# Thongs, Flip-flops, and Unintended Pregnancy: The Seduction of $P < 0.05$

Ruchi P. Moore, MD, Shelley L. Galvin, MA, Hytham M. Imseis, MD
Mountain Area Health Education Center, Department of Obstetrics and Gynecology Asheville, North Carolina

**Objective**: The objective of this project was to demonstrate the dangers of interpreting statistically significant findings from multiple analyses of databases (aka "data mining").
**Study Design**: A cross-sectional survey of personal preferences and obstetric outcomes was administered to consecutive patients presenting for routine obstetrical or postpartum care. Associations of preferences with primary outcomes including desire to be pregnant, route of delivery, and satisfaction with medical care were analyzed using Chi square, Spearman rho, and ANOVA, with $p < 0.05$ considered statistically significant.
**Results**: One hundred twenty-one women completed surveys for a response rate of 70%. We identified 26 significant associations from the 362 statistical analyses conducted; 18 were expected to be spurious based on $p < 0.05$. Unintended pregnancy was associated with preferences for scant clothing, brief communication, and sensual and poorly lit recreational activities.
**Conclusions**: Differentiating 8 meaningful relationships from the spurious findings revealed our biases and highlighted the true meaning of $p < 0.05$.

Key Words: Statistics, study design, data mining, resident education

## Introduction

All too often, research conclusions, like many rational decisions in obstetrics and gynecology (and in life itself), are based on probability. In medical practice weighing the risks of bad outcomes versus the chances of good outcomes is at the core of all interventions available in our healthcare repertoire: counseling patients, obtaining medical tests, performing procedures, and administering medical therapy. In performing research and drawing research conclusions, the risks inherent in accepting significant findings when $p < 0.05$ are not well understood and may contribute to the promulgation of misinformation. This in turn may feed belief in the old adage about "…lies, damned lies and statistics."

With acceptance of the conventional $p < 0.05$ for concluding statistical significance, it is tempting to ascribe importance to any statement accompanied by a $p < 0.05$. But it is important to remember that for every 100 comparisons or analyses completed, five Type I errors will be made. This means that 5 times out of 100 we will conclude that we have an important finding when in fact what we have found is spurious or due to random chance (false-positive result).

While the advent of electronic databases and statistical analysis software has made research less labor-intensive, they have propagated research based on mining datasets for statistical relationships. Using the multitude of variables within a dataset for data mining often yields to the identification of many spurious relationships. The concept of spurious relationships spawned from data mining has been discussed and exemplified in editorials[1] and tutorials[2-4] as well as demonstrated via the relationships of astrological signs to hospitalizations for gastrointestinal hemorrhage and humeral fractures.[5]

Outside of research circles, practicing physicians and resident physicians generally view statistics with great trepidation. In order to break down some of the anxiety generated by the concepts of statistics, we adopted an irreverent, tongue-in-cheek, case-based approach to teaching

these concepts to obstetrics and gynecology residents. This interactive approach has been very effective. Hence, this project was designed to demonstrate the seductive lure of $p < 0.05$ within the context of a resident research study analyzing the relationship between maternal personal preferences with obstetrical outcomes and with satisfaction with medical care.

## Materials and Methods

A cross-sectional design was used to survey consecutive pregnant and postpartum patients seeking care at an obstetrics and gynecology residency program clinic. Research packets consisting of an information letter, survey, and envelopes were offered to patients in exam rooms at either their 35 – 37 week prenatal visit or their 6-week postpartum visit. The survey asked for information on socio-demographics, the current or immediate postpartum pregnancy, delivery outcomes of postpartum women, and 13 healthcare satisfaction items rated on a 5-point Likert scale (not satisfied to very satisfied). Also included was a 41-item "Meyers-Briggs-esque" preference inventory;[6] a series of simplistic "either-or" preference questions fell into six categories: food, attire, activities, health and beauty, gender, and miscellaneous (e.g., plain or peanut M&Ms, panties or thongs, etc.)

Patients who chose to do so completed surveys voluntarily and confidentially and returned them sealed in envelopes to either their nurse or the front desk at check out. This project was exempted by the IRB of our local affiliated hospital.

We presented patients preferences and outcomes in frequencies and percentages. Preference selections could be skipped (if so desired by the participant) so frequencies did not always sum to 121; however, percentages were calculated on actual responses, and therefore, equal 100%, except where rounded up. The 13 satisfaction questions were averaged into one overall satisfaction with healthcare variable.

We examined relationships between the 41 preferences and 7 outcomes. These outcomes included intention to be pregnant at this time, pregnancy complications, gender of neonate, preferred feeding method for newborn, satisfaction with care, route of delivery and birth weight. When few patients ($< 15\%$) endorsed a specific preference, we excluded the pair from the analyses; we were more interested in the differences between patients than the similarities. Statistical analyses included Chi square or Mann-Whitney for categorical outcomes. Spearman correlations were conducted to look for relationships within categories of preferences, and we conducted a one-way analysis of variance (ANOVA) for each of the unrelated preferences and higher order ANOVAs for related preferences with birth weight as the dependent variable.

## Results

One hundred twenty-one women returned completed surveys for a response rate of 70% (121/179). Of these, 72 (65%) were postpartum [three sets of twins (2%); 54 (75%) term vs. 18 (25%) preterm] and 49 (35%) were pregnant [median weeks gestation = 37 (range, 35-40)]. Maternal age ranged from 13 to 40 years with a mean age of 22.8 ($SD = 5.2$) years. Most of the women were Caucasian,($n = 93$, 78%) and 19 (16%) were African American, 7 (6%) were Hispanic and 2 (2%) were "other." The majority had other children 92 (86%). Almost all were Medicaid recipients 109 (91%); 5 (4%) had private insurance, 3 (3%) had a combination of Medicaid and private insurance, and 4 (3%) were self-pay.

## Pregnancy Variables

With regard to intention to be pregnant at this time, only 36 (29%) indicated this was a good time to be pregnant. We termed these as "intended" pregnancies. Seventy-one (59%) would have preferred to have waited longer before getting pregnant, 8 (7%) never planned to be pregnant and 6 (5%) were not sure; these three response categories were added together and labeled "unintended pregnancies."[7]

Overall, 22 (18%) reported their pregnancy had been complicated by a medical diagnosis. These included diabetes [4 (3%) pre-gestational and 5 (4%) gestational], and hypertension [5 (4%) pre-gestational and 11 (10%) gestational], and pre-eclampsia [7 (7%)].

One hundred six women reported the sex of their 108 infants; 14 mothers (12%) indicated they did not know or left the question blank. Slightly more infants were boys than girls [56 (52%) vs. 52 (48%)].

Intention to breast feed among pregnant patients and the newborn feeding method among postpartum patients were assessed and responses collapsed into "preferred feeding method for newborns." A slight majority preferred bottle feeding to breast feeding [62 (53%) vs. 56 (48%)].

## Postpartum Variables

Among postpartum women the route of delivery was most often vaginal [37 (53%)] with another 3 (4%) reporting a vacuum assisted vaginal delivery. Thirty (43%) underwent Cesarean delivery [11 (16%) planned and 19 (27%) urgent]. Birth weights for 73 babies ranged from 765 to 4309 grams with a mean of 3050 (SD = 626) grams.

Patient Satisfaction

Overall satisfaction with their healthcare was very high with 64 (54%) reporting that they were "very satisfied." Another 41 (35%) reported "mostly satisfied" and only 13 (11%) reported being "little to somewhat satisfied." No respondents reported that they were dissatisfied.

Preferences

The percentage of women selecting each of the dichotomous preferences is shown in Table I. The vast majority of respondents (>85%) indicated a preference for female doctors over male doctors, sweet tea over unsweetened tea (our study was conducted in the South), shaved legs over unshaved legs, and long hair over short hair.

## Relationships

Significant relationships between preferences and intention to be pregnant at this time, sex of infant, and route of delivery are shown in Table II. Unintended pregnancy was more prevalent among women who preferred thongs, text messaging, Doritos®, silver, contact lens, watching a movie, beaches, and yoga. Female infants were more prevalent among women who preferred cars, chicken and girls; male infants were more frequent among those who preferred trucks, beef and boys. Cesarean deliveries were more prevalent among those who preferred contact lenses, butter, blue ink, and flip-flops. With regard to level of satisfaction with their healthcare, women were generally more satisfied when they preferred gold (75% very satisfied, 17% mostly satisfied, 8% little/somewhat satisfied) to silver (50% very satisfied, 40% mostly satisfied, 10% little/somewhat satisfied; p = 0.015) and spring (60% very satisfied, 25% mostly satisfied, 7% little/somewhat satisfied) to fall (44% very satisfied, 41% mostly satisfied, 15% little/somewhat satisfied; 0.022).

**Table I. Patients' Preferences**

| Option I | n (%) | Option 2 | n (%) |
|---|---|---|---|
| **Foods** | | | |
| Crunchy peanut butter | 38 (32%) | Creamy peanut butter | 81 (68%) |
| Hotdogs | 36 (30%) | Hamburgers | 83 (70%) |
| Chicken | 83 (70%) | Beef | 36 (30%) |
| Potato chips | 35 (29% | Doritos | 83 (70%) |
| Dominos | 42 (35%) | Papa John's | 77 (65%) |
| Wendy's | 66 (56%) | McDonald's | 51 (44%) |
| Coke | 45 (38%) | Pepsi | 73 (62%) |
| Coffee | 31 (26%) | Tea | 88 (74%) |
| Sweet tea | 107 (91%) | Unsweet tea | 11 (9%) |
| Plain M&Ms | 52 (44%) | Peanut M&Ms | 66 (56%) |
| Milk chocolate | 100 (83%) | Dark chocolate | 20 (17%) |
| Chocolate ice cream | 60 (51%) | Vanilla ice cream | 58 (49%) |
| Butter | 93 (76%) | Margarine | 27 (23%) |
| **Attire** | | | |
| Eye glasses | 36 (30%) | Contact lens | 83 (70%) |
| Gold | 57 (48%) | Silver | 63 (53%) |
| Thongs | 50 (42%) | Panties | 68 (58%) |
| Skirts/Dresses | 20 (17%) | Pants | 100 (83%) |
| Heels | 56 (47%) | Flats | 63 (53%) |
| Tennis shoes | 60 (50% | Flip-flops | 60 (50%) |
| **Activities** | | | |
| Rock and Roll | 70 (61%) | Country and Western | 44 (39%) |
| Read a book | 20 (17%) | Go to the movies | 100 (83%) |
| Yankees | 74 (68%) | Red Sox | 35 (32%) |
| Football | 74 (64%) | Basketball | 42 (36%) |
| Kickball | 85 (71%) | Dodge ball | 35 (28%) |
| Running | 82 (68%) | Yoga | 38 (32%) |
| Beach | 84 (71%) | Mountains | 35 (28%) |
| **Health and Beauty** | | | |
| Early to bed | 39 (32%) | Night owl | 82 (68%) |
| Bath | 62 (52%) | Shower | 58 (48%) |
| Pads | 37 (31%) | Tampons | 82 (69%) |
| Unpolished nails | 71 (59%) | Polished nails | 49 (41%) |
| Shaved legs | 110 (92%) | Unshaved legs | 10 (8%) |
| Long hair | 104 (87%) | Short hair | 16 (13%) |
| Make up | 69 (57%) | No make up | 52 (43%) |
| **Gender** | | | |
| Girls | 68 (55%) | Boys | 50 (42%) |
| Female doctor | 109 (92%) | Male doctor | 9 (8%) |
| **Miscellaneous** | | | |
| Paper | 24 (20%) | Plastic | 96 (80%) |
| Car | 84 (70%) | Truck | 36 (30%) |
| Black ink | 76 (63%) | Blue ink | 44 (37%) |
| Text message | 52 (44%) | Voice mail | 67 (54%) |
| Fall | 30 (25%) | Spring | 90 (75%) |

Notes. Preference selection could be skipped if unable to choose; N rarely = 121.
Percentages equal 100% except where rounded up.

**Table II. Significant Relationships between Preferences and Outcomes**

| Preferences | Frequency | Outcomes | | Significance |
|---|---|---|---|---|
| | n | Unintended pregnancy | Intended pregnancy | p* |
| Thongs | 50 | 82% | 18% | 0.034 |
| Panties | 67 | 64% | 36% | |
| Text messages | 52 | 83% | 17% | 0.022 |
| Voicemail | 66 | 64% | 36% | |
| Potato chips | 34 | 44% | 59% | 0.0001 |
| Doritos | 83 | 84% | 16% | |
| Read a book | 19 | 53% | 47% | 0.048 |
| Watch a movie | 100 | 75% | 74% | |
| Silver | 63 | 79% | 21% | 0.042 |
| Gold | 56 | 63% | 38% | |
| Beach | 84 | 81% | 19% | 0.0001 |
| Mountains | 34 | 47% | 53% | |
| Eye glasses | 36 | 56% | 44% | 0.013 |
| Contacts | 82 | 78% | 22% | |
| Running | 81 | 65% | 35% | 0.035 |
| Yoga | 38 | 84% | 16% | |
| **Preferences** | **n** | **Boy** | **Girl** | **p*** |
| Car | 73 | 43% | 58% | 0.009 |
| Truck | 33 | 70% | 30% | |
| Beef | 32 | 69% | 31% | 0.019 |
| Chicken | 73 | 44% | 56% | |
| Girls | 60 | 28% | 72% | 0.0001 |
| Boys | 44 | 84% | 16% | |
| **Preferences** | **n** | **Cesarean Delivery** | **Vaginal Delivery** | **p*** |
| Eye glasses | 19 | 21% | 79% | 0.025 |
| Contacts | 49 | 51% | 49% | |
| Butter | 58 | 48% | 52% | 0.016 |
| Margarine | 11 | 9% | 91% | |
| Blue ink | 21 | 67% | 33% | 0.006 |
| Black ink | 48 | 31% | 69% | |
| Tennis shoes | 38 | 26% | 74% | 0.003 |
| Flip-flops | 31 | 61% | 39% | |

Note. * Chi square analyses

**Table III. Significant Intra-category Preference Relationships**

| | Preferences | Spearman rho | p |
|---|---|---|---|
| **Food** | | | |
| M&Ms Plain/Peanut | Peanut butter Creamy/Crunchy | -0.205 | 0.027 |
| Butter/Margarine | Coke/Pepsi | 0.0220 | 0.017 |
| Wendy's/McDonald's | Beef/Chicken | 0.210 | 0.02 |
| Potato chips/Doritos | Beef/Chicken | 0.184 | 0.049 |
| **Attire** | | | |
| Tennis shoes/Flip-flops | Gold/Silver | -0.217 | 0.017 |
| **Activities** | | | |
| Beach/Mountains | Yankees/Red Soxs | -0.295 | 0.002 |
| Beach/Mountains | Watch a movie/Read a book | -0.252 | 0.006 |
| Rock & Roll/Country-Western | Yankees/Red Soxs | 0.201 | 0.039 |
| **Miscellaneous** | | | |
| Car/Truck | Cats/Dogs | 0.245 | 0.007 |

Note. Preferences listed in order of association, e.g., M&Ms plain and Creamy peanut butter, etc.

None of the preferences were significantly related to preferred feeding method for newborns or pregnancy complications. Significant intra-category relationships are shown Table III. Preferences significantly related to each other were entered together in higher order ANOVAs (e.g., fast food restaurants, chips, and meat preferences) to examine the associations of these clustered preferences on birth weights; we found no significant relationships.

## Comment

This project demonstrates the difficulties in interpreting significant statistical results from multiple analyses of important obstetric outcomes relative to individuals' preferences between choices available in their routine daily experiences. Thirty-seven Chi square analyses were conducted for five dichotomous outcomes including intention to be pregnant, baby's sex, any pregnancy complication, preferred feeding method for newborn, and route of delivery. Thirty-seven Mann-Whitney analyses were conducted with preferences and overall satisfaction with healthcare. Spearman correlations were conducted on variables within categories of preferences for a total of 116 comparisons. Seventeen one-way, four two-way, and one each of three-way, four-way and five-way ANOVAS were conducted with birth weight as the dependent variable. A total of 362 separate analyses were conducted.

With the overall null hypothesis of no significant relationships between any of the preferences and the seven outcomes and a significance level of $p < 0.05$, we would expect a total of 18 spurious significant findings or Type I errors (false positive results in 5 for every 100 analyses). Overall, we found 26 statistically significant associations of which 8 might be expected to be clinically meaningful, leaving us ponder which 8?

It might be hard to resist selecting those relationships that appear to fit obviously with what we believe about human behavior. The fact that women who are about to or just have delivered an infant of a certain sex were much more likely to prefer that gender or that crunchy peanut butter lovers prefer peanut M&Ms makes sense intuitively. Likewise, it just might make sense to some to note the propensity for women with unintended pregnancies to prefer thongs over panties, as well as beach vacations, both opportunities to be scantily clad! But Doritos®, texting, silver jewelry and contacts? Well, we could just dismiss those as the spurious associations. Should we counsel our flip-flop-wearing, butter-eating patients of an increased risk of cesarean section or our beach-loving, text-messaging, movie-going, thong-wearing, yogis to be more careful with birth control? Perhaps patients desiring boys should eat more beef and go buy a truck. Alas!

This study demonstrated some of the problems associated with multiple analyses and data mining in a humorous, experiential modality befitting adult learners. Simultaneously, we taught residents one methodology for learning about important obstetric outcomes in our local population (e.g., a 29% intended pregnancy rate), an ACGME competency of practice-based learning and improvement.

Although we had a good response rate for surveys, our sample size was extremely limited for the number of analyses conducted and not at all representative of our patients as no one who was unsatisfied with their obstetric care participated. Furthermore, the lack of a priori hypotheses led us on a relatively "successful" fishing expedition for significant associations in our newly generated database.

In conclusion, critically examining research articles rather than falling victim to blind acceptance of a $p < 0.05$ may spare you a Sirens' fate of crashing on a rocky statistical conclusion. Deciphering clinical significance out of the statistically significant may not always be easy and may rarely be free of biases inherent in human nature.

All of this begs the daunting question: Just how many studies published in the medical literature today use similar methodologies to make conclusions? When considering implementation of published findings in medical literature, it is important to peruse all components of the research including statistical methodology with a critical eye. Don't be enamored by a $p < 0.05$!

## References


1. Grimes DA. Epidemiologic research using administrative databases: garbage in, garbage out. Obstet Gynecol 2010;116(5):1018-9.
2. Grimes DA, Schulz KF. Clinical research in obstetrics and gynecology: more tips for busy clinicians. Obstet Gynecolo Surv 2005;60(9):S53-S69.
3. Helberg C. Pitfalls of data analysis (or how to avoid lies and damned lies). Workshop presented at the third International Applied Statistics in Industry Conference Dallas:TX;June 5-7, 1995. Available at http://www.execpc.com/~helberg/pitfalls/.
4. Lord SJ, Gebski VJ, Keech AC. Multiple analyses in clinical trials: sound science or data dredging? Med J Aust 2004;181(8):452-4.
5. Austin PC, Mamdani MM, Juurlink DN, Hux JE. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. J Clin Epidem 2006;59:964-9.
6. The Myers-Briggs Foundation. My MBTI personality type. 2011. Gainesville, FL: The Center for Applications of Psychological Types. Available at http://www.myersbriggs.org.
7. Center for Disease Control and Prevention. Pregnancy risk assessment monitoring system (PRAMS): PRAMS and unintended pregnancy. 2009. Atlanta, GA:DHHS CDC. Available at http://www.cdc.gov/prams/UP.htm.



**Disclosure**: None of the authors have a conflict of interest.
**Financial Support**: This study was underwritten by the MAHEC Department of Obstetrics and Gynecology

**Paper presented at the 2006 Annual Meeting of the North Carolina Obstetrical and Gynecological Society, Asheville, NC; May 6, 2006**

**Winner of the North Carolina Obstetrical and Gynecological Society Resident Research Award**



Reprints and Correspondence: Shelley L. Galvin, MA; MAHEC Department of Obstetrics and Gynecology, 119 Hendersonville Road, Asheville, NC 28803
828.771.5501 Email: Shelley.Galvin@mahec.net