National Children's Study Workshop
Expanding Methodologies for Capturing Day-Specific Probabilities of Conception
May 17–18, 2004
Doubletree Hotel
Rockville, MD

This meeting was held in conjunction with the National Children's Study, which is led by a consortium of federal agency partners: the U.S. Department of Health and Human Services (including the National Institute of Child Health and Human Development [NICHD] and the National Institute of Environmental Health Sciences [NIEHS], two parts of the National Institutes of Health, and the Centers for Disease Control and Prevention [CDC]) and the U.S. Environmental Protection Agency (EPA).

Welcome, Overview, Workshop Goals and Objectives

Joseph Stanford, M.D., M.S.P.H., University of Utah

Dr. Stanford welcomed participants. He said this workshop resulted from a request of the Fertility and Early Pregnancy Working Group of the National Children's Study (Study) Advisory Committee to gather cross-disciplinary information on approaches for a preconception cohort for the Study. The findings of this workshop will be published and disseminated to the larger scientific community, as well as to those involved in planning for the Study. Discussions on day-specific probability (DSP) of conception would cover both environmental exposures and genetic factors.

Related terminology was discussed:

- *Conception*: Conception has been defined in a variety of ways in different contexts. For example, medical dictionaries define conception as fertilization or implantation, while embryology textbooks define conception as fertilization. Obstetrics and gynecology organizations use conception to describe implantation. Historically, the word has been used in terms of clinically identified pregnancy. In this workshop, "conception" is used to mean fertilization, except where noted otherwise.
- *Fecundity*: the biologic capacity of men and women for reproduction
- *Fertility*: observed reproduction; can also denote fecundity
- *Fecund window*: also called the fertile window; period of days during a single menstrual cycle when there is maximum biologic capacity for conception to occur if intercourse frequency and timing of intercourse are optimized; anchored to the central event of ovulation.
- *Fecundability*: probability of conception in a single menstrual cycle; determines time to pregnancy; varies with couple or population; dependent on genetic factors (female and male) and environmental factors; also dependent on frequency and timing of intercourse in the fecund window
- *Maximum fecundability*: maximum attainable probability of conception in a single menstrual cycle, if timing of intercourse is optimized during fecund window; removes factor of sexual behavior; female and male factors; latent, not directly observable; derivable from day-specific probabilities

- *Day-specific probabilities (DSP) of conception*: the probability of conception in a given cycle if intercourse occurs only on the specific day during the fecund window; measured as a probability of conception for a given day, relative to ovulation; hard to measure directly: based on modeling; influenced by viability of sperm and ova, and associated environmental exposures, taking into account time between release and meeting of gametes.

Dr. Stanford said that workshop presentations would cover biologic, epidemiologic, and statistical concepts and methods of research in terms of early environmental exposures, reproductive physiology, and human developmental outcomes. The workshop was intended to serve as impetus for further development of methods described. The research reviewed at the workshop will inform Study decisions on various issues of preconception enrollment and conception, as well as assessments of practical application, feasibility, and utility of issues related to DSP. The goal would be to identify innovative designs for possible use in the Study.

Rationale for Pinpointing Conception and Overview of DSP of Conception

Germaine Buck, Ph.D., M.S., NICHD, NIH, DHHS

Dr. Buck said that few resources contain compilations of biological methods and statistics on DSP of conception. One workshop goal is to create a reference that contains discussion on the various workshop topics. She presented an overview of DSP and its importance in the study of environment and health.

There is general concern that efforts to pinpoint conception will be too expensive and that methods are not well developed. However, Dr. Buck countered that belief, stating that methods could be done in a longitudinal study and that workshop discussion will attempt to further refine current methods for application to the Study.

Dr. Buck reviewed the following items:

- The importance of getting both maternal and paternal data on environmental toxicants and the effects on reproduction and child development; including paternal data will provide a more complete picture of possible effects of paternal exposures to outcomes in children, and data on collective, lifelong exposures of both mother and father may eventually help couples achieve and maintain healthy pregnancies.
- Very little research exists on the impact of the environment on gamete viability and human reproductive health prior to conception. The ability to pinpoint conception may lead to a better understanding of the etiology of various disorders and critical windows of exposure in development.
- New research should focus on the period prior to organogenesis to reveal child outcomes, given various parental exposures.
- Identification of sensitive markers for pregnancy loss—both pre- and post-implantation—is also needed, especially to assess effects of couple-based exposures.
- Biological markers and longitudinal diaries containing details of menstrual cycle and intercourse are considered essential to more accurate estimations of DSP of conception for the Study.

- Monitoring fertility and detecting ovulation will yield additional information relating exposure to critical windows in development. Home pregnancy tests continue to improve, but are still not sensitive enough for reliable detection in the first week following implantation. Ultrasound will be an essential research tool, and infecund couples could be studied for potential toxicant exposure and higher developmental risk.

Clinical Reproductive Biology: Cervical Mucus and Other *In Vivo* Factors Affecting the Probability of Ovulation and Conception

Michael Zinaman, M.D., Loyola University of Chicago

Dr. Zinaman said that mucus studies could answer questions about timing of conception in large study groups. He described various conception-related issues:

- Cervical mucus studies are based on three biological events: ovulation, hormonal response, and the tertiary response seen in cervical mucus. (Primary—ovulation; secondary—ovarian hormones; tertiary—cervical mucus and other biomarkers.)
- Supersensitive assays are needed to detect minute hormone levels of conception biomarkers *in vivo*, but one study showed that a 1-ml saline and human chorionic gonadotropin (hCG) solution could be detected in urine a short time after being placed into the uterus.
- When in vitro fertilization (IVF) is performed, the time of fertilization is known and hCG testing can pinpoint the time of implantation.
- Cervical mucus is the primary biomarker of fertility awareness methods.
  - Calendar-based methods attempt to determine the fecund window of the menstrual cycle based on counting days from ovulation. Women must have relatively regular cycles to successfully use this method.
  - The standard days method identifies days 8–19 of the cycle as fertile days in women who have menstrual cycles between 26 and 32 days long. A necklace (called CycleBeads™) can be used to help a woman keep track of the days. This method can be used to avoid or plan pregnancy, but it overestimates the number of fertile days.
- Methodological decisions for any study should reflect how closely researchers want to come to determining exact day of conception.

Research based on cervical mucus has several advantages:

- It yields robust data and methods are well developed. Conception rates peak at 30 percent per cycle, corresponding with conception rates in hormonal studies, which indicates highly credible data.
- Methods used to identify the fertile period are effective. Mucus hydration monitors can be used to provide easy, objective measures of hydration, which is responsible for the changes in cervical mucus with ovulation.
- Cervical mucus studies have strong correlation with studies on LH, estrogen, and sperm penetrability.
- The scoring system for mucus observations is moderately easy for subjects to learn.
- Cervical mucus studies are suitable for women with regular or irregular menstrual cycles.

Saliva, basal body temperature (BBT), and other analyses have been less successful:

- Saliva analyses provided inconsistent results.
- BBT is only useful to delineate the infertile phase and to predict timing of menses.
- BBT could be used in tandem with hCG data to more closely pinpoint time of conception.
- OvaCue® (Zetek, Inc., Aurora, Colorado) monitors and similar methods that use vaginal electrical impedance have not been consistently reliable.

Dr. Zinaman said mucus studies are useful to pinpoint days of high probability of conception, algorithms are easy to learn and are suitable for large studies, and robust statistical models are already developed for analyses, based on existing datasets with mucus data.

Discussion. A participant commented that subjects would not wear beaded necklaces for standard days research. There was concern about the cost of mucus hydration monitors. Steven Schrader, Ph.D., CDC, DHHS, said that he successfully taught subjects to report mucus hydration without monitors by the second or third cycle. If pregnancy occurred before subjects could learn reporting techniques, data on timing of conception might be unreliable. Group teaching may help standardize reporting. Lifestyle differences such as alcohol use might affect mucus hydration reports.

Clinical Reproductive Biology: Markers of Conception (Fertilization) and Hormonal Measures and Correlates of Fecundity, Conception, Implantation, and Early Pregnancy Loss

Kenneth Campbell, Ph.D., University of Massachusetts, Boston

Research on the interval between ovulation and implantation focuses on events that occur in a relatively avascular space, where access to the products of these stages is limited. Ultrasound cannot distinguish oocytes, but can reveal the follicle, and well after implantation, can identify the implantation site. Limitations of current approaches are that they are based on indirect measures of maternal hormones or tissue responses to hormones.

- Estrogen and progesterone are measured by immunoassay, chromatography, or mass spectrometry.
- Follicle stimulating hormone (FSH), luteinizing hormone (LH), and hCG are measured by immunoassay, high-performance liquid chromatography, or by specialized forms of mass spectrometry in serum and saliva.

After implantation, the placenta is the contact between the embryo and the maternal blood supply, and mature uteroplacental circulation is developed by day 21 following conception. Clinical pregnancy cannot be confirmed until uteroplacental circulation is established; thus, trophoblastic products cannot reach maternal blood and urine prior to implantation.

Many physiological markers used in research today were also used 20 years ago. Combining data on LH, estradiol, and progesterone can define the fertile period with high accuracy, but many recently developed scientific techniques reveal new markers that merit further investigation. Recent advancements that may impact decisions include:

- Animal and human genome projects

- Development of DNA and protein databases and computerized search techniques
- Cloning
- Molecular monitoring
- Massive parallel DNA screening
- Instrumental methodologies
  - Portable ultrasound to monitor ovulation and early pregnancy
  - Commercial kits for serum, urinary and salivary hormone analysis
  - Software programs to track cervical mucus, BBT, and hormones
  - Screens that identify gene expression during placentation and early embryogenesis.

New methods identify new markers. Mass spectrometry will now allow analysis of proteins and small molecules, and microchip protein screening techniques might also identify new markers. Refined algorithms of urinary estrogen and progesterone ratios provide information on early pregnancy loss and impact of lifestyle and environment on fecundity and fertility. Gene expression analyses can now capture data on gene expression in trophoblast and decidual tissues. Other research focused on estrogen-driven apoptotic processes uses DNA from urine sediments to acquire data on apoptotic ladders related to steroid presentation in the menstrual cycle. These changes occur in urethral epithelium in response to estrogen and progesterone as well as in vaginal epithelium.

Dr. Campbell's recommendations for the Study include:
- Begin with sampling for existing markers, including those that are being used to monitor embryonic abnormalities early in pregnancy, such as alpha-fetoprotein and inhibin.
- Reserve samples should be kept for retrospective evaluation of urine sediment DNA.
  - Potentially useful methodologies remain to be validated.
  - New soluble markers will be defined over the 20-year term of the Study.
- A significant program of research to uncover new biomarkers should be undertaken, and new approaches using old methods and new genomic data may reveal additional markers that could be developed for use in the Study.

Discussion. Participants asked about novel measures. Dr. Campbell said most protein markers can be identified using antibodies. Many markers do not appear in urine or are new discoveries. When asked for suggestions on sample collection, Dr. Campbell said serum is most important, then urine, with saliva being least important. Vaginal fluid may also be worthwhile. Daily collection is best, but collection should be at least bi-weekly; there is typically lower compliance with longer intervals between samples. Dr. Zinaman conveyed concern about compliance with daily sampling schedules. Couples in Dr. Campbell's studies who were trying to conceive have been most compliant to such schedules. Much data can be collected using home kits.

Reproductive Epidemiology Studying Exposures to Both Sexes and Time to Pregnancy

Michael Joffe, M.D., Ph.D., Imperial College London

Dr. Joffe presented methods and analyses of various relevant studies to demonstrate the advantages and disadvantages of prospective and retrospective data in analysis of time to pregnancy (TTP). Both types of studies have unique advantages and disadvantages and can complement each other. Although prospective studies are currently proposed, retrospective studies may also be useful to fill informational gaps. Methods for human studies can be fine-tuned to help to identify subgroups and pinpoint and eliminate biases.

Prospective Studies. Dr. Joffe described potential advantages and disadvantages of prospective studies.

- Potential advantages:
  - Strength of biological and environmental data
  - Detailed, timed data on key biological events such as ovulation, implantation, and environmental exposures.
- Potential disadvantages:
  - Difficulty in identifying prospective sampling frames
  - Bias due to a strong volunteer element
    - Pregnancy: problems may arise if participants think their employers will learn of their intentions too early. This may hamper data collection.
    - Sensitive issues: fertility questionnaires have tended to have high non-response rates, and this affects the quality of the findings.
  - Compliance and data collection problems
  - Length of time and effort required to acquire data which can, in turn, increase attrition.

Retrospective Studies. Dr. Joffe described potential advantages and disadvantages of retrospective studies.

- Potential advantages:
  - Short time to collect data
  - Generally good participation and compliance.
- Potential disadvantages:
  - Fallibility of recollection and incorrect reporting
  - Biases
    - Potential participants may refuse questionnaires on sensitive issues. They are more likely to respond to sensitive questions embedded in larger, general questionnaires.
  - Depending on design, inability to assess many exposure covariates
    - Dr. Joffe's retrospective TTP studies corrected for various exposures by dividing data into groups such as smokers and non-smokers and manual and non-manual labor classes.

Issues of TTP testing for prospective and retrospective methods include:

- Biases: They can be problematic in both prospective and retrospective studies, but sensitivity analyses can identify problems to yield more objective results.
- Generalizability: Data should attempt to reflect general populations. Inclusion of deliberately planned pregnancies only may skew outcomes and may cause problems of external validity.
- Data analysis: Contraceptive failures, unsuccessful conception attempts, and planned versus unplanned pregnancy may not be clearly identified in TTP data.
- Risk populations: Some groups have higher risk of unplanned pregnancies or lower fecundability. Problem populations can be extracted from analyses, but this also may create bias.
- Accuracy in reporting: Women may not want to declare contraceptive failures and abortion.
  - Dr. Joffe advised against questions on abortion. Compared with public data, research data typically do not reflect the same statistical outcomes and indicate unwillingness to report.
- Participation: Very few people refuse questionnaires on TTP. In his experience, both men and woman typically respond enthusiastically.
- Recall in retrospective studies:
  - Dr. Joffe addressed this issue in retrospective studies on TTP in comparison with previous prospectively collected data. Although certain individuals made significant errors in recollection, general results showed high correspondence when compared to historical data.
  - Dr. Joffe's data were analyzed for two recall time frames. Recall accuracy was assessed for groups divided as 0 to 14-year recall, and from 14 to 20-year recall. Recall for both groups was virtually the same. Amount of education was not a factor in response validity.
- Smoking: Although smoking has been shown to lower fecundability, smokers may be prone to risk-taking behaviors and may be more likely to have unplanned pregnancies.
- Age: Data analysis of age and fecundability may be more accurately conveyed when contraceptive information is included.
- Sensitivity analyses can help ascertain quality of data and justify methodologies.

Discussion. Rafael Mikolajczyk, M.D., University of Bielefeld, asked if lower fecundity might be caused by underreporting of abortions, and added this would be a significant factor in certain age and socioeconomic populations. Dr. Joffe said this is an unsolved problem in TTP studies. He said that national figures might be used in place of reported data to produce more realistic data results. He pointed out that women who have abortions might actually be more fecund than women in the general population and added that the best studies reflect considerations of all potential variables. Dr. Buck commented that retrospective studies do not include people who are attempting pregnancy for the first time. Dr. Joffe said that much discussion is needed to refine methodology. Sherry Selevan, Ph.D., EPA, asked about digit preference in retrospective studies. Dr. Joffe said that was a factor in his data, and there was less sensitivity in longer intervals as people stopped reflecting TTP in months and instead used year intervals in reporting. He added that data were quite irregular for TTP periods of longer than one year and that this likely reflected defective reporting. Dr. Selevan pointed out that this would be important in context of environmental exposures and could skew outcomes.

Reproductive and Perinatal Epidemiology: Current Knowledge About the Relations Between Delayed TTP and Perinatal and Later Outcomes

Anne Sweeney, Ph.D., Texas A&M University System Health Science Center

Dr. Sweeney said that research information on TTP is limited. Studies are generally concerned with either timed or planned pregnancy or with subfertility. It is important to look at differences in these two approaches in determining perinatal outcomes. More specifically, it is important to understand the relationship between exposures and delayed TTP, as well as the TTP and its relationship to later outcomes.

Both retrospective and prospective studies have looked at probability of conception per cycle and subfertility. In most studies, subfertility is defined as inability to conceive for one year or more. Data collection typically required for TTP studies includes information on frequency and timing of intercourse; contraceptive history; and previous information on planned pregnancies, unplanned pregnancies, and contraceptive failures. Dr. Sweeny reinforced Dr. Joffe's previous comment that women are not likely to report contraceptive failure. She advised that studies should attempt to limit the number of variables and keep participant burden to a minimum. She pointed out potential problems of research in this area:

- Retrospective studies do not include sterile couples and may not be able to assess relevant exposures.
- Prospective studies tend to segregate data of "planners" and "nonplanners."
  - Because planners tend to be healthier, less likely to smoke or drink alcohol, and have better nutrition, separating the planners and nonplanners does not result in generalizable data, especially since national data indicates approximately half of all pregnancies in the U.S. are not planned.
- Populations that are likely to be exposed to environmental contaminants will not fit the desired subject profiles of couples trying to conceive, and thus these populations will also be excluded.

Dr. Sweeney said that more research is needed on risks associated with delayed TTP. She outlined various aspects of previous research on the topic:

- Subfertility
  - Subfertility is the inability to conceive for more than one year after discontinuing contraception. (Other investigators and clinicians use the term "infertility.")
  - Subfertile women are at higher risk of spontaneous abortion once they conceive.
  - Among women under 30 years, younger populations with longer TTP have higher risk for spontaneous abortion.
- Preterm births
  - Women whose pregnancies ended in preterm births required approximately 15 percent longer to conceive than women whose pregnancies went full term.
  - Risk of preterm birth increases as TTP increases.
- Low birthweight
  - TTP of more than 6 months is associated with low birthweight and shorter birth length.

- Subfertile women are at higher risk for low birthweight babies.

Dr. Sweeney discussed the experience of a Chinese occupational cohort study and how it may affect planning of other efforts:

Eligibility criteria were very specific in the Chinese study.
  - The Study should have eligibility criteria that are broad enough to get a representative sample and yield generalizable data.
- The Chinese study had a 21 percent dropout rate and overall a 45.6 percent exclusion rate.
  - Data collection must be relatively easy. Burden on subjects must be low enough that it does not create potential for dropout.
  - Primary reasons for subject exclusion were refusal to follow the protocol, continuing contraception, pregnancy during use of contraception (birth control failure), insufficient diary, insufficient hCG data, prior IUD use, and personal reasons.
- The Chinese cohort did not consider paternal exposure data.
  - Paternal data will be necessary to determine whether they may contribute to health risks and outcomes.
- Conception rate declined as TTP increased.
  - Pinpointing time of conception will be fundamental to obtaining conception data and early pregnancy loss data.
- The Chinese study found that early unrecognized pregnancy loss was not a risk factor for delayed time to pregnancy or miscarriage.

Reproductive Biology and Toxicology: Viability Span of Female Gametes and Fertilization, and Effects of Environmental Exposures

Jodie Flaws, Ph.D., University of Maryland

Dr. Flaws said the female reproductive organs are potential targets for toxicity. Several organs make up the female reproductive system, and the consequences of environmental exposures for reproduction are varied. Dr. Flaws described the system:

- *Vagina and cervix*: These organs play a key role prior to conception. They must be normal and healthy for intercourse to take place. Secretions are important in intercourse and also help fight infection. If an infection targets these tissues, dryness may cause pain during sexual intercourse. Structural abnormalities may make intercourse impossible. There are environmental exposures that cause excessive proliferation of cells in these tissues and often lead to increased risk of cancer.
- *Ovary*: Females are born with a finite number of primordial follicles that contain eggs, each of which are surrounded by a single layer of somatic cells. Follicles grow and develop into preovulatory follicles, which have the capacity to ovulate. During ovulation, follicles rupture and release the egg into the oviduct. The remaining somatic cells form the corpus luteum, which produces progesterone during the luteal phase, and longer if pregnancy occurs. A variety of environmental exposures target the ovary and follicles. Exposures can stunt development of follicles and reduce fertility. If exposures kill follicles, early menopause may result. Current research suggests that some exposures can target the developing egg, resulting

in increased risk of birth defects. Some environmental exposures increase the risk of ovarian cancer.

- *Oviduct*: Events that occur in the oviduct are not well researched. The egg is released into the oviduct, where fertilization occurs. The fertilized egg performs cell division and develops into a zygote; further division leads to development of the morula; and eventually, a blastocyst is formed about 7 days after ovulation. Implantation occurs in the uterus where the embryo develops. Exposures in the oviduct target the site of fertilization, and thus, they can target the zygote or interfere with implantation and development of the blastocyst.
- *Uterus*: The uterus maintains and develops the pregnancy. Uterine exposures can impair implantation or reduce the ability to maintain, hold, or deliver a pregnancy. Exposures can also increase the risk of cancer and endometriosis and can significantly impact fertility.

There is currently no easy way to detect exposures and related problems in the female. Exposures may impair transport of gametes or prevent the ovary from releasing the egg, or problems may occur at the cervix or vagina, impeding sperm transport. If the egg and sperm cannot reach the oviduct, fertilization cannot take place. If exposures interfere with development of the embryo, the blastocyst cannot implant properly into the uterus.

Environmental chemicals, stress, demographic factors, nutrition and genetic factors can affect the ability to conceive. There are five categories of environmental chemicals associated with reproductive toxicity.

- Cigarette smoke is the most common toxicant evaluated in epidemiological studies. In the United States, 28 million men and 23 million women smoke. There are 4,000 chemicals in cigarette smoke, including 43 carcinogens, and 300 polyaromatic hydrocarbons that can destroy follicles in the ovary, reduce fertility, and trigger early menopause. A recent study of smokers showed smokers have 1.6 greater risk of being infertile compared to nonsmokers.
- Alcohol studies yield controversial results. Research indicates that a dose of more than two drinks per day is associated with increased risk of spontaneous abortion and birth defects. The effects of small doses are still unclear.
- Medical agents, such as ethylene oxide, a sterilizing agent used in dental and medical offices, are associated with increased risk of spontaneous abortion in health workers. Nitrous oxide, an anesthetic, is associated with spontaneous abortion and inability to conceive.
- Little research has been done on occupational and environmental chemicals. Perchloroethylene, a dry cleaning solvent, is associated with increased time to conception and spontaneous abortion. Toluene, a color printing solvent, is associated with reduced fecundity. Organic solvents, particularly in hairstyling products, are associated with increased of abnormal menstruation and spontaneous abortion. Paints and varnishes are associated with increased risk of spontaneous abortion.
- Environmental contaminants are chemicals that have been banned from use, but they do not biodegrade. They can also be byproducts of the breakdown of other chemicals. Polychlorinated biphenyls (PCBs) were banned in the 1970s due to toxicity. These were used in electrical transformers, capacitators, hydraulic fluids, and adhesives. These chemicals persist in the environment and are now found in high concentrations in fish and wildlife. These may cause increased TTP. When humans eat these animals, the chemicals are

absorbed into the system. Pesticides have been associated with spontaneous abortion, endometriosis, and low conception rates. Higher blood levels of hexocychlorohexane have been associated with miscarriage. Dioxins, used in paper bleaching and pesticide manufacturing, have been associated with increased spontaneous abortion, endometriosis, and low male offspring secondary sex ratios.

It is clear that some environmental exposures increase toxicity in female reproductive organs. There are several targets and many consequences. Future studies with large sample sizes are needed to confirm findings. Direct biologic measures are needed.

Discussion. Dr. Joffe asked how difficult it is to find relationships between health outcomes and environmental exposures, even when they may be strong and affect a large portion of the population. Dr. Flaws said that occupational samples are quite difficult. Current research primarily focuses on early life and development, but long-term outcomes of exposure are rarely evaluated in research. Animal models are easier because researchers are able to control the environment and the number of exposures, but humans may be exposed to any number of different toxins, making research more difficult in human populations. Dr. Campbell commented that considerations of male exposures compound the difficulty of pinpointing problematic environmental exposures and outcomes for the offspring. He added that dietary exposures should also be investigated.

Reproductive Biology and Toxicology: Viability Span of Male Gametes and Fertilization, and Effects of Environmental Exposures

Steven Schrader, Ph.D., National Institute for Occupational Safety and Health, CDC, DHHS

Dr. Schrader said there are several sites of possible toxicant effects in the male reproductive system including the endocrine system, the testes, and accessory sex glands. Sexual function can also be affected. Male hormones can be measured in blood serum or first morning urine; shift work could cause urine profile variations. He said that several occupational exposures affect the endocrine system in men, including:

- Lead, which is associated with increased LH and lower testosterone
- Kepone, an estrogen compound associated with changes in the overall hormone profile; men had enlarged breasts (gynecomastia)
- Stilbene, an optical whitener, has been associated with decreased testosterone and sexual function.

Genetic structure and sperm production can be affected by testicular exposure. In the testes, structures called seminiferous tubules produce sperm. Primary spermatagonia go through mitosis and meiosis and result in sperm. Type A spermatagonia replicate, allowing spermatogenesis to continue. If spermatagonia are destroyed, mature sperm are not produced. Animal studies have evaluated spermatogenesis inhibition following exposure to toxicants. Sperm evaluation is difficult in humans because sperm are produced in assembly line fashion, and thus, there are sperm of various stages of maturity existing inside the male body at all times.

Examples of environmental toxins associated with changes in male reproduction include:
- Dibromochloropropane (DBCP): the most widely known male reproductive toxicant that kills spermatagonia. This effect can be irreversible.
- 2-methoxyethanol: attacks spermatocytes. Sperm production eventually resumes.
- Lead: affects the release of sperm into the lumen of seminiferous tubules in animals.
- 2,5-hexanendoine: affects Sertoli cells, which are nurse cells that follow sperm as they go through spermatogenesis. If these are destroyed, reproduction is impossible.

In a single seminiferous tubule, sperm in various stages of development can be seen. Thus, if sperm are destroyed in one stage of development, sperm in other stages of development may still be viable and able to fertilize an egg. To assess sperm production, researchers can evaluate:
- Testicular size
- Testicular weight
- Testicular biopsy
- Sperm morphology and morphometry: abnormality in shape or size. Researcher can do subjective classification, or computer analysis can measure and categorize sperm.
- Sperm count: this assessment is used most often. Although there is some variability, sperm counts are relatively consistent in a particular male over time.

Other chemicals that affect sperm production are:
- Dibromochloropropane (DBCP): sperm count decreases
- Ethylene dibromide: sperm head size decreases and sperm count is low
- Glycol ether: reduced sperm count.

Assessing actual genetic mutations and potential for altered offspring development is difficult. Several occupational studies have shown an association between paternal exposures and adverse outcomes in offspring.
- Karyotyping is a method used to assess chromosomal makeup during mitosis and can be used to analyze genetic material in sperm.
  - The physical structure of sperm makes karyotyping both time-consuming and difficult.
  - Karyotyping is usually performed after sperm have fertilized an egg.
- Sperm comet assay is a type of electrophoresis that assesses chromatin.
  - Chromatin reveals the stability of DNA strands.
  - Studies clearly indicate that sperm chromatin is affected by air pollution particulates.

Dr. Schrader's list of accessory sex glands included the epididymis, prostate, and seminal vesicles. Ways to assess status and function of these glands include:
- Gland palpation
- Biochemistry: examine the sperm. If sperm die after they depart the testes, the accessory sex glands are most likely the problem.
- Semen volume: gives some idea if all accessory sex glands are functioning properly
- Semen pH: can be assessed at same time volume checked
- Vital stain: allows viewing to see whether sperm are alive or dead
- Hypo-osmotic swelling: assesses osmotic membrane stability involved in fertilization

- Sperm motility: sperm must be able to swim to the oviduct and penetrate the cervical mucus.

Ethylene dibromide is associated with low semen volume, low percent of motile sperm, low number of living sperm, higher semen pH, and low fructose. This suggests some vesicles were affected; hydrocarbons were found in the semen itself.

Sexual function is divided into three elements:
- Libido: the most difficult to assess because evaluation uses questionnaires, and results may be unreliable due to the sensitive nature of these questions.
- Ejaculation: easy to assess; if semen is present, ejaculation occurred.
- Erection: In recent years, erectile dysfunction has attained a high degree of social importance. Rigiscan monitors are computerized units that are placed on the penis at night. On any given night, the average male has three to five erections in his sleep, or about 40 percent of his sleep time. The device is easily transported, and data can be downloaded in the lab.

Several occupational studies have demonstrated sexual dysfunction in men involving chemical exposures, including lead, carbon disulfide, stilbene, and cadmium, as well as other occupational exposures (for example, bicycling police officers).

Determining sperm lifespan is also difficult. Normal ejaculate contains millions of sperm, but relatively few will reach the oviduct and have the opportunity to fertilize the egg. Most estimates indicate a 72-hour life span for sperm, but anecdotal evidence sometimes suggests sometimes survival may be longer.

Discussion. Dr. Joffe stated that epidemiological trends indicate increasing dysfunction of the human male reproductive system as evidenced by increasing testicular cancer and falling sperm count. He advised that much recent research has shown some evidence of genetic linkage in testicular cancer, and research trends are focusing more heavily on male infertility. Dr. Schrader said that assessing genetic damage in sperm has been attempted for quite some time, but the difficult and time-consuming techniques are primary drawbacks for most projects. He said that because there are millions of sperm per ejaculation, the best that can be done is a subsample assessment that may or may not contain a mutation. New analyses must be developed to efficiently assess the health of sperm. Dr. Zinaman commented that, in recent years, researchers have found many Y-chromosome mutations or autosome mutations associated with Y-chromosome function that lead to poor quality sperm. He said experts are now looking at autosomes that interact to find new mutations. A participant asked if implantation bleeding might be misinterpreted as menstruation, thus leading females to believe that they had conceived in a "cycle" (after the misinterpreted bleeding episode) when they had not had intercourse. Dr. Schrader said that an example case in Australia was unexplained, but that the participant's explanation was plausible in that case. It will be important in the Study to avoid confusion of menstruation with implantation bleeding. Dr. Zinaman commented that research has been attempting to explain reports of falling sperm counts for over the last 40 or more years. Dr. Schrader said research data are unclear and added that some studies show no overall change in

population sperm counts. However, reports indicating falling counts far outweigh those that have shown increasing sperm counts. He said that as new methods are discovered to assess sperm quality, the answer might become clearer.

Statistical Models and Issues for Day-Specific Probability of Conception: History and Development

Haibo Zhou, Ph.D., University of North Carolina

Dr. Zhou presented statistical modeling that may be used to determine DSP of conception in data analyses. He described how models have been modified over time.

In the Barrett and Marshall model, variables include:

- Conception
- Intercourse: timing and frequency
- Probability of conception.

Issues with this model are:

- It assumes that sperm batches act independently.
- It also assumes menstrual cycles will be homogenous, both between women and within the same woman. Daily exposures, sporadic use of barrier methods of contraception, and cervical mucus may have an effect.
- In extreme cases of repeated intercourse during the fecund window, it results in an implausibly high probability of conception.

In response to the latter concern, Schwartz modified the Barrett and Marshall model:

- Conception not only depends on timing and frequency, but also a third factor that is a multiplicative to DSP: the viability factor.
  - Viability factor: represents biological and environmental factors for both males and females.
  - All factors other than timing and frequency are grouped together in covariate, which Schwartz called "A."
  - Some use of linear modeling for age might show an effect on "A."

In 1996, Zhou and Weinberg outlined other limitations to the Schwartz model:

- Cycles for the same couple are modeled as independent.
- Cycles could be heterogeneous within or between participants.

The Weinberg paper in 1996 in *Biometrics* modeled this P_k as a function of covariates. This incorporates a whole array of other variables and simplifies the three-level structure. For every cycle observed, the following parameters can be used:

- I = individual
- J = cycle
- K = day

Day-specific factors that could be accounted for include: caffeine consumption, smoking, and alcohol use.

There are also binary indicators of intercourse and whether it leads to conception.

Another limitation of the model was that every day in a cycle had the same potential for success or failure. This was an issue in that, if conception did not occur, all days were failures. But if conception did occur, on which day did the success occur? The model does not indicate the day of "successful" conception. It is not a 1-day success; it is rather that there was success on at least one of the days with intercourse. These covariates can be teased out individually. Theta, another exposure variable corresponding with P_k, can represent exposures of interest. Here, the inference is on conception likelihood for a given day, and it can be easily applied to get theta values.

An aggregated model further reduced the model to a simpler structure. Latent factors affect the probability of conception, and the outcome of these is the cycle viability. There are two sets of covariables that can be teased out individually: exposure and viability. Variables are viewed as a heterogeneous group of data, and adjustment is made for the cluster effect with a standard error formula. The point estimate for the effect is the same as before, but the standard error is now going to be different.

The model has several assumptions:

- The reference day (estimated day of ovulation, designated as day zero) has to be known clearly, classified, and defined.
- Each batch of sperm contributes independently.
- Each cycle's contribution is independent for the same couple.

A random effect model can be used to model individuals from their unique data. But days still do not contribute independently. Random effect models can be done both parametrically and nonparametrically. Conditional conception probability then attempts to separate out daily parameters by providing values to the 6 days prior to conception, with the day of conception being the zero reference point.

Cycle viability probability modeling looks at exposure variables such as smoking. The Bayesian approach can be used to model cycle viability into two components. Obviously, sterile populations will not fit this model.

Discussion. Dr. Holman asked how the mixture model used the control sterile subgroup versus the sterile subfraction of the potentially fecund model. Dr. Zhou stated that the groups were both included in calculations and that they were indistinguishable. Dr. Joffe asked if the age-related decline in cycle variability could be related to menopause. Dr. Zhou explained that implications were difficult to assess. Dr. Zinaman asked how data were assessed for multiple daily intercourse acts, and there was general discussion about potential for lower sperm count in these groups. Dr. Schrader said clinical trials indicated that when several ejaculate samples were collected in short succession, the second sample had the highest sperm count. He stated that normal males could

have intercourse on a daily basis without any loss in fecundity and added that sperm counts remain relatively stable over time for individuals. Dr. Dunson said the data did not contain information on number of intercourse acts in a day, but intercourse frequency was very high for groups assessed because the sample consisted of couples attempting pregnancy.

New Statistical Methods for Studying Predictors of Day-Specific Conception Probabilities

David B. Dunson, Ph.D., NIEHS, NIH, DHHS

Dr. Dunson described the Schwartz model, which can be used in analysis of cycle viability and for day-specific conception probabilities in a viable cycle. Day-specific data can include hormone levels, cervical mucus data, various exposures, and intercourse data. Age and environmental exposure information for both males and females can be included, and generalization can allow for covariate effects and heterogeneity among couples. Dr. Dunson has used the model extensively and has found some limitations with it.

Because he was interested in predictors and needed to account for heterogeneity, Dr. Dunson developed an alternative strategy that extended the Schwartz model but added constraints to improve stability. He set the maximum of the DSP as equal to one, viewing it as the probability of conception on the most fertile day of the menstrual cycle. He also constrained the day-specific possibilities to follow an umbrella ordering for the curve of DSP. Dr. Dunson said that if there are no constraints on DSPs, extremely wide confidence intervals result. Constraining the DSPs can yield more efficient inferences. Under selected constraints, he incorporated separate covariates and random effects for the level of the peak and day-specific ratios. Some covariates pushed up the level of fertility on the most fertile day, and others caused the fertile interval to shrink or expand.

Using this model, he performed analysis of caffeine data from the North Carolina Early Pregnancy Study and found an interaction between the caffeine effect and the timing in the fertile interval. The caffeine effect was strongest on the most fertile day, but closer to the boundaries of the fertile interval, there was no caffeine effect.

Another study on the caffeine effect in the late 1980s was a standard TTP analysis that did not incorporate any information on the frequency or timing of intercourse. In this study, Dr. Dunson found an effect, but compared with the DSP analysis using a simpler model with caffeine as a cycle-specific predictor affecting cycle viability, the effect was significantly underestimated due to the interaction of variables. He emphasized the importance of recognizing interaction effects.

This model has also been used to study declines in fertility in males and females with regard to age. Dr. Dunson said this approach is very computationally intensive, which prevents routine implementation, and simpler, robust models are needed. Databases containing mucus information on each day of the menstrual cycle could be used to predict the fertile window of the menstrual cycle. The challenge is to find a way to model the distribution of the fertile interval, because the fertile interval changes along with covariates.

Dr. Dunson said he wants to begin to look at models that can investigate order trends and improve efficiency. Epidemiologic studies often have ordered categorical predictors and levels of covariates. New methodology is needed that will simplify calculation. He is currently working on an alternative data augmentation approach that is very computationally efficient; he hopes to publish methods in *Biometrics* relatively soon. [Note: This paper was published in *Biometrics*, 2005;61:126-33.] His model is a hierarchical generalization of the Barrett and Marshall model that includes random effects, but attempts to retain relative ease of analysis. More specifically, this is a model for the conditional probability of conception in a cycle based on a couple's characteristics, and with a random effect term that essentially assesses the unmeasured factors within that couple. It has a simple, closed mathematical form that is an extension of the logistic regression model. Another key implementation factor is an auxiliary variable specification that introduces a latent variable for computational purposes without changing anything about the model interpretation. Simple conjugate priors can be chosen in this auxiliary variable structure, which makes it easier to normalize constants and creates efficiency, while also providing flexibility. Dr. Dunson talked about his plans for future research. He emphasized the importance of mucus data over hormone data because of ease of collection and their high information value. He presented an analysis from a European study showing a strong effect of mucus quality on probability of conception—stronger than timing relative to ovulation. Future analyses with this model will begin to look at interactions of sperm, mucus, and age.

Discussion. Dr. Campbell asked if some of the models used redundant parameters or could be restructured to use redundant parameters such as mucus, electrolytes, or saliva. Dr. Dunson clarified and stated that redundant parameters are measures repeated across cycles. He said this could be done, but it is not a statistical necessity. A more simplified approach can yield desired results about underlying, complicated processes, and a better approach for summarizing complicated information. Dr. Joffe expressed interest in couple-specific frailty analysis. He said variable W (cycle viability) seemed to have dropped from the equation. Dr. Dunson said that if W was included, the model would be overparameterized and added that closely interrelated variables could not be segregated. He eliminated W from his model, and the effect was higher stability with regard to random effects. Dr. Zhou has worked with frailty distributions, which are essentially fecundability distributions. When distributions are plotted, there is residual heterogeneity that is not accounted for by the predictors of that model. By knowing the population distribution of those predictors, analysis can yield a real distribution of the heterogeneity in fecundability in the population. The more terms that are added to the model, the larger is the sample size needed to yield estimations. Dr. Zhou commented that the model was very reliable. He said a converging Bayesian approach could be more troublesome.

Statistical Models for Prediction of Conception Using Mucus If There Is No Ovulation Marker

Bruno Scarpa, Ph.D., University of Pavia, Italy

Dr. Scarpa said his research has developed statistical methods that describe the relationship between probability of conception and type of cervical mucus. His method was based on an

Italian study of couples that were using the Billings ovulation method. The Italian study employed a five level mucus observation classification scheme, but because the last two levels were so similar, Dr. Scarpa's research combined them and used only a four level mucus classification.

In each cycle, there are on average about 6 days in which the mucus is most conducive to pregnancy. The variability of results in mucus category four was quite high and may be due to differences in fertile intervals. Out of 2,755 cycles, data were discarded for approximately 219 cycles for various reasons, including:

- Incomplete or missing information about intercourse
- Incomplete mucus records.

Data clearly revealed a relationship between mucus type and day of the cycle. A random effects model was used to perform woman-specific analyses, and Bayesian estimators were also used with some constraint of the parameters. Mucus effects were consistently ordered, with type four having the largest effect, but the method did allow for the possibility that mucus had no effect. A weak prior distribution was chosen on the basis of previous work with this method. Mucus effects were operative during a "middle" time window of the menstrual cycle (corresponding to the fertile window), but not at the very beginning or very end of the cycle.

Dr. Scarpa said that a possible continuation of this work is observing that after one point in the cycle, and before another point in the cycle, the probability of conception does not depend on mucus. One idea is to divide the cycle in three parts with the middle part allowing for mucus effect and the other two parts having baseline parameters. Dr. Scarpa's other work focuses on developing a framework to define an optimal rule for defining days of fertility and infertility for couples who wish to avoid pregnancy based on periodic abstinence. He added that a Schwartz-type model can also be used to estimate the day of ovulation using mucus classification.

Discussion. Dr. Schrader asked about training required for identifying mucus by study participants and asked how quickly they could learn the mucus classifications. He was concerned about pregnancy prior to learning the classifications. Dr. Zinaman said subjects could not be trained while using oral contraceptives because of differences in mucus caused by these contraceptive methods. He added that most gynecologists counsel that women not conceive until their second natural menses after discontinuing oral contraceptives, because of the increased risk of spontaneous abortion within the first cycle. Training should take place during those first two cycles for women discontinuing birth control.

Issues in Longitudinal Modeling of Outcomes for Periconceptional Exposures and Subsequent Outcomes

Patrick Heagerty, Ph.D., University of Washington

Dr. Heagerty stated that timing is critical to the analysis of longitudinal data. Because of the numerous processes related to conception and DSP, it will be necessary to think carefully about data analysis and integration, as well as effects of dropout or participant death in the 20-year

Study. He said that analysis with time-varying exposures may be very complicated. Nevertheless, longitudinal data can be assessed precisely for the threats to human reproduction and can help determine the effects of very early exposures that go beyond conception and cover issues of impaired gestation and growth.

Dr. Heagerty discussed the Schwartz model, which is a variation of the Barrett and Marshall analytical model. This model identifies each act of coitus as an independent event with its own probability of conception. It is important to consider the role and nature of variables when making decisions on data analysis of periconceptional exposures and toxicants. Intercourse pattern may play one of three roles in the Study:

- It is a good independent predictor of conceptional outcome, so it may offer precision.
- It may have confounding properties.
- It can modify effects of the exposures.

The Schwartz model is a type of generalized linear regression that can be used to estimate DSP. This model is flexible and allows for statistical adjustments to avoid selection bias and confounding. Function of the expected outcome is linearly related to exposures that are ordered by time.

General logistic regression can be adjusted to yield high-dimensional predictors of time-specific exposures. Timing is not generally a central feature in logistic regression, but timing can lend precision in the analysis. Dr. Heagerty commented that this general regression could be useful for evaluating conception and subsequent developmental outcomes. It will also be important to think about models for coefficient functions and desired parameters.

In one analysis, Dr. Heagerty used a series of simulations to structure a simple model that assumed a single intercourse event in the window of opportunity. Each woman contributed a single cycle with coitus on one day. Day was not associated with exposure, but was a dichotomous variable. This allowed for a crude test of the influence of timing of intercourse versus exposure on the probability of conception. This was an attempt to assess the level of precision added by information about the timing of intercourse. With an assumed exposure odds ratio of 0.50, half of the sample exposed, a sample size of 1,000, adjusting for the timing of intercourse increased the power from 0.657 to 0.691. Thus, precision was gained, but it was not a dramatic effect.

Another important issue is if the timing of intercourse may be an effect modifier for exposure. Calculations should be done to assess sample size requirements for this analysis. In general, very large sample sizes are required to get the power to detect these types of interactions. Varying coefficient methods may be a potential remedy to handle high-dimensional stratifying variable issues and accommodate high-dimensional vectors. Varied coefficient models are straightforward statistical analyses that can be done in standard regression.

Other methodological issues arise as studies move beyond conception and look at TTP. In these analyses, couples contribute multiple attempts until conception results. Trial time scales can be important predictors in these analyses.

Dr. Heagerty said that although periconceptional data may be difficult to analyze, the good news for post-conceptional outcomes is that exposures of baseline variables, such as timing variance, are very good for longitudinal analysis. He said many longitudinal models yield good results and are well understood, but it will be important to recognize secondary processes and patterns of outcomes, and to consider why outcomes happen when they do.

Discussion. Dr. Dunson asked if the simple generalized estimating equations (GEE) model allowed for cluster size where the number of observations might be informative about the outcome or exposure effect. Dr. Heagerty said there are a number of difficulties at play. The GEE method pools datasets, but a linear mixed model can provide valid inferences under more stringent missing data mechanisms. These methods impute the missing data for people with partial outcomes. Dr. Dunson asked about the probability of conception if there were no intercourse acts and how this would affect sample size determinations. Dr. Heagerty said the reason for the interaction assessment was to determine realistic numbers required to get significant results. He said that these determinations can be difficult for the study of intercourse timing, exposures, and conception. The Schwartz model allows for a probability of zero and simulations in which there is only one event taking place in a particular window. Dr. Zhou commented that there is no immediate or obvious way to incorporate timing and frequency into data. Dr. Heagerty concurred and said it is a high-dimensional vector represented by a string of zeros and ones, but he added that there are many ways to incorporate that into analyses. One way would be to give each element its own parameter.

Issues with Clustering and Dependency in Prospective Pregnancy Studies

Vanja Dukic, Ph.D., University of Chicago

Dr. Dukic said that structure of prospectively collected pregnancy-related data is quite complicated, but it can be viewed in a hierarchical or multi-level way. Several processes are nested and embedded in each other, and they are elements of this complex hierarchical structure. Levels of prospective pregnancy data can include information on:

- Couples, including data on environmental exposures such as lead in drinking water
- Individual data on both men and women
- Menstrual cycles, including various related parameters such as cervical mucus and ovulation
- Days of the cycle; days in the fecund window have a high probability of conception.

Such datasets represent levels of informational clusters that have unique issues:

- Pregnancy data contain complicated dependency patterns that are induced by this multilevel structure.
- Assessment of the dependency patterns needs to be done to formulate the best study design and analysis.
- Ignoring these patterns can lead to loss of statistical power and efficiency, as well as bias.

Dr. Dukic discussed ways to manage dependencies, including:

- Design stage: units of analysis must be determined, as well as an appropriate power analysis based on characteristics of the chosen unit
- Analysis stage: there are several model options
  - GEE: analysis with Stata, SAS, and other software
  - Mixed or Bayesian models: analysis with SAS, WinBUGS, HLM, or MIXOR software.

There are several options for analysis of dependency patterns:

- They have traditionally been ignored.
- They can be adjusted using Huber White or another robust standard error correction.
- A generalized approach can be taken, such as using generalized estimating equations that can be implemented as datasets.
- Going one step further with modeling, hierarchical or Bayesian models or mixed models can be used.

Dr. Dukic gave a general description of the uses of various methods and discussed advantages and disadvantages:

- Huber White is a clustering correction.
  - Advantages: this method is robust, less sensitive to model misspecifications, and is easy to implement.
  - Disadvantages: this method does not allow for individual predictions or heterogeneity estimates. It is not very efficient, and strong assumptions are needed for missing data.
- GEE specifies the longitudinal correlation that weights cluster data.
  - Advantages: it is valid under misspecification of longitudinal correlation and makes estimation more efficient.
  - Disadvantages: like the Huber White, the GEE does not allow for individual prediction or heterogeneity estimates.
- Bayesian models are mixed, multilevel, hierarchical models.
  - Advantages: They lend themselves well to elaborate biological models and allow for extensive modeling of heterogeneous outcomes.
  - They are more complicated and model every process separately, yet also link processes together.

Within models, there are processes and subprocesses. Biological models may have several parameters that cause instability and require incorporation of prior information. Information from experts or prior studies can be embedded into these models, making them particularly useful for clinical predictions and patient counseling. However, level-specific parameters or random effects must be unrelated to the covariates that are already in the model. Bayesian methods are also extendable, in that they can accommodate a variety of sources of information, including combinations of prospective and retrospective studies, and this allows for meta-analysis.

Dr. Dukic described a study on smoking effects on pregnancy outcomes in a dataset where many women contributed repeat pregnancies. Multiple pregnancy data had complicated dependency patterns. GEE and a mixed model with random intercept were useful in analysis of outcomes.

Concluding remarks:
- Pregnancy outcomes are strongly correlated.
- Single pregnancy models are inefficient and may overlook important effects such as age at first pregnancy.
- GEE analysis with independence working correlation and mixed models are useful approaches to modeling pregnancy outcomes, but they answer fundamentally different questions**.**

Discussion. Dr. Dunson commented that research has shown the GEE approach can effectively underestimate results because the less fertile women contribute more cycles to the dataset. Dr. Dukic said that characteristics of the biological process of interest should determine the model and that the model should be tailored to that bioprocess as much as possible.

Access and Dropout Bias

Joseph Hogan, Sc.D., Brown University

Dr. Hogan said his goal was to show how two different commonly used models give rise to very different results when used to analyze the same data. To avoid this problem, when designing methods of analysis, researchers must consider two issues:
- Models must be designed appropriately for the hypothesis.
- How dropout rates are used can give different results.

He used data from a prospective study on tubal disease in women undergoing IVF to exemplify problems with subject attrition. The main objective of data analysis was to estimate per embryo implantation probability. The model allowed for covariates that could differ from woman to woman, and from cycle to cycle. Covariates used were cycle number, age of the woman, and number of the embryos that were not transferred. A key point is that in IVF, several embryos can be fertilized in vitro, but not all are transferred. The number not transferred is an important marker of eventual success because the embryologist is allowed to choose the best embryos, thus raising the chance of successful implantation.

Pregnancy and failure of IVF led to dropout of study couples.

Problems for analysis are similar between this study and the study of DSP of conception:
- When more than one embryo is transferred, there is no way of identifying which embryo implanted.
- Precise day of conception cannot be determined if there are multiple intercourse events in a single cycle.

There are several potential ways to estimate conception probabilities or even day of conception from a longitudinal cohort study. Dr. Hogan illustrated formulation of marginal and conditional models, and spoke about embryo implant study structure, describing the implications for timing of conception studies. He described the various features of the models:

- Marginal models:
  - These models compare different slices within populations and make inferences about between-subject effects.
  - Estimation methods are freely available for these models.
- Conditional models:
  - These models can be constructed with, or without random effects.
  - They can make inferences about within-subject effects.
  - Attributes of the embryo can be measured.

Because these models have decidedly different interpretations in the context of repeated cycles, careful consideration must be given to methods used. GEE is a very popular method to estimate marginal models, and random effects are very popular specifications of conditional models.

Dr. Hogan described in further detail the problem of attrition in his IVF study. There was an 80 percent implantation rate among women in their first two cycles. Population turnover was a factor in the sense that the population that conceived would leave the study, and new subjects would subsequently be added. The bias resulted primarily because the people most likely to conceive were selected out of the sample in the first 2 months. Subgroups for which IVF was not successful would also depart, and researchers were not always informed as to subjects' reasons for leaving the study. Thus, the mixture of those two populations, and particularly the latter population, contributed to a decline in the per cycle implantation rate. Another problem was that the effects of covariates that fluctuated within cycles also apply to these cycle-specific populations. In other words, the aggregated effect of covariates that varied across a cycle were more important for those couples who were followed longer compared to those who quit after becoming pregnant.

Dr. Hogan said that he used a very simple random effects model. His model contained a variable that characterizes a host of unobserved characteristics that could potentially be identified by the outcomes such as embryo implantations. The variable reflects the couple-specific propensity to succeed at IVF and conditional effects represented a within-couple or within-woman effect. Between-subject effects were seen in the marginal model. The within-subject analysis attempted to ascertain whether a woman became less likely to succeed at IVF as she attempted more cycles, while the marginal model dealt with the probability of implantation in populations starting in various cycles.

In Dr. Hogan's model, the random variable U used information from the observed outcomes of implantation to identify a distribution. Using a likelihood model, either Bayesian or maximum likelihood, bias due to dropout from observed events such as pregnancy are adjusted for, but dropout with no follow-up report could not be incorporated.

The result of Dr. Hogan's analysis was a cycle effect in the opposite direction of the marginal model. The marginal model found that in a cohort of women with repeated IVF cycles over time, as women drop out, the likelihood of success diminished in subsequent cycles. However, the random effects model found that as the same woman attempted more and more cycles, her likelihood of implantation increased. The latter result could have been influenced by bias due to women with a lower chance of success with IVF dropping out of the cohort.

Discussion. Dr. Dunson asked how the variance of uterine implantation was identified. Dr. Hogan said there were repeated clusters of binary outcomes, and the variance of uterine implantation was identified using data from single embryo IVFs assessed using an 8-week ultrasound to confirm implantation. Dr. Heagerty inquired about data filtering and selective removal of subjects on the basis of their fertility status. Dr. Hogan said that, because his model makes the reasonable assumption that a couple's fertility does not necessarily vary with cycle, there should be a concrete reason for modifying the assumption in the model. Dr. Joffe said a recent paper (Basso et al., *Int J Epidemiol* 2000 29(5):856-61) covers the phenomenon of age-related issues of subject attrition.

Simulating DSP of Conception and Rates of Early Pregnancy Loss in the Presence of Ovulation Information But Absence of Daily Coital Information

Darryl J. Holman, Ph.D., University of Washington

Dr. Holman explained that his approach to research is from an anthropologic or demographic perspective. Analytical tools used in these disciplines are very similar to those statisticians use, but his focus was more on modeling aspects of fertility processes. He summarized several decades of demographic work on age-specific apparent fecundability and age-specific apparent pregnancy loss. Results from many different studies were adjusted to a common fecundability equal to one at age 22 and showed an increasing pattern of risk for pregnancy loss over time. He said that the statistical curves shown were not exactly correct, as fecundability and pregnancy loss are related in a number of important ways. Researchers generally believe that pregnancy loss is much higher than results indicate, because most of these studies are done using self reports of pregnancy loss rather than using hCG assays.

Dr. Holman provided the following definitions:

- Total fecundability: the probability of conception in a menstrual cycle. This is not dependent on the ability to detect the pregnancy.
- Total pregnancy loss: the probability per conception that the pregnancy will end in a loss. This is not conditional on being detected.

Fecundability studies cannot be compared unless the same methods are used for detecting pregnancy in each study, and the same is true for pregnancy loss. The ultimate fecundability study would detect pregnancies at conception. Menstrual cycles would be recorded, as well as conception, and these pregnancies would be tracked for 9 months. Total fecundability could be assessed by taking total pregnancies divided by total number of pregnancy losses. Because there is no assay to detect the moment of conception, conception is determined after a certain

gestational period has passed. Thus, a limitation of these studies is that, if pregnancy loss occurs before it can be detected, data on both the pregnancy and the pregnancy loss are lost.

To correct for this problem, demographers use hazard and survival models. Although these models do not yield products of actual measures, they are based on simple biological principles that can be quantified as a parametric model. This involves data extrapolation based on recognized scientific theory and provides an estimate of data that cannot be empirically observed.

The Wood-Boklage model of fetal loss divides the products of conception into two risk groups: high-risk conceptuses (chromosomally abnormal) and low-risk conceptuses (chromosomally normal). This type of model is known in the field of demography as a two-point frailty model. Statisticians would generally call it a mixture model, and more specifically, a hyperexponential model. Because most abnormal pregnancies will be lost early in pregnancy, risk drops rapidly at the beginning of the process, then levels out as the high-risk conceptuses are selected out and an increasing level of normal conceptuses remain until the end of pregnancy. Covariates on hazards and the mixing fraction are easy to model in this method.

Dr. Holman discussed sensitivity and specificity of assays as a significant issue for the Study. In the Study, thousands of pregnancy tests will be done, and most of them will be negative either because it is too soon to detect pregnancy, or pregnancy loss has occurred, or conversely, a false positive may occur. But if the sensitivity and specificity of the assay are known, probability of pregnancy can be determined.

Dr. Holman said that his research has focused on trying to estimate total fecundability, age-specific total fecundability, and age-specific pregnancy loss. He described a field study in Bangladesh that yielded estimation parameters and showed that the only effect of a woman's age was higher risk of early pregnancy loss. While the probability of conception remained high at higher ages (up to 48), the probability of unrecognized early pregnancy loss increased steadily with age. Levels of early pregnancy loss were surprisingly high, ranging from approximately 51 percent at age 20 to approximately 96 percent at age 40. Fecundability was high and nearly constant across most of the reproductive span. Most of the age-related decrease in fecundability was due to increased fetal loss. Dr. Holman said his model did not evaluate day-specific fecundability, and there were no data on coital frequency. He said that with a larger sample size and coital information, the model may be useful for day-specific fecundability estimates. He presented simulation results suggesting that his model could be applied to DSP of conception.

Comparing DSP of Conception in Datasets with Different Markers of Ovulation

Rafael Mikolajczyk, M.D., University of Bielefeld, Germany

Dr. Mikolajczyk said there were various ways to use available datasets to generate hypotheses. His objective was to develop a simple method to estimate the effects of covariates on fecundity using a minimum of information—cycle length and records of intercourse. The method was based on the ratio of observed to expected pregnancies and analysis of couple and cycle

covariates. The expected pregnancies were estimated by developing a set of estimates for DSP of conception referenced to the last day of the menstrual cycle.

Considerations when adjusting for the frequency and timing of intercourse include:
- Is it sufficient to adjust only for the frequency of intercourse?
- Is it sufficient to know how many acts of intercourse there were in cycles?

If a woman gets pregnant in a particular cycle, the preceding cycle can be used as a proxy for the standard length of the cycle in calculations of expected versus observed pregnancies.

Dr. Mikolajczyk performed analysis of three datasets from mucus, follicular ultrasound, and basal body temperature (BBT) databases, along with fecundity information from a European fecundity study. Data were available on follicular rupture by ultrasound, which is an objective marker of ovulation. The mucus database contained information on over 1,500 cycles, the ultrasound database had about 270 cycles, and the BBT database contained nearly 30,000 cycles.

Assessing for the possibility that women in the natural family planning studies of mucus and BBT stayed longer and contributed more regular cycles, Dr. Mikolajczyk checked for bias in the sample, but analysis revealed no significant difference when only one cycle per woman or all cycles were included in the analysis of the distribution of ovulation in reference to the last day of the menstrual cycle. The distributions of the ovulation markers of BBT, mucus, and ultrasound were referenced to the last day of the cycle. He merged the information on distribution of fecundity markers to generate daily fecundity estimates. The window of fecundity was modeled to be between day 10 and 23, counting backwards from the end of the cycle. Before day 23, the probabilities were miniscule, and after day 10, the physiologic assumption was made that the short luteal phase would not allow for the maintenance of pregnancy. The distribution of DSP of conception was quite similar between the different datasets used, despite the fact that they used different markers of ovulation.

Dr. Mikolajczyk described how he calculated expected pregnancies. The first analysis used cycle-level covariates, and the other used the woman-level, or individual-level, covariates that indicated cumulative effects over several cycles. He illustrated the process with data from three cycles of one woman that included days of intercourse and cycle length. Intercourse in the fecund window occurred in all three cycles, and pregnancy occurred in the third cycle. For each day of intercourse, the DSP of Conception (relative to the last day of the cycle) yielded a probability or fraction of an "expected pregnancy." For this woman, the cycle-specific probabilities were 0.087, 0.209, and 0.203 for the three cycles, respectively. These probabilities accumulated across cycles to yield a final expected probability for this woman over three cycles of 0.42. In the third cycle, pregnancy occurred, so there was no cycle length, but previous cycle data were used as proxy for cycle length. The end result was a table that indicated the days of the cycle and probability of pregnancy.

Using the various datasets, Dr. Mikolajczyk demonstrated that expected pregnancies fit the observed pregnancies well for the entire samples, but deviated substantially for some subgroups such as women with previous births, who demonstrated a much higher observed fecundity using

this analysis, even though the actual proportion of women with previous birth who conceived was substantially less. Thus, the observed versus expected fecundity ratio analysis demonstrated that women with previous births had higher fecundity (pregnancy when intercourse occurred during the fecund window) despite a lower observed fertility (absolute pregnancy rate), because of much less intercourse during the fecund window.

Advantages of this method:

- Generally, this method is good for clinical use.
- It is a simple way to standardize observed pregnancies across different subgroups.
- It can be used to compare fecundity estimates between studies that have used different markers of ovulation.
- It can be used to generate hypotheses about effects of covariates in studies with limited information. This is because when exposures are rare, larger sample sizes are needed, and in larger samples, it may be easiest to collect more limited data (that is, intercourse timing and cycle length are sufficient).

Conclusions of the presentation were:

- DSP of conception referenced to the last day of the cycle is a valid and reliable measure of fecundity.
- This method is suitable for comparison on the group level as the ratio between observed and expected pregnancies.
- This method is suitable for measuring effects of covariates either at the woman/couple level or at the cycle level.
- Less information is necessary, and therefore, the method is easier logistically to apply in large studies.
- This method is only suitable for evaluation of aggregate data, and not individual data; the methods are also not practical for day-level covariates.
- Further validation in other datasets is needed.

Discussion. Dr. Dunson asked if the method looked at effects related to follicular phase length associated with fecundability. Dr. Mikolajczyk said follicular phase was not a consideration because its length is unknown. He suggested that this serve as a point of departure for development of more elaborate analyses. Dr. Buck commented that too much trust is put in lab results. She advised that important information on chromosomal anomalies may be missed because false positive pregnancy results are not investigated further. Dr. Holman said literature indicates assays are becoming more reliable and able to determine pregnancy earlier. However, because conception is still not detectable and there is a certain amount of gestation that must take place before pregnancy can be detected, pregnancy losses may also go undetected, especially pre-implantation pregnancies. This information would be essential to identifying exposures that increase rates of very early pregnancy loss, increasing aneuploidies, or subfecundity. Dr. Zinaman pointed out that supersensitive hCG assays may be more likely to give false positives. Dr. Buck said the real issue concerns limits of detection and weights of criteria. She said that serial hCG measures might be a more reliable choice. Dr. Zinaman reiterated Dr. Campbell's suggestion that samples should be frozen and said that new tests will eventually be developed

that will allow for better analysis. He added that substrates tested must be stable. Dr. Schrader asked if single sperm samples would be considered for the Study. There was general concern that subjects would refuse such analyses, and Dr. Joffe said that many of his European colleagues found it difficult to find subjects for such studies. Dr. Schrader said his couple studies were very successful in finding participants, but samples were collected in plastic condoms that were used during intercourse. Sperm counts tend to stay high or low over time.

Existing Data and Analyses: Presentations of Existing Studies

On the second day of this workshop, speakers were invited to formally present information on existing studies (for which they either served as principal investigators or performed analyses) of DSPs of conception. For each study, the speaker was asked to address the following key questions, to the extent applicable:

- What is the primary purpose of the study? Describe hypotheses and research questions.
- What is the target population and how was the sample selected? Describe the sampling framework.
- What is the unit of analysis (for example, couples, women, cycles)?
- What analyses have been done regarding DSPs? What statistical models were used? What were the main results?
- What future analyses are planned or possible with the data regarding DSPs of conception?
- What are the main methodological issues that have been identified with regard to DSPs of conception?

NIEHS North Carolina Early Pregnancy Study

David B. Dunson, Ph.D., NIEHS, NIH, DHHS

Dr. Dunson began his presentation by describing the focus of the workshop activities:

- To gather the best insights from the workshop participants
- To identify the best procedures for collecting data on a preconception cohort

Dr. Dunson then reviewed the NIEHS North Carolina Early Pregnancy Study (EPS). The purpose of this epidemiologic study was to:

- Develop methods to study incidence of early pregnancy loss
- Estimate incidence of early pregnancy loss
- Assess basic risk factors (environmental, biological, behavioral, and so on).

According to Dr. Dunson, the study design was motivated by efficiency considerations in studying pregnancy loss. He noted that describing EPS as a "validity" study would be inaccurate because there is no standard against which to assess the measures.

Dr. Dunson characterized the study's participants and recruitment as follows:

- Women were recruited through newspapers, clinics, and other means (convenience sample).
- Women were given initial information, with sign-up occurring later when couples committed to attempting to conceive.

- The study purposely selected highly motivated women because of data collection demands.
- There was no real sampling frame other than recruiting pregnancy planners.

Dr. Dunson commented that the fecundability and incidence of miscarriage in the EPS women were the same as those generally reported in the literature, supporting the assumption that the sample was not skewed with regard to biological characteristics. One-third of the women had never been pregnant, and so they were unselected with regard to previous outcome.

Dr. Dunson characterized enrollment and data collection as follows:
- Women were enrolled before going off birth control.
- Current attempters were not eligible because they were considered to be potentially less fertile.
- Women were given an initial questionnaire, with follow-up questionnaires at 8 weeks and at 3, 6, and 12 months.
- From 1982 to 1985, 221 healthy women were enrolled in EPS.
- Women were followed prospectively for 6 months or until conceiving a self-identified pregnancy.
- Women provided daily urine samples and stored them in freezers at home; samples were collected weekly for subsequent analysis.
- Women completed daily diaries for menstrual bleeding and intercourse.
- All women who became pregnant were later contacted about pregnancy outcome.
- Women who did not conceive were contacted and followed for up to 2 years.
- Clinically recognized pregnancies that occurred after 6 months were recorded.

Daily urine samples were assayed for progesterone and estrogen metabolites. Ovulation was estimated for each menstrual cycle using the Baird et al. (1991) approach. Pregnancies were detected using a highly sensitive and specific hCG assay, which was a polyclonal assay from one rabbit that gave an especially good response to the antigen. Once the rabbit died, the source was gone. The highly sensitive hCG assay resulted in a high-quality study. Although no subsequent study has had the benefit of this particular assay, subsequent monoclonal assays are approaching the sensitivity and specificity of this particular assay.

Dr. Dunson summarized the EPS data:
- Urines were more than 97 percent complete.
- Of the 199 conceptions, 48 ended in early loss (within 6 weeks of last menstrual period [LMP]).
- There were 699 menstrual cycles recorded from 219 women.
- Out of these cycles, there were 192 conceptions.

EPS produced the following main results:
- The rate of early pregnancy loss was estimated as 24 percent.
- There was increased TTP with prenatal exposure to caffeine.
- There was increased TTP with prenatal exposure to smoking.
- Early pregnancy loss was found to have seasonal variation.

- The sex of the baby was associated with follicular phase length.
- Day-specific conception probabilities were estimated.
- Postovulatory aging of oocyte increased early loss.
- Timing of implantation after ovulation was determined.
- There was variability in the timing of the fertile interval.
- Biological factors promoted intercourse during fertile days.

Highlights of the EPS statistical analyses include:
- EPS created a major test database for development of statistical methods.
- Early results in the late 1980s and early 1990s were based on simple TTP analyses.
- Researchers realized that it was important to analyze day-specific conception probabilities.
- The study stimulated considerable methods research, including improvements in modeling.

The major results of EPS have been published. The main focus at this time is on using the database for further methods development. A remaining issue is establishing methods to analyze hormonal data.

Dr. Dunson concluded by listing the main methodological issues to analyze complex hormonal profiles (daily hormone values):
- The data structure is nonstandard and existing software cannot be used.
- Classical nonlinear models are weakly identified and highly nonstable.
- There is within-couple correlation.
- There are high dimensional day-specific covariates.

Discussion. Dr. Dunson commented that the EPS dataset did not include observations on cervical mucus. In response to a question about whether clumping of intercourse implies knowledge of natural family planning (NFP), Dr. Dunson explained that studies that excluded NFP-knowledgeable women produced similar results. Dr. Schrader asked about EPS's excellent compliance with urine collections. EPS paid women for their urine samples, Dr. Dunson said. Although the payments were modest, women believed that payment indicated the importance of the study. In response to a question about whether some women conceived more than once, Dr. Dunson answered "yes." Early loss with later live births yields a higher subclinical pregnancy rate. Because early pregnancy loss is a marker of fecundity/fertility, there may be a higher rate of early loss in highly fertile women. Dr. Dunson mentioned a postovulatory effect in which an egg "sits around" prior to fertilization when intercourse occurs on the day of ovulation. This may result in a higher rate of early pregnancy loss.

Multinational Fecundability Study

Bruno Scarpa, Ph.D., University of Pavia, Italy

Dr. Scarpa explained that the primary purpose of the Multinational Fecundability Study (MFS) was to:
- Predict the fertile phase in a woman's menstrual cycle using BBT and cervical mucus symptom (CMS)

- Characterize the CMS marker and determine whether its relationship with intercourse behavior could be used to identify levels of daily fecundability.

Dr. Scarpa noted that because previous studies relating CMS information to fecundability were limited by underreporting of intercourse, and because detailed information on the timing of intercourse relative to a marker of ovulation was lacking, there was a clear need for establishing a new, more reliable database.

From 1992 to 1996, 782 women were recruited from seven European centers that provided services on fertility awareness and natural family planning (NFP). The sample for this prospective cohort study was determined by the following entry criteria:

- Women must be experienced in the use of NFP.
- Women must be married or in a stable relationship.
- Women must be between 18 and 40 years of age at admission.
- Women had at least one menses after cessation of breastfeeding or after delivery (or miscarriage).
- Women were not taking hormonal medication or drugs affecting fertility.
- Neither partner could be permanently infertile; both must be free from any illness that might cause subfertility.
- It was strictly required that couples did not have the habit of mixing unprotected with protected intercourse.
- Women were excluded if any one of these criteria was not fulfilled.

Another 99 women were included retrospectively from a prospective investigation carried out in Auckland, New Zealand, from 1979 to 1985. This study recruited from couples of proven fertility who were contemplating a further pregnancy. Because the study design restricted the couples to only one act of intercourse during the fertile phase of the cycle, subjects frequently dropped out of the study if they had not achieved a pregnancy after three or four cycles of trying. The resulting short observational period of sexually active nonconception cycles is a plausible source of positive bias in the estimate of the level of daily fecundability. Although the Auckland data are of significant value to other aspects of the study, only results from the seven European centers were used to determine daily probabilities of conception.

At entry into the MFS, the following information was collected:

- Month and year of birth of the woman and of her partner
- Number of previous pregnancies, if any
- Date of the last delivery (or miscarriage) and of the end of breastfeeding, if relevant
- Date of last contraceptive pill taken, if relevant
- Date of marriage and sex of the baby born (only after the collection of data had begun).

In each menstrual cycle, the subject was asked to chart the following information:

- Days of menses
- Any disturbances (such as illness, broken sleep)
- BBT until clear detection of BBT rise

- Daily CMS
- Every daily act of intercourse together with specification whether it was unprotected or protected (barrier methods, withdrawal, and so on).

In the analysis of DSP of conception in the MFS:
- Cycles in which even a single act of protected intercourse or a simple genital contact occurred were excluded from the analysis
- The "three over six rule" was used to determine the BBT shift.
- CMS was coded by women and instructors into one of five categories (as described in Dr. Scarpa's prior presentation).
- A *menstrual cycle* was defined as the interval from the beginning of one period of vaginal bleeding until the commencement of the next, when day one was the first day of fresh red bleeding, excluding any preceding day with spotting.
- A *pregnancy* was assumed in the presence of amenorrhea continuing at 60 days from the onset of the last menses, or when, before that term, a miscarriage was clinically detected.

Dr. Scarpa presented the detailed compiled data that were collected from the eight MFS centers and listed the nine published analyses that were performed on the data. He described the following methodological issues:
- Study design problems included:
  - BBT codification: 3 over 6, hand management of the paper versus an automatic algorithm
  - Mucus codification: different coding between centers
  - Strong participation of the couples based on the purposes of the study
  - First period of data collection (selection of a nonrandom sample).
- Data analysis problems included:
  - Small numbers of cycles and of conceptions in the data had incomplete data
  - Schwartz (and other) models need much large numbers of observations, which limited the number of covariates that could be analyzed in reference to DSP of conception.

Discussion. A participant asked whether it took longer to conceive in some of the studies. Dr. Scarpa replied that it had taken longer to conceive in some studies, because although some couples use NFP, they may not be fully intent upon achieving conception.

Creighton Model Multicenter Fecundity Study (USA)

Joseph B. Stanford, M.D., M.S.P.H., University of Utah

Dr. Stanford said that the purpose of the Creighton Model Multicenter Fecundity Study was to:
- Examine DSPs of conception (clinical pregnancy) by vaginal mucus observations
- Use of standardized vaginal discharge monitoring (Creighton Model Fertility Care System [CrM])
- Comparison of DSP of conception for normal fertility and known subfertility.

For the purposes of fecundity studies, vaginal discharge:
- Provides good correlation with ovulation

- Prospectively indicates ovulation
- Gives information about sperm survival
- Correlates with ovarian function
- Correlates with endocervical mucus sampling
- Is easily learned by women
- Is inexpensive.

According to Dr. Stanford, the CrM vaginal discharge recording system can be taught over several sessions with a picture dictionary and provides standardized descriptions of:
- Bleeding
- Mucus stretch
- Mucus color
- Lubrication (presence or absence)
- Coitus or genital contact.

The Creighton Model Multicenter Fecundity Study was a retrospective cohort study of women from six U.S. CrM centers. The study included women with subfertility, used record abstraction, and had DSPs determined by the Dunson model (2001). Dr. Stanford characterized the sample as women (couples) who were:
- 18–39 years of age
- Not breastfeeding
- Not pregnant
- Sexually active
- Attended at least four follow-ups (2 months)
- Began use of CrM between 1990 and 1996.

The study results were based on information from 426 eligible couples. Of these couples, 309 had apparently normal fecundity, with 1,681 eligible cycles (579 with intercourse during fecund interval) and 81 conceptions. Subfecundity was determined in 117 couples who had 373 eligible cycles and 30 conceptions. For all women, Dr. Stanford listed demographic and reproductive characteristics. Probability of pregnancy and CrM charting (intercourse day and mucus peak) were compared between normal fertility and subfertility couples.

The quality of mucus discharge was determined using the mucus cycle score of Hilgers, with the following characteristics:
- The score is based on color, stretch, and lubrication.
- Daily scores range from 0 to 16.
- Cycle score is mean of peak day and previous 5 days (0–16).
- Score correlates with serum estrogen and follicular size on ultrasound (Hilgers).

Dr. Stanford provided the following study conclusions:
- Vaginal discharge observations identify the fecund window.
- These observations apply for couples of both normal fecundity and subfecundity.
- Peak day provides a good reference point for the fecund window.

- Vaginal discharge quality of cycle correlates with cycle fecundity in the absence of couple subfertility.

Study of Time to Pregnancy in Normal Fertility

Joseph B. Stanford, M.D., M.S.P.H., University of Utah

The purpose of this study was to examine TTP in couples of proven fertility, with and without explicit awareness of the fecund window. The study design was a randomized trial of the Creighton Model Fertility Care System (CrM) to conceive. All couples used a modified (blinded) version of the ClearPlan (ClearBlue) Easy Fertility Monitor to independently identify the time of ovulation (LH). The current enrollment is approximately 50 women, with a target enrollment of 200.

Dr. Stanford listed the methodological issues for this study:
- Obtaining good daily diaries from women cannot be taken for granted.
- Many couples do not want to wait to try to conceive, even 1–2 months.
- Compensation schemes can affect outcomes.
- The study had to abandon a pay-by-cycle scheme when it became clear that some women were delaying conception to increase their remuneration in the study.
- There are a number of technical details with the ClearPlan Fertility Monitor.

Discussion. A brief initial discussion focused on:
- Hormonal data from single cycle charts
- Patterns of intercourse activity
- Fertility awareness affecting sexual behavior.

Dr. Stanford explained that this study paid participants $140, asked them to abstain from intercourse or use nonhormonal contraception for 1 month (during the instruction period), and informed them about fertile days.

In response to a question about the Creighton Model Multicenter Fecundity Study (immediately prior presentation), Dr. Stanford said that a "major puzzle" of the study results was the similarity of mucus scores for women of higher and lower fecundity; that basically, there were no differences in scores for fertile and infertile women. A participant asked whether the couples did the scoring. Dr. Stanford said that the couples charted the observations but that the study investigators assigned numerical scores based on the charts. A participant noted that men can be involved in instruction sessions (with up to 50 percent participation) and can assist with charting observations.

The LIFE Study: Longitudinal Investigation of Fertility and the Environment

Germaine M. Buck, Ph.D., M.S., Division of Epidemiology, Statistics, and Prevention Research, NICHD, NIH, DHHS

Dr. Buck listed the participating agencies and organizations for the soon-to-be-implemented Longitudinal Investigation of Fertility and the Environment (LIFE) Study. She stated that the purpose of the LIFE Study is to assess the potential reproductive and developmental toxicity of persistent environmental chemicals in the context of lifestyle among couples at risk for pregnancy.

LIFE Study exposures will include:
- Chemicals
  - PCBs, PFOA, PFOS, pesticides
  - Heavy metals
  - Phytoestrogens
- Lifestyle
  - Smoking
  - Alcohol
  - Caffeine
  - Exercise
  - Vitamins
  - Stress.

There are five LIFE Study hypotheses:
- Exposure will increase TTP, infertility, and pregnancy loss.
- Exposure will decrease gestation time and birthweight.

Eligibility criteria include:
- Women must be 18–40 years of age.
- Women must be planning pregnancy or sexually active for < 2 months without contraception.
- Neither partner is sterile as diagnosed by a physician.

The LIFE Study will recruit study participants using fishing license registries in three states (Maryland, North Carolina, and Texas). Assumptions concerning this study cohort include:
- Representative sampling
- Lower socioeconomic status (including Whites)
- Underserved populations such as Hispanics, African Americans, Vietnamese, and other minorities
- Higher exposure levels.

Baseline data collection will include an interview and collection of biospecimens including blood, urine, semen, and saliva. During the time of attempting pregnancy, data collection will include daily journals; semen, urine, and saliva; fertility monitors; and pregnancy tests. In addition to intercourse activity, data on ejaculations will be collected. Study investigators hope to capture information on critical exposures around the time of conception. At hCG pregnancy or after 12 months, data collection will include monthly journals and urine specimens. Birth data will include gestation and birth size. The study participants will conduct home pregnancy tests

(ClearBlue Easy Fertility Monitor, measuring E3G and LH) and maintain diaries (responses to specific questions).

The LIFE Study will begin enrollment in September 2004 and end enrollment in September 2006, with all pregnant women delivering by June 2007.

Discussion. In response to Dr. Dunson's question about how the LIFE Study will recruit couples, Dr. Buck explained that study investigators will review fishing license registry for age-based eligibility (both men and women), send a letter (with a reply postcard) to candidates explaining the study, and follow up with telephone calls. Dr. Buck commented on the study's intent to include participants who may be infertile, who are currently using contraceptives, and who are planning to conceive. Study investigators would like to capture information on chemical exposures and lifestyle issues during the 2-month interval before couples say they are trying to conceive. The study will, for example, measure PCB levels in blood, urine, saliva, and semen. Diary data will include information on smoking. The second semen specimen will be to specifically look for the presence of toxicants. The study will not explicitly ask about the use of illicit drugs.

The Georgetown Study

Sherry G. Selevan, Ph.D., Office of Research and Development, EPA

On behalf of Dr. Zinaman (the Georgetown Study's principal investigator), Dr. Selevan gave an informal presentation of the Georgetown Study and asked participants whether such a study would be appropriate for assessing DSP of conception. According to Dr. Selevan, this study focused on semen quality, not on women's fertility.

The purpose of the study was to perform an assessment of semen quality. Couples were recruited, and women were excluded if they had recognized subfertility. Study participants were highly motivated and predominantly White, higher socioeconomic status, educated, and from the Washington, DC, area. Study participants were actively followed for 3 months, with follow-up phone calls to 1 year.

Data were collected on women's histories of menstruation and ovulation and included one postcoital sperm in cervical mucus test on "the morning after." Probably fertile periods were identified using calendar calculations of ovulation, and couples were instructed to abstain from sex for 2 days prior to semen collection and when to have intercourse. Semen samples were collected for 2–3 cycles, with one sample collected at home and frozen, plus one sample collected at the clinic to allow computer assisted semen analysis. Daily urine specimens were collected from the women. If a woman's menstruation was delayed, she came to the clinic for a pregnancy test.

Study participants were actively followed for three menstrual cycles, and follow-up information was collected through telephone interviews up to 1 year, or the end of the pregnancy, whichever was later. Although 210 couples volunteered for the study, 200 were enrolled. Participants were

paid at protocol completion. Dr. Selevan concluded by explaining that the Georgetown study's focus was on the association of sperm with subsequent fertility of participants.

German Natural Family Planning Database

Joseph B. Stanford, M.D., M.S.P.H., University of Utah

Dr. Stanford presented on behalf of study investigators Christian Gnoth, Petra Frank-Herrmann, Erhard Godehardt, Günter Freundl, and Heinrich Heine of the University of Düsseldorf, Germany. The German Natural Family Planning Cycle Database 1982–2004 is a prospective follow-up of NFP users in Germany. The purpose of the database is:

- Research and quality control in using and teaching the symptothermal method (STM) of NFP
- Development and evaluation of STM rules
- Calculations of efficacy figures for general use, post pill, after breastfeeding, postpartum, postabortum, early menopause, and puberty
- Calculations of conception probabilities (German Fertility Survey)
- Assessment of acceptability of NFP in Germany
- NFP and management of infertility.

Dr. Stanford characterized the target population and data collection as follows:

- All couples practicing NFP either for contraception or for achieving a pregnancy were included.
- All couples were trained in the STM by educated teachers (one basic 12-hour course).
- About 900 NFP teachers in Germany asked their clients to participate in the Düsseldorf study after informed consent.
- All clients directly send their cycle charts on a monthly basis to the study center in Düsseldorf, with an automatic follow-up of at least 3 months.
- Data input in NFPDAT 1.5 is performed and supervised by a trained study coordinator.
- There is continuous supervision by a scientific committee.

Variables and features of the study include:

- Data sheet with 117 different variables (age, medical history, socio-demographic data)
- Cycle sheet with 277 different variables (temperature, intercourse, day of ovulation, mucus)
- Dropout sheet with 50 different variables (reasons for dropout, pregnancy data)
- More than 200 internal error formulas
- Scans of all stored cycles
- Report generator (logical, mathematical, and Boolean operations)
- Units of analysis: women or cycles.

As of April 2004, the study featured:

- 1,578 clients
- 34,772 cycles
- Longest continuous observation: > 120 subsequent cycles
- Five unintended pregnancies (Pearl Index [method failure]: 0.2 percent)
- 450 total documented pregnancies.

This database was used for a prospective study on daily probabilities of conception (Bremme, 1991 [thesis]). This study's methods included closest intercourse to ovulation in conception cycles/conception plus nonconception cycles. Its outcomes included clinical pregnancy (109 conception cycles). The DSP of conception in this study were similar to other studies, as summarized by Colombo (2000).

The German database was used to conduct a prospective study of TTP among couples with normal fertility and use of fertility awareness (symptothermal NFP) to conceive (Gnoth, 2003, *Human Reproduction*).

Important features of the German study on the use of NFP are:
- It is prospective (retrospective studies are biased because of exclusion of infertile couples; selection bias in determining the age dependence of waiting TTP [Jenssen, 2000]).
- All women were observed from their first cycle onward attempting a pregnancy.

The deadline for data inclusion for the German prospective NFP study was March 31, 2001. The database contains information on:
- 346 women trying to achieve a pregnancy
- 310 pregnancies
- 36 infertile women (10.4 percent).

Dr. Gnoth and colleagues provided the following conclusions for the prospective German TTP study:
- Most of the couples conceived early (80 percent in the first six cycles) by timed intercourse. Their cumulative probabilities of conception decreased slightly with age, but this was not statistically significant.
- Of all couples who eventually conceived over long-term follow-up, 90 percent of couples conceived in the first six cycles. Their cumulative probabilities of conception are probably age-independent.
- Subfertile couples should be investigated after six unsuccessful cycles for infertility, rather than waiting until 12 months of trying to conceive, regardless of age.
- Couples with good prognosis after investigation (especially unexplained infertility) may be advised to wait because they still have a reasonable chance (> 60 percent) to conceive in the next 36 months.
- In this study, there was no significant decrease of women's fertility of a woman does not decline gradually with advancing age.

According to Dr. Gnoth, areas of future research include:
- Cumulative pregnancy rates as a marker of human fertility potential (age effects, body weight, regional differences, seasonal differences, and so on)
- Management of infertility (role of NFP, early workup of infertile patients, self-diagnosis, self-treatment, treatment independent pregnancy rates in NFP users, and so on)

- New technologies for assessment of human fertility (mobile computing of symptoms of self-observation and calculation of daily conception probabilities).

Discussion. Participants expressed concerns with the following issues of the German NFP study:
- Stratification on women who eventually got pregnant, from 6 months to 3 years
- Imprecise definition of "infertile"
- Use of paper charts (but computerized database)
- Interpretation and analysis of data.

Dr. Buck commented that a woman's past reproductive history and biological data are important and need to be considered in fertility studies. Hanna Klaus, M.D., National Family Planning Center of Washington, DC, noted that pregnancy affects the squamous epithelium of the cervix.

Intercourse Data from WHO Study 90905 (New Zealand, Australia, and Chile)

Joseph B. Stanford, M.D., M.S.P.H., University of Utah

Dr. Stanford explained that this presentation, made on behalf of Len Blackwell, Ph.D., describes some of the available data on intercourse in relation to hormone levels recorded during WHO Study 90905 on the ovarian monitor (OM). The dates of the study are approximately 1997–1998, whereas the other studies may be older. The primary purpose of Study 90905 was to:
- Use of a simple home device for measuring E1G and PdG, urinary metabolites of estradiol and progesterone
- Determine the phases of potential fertility and infertility in the human menstrual cycle
- Compare the results with the NFP symptoms using OM and symptothermal methods.

The hypothesis of the study was that the women could determine:
- The beginning of their fertile period by the first E1G rise
- The most fertile day from the E1G peak day
- The end of the fertile period from the PdG cut-off value.

The sample population was 20 women from 3 centers (Santiago, Chile; Sydney, Australia, Palmerston North [PN], New Zealand). All women were using recognized methods of NFP; all women were considered fertile, as judged by previous pregnancy and at least three previous regular menstrual cycles; and all women were intending to avoid a pregnancy during the study. The sample was recruited from NFP clinics and by advertising in church newspapers and notice boards. The focus of analysis was the daily urinary steroid glucuronide values in a cycle.

DSPs of conception were not calculated because this was not an objective of this study. The intention was to examine intercourse patterns and how they related to the women's perceptions of their fertility. The numbers of acts of intercourse were recorded across three phases of the menstrual cycle:
- Preovulatory
- Fertile period
- Postovulatory.

For the purpose of this study, the beginning of fertile period was defined as beginning on LH peak day –5 and ending on the day before PdG cut-off day.

Dr. Stanford presented data on acts of intercourse, including preovulatory acts, acts using condoms, and acts without condom use.

Dr. Stanford commented that it is not clear from the analysis by the World Health Organization (WHO) and the information given in these slides that all acts of intercourse were recorded. Dr. Stanford summarized data on pregnancies and displayed histograms for the start of fertile period and for the difference between E1G peak day and LH peak day. He noted that, based on such data, it is sometimes difficult to define the beginning of the fertile period.

Agreement of LH and E1G peaks was summarized as follows:
- There is generally good agreement between LH, E1G, and mucus peaks (± 2 days).
- However, there were two cycles where LH peak was too early compared with E1G peak and subsequent PdG rise (one seriously so).
- There were 20 cycles in which LH peak was absent or too low or too early.
- All cycles showed an E1G peak.

Open Forum

The Oxford Conception Study

Cecilia Pyper, M.B., B.S., Oxford University

In this first open forum presentation, Ms. Pyper shared some highlights of the recently implemented Oxford Conception Study, which is a randomized controlled trial to determine whether daily information about potential fertility from a fertility-monitoring device will increase the conception rate in women wishing to achieve a pregnancy. It is a three-arm randomised controlled trial with two intervention arms and a control arm.

The volunteers are randomized into one of three groups; each group receives different information from the fertility monitor:
- Late fertile time group—monitor shows high fertility from the first appearance of LH and for the next 2 days, and then shows low fertility until the end of the menstrual cycle
- Early fertile time group—monitor shows high fertility from the first appearance of E3G and low fertility from the appearance of LH until the end of the menstrual cycle
- Control group—monitor reveals no information about the fertility status.

The sample size calculation is based on the primary study objective, which is to compare the cumulative three-cycle pregnancy rate between women from the late fertile time group and the control group. For 80 percent power and the ability to detect a 10 percent difference in three-cycle pregnancy rate between the "late fertile time group" and the control group, a sample size of 450 women per group (1,350 women total) will be required. A 15 percent nonpregnancy dropout

rate has been estimated. Additional women will be recruited to replace any women who are found to have been pregnant at the start of the trial (a positive pregnancy result between day 7 and day 14 of the first cycle).

Ms. Pyper described the primary research question for this study as:

- Is the pregnancy rate (assessed in terms of 3-month and 6-month cumulative pregnancy rates) in couples given information about the woman's potential fertility higher than in those that do not receive the information?
- Secondary research questions include:
  - Are the early fertile days, as identified by a fertility-monitoring device, better days for conception than the later fertile days?
  - Does information from fertility monitoring devices help couples to target sexual intercourse more effectively than no information?
  - Does monitoring the fertile time alter women's "stress" levels (hospital anxiety and depression score—HAD—score)?
  - Is there an association between the following variables on: conception rates; miscarriage; pregnancy outcome and early child development
    - Age of woman
    - Age of partner
    - Age of the gametes, (intercourse at the outer limits of the fertile time)
    - Intercourse frequency
    - Cycle length variability.

Ms. Pyper presented detailed information on the following aspects of the study:

- Feasibility of a full conception probability study
- Probability of conception
- Randomization
- Protection against bias
- Recruitment
- Screening
- Admission interview
- Frequency and duration of follow-up.

As of May 18, 2004, the study has recruited 570 volunteers. Of these 200 are active in study. None have yet been lost to follow-up. Other details include:

- 191 total pregnancies
- 31 trial miscarriages
- 35 births.

Discussion. A participant asked whether information about fertile time affects pregnancy. Ms. Pyper explained that the rationale of the fertility monitor in the control arm of the study was to have consistency across groups in an effort to collect independent data. Another participant asked about the timing of intercourse information. In response, Ms. Pyper noted that, although

there is a fertile window, the fertility-monitoring device used in this study does not identify the entire fertile window. In one arm, it identifies the "early" fertile window, in one arm the "late" fertile window, and in the third arm it gives no indication of the fertile window. The participants are asked to target intercourse around the fertile window, either early or late; but they will not have specific information on the total fertile window.

Mucus Observations and Fertility

Jamie L. Bigelow, M.S., University of North Carolina

Ms. Bigelow began her presentation by defining the *fertile window* as the time interval relative to ovulation when pregnancy is most likely to occur. The purpose of her talk was to describe how cervical mucus is related to the pregnancy probabilities on these days. Cervical mucus has a dual role in fertility:

- Marking the fertile window
- Regulating sperm survival and transport.

Other details about cervical mucus and the fertile window include:

- Pregnancies result from intercourse during a 6-day fertile window.
- Estrogen levels rise near start of the window.
- Changes in cervical mucus occur just prior to the fertile window.
- "Estrogenic" mucus tends to be stretchy and slippery, and occurs in larger quantities during the fertile window.
- First use for monitoring cervical mucus was for marking the fertile interval.
- Vaginal discharge reflects the nature of cervical mucus.
- Cervical mucus is useful in identifying most fertile days of the cycle.
- Changes in cervical mucus occur before LH surge and BBT rise.
- Cervical mucus is a water-based gel made up of mucins.

Ms. Bigelow briefly reviewed the European Study of Daily Fecundability (also reviewed above), which was a multinational prospective study of couples in seven natural family planning centers. The sample included women 18–40 years of age, who were married or in a stable relationship. The women recorded on a daily basis:

- BBT
- Intercourse
- Menstrual bleeding
- Cervical mucus symptoms based on a vaginal discharge score (1–4 scale):
  1. Dry feeling
  2. Damp feeling
  3. Thick or creamy discharge
  4. Slippery, wet, stretchy discharge.

Data were recorded for 1,473 cycles from 516 women. There were 353 pregnancies in the study.

The goal of the study's statistical analysis was to estimate pregnancy probabilities for each day in the fertile window based on mucus score. Some complicating factors were that:

- Data are not independent, but are hierarchical.
- There may be multiple acts of intercourse during the fertile window.

The statistical analysis used a Bayesian hierarchical modeling approach.

The study results indicated that cervical mucus provides important information beyond marking the fertile interval. Ms. Bigelow explained that if all mucus did was mark the fertile window, there would not be discrepancies in probabilities within the fertile window. Cervical mucus provides information about fertility even within the window. Specifically, estrogenic-type discharge is indicative of higher pregnancy probability.

Analyses of probabilities of conception were performed to test the following hypothesis: Sperm from older men are more sensitive to cervical mucus. Ms. Bigelow summarized the results of these analyses as:

- Inside the fertile window, estrogenic mucus indicates high pregnancy probability.
- Declining male fertility in late 30s may be due to lowered sperm motility, specifically in nonestrogenic mucus.

Ms. Bigelow listed the following implications for cervical mucus and fertility:

- Mucus is a useful clinical marker of days with high conception probabilities.
- Within the fertile window, estrogenic mucus predicts higher fertility.
- Mucus monitoring may be especially helpful for older couples trying to achieve pregnancy.

Discussion. Ms. Bigelow noted that the day of ovulation was estimated by BBT in this study.

Hierarchical Models for the Probabilities of Conception

Cuirong Ren, South Dakota State University

Mr. Ren began his presentation by noting the following:

- Finding the probability of pregnancy due to unprotected or protected intercourse on a particular day of a menstrual cycle is of interest to reproductive scientists.
- The earliest model is due to Barrett and Marshall (1969). The model has been extended since then.
- Zhou and Weinberg (1996) claimed that their model could be used to account for the timing of intercourse relative to ovulation, heterogeneity among fecund couples, and multiple covariates.
- Ren and colleagues modified the method of Zhou and Weinberg by modeling the ratio of the conception probability for protected intercourse to the conception probability for unprotected intercourse on day, denoted by *s*.

Mr. Ren defined the variables for and elaborated on:

- A mixed model of fecundity

- A Bayesian hierarchical model
- Full conditional distributions.

Mr. Ren then presented the following details on his analysis of California Study Data:
- 372 women remained and 31 dropped out
- 1,387 cycles "at risk" (if ≥1 unprotected or protected intercourse)
- 441 cycles with ≥1 unprotected intercourse and 54 conceptions
- 946 cycles with protected intercourse only and six conceptions
- Three covariates: TT (water contaminants), MINPDG (minimum progesterone), and MAXPDG (maximum progesterone)
- Three priors are chosen based on previous study.
- Run 40,000 iterations including 10,000 burn-in; last 30,000 were used.

Mr. Ren presented two tables that listed the posterior means and standard deviations under the three priors with the unimodality assumption. He described the details of the unimodality assumption.

Discussion. In response to a question from Dr. Dunson, Mr. Ren explained that TT is the covariate for water contaminants. He noted that day 0 was specified as the day of ovulation.

Day-Specific Pregnancy Probability Estimation in the Efficacy Trial of Spermicidal Agents

Pai-Lien Chen, Family Health International

Mr. Chen provided the following details on the Efficacy Trial of Spermicidal Agents (ETSA):
- Randomized trial (14 sites in United States)
- Five nonoxynol-9 spermicides
- Targeted 1,800 participants
- Each participant followed for 30 weeks
- Outcomes included:
  - Pregnancy
  - Acceptability
  - Safety
  - Product use.

The main objectives of ETSA were to estimate the absolute and relative risk of pregnancy over 6 months/cycles of use for five spermicide products. The study used a per-act analysis to assess the level of protection provided by the five spermicides. The analysis also estimated day-specific pregnancy probabilities. Investigators used a per-act analysis instead of per-day analysis because:
- Multiple acts of intercourse may occur on a single day (leading to multiple opportunities of pregnancy).
- Values of important covariates may vary across acts.

Mr. Chen described the targeted study sample:

- 1,800 patients total (360 per group)
- Based on financial, logistical, statistical considerations
- Statistical considerations included:
  - Precision of 6-month probability estimates within group
  - Power to detect differences between groups.

Characteristics of the target population comprised the following:
- Age 18–40
- No known infertility
- Wants to use spermicide as only method of contraception for 7 months (six cycles)
- Willing to accept a moderate risk of pregnancy
- No contraindications to pregnancy
- At low risk for sexually transmitted infections.

Pregnancy and coital data collection included:
- Pregnancy data
  - Admission visit
  - Three scheduled follow-up visits (4, 17, and 30 weeks)
  - Three home pregnancy tests (2, 10, and 23 weeks)
- Coital diaries
  - Time of intercourse
  - Method use
  - Menstrual bleeding.

From June 1998 to January 2002, 1,536 participants were enrolled in the study. These participants generated the following information:
- There were 952 (62 percent) subjects with at least one cycle.
- The subjects produced a total of 3,764 cycles.
- There were 32,235 total acts of intercourse.
- There were 28,512 total days with acts.
- There were 3,461 (12 percent) days with multiple acts.

Details on intercourse and spermicide use include:
- There was a median of eight acts of intercourse per cycle.
- Of all acts of intercourse, 91 percent used spermicide alone.
- Of all acts of intercourse, 4 percent were without protection.
- Twenty-eight percent of the cycles had at least one unprotected act of intercourse.

Data were analyzed using an extension of Schwartz-Barrett-Marshall model (Dominik, Zhou, and Cai, 2001). Modifications to the model included:
- Replace the daily pregnancy probability function with a per act probability function
- Allow a wider fertile window through parametric modeling of the day of act
- No precise benchmark for ovulation

- Estimate probability by cycle day
- Able to include covariates at the per act level
- Estimate parameters by a pseudo likelihood approach/Bayesian methods.

Mr. Chen presented study findings in two graphs:
- Probability of recognizable pregnancy by cycle day Femcap trial (using the same analytic approach in a previous study); DSP of conception referenced to first day of menses; distribution similar to that found by Wilcox et al., 2001
- Probability of recognizable pregnancy by cycle day ETSA for the different spermicide presentations; DSP of conception referenced to first day of menses.

Mr. Chen listed the following topics for future research:
- Integrate couple level and cycle level exposure variables into models
- Evaluate the impact of errors in coital activity or cycle data
- Extend the per act probability function to incorporate contraceptive effects that might last more than one act.

Overview of the National Children's Study

Sherry G. Selevan, Ph.D., Office of Research and Development, EPA

In this presentation, Dr. Selevan outlined a brief history of the Study. The idea of a national study was conceived by the President's Task Force on Environmental Health Risks and Safety Risks to Children. The Children's Health Act of 2000 (PL 106-310) authorized the Director of NICHD, together with representatives from EPA, CDC, and (later) NIEHS to:
- Plan and implement a national longitudinal study of environmental influences on children's health and development
- Investigate basic mechanisms of developmental disorders and environmental factors, both risk and protective.

In this context, *environment* is defined broadly to include physical, chemical, biological, and psychosocial factors.

The rationale for implementing the Study included the following issues:
- Increasing awareness of potential effects of environment on development and later life—life course design allows assessment of what is harmful, harmless, and helpful to child development
- Recent (and future) developments in biomarkers, biotechnology, and information technology
- No large longitudinal study of development in the United States since 1960s.

Dr. Selevan explained that the relevance of public health issues is reflected in the prevalence (number per 1,000 juvenile U.S. population) of the following conditions in children:
- Asthma 53
- Autism 2.5
- Diabetes 1.9

- Injury 225
- Obesity 130
- Schizophrenia 3

The estimated annual cost to United States from disease burden (persons < 20 years of age) for priority disease outcomes of the Study are:
- Asthma $393 billion
- Autism $117 billion
- Diabetes $98 billion
- Injury $13 billion
- Obesity $12 billion
- Schizophrenia $9 billion.

Dr. Selevan listed the following Study concepts:
- High-quality longitudinal study of children, their families, and their environment
- National in scope
- Environment defined broadly (chemical, physical, behavioral, social, and cultural factors)
- Study common range of "environmental" exposures and less common outcomes
- Environment and genetic expression
- State-of-the-art technology (tracking, measurement, data management, and so on)
- National resource for future studies.

Criteria for core hypotheses of the Study include:
- No single hypothesis
- Hypotheses are required for costly elements
- Important for child health and development (prevalence, severity, morbidity, mortality, disability, cost, public health significance)
- Scientific rationale
- Requires a large sample size (approximately 100,000 children)
- Measurable with study of this size
- Requires longitudinal follow-up.

Priority health and disease outcomes include:
- Pregnancy outcomes
- Neurodevelopment and behavior
- Childhood injury
- Asthma
- Obesity and physical development.

Priority environmental exposures and other factors include:
- Physical environment
- Chemical exposures
- Biological environment
- Psychosocial exposures

- Genetics.

Organization of planning phase of the Study involves the following individuals and groups:
- Institute and Study Director—provide overall guidance and strategic decisions
- Interagency Coordinating Committee—appointed program staff of lead agencies (HHS/NICHD, NIEHS, CDC, and EPA); responsible for operational decisions
- Program Office at NICHD—Federal staff; responsible for operations
- Chartered Federal Advisory Committee—provides advice; manage working groups
- Working Groups (22)—non-Federal and Federal scientist and investigators (approximately 250); responsible for hypotheses, design, measures, consultation and findings
- Federal Consortium—representatives of Federal agencies; provide strategic guidance
- Study Assembly—all interested parties (more than 3,000 individuals); provide diverse input.

Dr. Selevan reminded participants that the purpose of the workshop was to address the question: How does preconception enrollment fit within the Study? This approach has been recommended by a number of sources, including the Fertility and Early Pregnancy Working Group, the Birth Defects Working Group, and the Early Origins of Adult Disease Working Group. This approach was discussed and examined at the Sampling Design Workshop in March 2004. Dr. Selevan noted the following concerns for preconception enrollment:
- Costs
- Feasibility
- Selection.

Discussion. Dr. Joffe asked if the intent of the Study is to recruit 100,000 children preconception or whether a prospective cohort would be nested within the larger sample. Dr. Selevan replied that these issues are under discussion and depend on a number of variables such as funding, resources, and implementation issues. Ruth A. Brenner, M.D., M.P.H., NICHD, NIH, commented that feasibility and costs are major issues for the Study. Other concerns are internal validity and trade-offs of a preconception cohort. Dr. Selevan commented that study planners hope that at least part of the study sample will involve preconception enrollment. According to Dr. Brenner, early exposure measures are a key issue for some groups and some questions. Dr. Selevan noted that early exposure measures are the biggest driving central theme for the inclusion of preconception data. Dr. Stanford asked whether the Study would enroll potentially fertile subjects versus planners. Dr. Brenner replied that planners are different from nonplanners and that the Study would consider various groups from the broadest (most representative) to the narrowest, depending on feasibility. A participant suggested that children enrolled in the study receive a reproductive assessment prior to their leaving the Study (that is, prior to their 21st birthdays). The group discussed the difficulty, intensity of labor, and expense involved with a preconception cohort. Cost per yield in such Study efforts should be considered. A participant explained that fetal death is not an end point of the Study, but that pregnancy loss should be considered an outcome. Dr. Joffe noted that gamete quality, which could be measured indirectly, could be an important part of the Study. Dr. Joffe suggested that a smaller cohort (approximately 10,000 subjects) could be involved in a more intensive arm of the Study, while a larger cohort

(approximately 20,000) could be involved in a less intensive arm. Subjects could opt in or out of the different Study arms.

Large Group Discussion

In this final session, workshop participants were asked to discuss in "big picture, brainstorming" fashion and attempt to answer specific questions about a preconception cohort for the Study. The intent was to summarize this discussion as workshop findings for consideration by the Study.

The discussion during the last session of the meeting covered practical aspects in the collection of these data in a longitudinal study. The following describes four major themes discussed by the panel and suggestions made for each theme: (1) differences between pregnancy planners and non-planners; (2) a "repeat pregnancy" approach to obtaining data on early pregnancy; (3) a tiered approach to data collection; and (4) preconception data gaps.

Differences Between Pregnancy Planners and Non-Planners. The challenges in collecting data before conception include possible difficulties in recruiting women (and their partners) prior to pregnancy, and retaining them long-term (Buck et al., 2004). The differences in recruitment and retention could be most marked when comparing those planning pregnancy to those who might become pregnant due to contraceptive failure, for example. To be able to truly understand relationships between exposures and outcomes, both groups should be evaluated. The two groups are likely to have different levels of motivation for participating in the study, especially with the level of effort for participation. Follow-up of pregnancy planners would result in a higher rate of pregnancy than non-planners. Those potentially at risk, but not planning pregnancy are less likely to become pregnant, so additional follow-up can add to costs for the study. However, the group felt that these costs would be justified for the scientific information returned, at least for a subgroup of the study.

Some of the potential differences between planners and non-planners include changes in behavior for those who are planning pregnancy (for example, vitamin usage and personal habits such as cigarette smoking and alcohol consumption). However, the panel was unaware of any empirical data supporting differences in chemical exposure profiles between couples who do, or do not, plan pregnancies. This remains a critical data gap, one relevant for the interpretation of findings. This discussion also raised the concern that some particular subgroups might have different cultural norms that might affect participation and data collection before conception or during early pregnancy, which need to be addressed. In addition to recruitment issues, the potential for differential rates of retention should be assessed to evaluate whether bias may be introduced.

Repeat Pregnancy Approach to Obtaining Data on Early Pregnancy. Another topic discussed was the potential for using "proxy" data to try to assess exposures early in pregnancy. In this context, proxy data referred to collection of early pregnancy information on subsequent pregnancies to women who joined the study during an earlier pregnancy. The preconception and very early pregnancy exposures assessed in the subsequent pregnancies could then serve as

"proxy" exposure data for the initial pregnancy that did not have preconception and very early pregnancy data. This is probably a simpler approach than initial preconception recruitment, since they have already been involved in the study and probably would be more willing to add collection of preconception data to their participation. The workshop members were concerned that this might be the only approach used in the Study, and emphasized that it is not possible to get the whole picture using this approach. The key concern is that this approach is restricted to fecund women who have already achieved at least one pregnancy. The first obvious problem is that none of these subsequent pregnancies could, by definition, be first pregnancies, potentially biasing the data by excluding this portion of the reproductive continuum. Second, exposure differences may occur over time with either short or long half life chemicals; in addition, for chemicals with long half-lives (persistent chemicals), greater age potentially means a greater body burden. Alternatively, a mother who breastfeeds would have a lower body burden for the second pregnancy in the study than for the first (LaKind et al., 2001). Moreover, statistical analysis of exposures on pregnancy outcomes would need to address issues pertaining to the lack of independence of pregnancies within a given woman.

Tiered Approach to Data Collection. The workshop panel suggested a tiered approach to allow data collection for all women at-risk for pregnancy. An "intensive" preconception tier for those planning pregnancy would involve techniques commonly used in "day-specific probability of conception" studies, including using a fertility monitor, a daily diary, or other methods. This approach would be practical since the majority of them are likely to be pregnant within 3 months; the workshop suggested following women at the intensive level for up to 6 months. Monitors can be programmed to collect the specific measurements of interest for the Study, can hold up to 3 months of data, and ultimately can be cleaned and reprogrammed for use by other participants. A less intensive "observational preconception" tier would include those women not planning pregnancy but still at appreciable risk for pregnancy, for example, those using contraception methods with appreciable failure rates (FDA, 1997). These women would be given pregnancy test kits and followed up with periodic contacts from the Study every few months. Whenever a participant uses a pregnancy test kit, she can request additional ones from the Study. For both tiers, baseline data and samples would be collected on exposures, health history, and so on with updates when pregnancy is detected.

Preconception Data Gaps. Finally, the workshop participants discussed the data collection recommended for women or couples recruited prior to conception. The following table includes all items for consideration.

	Female	Male
At enrollment	Intensive Tier ▪ Questionnaire ▪ Serum ▪ Urine ▪ Saliva ▪ Environmental samples	Intensive Tier ▪ Questionnaire ▪ Serum ▪ Urine ▪ Saliva ▪ Semen ▪ Environmental samples
	Observational Tier ▪ Questionnaire ▪ Serum ▪ Urine ▪ Saliva ▪ Environmental samples	Observational Tier ▪ Questionnaire ▪ Serum ▪ Urine ▪ Saliva ▪ Semen ▪ Environmental samples
Preconception	Intensive Tier ▪ Daily diary (dietary, intercourse, personal habits, menstrual cycle data) ▪ Serum ▪ Urine ▪ Fertility monitor (daily urine testing) ▪ Pregnancy test kits ▪ Mucus observation ▪ Menstrual flow observation ▪ Changes in environmental and occupational exposures	Intensive Tier ▪ Daily diary ▪ Serum ▪ Urine ▪ Semen ▪ Changes in environmental and occupational exposures
	Observational Tier ▪ Daily diary (dietary, contraception, intercourse, personal habits, menstrual cycle data) ▪ Urine ▪ Pregnancy test kits ▪ Changes in environmental and occupational exposures	Observational Tier ▪ Daily diary ▪ Urine

At pregnancy	Intensive Tier ▪ Daily or weekly pregnancy journal ▪ Updated questionnaire ▪ Serum ▪ Urine ▪ Saliva ▪ Environmental samples	Intensive Tier ▪ Updated questionnaires ▪ Serum ▪ Semen ▪ Urine ▪ Environmental samples
	Observational Tier ▪ Monthly pregnancy journal ▪ Updated questionnaire ▪ Serum ▪ Urine ▪ Environmental samples	Observational Tier ▪ Updated questionnaire ▪ Serum ▪ Urine ▪ Semen ▪ Environmental samples

References:

Buck GM, Lynch CD, Stanford JB, Sweeney AM, Schieve LA, Rockett JC, Selevan SG, Schrader SM. Prospective pregnancy study designs for assessing reproductive and developmental toxicants. *Environ Health Perspect* 2004 112(1):79-86.

FDA, 1997. http://www.fda.gov/fdac/features/1997/babyguide2.pdf

LaKind JS, Berlin CM, Naiman DQ. Infant Exposure to Chemicals in Breast Milk in the United States: What We Need to Learn From a Breast Milk Monitoring Program. *Environ Health Perspect* 109:75–88 (2001).

Participants

Jamie L. Bigelow, M.S., University of North Carolina, Chapel Hill
Ruth A. Brenner, M.D., M.P.H., NICHD, NIH, DHHS
Germaine M. Buck, Ph.D., M.S., NICHD, NIH, DHHS
Kenneth L. Campbell, Ph.D., University of Massachusetts, Boston
Pau-Lien Chen, Ph.D., Institute for Family Health
Vanja Dukic, Ph.D., University of Chicago
David B. Dunson, Ph.D., NIEHS, NIH, DHHS
Jodi Anne Flaws, Ph.D., University of Maryland
Warren Galke, Ph.D., NICHD, NIH, DHHS
Patrick J. Heagerty, Ph.D., University of Washington
Joseph Hogan, Sc.D., Brown University
Darryl J. Holman, Ph.D., University of Washington
Leila W. Jackson, Ph.D., M.P.H., NICHD, NIH, DHHS

Michael Joffe, M.D., Ph.D., Imperial College London
Michele Kiely, Dr.P.H., NICHD, NIH, DHHS
Hanna Klaus, M.D., FACOG, National Family Planning Center of Washington, DC
Aiyi Liu, Ph.D., NICHD, NIH, DHHS
Courtney D. Lynch, Ph.D., M.P.H., NICHD, NIH, DHHS
Christopher McClure, M.P.H., RTI International
Rafael T. Mikolajczyk, M.D., University of Bielefeld, Germany
Cecilia Pyper, M.B., University of Oxford
Cuirong Ren, South Dakota State University
Bruno Scarpa, Ph.D., University of Pavia, Italy
Enrique Schisterman, Ph.D., NICHD, NIH, DHHS
Steven M. Schrader, Ph.D., National Institute for Occupational Safety and Health, CDC, DHHS
Sherry G. Selevan, Ph.D., Office of Research and Development, EPA
Irit Sinai, Ph.D., Georgetown University
Joseph B. Stanford, M.D., M.S.P.H., University of Utah
Anne M. Sweeney, Ph.D., Texas A&M University System
Anjel Vahratian, Ph.D., NICHD, NIH, DHHS
Haibo Zhou, Ph.D., University of North Carolina, Chapel Hill
Michael Zinaman, M.D., Loyola University of Chicago