds ratios in logistic regression) are difficult to translate into clinical practice and policy.[16]

Classification tree analysis, also called recursive partitioning, avoids the limitations associated with logistic regression and other standard methods. The technique uses a set of predictors to create subgroups of individuals that vary in risk for the main outcome of interest, ultimately producing a user-friendly classification tree that can be employed to guide care, treatment and prevention efforts in clinical settings. For example, in a study of young adults at an STD clinic, several subgroups were classified as having a high STD risk because of particular behavioral and emotional factors (e.g., one group felt bad about themselves after having sex and had sex to relieve tension; one group fell bad about themselves after having sex and had sex to get back at someone).[17] Use of classification tree analysis has been documented in the health care and managed care literature,[18] and has been used in many clinical and medical settings to identify high-risk patients and to help create clinical guidelines.[19–23]

Clinicians often use informal decision trees when making clinical decisions. On the basis of a complex interaction of medical, demographic and behavioral factors, clinicians determine the likelihood of disease outcomes and often use these patient characteristics to guide care and treatment decisions (e.g., if a woman is having risky sex and is younger than 25, the clinician

may screen for STDs because past experience suggests that the patient is at risk). The purpose of this study is to use individual, dyad, and family- and community-level characteristics to develop a clinically relevant classification tree that will help identify women who are at risk of acquiring an STD during pregnancy and will provide a formal decision-making structure for STD prevention, treatment and care.

METHODS

Sample

Data for this study come from baseline and follow-up interviews with pregnant women aged 14−25 enrolled in a randomized controlled trial of an HIV prevention intervention.[24,25] Between September 2001 and December 2004, young women attending their first or second prenatal care visit at obstetrics and gynecology clinics in two university-affiliated hospitals, in New Haven, Connecticut, and Atlanta, were referred by a health care provider or approached directly by research staff. The centers represent the largest obstetric and gynecologic care providers in their respective cities. Black and Latina women with limited financial resources are overreprescntcd in the sample, reflecting the population of prenatal patients using these clinics.

Inclusion and exclusion criteria pertain to the larger randomized controlled trial and include being pregnant at less than 24 weeks' gestation; being no older than 25 at last birthday; having no severe medical problem requiring indivdualized assessment and tracking as "a high-risk pregnancy" (e.g., diabetes, hypertension, HIV); being able to speak English or Spanish; and having a willingness to be randomized. Potential participants were screened, and research staff explained the study in detail to eligible women, answered any questions they had and obtained informed consent.

Of the 1,542 eligible young women, 1,047 enrolled in the randomized controlled trial (participation rate, 68%). Those who agreed to participate were more likely to be black, were older and were further along in their pregnancies at screening than those who refused to participate. Because we are not assessing the effect of the intervention, data are drawn from the 729 women from the trial's two control groups. These women did not differ significantly from the women in the intervention group in terms of demographic characteristics and sexual risk behavior at baseline (e.g., age, education, parity, condom use, number of partners and STD history) or any other of the main study variables.

Women completed structured interviews via audio computer assisted self-interviewing (audio-CASI) upon study entry. Audio-CASI allows respondents to listen through headphones to spoken questions that have been digitally recorded and stored on a computer, as well as to see the questions displayed on the computer's screen. This technology improves the validity and accuracy of reporting of high-risk and stigmatizing behaviors.[26–28]

Women completed their baseline interviews during their second trimester (average gestation, 18 weeks; standard deviation, 3.3). Follow-up interviews were completed in the third trimester (average gestation, 35 weeks; standard deviation, 3.1). Participants were paid $25 for each interview. Of the 729 women in the control groups, 647 (89%) completed the follow-up interview. There were no significant differences between those who completed the follow-up interview and those who did not. All procedures were approved by the Yale University Human Investigation Committee and by an institutional review board at each study clinic.

Measures

•**Incident STD**—At each interview, women gave a urine sample, which was tested at a central laboratory for chlamydia and gonorrhea using ligation-based strand displacement

amplification. In addition, at follow-up, women were asked if they had received a diagnosis of an STD (chlamydia, trichomonas, herpes, genital warts, gonorrhea, syphilis or HIV) since their last interview; for each reported STD, women were asked the date of their most recent diagnosis. Women were classified as having an incident STD if they tested positive for chlamydia or gonorrhea at follow-up or if they reported having received an STD diagnosis after baseline. Using a combination of biological testing and self-reports was done to capture STDs that may have been detected and treated between biological assessments.

•**Individual-level predictors**—Individual-level predictors included social and demographic characteristics (race and ethnicity, age, education, number of previous children and current relationship status) and sexual behavior and history characteristics (number of partners in the past six months, condom use in the past six months, lifetime number of partners, unprotected sex with a risky partner in the past six months and history of an STD). Condom use, a continuous measure (range, 0−100%), was the average estimated percentage of time women had used condoms with all partners. Unprotected sex with a risky partner was defined as less than consistent condom use with a partner who was an injection-drug user, was HIV-positive, had had an STD or had had sex with another person in the past six months. Participants were categorized as having an STD history if they reported ever having had chlamydia, trichomonas, genital herpes, genital warts or human papillomavirus, gonorrhea or syphilis. These measures have demonstrated predictive and construct validity in past studies with young women.[6,29–32]

In addition, we included several psychological characteristics as individual-level predictors. Depression was assessed by the 15 cognitive-affective items of the Center for Epidemiological Studies-Depression Scale.[33] Respondents reported the frequency with which they had experienced various symptoms (e.g., sadness, crying and hopelessness) over the last seven days. The scale had a range of 0−45, with higher scores indicating more depressive symptoms; results showed good internal consistency (Cronbach's alpha, 0.85). Stress was assessed by the 10-item Perceived Stress Scale.[34] For example, respondents were asked how often they were "on top of things"; a five-point Likert scale for each item ranged from "never" to "very often." The scale had a range of 0−40, with higher scores indicating more perceived stress; results showed good internal consistency (Cronbach's alpha, 0.81). Women's barriers to condom use were assessed through their level of agreement with two items: "Sex is not as good with a condom" and "Using condoms means you don't trust your partner." A four-point Likert scale for each item ranged from "strongly disagree" to "strongly agree."[35] Perceived risk of STD was assessed using a single item, in which women estimated their chances of acquiring an STD in the next year; four possible response options ranged from "no chance" to "good chance."[30,31] Knowledge of HIV and other STD risk was assessed by an 11-item measure (e.g., "most people who carry sexually transmitted diseases or the AIDS virus look and feel healthy");[35] each item was assessed on a five-point scale that ranged from "definitely false" to "definitely true." The measure had a range of 0−44, with higher scores indicating more knowledge; results showed adequate internal consistency (Cronbach's alpha, 0.67).

•**Dyad-level predictors**—Dyad-level characteristics included whether women were currently in a relationship or living with the partner with whom they conceived, relationship duration, partner's age and partnership commitment (measured on a four-point scale, ranging from "not at all committed" to "totally committed").

•**Family- and community-level predictors**—Women's level of social support was assessed using a seven-item subscale of the Social Relationship Scale. For example, respondents were asked "would people in your personal life be available to talk to you if you were upset, nervous or depressed?" The scale had a range of 7−35, with higher scores indicating more social support, and showed good internal consistency (Cronbach's alpha, 0.90).[36] Peer

norms of condom use were assessed using a five-item scale, which had a range of 5−20, with higher scores indicating more positive peer norms about condom use. The measure showed good internal consistency (Cronbach's alpha, 0.77).[35]

Data Analysis

Classification tree analysis is a nonparametric method based on repeated partitioning of a sample into subgroups. The analysis is conducted in a stepwise fashion in which the most significant predictor at each step is used to split the sample into subgroups. This process continues until differences are no longer statistically significant.[37] Results are presented as classification, or decision, trees that require no calculations for their use, but identify groups that are expected to experience a given outcome.

The starting group (e.g., the entire sample) is referred to as the root node, and each subsequent split creates 2−3 nodes. We used the exhaustive chi-square automatic interaction detector algorithm, which selects predictors on the basis of an adjusted chi-square statistic. The algorithm searches through a set of predictors to detect the variable, along with the cut point on that variable (if the predictor is continuous) or the combination of categories (if it is categorical), that maximizes homogeneity within nodes (i.e., the similarity of subgroup members with respect to the outcome variable).[16,38] The sample is split at this cut point, and nodes are created that maximize group differences on the outcome. This process is continued for all resulting nodes until splits no longer have a significant adjusted chi-square value or nodes are too small to be considered stable. (We used a minimum node size of 30, approximately 5% of the total sample, which is consistent with approaches in other studies using recursive partitioning.[16]) A node in which further splits fail to improve prediction forms a terminal node. The end result of the analysis is a set of terminal nodes, which define the subgroups.

Because this procedure is data-driven, cross-validating the results is critical. Therefore, we performed a 10-fold cross-validation procedure in which the sample was randomly divided into 10 subsamples. Each subsample was then left out and tested using the tree developed for the nine others. This process is conducted iteratively, and an average classification rate is obtained for the 10 subsamples.[37] This procedure is the most commonly used approach for validating tree structures for small samples and has demonstrated relatively little bias in model selection, compared with other cross-validation techniques.[37,39]

By comparing the classification rate of the entire sample to the cross-validated classification rate, we can assess the generalizability and stability of the classification tree. In addition, we hand-pruned the tree by comparing cross-validation coefficients of all possible subtrees. The largest subtree with the smallest difference in the cross-validated classification rate was used as the final tree. Descriptive analyses, analyses of variance and chi-square tests were used to characterize the groups on key demographic and sexual risk variables.

Finally, to evaluate the utility of the classification tree analysis, we conducted two sets of logistic regression analyses. First, we wanted to test the possibility that the classification tree groups are merely a proxy for demographic and sexual risk variables, instead of a unique set of variables with independent predictive power. Therefore, we conducted logistic regression to assess whether the groups significantly predicted STD at 35 weeks' gestation, controlling for demographic and sexual risk variables commonly used in clinical settings to identify high-risk individuals. Second, we used logistic regression to directly compare the predictability of the classification tree groups with the predictability of the six common demographic and sexual risk variables. We measured the sensitivity and specificity and the amount of unique variability predicted using the Nagelkerke R-square.[40] All data analyses were performed using Answer Tree and SPSS 11.5 software.

RESULTS

Of the sample of 647 women, 75% were black, 16% Latina and 9% white or of other racial or ethnic groups (Table 1, page 143); 49% were between the ages of 14 and 19, and 51% were 20−25 years old. Half had at least a high school education, and two-thirds had had no previous births. The vast majority of women were in a current relationship (80%), had had only one partner in the last six months (83%) and had not had unprotected sex with a risky partner during that period (77%). On average, women reported using condoms about one-third of the time they had sex (36%; standard deviation, 37-not shown). Half of women had a history of STDs.

Classification Tree Analysis

Figure 1 depicts the results from classification tree analysis, including the classification variable and the cut point for each split. Within each node is shown the proportion of women who had received an STD diagnosis between the baseline and follow-up assessments.

The first split was based on whether the participant lived with the partner with whom she had conceived. The node representing women living with their partner did not split again; the node representing those not living with their partner split into three groups on the basis of depression score: low (score of 11 or lower), moderate (12−19) and high (greater than 19). Three more splits took place. The node representing women with high levels of depression was further split by their history of an STD, and the node representing those with moderate levels of depression was further split by perceived susceptibility to STDs. Finally, the node comprising women who thought they had no chance of acquiring an STD was split by social support. Therefore, the analysis resulted in seven terminal nodes.

The STD infection rate for the sample overall (node 0) was 19%; the 95% confidence interval around that rate was 15−22. We categorized the rate in each terminal node as high if it exceeded the upper limit of that confidence interval, as moderate if it was within the confidence interval and as low if it was below the lower limit. Thus, nodes 7, 8 and 10 represented women with a high incidence rate (33−61%); node 3 represented women with a moderate rate (16%); and nodes 2, 9 and 11 represented women with a low rate (6−11%). Women are therefore classified as being at high risk if they did not live with their partner, had a moderate level of depression and perceived some chance of getting an STD (node 7); they did not live with their partner, had a high level of depression and had a history of an STD (node 8); or they did not live with their partner, had a moderate level of depression, perceived no chance of getting an STD and had low social support (node 10). These classifications form the basis of a classification tree tool, which clinicians could use to identify individuals at high risk of STD during pregnancy (Figure 2).

The tree achieved adequate cross-validation. There was no marked difference between the misclassification rates of the entire sample and of the cross-validated estimate (17% vs. 20%).

Subgroup Characteristics

The groups defined by the seven terminal nodes differed with respect to some of the key demographic and sexual risk variables: multiple partnerships, unprotected sex with a risky partner, STD history, condom use and relationship status (Table 2). However, the relationships were not linear. For example, women in node 7 and node 2 (high- and low-risk groups, respectively) had the lowest levels of condom use at baseline. In addition, women in the low-risk groups demonstrated high-risk behaviors. For example, 30% of women in node 11 reported having unprotected sex with a risky partner–a greater proportion than of women in the moderate-risk node 3 or the high-risk node 10.

Comparison Analyses

In the first set of logistic regression analyses (Table 3, page 146), demographic and sexual risk variables significantly predicted STD incidence at 35 weeks' gestation. Individual variables significantly associated with STDs included younger age (odds ratio, 0.9) having had multiple partners (1.9), greater mean proportion of condom use (1.02) and STD history (2.0). Adding the classification tree groups to the analyses accounted for 13% of variance of STD incidence above and beyond that of the demographic and sexual risk variables (chi-square, 57.18; p<.01; Nagelkerke R^2 change, 0.129–not shown). This suggests that the classification tree groups are important and unique predictors of STD incidence and demonstrates the utility of using this classification tree analysis to identify high-risk STD groups.

In the second set of logistic regression analyses, a model using only the classification tree groups accounted for more variance than the previous model (17% vs. 10% – Table 3). When we used the logistic regression models to calculate sensitivity and specificity, assuming we would target women in moderate- and high-risk groups for screening and prevention, the model using the classification tree groups had equivalent sensitivity (74% vs. 74%) but better specificity (54% vs. 47%) compared with the model with the individual demographic and sexual risk predictors.

DISCUSSION

In this study, variables at each level of the ecological model–individual, dyad, and family and community–played a role in predicting STD incidence during pregnancy. The dyad-level factor of whether the woman lived with the partner with whom she conceived had the largest initial association with STD incidence, showing the importance of understanding the relationship and family context of young pregnant women in addition to their medical and sexual risk history. These results support other work suggesting that it is necessary to look beyond the individual to understand health and health behavior.[3,10,11] In addition, of the variables found to be associated with STD incidence, only STD history is routinely assessed by clinicians. Thus, our results suggest that it is important to go beyond the demographic, medical and sexual risk behaviors typically assessed when trying to understand and predict negative reproductive health outcomes.

One potential barrier of expanding initial assessments in clinical settings, however, is that the additional questions that would need to be asked during visits may increase initial burden in terms of cost and time for patients and clinicians. Guidelines of the American College of Obstetricians and Gynecologists recommend that all women be screened for depression during pregnancy,[41] and several short and externally validated scales have been identified.[42] Therefore, only eight additional questions (using the scales from our study) would be needed, to assess a woman's social support, perceived risk and partner status. The potential usefulness of the information in implementing prevention and treatment protocols may well outweigh the additional cost and time.

The classification tree analysis created subgroups that had better predictive ability, equivalent sensitivity and better specificity, compared with traditional logistic regression models with common demographic and sexual risk predictors. This demonstrates the utility of using not only predictors across levels of the ecological model but also nonparametric statistical analytic techniques that can identify complex patterns of relationships among predictors (i.e., higher order interactions). Although one can model interactions within logistic regression, such calculations are not commonly done and can be statistically cumbersome when there are many predictors (and therefore many possible interaction terms). It is doubtful that in real-world settings, variables have straightforward linear relationships with important health outcomes; nonparametric techniques like classification tree analysis are better able to identify these

complex relationships. The classification tree tool created in this study may help clinicians identify subgroups of women at high risk of STDs, allowing those women to receive additional clinical observation and referrals to appropriate prevention programs.

Study Limitations and Strengths

This study had several limitations. First, all the subgroups created by the classification tree analysis, even those labeled as low-risk, had relatively high rates of STDs (6−11%). Therefore, all groups should be screened and provided with prevention and care information. However, the nature and magnitude of focus may differ depending on the overall level of risk of that subgroup.

Second, we studied a high-risk, urban, clinic-based sample of pregnant women, who had a high STD rate during the third trimester of pregnancy (19%). This sample may be behaviorally different than a random sample of pregnant women; therefore, our results should be applied and generalized primarily to high-risk urban samples. Such women may benefit most from a tool to aid in targeted care, treatment and prevention, but similar studies are needed to determine if the classification tree functions differently in low-risk populations.

Third, because we relied on self-reported data from women participating in a large randomized control trial, our results may have been subject to recall or selection bias. Finally, other possible predictors (e.g., domestic violence) not included in the analysis may have changed the structure of the classification tree.

Despite the limitations, this study had several strengths. Research to develop classification trees for predicting STDs in pregnancy has been limited, and few studies have included predictors from a variety of levels of the ecological model. In addition, this study included a large, ethnically diverse sample of pregnant adolescents and young adults. This population is a group in need of intervention, and results from this study can be applied directly to program designs for accessible and at-risk populations.

Conclusions

Clinical settings are ideal for integrating prevention because they provide frequent access to individuals at risk of a variety of reproductive health outcomes and because they provide a structured, credible environment that helps to facilitate learning and change.[7–9] Therefore, clinicians could team up with public health professionals to implement STD interventions that best fit the characteristics of high- and moderate-risk groups. For example, the CDC has identified 18 interventions that have demonstrated effectiveness in heterosexual and youth populations.[43] Clinicians could work with public health practitioners to choose and implement the intervention from that list that best fits with clinic practice and best suits the characteristics and risk factors of the subgroups. By working with public health professionals, clinicians can implement prevention programs that complement and extend clinical care to include prevention of future negative reproductive health outcomes.

REFERENCES

1. Centers for Disease Control and Prevention (CDC). STD Surveillance Report. CDC; Rockville, MD: 2003. 2003.
2. Niccolai LM, et al. Pregnant adolescents at risk: sexual behaviors and sexually transmitted disease prevalence. American Journal of Obstetrics and Cynecology 2003;188(1):63–70.

3. Meade CS, Ickovics JR. Systematic review of sexual risk among pregnant and mothering teens in the USA: pregnancy as an opportunity for integrated prevention of STD and repeat pregnancy. Social Science & Medicine 2005;60(4):661–678. [PubMed: 15571886]

4. CDC. Guidelines for Treatment of Sexually Transmitted Disease. CDC; Atlanta: 1998.

5. Barnes RS, et al. Georgia prenatal care providers' perceptions of barriers to sexually transmitted disease screening. Southern Medical Journal 2003;96(9):845–849. [PubMed: 14513977]

6. Kershaw TS, et al. Sexual risk following a sexually transmitted disease diagnosis: the more things change the more they stay the same. Behavioral Medicine 2004;27(5):445–46l.

7. Bull SS, et al. Acceptability and feasibility of urine screening for chlamydia and gonorrhea in community organizations: perspectives from Denver and St. Louis. American Journal of Public Health 2000;90(2):285–286. [PubMed: 10667194]

8. Fisher JD, et al. Clinician-initiated HIV risk reduction intervention for HIV-positive persons: formative research, acceptability, and fidelity of the Options Project. Journal of Acquired Immune Deficiency Syndrome 2004;37(Suppl 2):S78–S87.

9. Fisher JD, et al. Clinician-delivered intervention during routine clinical care reduces unprotected sexual behavior among HIV-infected patients. Journal of Acquired Immune Deficiency Syndrome 2006;41 (l):44–52.

10. DiClemente R, et al. Prevention and control of sexually transmitted infections among adolescents: the importance of a socioecological perspeciivc–a commentary. Public Health 2005;119(9):1–12. [PubMed: 15560895]

11. Institute of Medicine. The Future of the Public's Health in the 21st Century. National Academy Press; Washington, DC: 2002.

12. Bronfenbrenner, U. Ecological systems theory. In: Vasta, R., editor. Annals of Child Development. JAI Press; Greenwich, CT: 1989. p. 187-249.

13. Mandara J, Murray CB, Bangi AK. Predictors of African American adolescent sexual activity: an ecological framework. Journal of Black Psychology 2003;29(3):337–356.

14. Resnick MD, et al. Protecting adolescents from harm: findings from the National Longitudinal Study on Adolescent Health. Journal of the American Medical Association 1997;278(10):823–832. [PubMed: 9293990]

15. Small SA, Luster T. Adolescent sexual activity: an ecological, risk-factor approach. Journal of Marriage and the Family 1994;56(1):181–192.

16. Lemon SC, et al. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. Annals of Behavioral Medicine 2003;26(3):172–181. [PubMed: 14644693]

17. Whitten K, et al. The emotional experience of intercourse and sexually transmitted discases: a decision tree analysis. Sexually Tmnsmitted Diseases 2003;30(4):348–356.

18. Anderson R, Balkrishnan R, Camacho F. Risk classification of Medicare HMO cnrollee cost levels using a decision-tree approach. American Journal of Managed Care 2004;10(2):89–98. [PubMed: 15011809]

19. Barnes GM, Welle JW, Dintcheff B. Drinking among subgroups in the adult population of New York State: a classification analysis using CART. Journal of Studies on Alcohol 1989;52(4):338–344. [PubMed: 1875707]

20. Choi SC, et al. Prediction tree for severely head-injured patients. Journal of Neurosurgery 1991;75 (2):251–255. [PubMed: 2072162]

21. Gregory KD, Korst LM, Platt LD. Variation in elective primary cesarean delivery by patient and hospital factors. American Journal of Obstetrics and Gynecology 2001;184(7):1521–1534. [PubMed: 11408876]

22. Zhang H, Bracken MB. Tree based, two-stage risk factor analysis for spontaneous abortion. American Journal of Epidemiology 1996;144(10):989–995. [PubMed: 8916510]

23. Levey DE, et al. Predicting outcome from hypoxic-ischemic coma. Journal of the American Medical Association 1985;253(10):1420–1426. [PubMed: 3968772]

24. Kershaw T, et al. Avoidance, anxiety, and sex: the influence of romantic attachment on. HIV-risk among pregnant women. AIDS and Behavior 2006;10(3):309–318. [PubMed: 16482404]

25. Ickovics J, et al. Group prenatal care reduces preterm birth: results from a multi-site randomized controlled trial. Obstetrics & Gynecology. 2007(forthcoming)

26. C'de Baca J, et al. Use of computer interview data to test associations between risk factors and pregnancy outcomes. Computers and Biomedicai Research 1997;30(3):232–243.

27. Des Jaríais DC, Turner CV. Audio-CASI versus interviewer obtained survey from participants in syringe exchange programs, paper presented at the North American Syringe Exchange Convention. Baltimore, Apr 1998;25

28. Turner CF, et al. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. Science 1998;280(8):867–873. [PubMed: 9572724]

29. Kershaw TS, et al. Mispcrceived risk among female adolescents: social and psychological factors associated with sexual risk accuracy. Health Psychology 2003;22(5):523–532. [PubMed: 14570536]

30. Kershaw TS, et al. The influence of pregnancy, STD, and HIV susceptibility patterns on sexual risk reduction for adolescent females. Journal of Community Psychology 2005;33(3):313–331.

31. Kershaw TS, et al. Perceived susceptibility to pregnancy and sexually transmitted disease among pregnant and nonpregnant adolescents. Journal of Community Psychology 2003;31(4):419–434.

32. Kershaw TS, et al. Short and long-term impact of adolescent pregnancy on postpartum contraceptive use: implications for prevention of repeat pregnancy. Journal of Adolescent Health 2003;33(5):359–368. [PubMed: 14596957]

33. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Applied Psychological Measurement 1977;1(3):385–401.

34. Cohen, S.; Williamson, GM. Perceived stress in a probability sample of the United States. In: Spacapan, S.; Oskamp, S., editors. The Social Psychology of Health. Sage; Newbury Park, CA: 1988.

35. Sikkema KJ, et al. HIV risk behaviors among women living in low-income, inner-city housing developments. American Journal of Public Health 1996;86(8):1123–1128. [PubMed: 8712272]

36. O'Brien K, et al. Social relationships of men at risk for AIDS. Social Science & Medicine 1993;36(9):1161–1167. [PubMed: 8511645]

37. Zhang, H.; Singer, B. Recursive Partitioning in the Health Sciences. Springer; New York: 1999.

38. Dudley WN, DiIorio C, Soet J. Detecting and explicating interactions in categorical data. Nursing Research 2000;49(1):53–56. [PubMed: 10667629]

39. Weiss S, Indurkhya N. Selecting the right-size model for prediction. Applied Intelligence 1996;6(4):261–273.

40. Nagelkerke N. A note on the general definition of the coefficient of determination. Biomctrika 1991;78(l):691–692.

41. American College of Obstetricians and Gynecologists (ACOG). Guidelines for Women's Health Care. ACOG; Washington, DC: 2002.

42. Williams JW, et al. Is this patient clinically depressed? Journal of the American Medical Association 2002;287(9):1160–1170. [PubMed: 11879114]

43. CDC. Compendium of HIV Prevention Interventions with Evidence of Effectiveness. CDC; Atlanta: 2001.
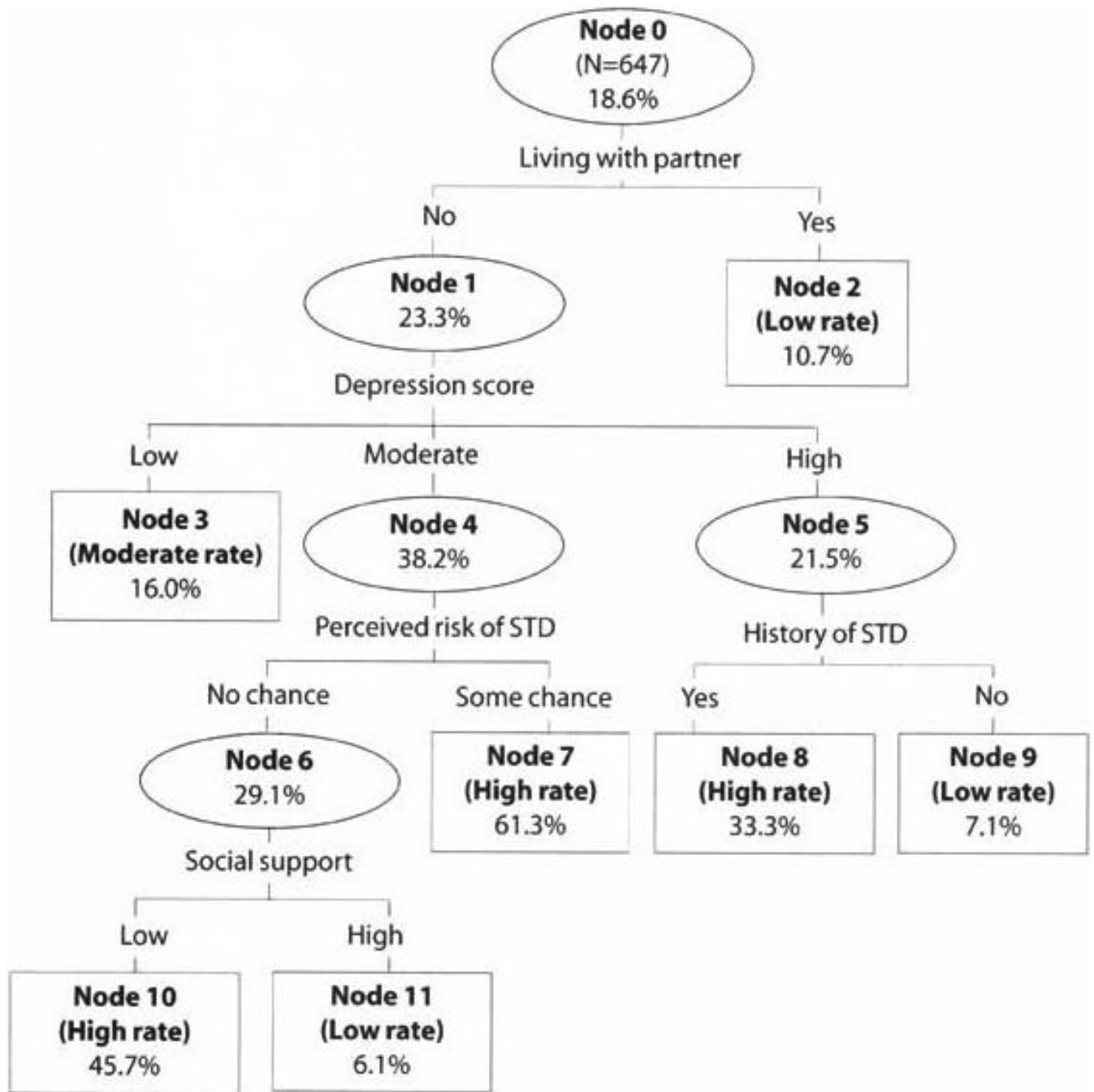
FIGURE 1.
Subgroups of women identified through classification tree analysis and their Incident STD rates by 35 weeks' gestation
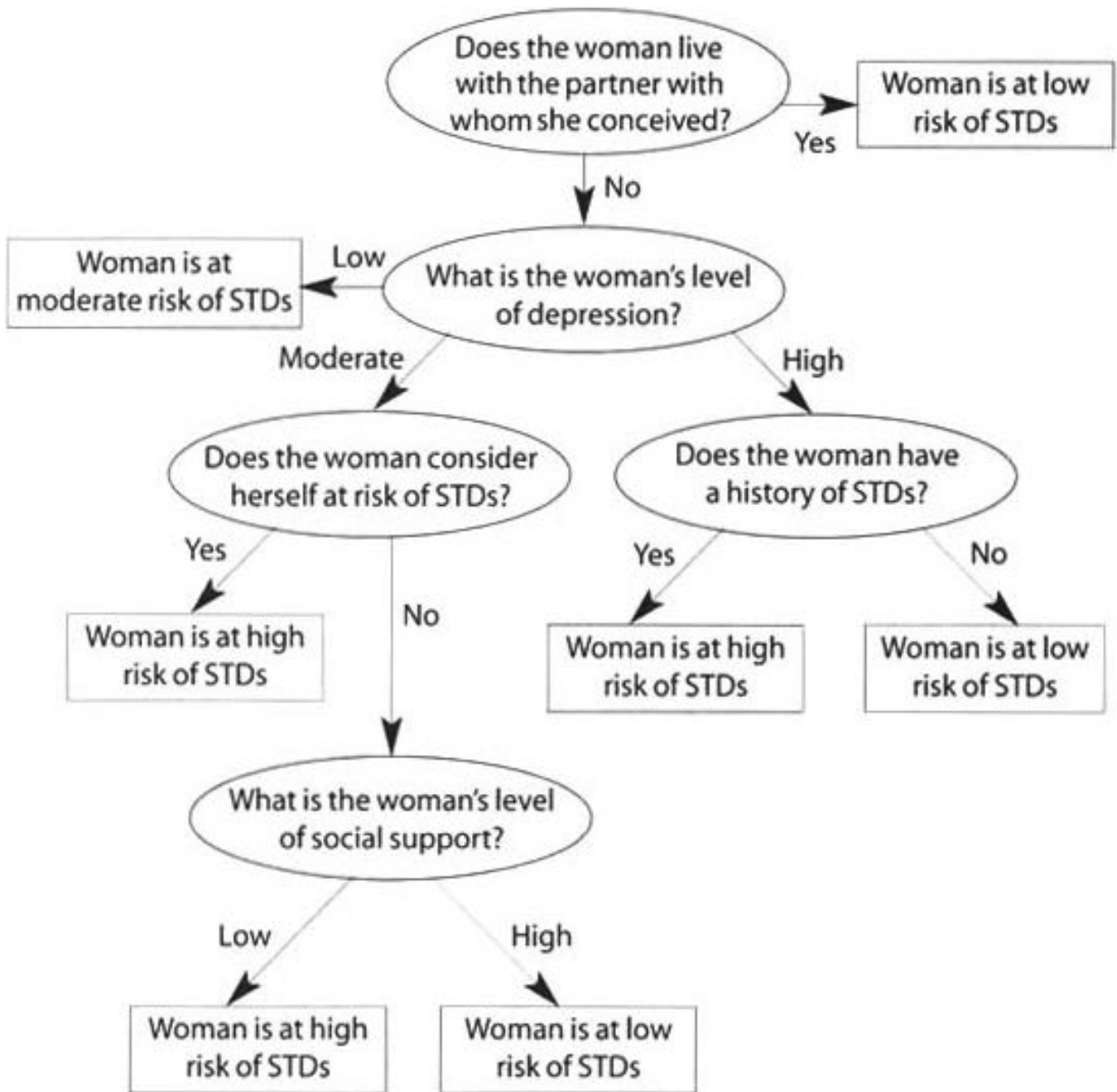
FIGURE 2.
Classification tree for assessing a woman's likelihood of acquiring an STD during pregnancy

TABLE 1

Percentage distribution of pregnant women in the control group of an HIV prevention intervention study conducted in two urban clinics, by selected characteristics

Characteristic	% (N=647)
Race/ethnicity	
Black	75
Latina	16
White/other	9
Age	
14−19	49
20−25	51
Education	
<high school	49
High school	34
>high school	17
No. of previous births	
0	64
≥1	36
In a current relationship	
No	20
Yes	80
No. of partners in past six months	
1	83
≥2	17
Unprotected sex with a risky partner in past six months[†]	
No	77
Yes	23
History of STD	
No	49
Yes	51
Total	100

[†] Unprotected sex with a risky partner was defined as not always using a condom with a partner who was an injection-drug user, was HIV-positive, had had an STD or had had sex with another person in the past six months.

TABLE 2

Percentage of pregnant young women reporting selected characteristics related to STD risk, by risk level and classification tree subgroup

Classification tree subgroup	Multiple partners	Unprotected sex with risky partner	STD history	Condom use (mean %)	In committed relationship
High risk					
Node 7	29	39	61	28 (31)	68
Node 8	39	43	100	31 (32)	45
Node 10	28	26	50	59 (40)	52
Moderate risk					
Node 3	18	19	49	42 (38)	69
Low risk					
Node 2	8	19	50	28 (35)	89
Node 9	21	24	0	41 (37)	45
Node 11	15	30	46	43 (44)	64
χ^2 or F	38.96**	21.33**	94.83**	6.08**	76.47**

** p<.01. *Notes*: Data are percentages unless otherwise noted. Figures in parentheses are standard deviations.

TABLE 3

Odds ratios (and 95% confidence intervals) from logistic regression analyses assessing women's likelihood of acquiring an STD during pregnancy, by variable type

Variable	Odds ratio
Demographic/sexual characteristics	
Parity	1.14 (0.78−1.65)
Age	0.86 (0.77−0.95)[*]
Highest grade completed	1.06 (0.90−1.24)
Multiple partners	1.89 (1.16−3.08)[*]
Condom use	1.02 (1.01−1.03)[*]
STD history	2.23 (1.44−3.45)[*]
χ^2	*40.34*
Nagelkerke R^2	*.099*
Classification tree subgroups	
Node 7	8.31 (3.60−18.79)[*]
Node 8	2.63 (1.31−5.26)[*]
Node 10	4.41 (2.21−8.81)[*]
Node 3 (ref)	1.00
Node 2	0.63 (0.36−1.09)
Node 9	0.40 (0.12−1.39)
Node 11	0.34 (0.08−1.49)
χ^2	*86.73*
Nagelkerke R^2	*.172*

[*] p<.05. *Note:* ref=reference group.