Department of Population Health, Medical College of Wisconsin

MEDICAL COLLEGE OF WISCONSIN

DATUM

NEWSLETTER OF THE BIOSTATISTICS & EPIDEMIOLOGY KEY FUNCTION OF THE CTSI

It Figures Aniko Szabo, Ph.D. Division of Biostatistics

Spaghetti is good for more than just eating

The previous "It Figures" article (*Datum*, Volume 15, Number 3) discussed how showing the underlying data values using dot plots can enhance a plot. *Spaghetti plots*, also called *profile plots,* are an extension of the same idea for longitudinal data. Longitudinal data is commonly plotted by showing the mean and standard error at each time-point. Such plots have several problems:

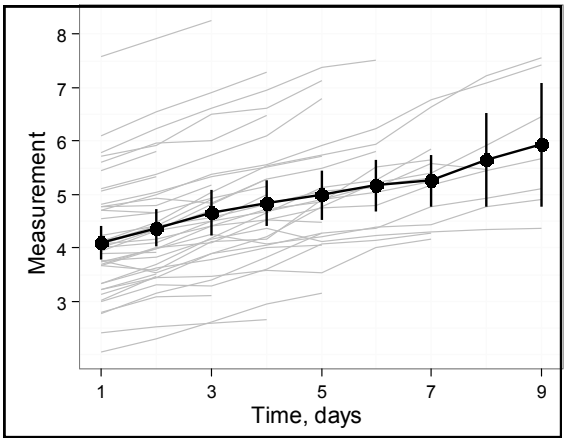
1) The average of the response trajectories is often not the "average" or typical trajectory. In fact, it might not resemble any of the actual trajectories.

2) The error bars don't show the nature of the variability around the mean. Is an increase in the mean typical among all subjects, or is it just driven by a few outliers? Perhaps some subjects have consistently high/low values, but the time-trend is very similar for everybody, or, conversely, the values keep varying around the average for everybody?

3) If the timing of measurement varies between subjects, rounding or approximation might be needed to collect enough values at each time-point for averaging.

In a spaghetti plot the trajectory of the measurements for all the individuals are shown. The result often looks like a tangle of spaghetti strands, hence the name. Figure 1 uses a spaghetti plot to show the recorded measurements of 50 individuals over 9 days (artificial data). The gray lines are the individual trajectories; the day-specific means with standard error bars are overlaid in black. It is clear that in this case the average trend does capture a typical trajectory: the measurements of all subjects tend to increase over time. Actually, most of the uncertainty in the error bars comes from the different starting values of the subjects – there is much less uncertainty about the trend itself.

Figure 1



(Continued on page 2)

Volume 15, Number 4

September/October 2009

In this issue:

CLINICAL & TRANSLATIONAL SCIENCE INSTITUTE

New Release

Wisconsin Public Health Profiles, 2007 are now available! Published annually, these profiles contain health and demographic information for each county in Wisconsin, as well as for Division of Public Health and Perinatal Regions in the state.

To view the profiles, visit **http://dhs.wisconsin.gov/localdata/pubhlthprofiles.htm**.

(Continued from page 1)

The story told by the spaghetti plot in Figure 2 is somewhat different, even though the mean/standard error line is exactly the same. Here the within-individual variability is much larger, and the presence of the trend within each subject is less clear.

In Figure 3 the average trend-line is again similar to that in the previous two figures, however the underlying situation is very different. Here the average is not representative of any of the actual trajectories, which are fairly constant. The apparent increase is only an artifact of the earlier truncation of the low measurements. Unfortunately, we would not be able to tell this without seeing the actual trajectories.

In summary, spaghetti plots can add clarity and insight to plots of longitudinal data. When drawn using lighter colors, they stay in the background and support the main message of the plot. The presence of a notable feature such as a trend, or a sudden change is much more persuasive when it can be seen in most trajectories and not just the average. An additional benefit is that measurements taken at irregular and/or inconsistent time points can be easily shown. Even when not shown in the final version of a plot, individual trajectories should be examined during data analysis to ensure that the average is indeed representative and does not misrepresent the underlying data.

Figure 2

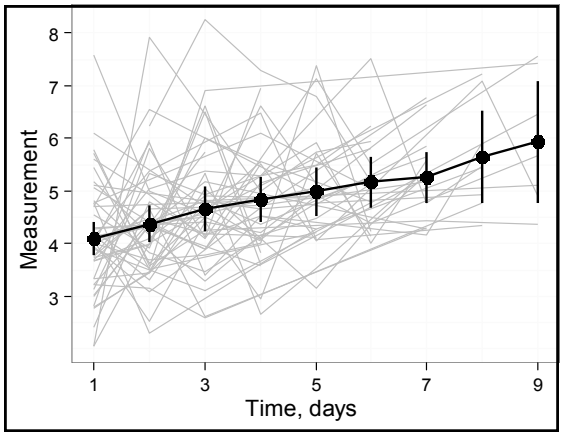
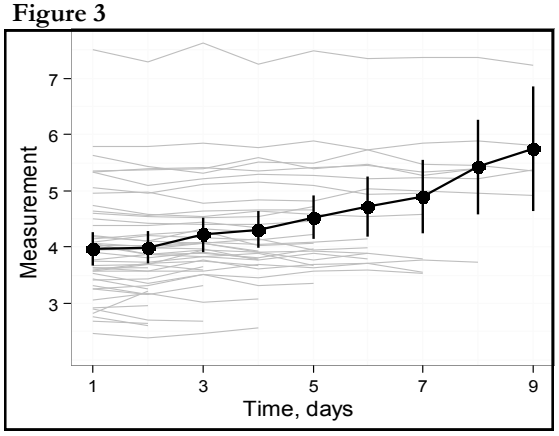


Figure 3



DATUM

Editors: Aniko Szabo
Emily McGinley

Department of Population Health
Medical College of Wisconsin

Datum is published by the Department of Population Health and Biostatistics & Epidemiology Key Function of the CTSI. It is available online at **www.mcw.edu/edsc/datum**.

We welcome your newsletter questions, comments, and suggestions. Please contact us with address updates or to be added to or removed from our mailing list.

Please direct all newsletter correspondence to:

Datum Newsletter
Department of Population Health
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI 53226—0509

Phone: 414-456-4255
Fax: 414-456-6520
E-mail: emcginley@mcw.edu

On the Map

A geographic information system, or GIS, is a computer technology that allows data to be analyzed within a geographic context. Here, we provide information that is of interest to health GIS users:

DATA

- Looking for spatial data for Wisconsin? Check out the Data Resources page on the State Geographic Information Office website, **http://gio.wi.gov/Resources/Data/tabid/253/Default.asp**x. They list links with descriptions and examples of the data available at each. While you're there check out the Projects tab, for updates on the **Wisconsin Spatial Data Repository** Project. And if you have questions or ideas, they welcome feedback!

TRAINING & EVENTS

- Mark your calendar for **GIS Day 2009**! As part of National Geography Awareness Week, GIS Day is an international day of education with local events including workshops, demonstrations, open houses and more! Again this year, there are two University of Wisconsin opportunities to experience GIS Day. Both are free and open to the public. There is something for everyone, experienced and novice!
  - UW-Milwaukee will host GIS Day activities on Wednesday, **November 18th**. In addition to the map gallery and organization tables, this year's agenda includes a session on GIS Apps in Public Health, as well as an Intro to GIS for those who want to learn the basics using ArcGIS. Luncheon speaker, State Cartographer, Howard Veregin, will be presenting *Geo-Enabled Cartography*. Find the latest information, or register for sessions at **www4.uwm.edu/gis/gisday/**.
  - On Friday, **November 20th**, UW-Madison will host their GIS Day Expo event. This year's theme is "History: Education: Research: Application". Find details at **www.geography.wisc.edu/GISDay/**
- New to Geographic Information Systems? ESRI offers a **free online course** for those with no GIS background or experience. *Getting Started with GIS* covers the basic features of GIS and a geographic approach to solving problems. Find out more at **http://training.esri.com/acb2000/showdetl.cfm?DID=6&Product_ID=915**.
- And remember, if you use GIS in your research at MCW and would like to connect with others who do too, join the **MCW GIS User Group**. Contact Emily McGinley at emcginley@hpi.mcw.edu or 456-4255 for more information.

Would you like to know more about GIS? Visit the EDSC's GIS resource website at ***www.mcw.edu/edsc/gis.htm.***

You Asked…

"I'm having trouble extracting the shapefiles downloaded from the U.S. Census Bureau's website. Microsoft Windows keeps giving me an error. Is there another way to retrieve these boundary files?"

The issue with the shapefile format boundaries available for download from the Census website is the multiple levels of folders (some can be 5+ deep). Microsoft Windows sees this as a possible security threat and will not extract the files. Try using the open source software, 7-Zip. This free software is available at, **www.7-zip.org**. Once extracted, you can move the files to the desired location on your system.

Do you have a question for us? Send it to Datum *at emcginley@mcw.edu, and we'll answer it in a future issue.*

Statistical Classroom Brent Logan, Ph.D., Division of Biostatistics

Multiple Comparisons, or Torturing Your Data Until They Confess

Researchers are often interested in exploring a number of research questions on the same dataset. These situations can come in many forms. An investigator may be interested in analyzing the effect of a treatment on each of several outcome variables. They may be comparing several different types of treatments or multiple doses of a drug, to determine whether one treatment or dose works better than the others. The researcher may be interested in studying whether the treatment effect is consistent across different subgroups of patients. In some cases, the number of questions being considered can get very large, when, for example, an investigator examines a large number of genes (possibly thousands) for association with an outcome in a microarray study. Each of these research questions will often be tested using a statistical hypothesis test.

In each of these cases, it is important to consider the impact that the number of tests done can have on how you interpret the results of that hypothesis test. As an example, let's consider a panic disorder study to measure the effectiveness of a new drug treatment against a control group on each of four outcomes: 1) severity of anticipatory anxiety, 2) total number of panic attacks, 3) severity of phobic avoidance, and 4) global assessment of the patient. For each of the four outcomes, we can test a null hypothesis that there is no effect of the drug on that outcome. Each of these hypothesis tests results in a p-value (0.04, 0.10, 0.72, and 0.38 respectively for the four outcomes). Often investigators would compare these p-values to a significance level of 0.05, and conclude that because the p-value for severity of anticipatory anxiety is $0.04<0.05$, the treatment has a statistically significant effect on this outcome. Here the 0.05 threshold refers to the type I error rate.

The problem with basing our conclusion on comparing the p-value to 0.05 is that we did four hypothesis tests. Any time you perform multiple tests, you increase the chance of obtaining at least one false significant finding. This raises the question of whether the impact of treatment on severity of anticipatory anxiety is real, or whether it is an artifact of performing multiple tests. As an extreme case, one could keep performing more and more hypothesis tests until you found a significant finding (e.g. torturing your data until they confess). If one performed 10 independent hypothesis tests, the chance of making at least one type I error would be 40%, and if one performed 50 tests, the chance of making at least one type I error would be over 90%.

Key Concepts in Multiple Comparisons

Type I error rate: Probability of incorrectly rejecting a single null hypothesis

Familywise Error Rate: Probability of incorrectly rejecting at least one null hypothesis out of all tests being performed

False Discovery Rate: Expected proportion of false discoveries (incorrectly rejected null hypotheses) out of the total number of significant findings

There are a number of strategies which have been proposed for dealing with the multiple testing problem. Some of them are conceptual, and others statistical. First, the easiest way to reduce the multiplicity problem is to focus or prioritize your research questions and do fewer tests. This is one of the reasons that in phase III clinical trials, one often prespecifies a single primary outcome or endpoint, and refers to the remaining outcomes as secondary endpoints. Often this primary endpoint is a composite outcome reflecting multiple endpoints, such as disease-free survival.

Performing multiple tests without adjustment is most often done in exploratory or hypothesis-generating research, where there are concerns about loss of power due to multiplicity adjustment. Sometimes an ad-hoc adjustment is used, such as employing a significance level of 1% instead of the usual 5%, to make it more difficult to reject the null hypothesis and therefore less likely to make a type I error. When multiple tests are performed without adjustment, you should be transparent about how many tests were performed and be very cautious in your interpretation.

The statistical strategy to deal with the multiplicity problem is to control a different error rate which incorporates the number of tests performed. The Familywise Error Rate (FWE) is often used in a confirmatory

(Classroom, continued on page 5)

Software Corner Brent Logan, Ph.D., Division of Biostatistics

Software Options for Multiple Comparisons

The two simple and popular procedures (Bonferroni and Benjamini-Hochberg) mentioned in this issue's "Statistical Classroom" article (page X) are based on a set of p-values for virtually any set of hypotheses of interest. As such, they are fairly straightforward to apply using hand calculations or a spreadsheet. Alternatively, SAS has a procedure (PROC MULTTEST) which will perform these adjustments using just a dataset containing p-values.

Other than the above, multiple comparison procedures are usually built into the statistical software package for specific types of analysis. The most common built-in application of multiple testing methods is pairwise comparisons among means of multiple groups. This is typically analyzed using analysis of variance (ANOVA) techniques, or a general linear model (GLM) framework to adjust for covariates. Minitab software includes Bonferroni, Tukey, and Dunnett multiple comparison adjustments in its general linear model analysis. Stata software includes the Bonferroni and Holm procedures in its comparison of means after ANOVA or MANOVA, as well as in a few other analysis procedures. SAS includes adjustment options for pairwise comparisons among means in its PROC ANOVA and PROC GLM procedures, including the Bonferroni procedure and a resampling-based adjustment which is probably the most powerful general option for the pairwise comparisons setting. There are also SAS macros available as described in Westfall et al. (1999) which can be used to perform powerful multiplicity adjustment from output of a variety of other analyses.

References:

Westfall PH, Tobias RD, Rom D, Wolfinger RD, Hochberg Y (1999) Multiple comparisons and multiple tests using the SAS system. SAS Institute, Cary, NC.

(Classroom, continued from page 4)

research setting, where strict control of any type I errors is important. False Discovery Rate (FDR) controlling procedures have less stringent control over type I errors, but maintain higher power to detect true differences. They are used in more exploratory settings, especially for large scale multiple testing problems, such as genetics or imaging. Either way, explicit statistical correction for the number of hypothesis tests performed strengthens the evidence for significant research findings. However, there is a penalty in the sense that multiplicity adjustment makes it harder to detect real treatment effects (i.e. loss of power). This penalty gets worse as you adjust for more hypotheses, so it is important to focus your adjustment on the most important hypotheses if possible .

There are two simple, common strategies to control the FWE. The first is to perform an overall test of all the null hypotheses at once, and only look at the individual hypotheses if the overall test is significant. In our panic disorder study, the p-value is 0.20 for the test of the overall null hypothesis that there is no effect of drug on ANY of the outcomes. This indicates that there is no evidence of a drug effect, and we should not be looking individually at each outcome. Another simple option is a Bonferroni adjustment, in which we use a type I error rate for each comparison which is divided by the number of hypothesis tests being performed. Here we have four outcomes, so each p-value should be compared to a significance level of 0.05/4=0.0125. With this adjusted significance level, our conclusion would be that there is no effect of drug on any of the outcomes since none of the p-values are below 0.0125. There are many other more efficient ways of adjusting for multiple testing but they are generally more complicated and specific to a particular scenario, while these two strategies are simple and applicable to a large number of settings.

To control the FDR, the most common procedure is the Benjamini-Hochberg procedure, which can be applied simply to a set of p-values. There are several more efficient variations have been recently developed for application in settings such as microarray data analysis.

Multiple testing is important to discuss with a consulting statistician, both in the design and analysis phase of a study. A statistician can help you refine your family of hypotheses, choose an appropriate adjustment technique, and interpret the results. Explicit multiplicity adjustment should also be considered in the determination of sample size. For an introduction to available software options for multiple comparisons, see the "Software Corner" article in this issue (above).

Consultant's Forum Mei-Jie Zhang, Ph.D. & Scott Jackson, M.S. Division of Biostatistics

Effect of Total Hip Arthroplasty on Bone Structure

In 2008, Dr. James Stiehl, an orthopedic surgeon at Columbia St. Mary's Hospital, came to the consulting service with data on patients who had total hip arthroplasty performed. Of interest was the change in bone mineral content (BMC), measured no more than 15 months prior to surgery and again within 13 years after surgery. In 15 randomly selected patients, BMC was measured in seven "Gruen" zones in the periprosthetic bone (Figure 1). Dual-energy xray absorptiometry (DEXA) was used to measure BMC pre- and post-surgery. In patients with an unaffected leg, measurements were also taken in the unaffected legs to use as a control.

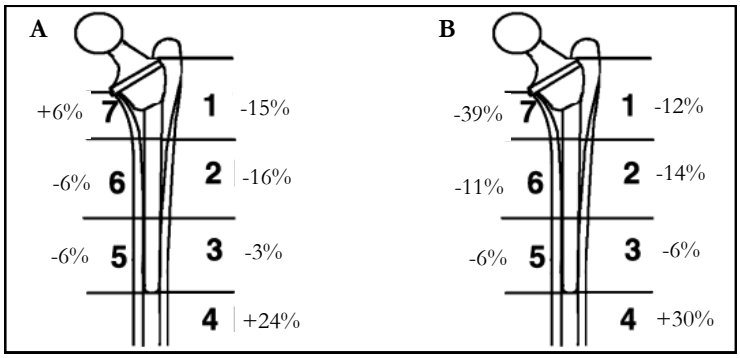


Figure 1. Gruen zones in periprosthetic bone showing distribution of BMC changes in unaffected (A) and affected (B) limbs.

Since these data contain repeated measures (i.e. more than one measurement per patient), we must use a model which can account for the correlation between measurements on the same patient. In this case, PROC MIXED was used in SAS to account for these correlations. Our model identified a unique observation in the dataset as one with the same affected status, measurement time point and zone. SAS results of post-op versus pre-op comparisons for each zone are given in Table 1.

Gruen Zone	Estimated BMC Change (g)	95% Confidence Interval	p-value
1	1.874	(0.35, 3.39)	0.022
2	1.894	(-1.25, 5.04)	0.24
3	0.243	(-2.29, 2.77)	0.85
4	0.985	(-0.85, 2.82)	0.30
5	1.086	(-3.32, 5.49)	0.63
6	0.556	(-1.95, 3.07)	0.66
7	-3.575	(-6.33, -0.81)	0.017

Table 1. SAS results of post-op versus pre-op comparisons

In zone 1, we found a significant increase in BMC (1.874, p=0.022). Zones 2-6 showed a positive, but non-significant change. Only in zone 7 did BMC decrease, which was significant at the 5% level (-3.575, p=0.017). The full analysis result has been published in *Clinical Orthopaedics and Related Research* 2009; 467(9):2356-61.

Biostatistics Lecture Series

Wednesday, November 18, 2009. 8:50-9:50 am - Location To Be Announced
Brent Logan, Ph.D.
Designing Clinical Trials

Friday, December 11, 2009. 7:00-8:00 am - Froedtert NT2209
Aniko Szabo, Ph.D.
Multiple Comparisons, or Torturing Your Data Until It Confesses

Wednesday, December 16, 2009. 8:50-9:50 am - Location To Be Announced
Rute Bajorunaite, Ph.D.
What Should We Do if Data Come in Pairs?

CME credit is available. Slides from past lectures are available on the Biostatistics website,
www.mcw.edu/biostatistics/CalendarCurrentEvents/SeminarSeries.htm

Feel like you missed something?

Past lecture series presentations are now available at the MCW Library for check-out and viewing!

BIOSTATISTICS & EPIDEMIOLOGY KEY FUNCTION CONTACT INFORMATION

CLINICAL & TRANSLATIONAL SCIENCE INSTITUTE

Medical College of Wisconsin
Department of Population Health
8701 Watertown Plank Road
Milwaukee, WI 53226-0509

Biostatistics Consulting Service
Kellie Bialzik
Health Research Center, Suite H2400
Phone: 414-456-8280
Fax: 414-456-6513
Email: kbialzik@mcw.edu
Web: www.mcw.edu/biostatsconsult.htm

Epidemiology Data Service Center
Emily McGinley
Health Research Center, Suite H2100
Phone: 414-456-4255
Fax: 414-456-6520
Email: emcginley@mcw.edu
Web: www.mcw.edu/edsc.htm

RECENT PUBLICATIONS

Bajaj JS, Ananthakrishnan AN, Hafeezullah M, Zadvornova Y, **Dye A**, **McGinley EL**, Saeian K, Heuman D, Sanyal AJ, Hoffmann RG. *Clostridium difficile* Is Associated with Poor Outcomes in Patients with Cirrhosis: A National and Tertiary Center Perspective. *American Journal of Gastroenterology*, 2009 Oct 20. [Epub ahead of print].

Collins NH, Gee AP, Durett AG, Kan F, **Zhang MJ**, Champlin RE, Confer D, Eapen M, Howard A, King R, Laughlin MJ, Plante RJ, Setterholm M, Spellman S, Keever-Taylor C, Wagner JE, and Weisdorf DJ, The effect of the composition of unrelated donor bone marrow and peripheral blood progenitor cell grafts on transplantation outcomes, *Biology of Blood and Marrow Transplantation*; Epub 2009 Oct 9.

Ananthakrishnan AN, **McGinley EL**, Saeian K, Binion DG. Trends in Ambulatory and Emergency Room Visits for Inflammatory Bowel Disease in the United States: 1994-2005. *American Journal of Gastroenterology*, 2009 Oct 6. [Epub ahead of print].

Ellinas EH, **Eastwood DC**, Patel SN, and Maitra-D'Cruze AM, The effect of obesity on neuraxial anesthesia difficulty in pregnant patients: A prospective, observational study, *Anesthesia & Analgesia*; 109(4):1225-31, 2009.

Liu Q, Wong-Riley MTT, Postnatal changes in the expressions of serotonin 1A, 1B, and 2A receptors in ten brain stem nuclei of the rat: Implication for a sensitive period, *Neuroscience*; 2009 Oct 2. [Epub ahead of print].

Marks DI, Ballen KK, **Logan BR**, Wang ZW, Sobocinski KA, Bacigalupo A, Burns LJ, Gupta V, Ho V, McCarthy P, Ringden O, Schouten HC, Seftel M, and Rizzo JD, The effect of smoking on allogenic transplant outcomes, *Biol Blood Marrow Transplant*; 15(10):1277-87, 2009.

Migrino RQ, **Ahn KW**, Brahmbhatt T, Harmann L, Jurva J, and Pajewski N, Usefulness of two-dimensional strain echo-cardiography to predict segmental viability following acute myocardial infarction and optimization using bayesian logistic spatial modeling, *American Journal of Cardiology*; 104(8):1023-9, 2009.

Graf A, Hassani S, Krzak J, Caudill A, Flanagan A, **Bajorunaite R**, Harris G, and Smith P, Gait Characteristics and Functional Assessment of Children with Type I Osteogenesis Imperfecta, *Journal of Orthopaedic Research*; (9):1182-90, 2009.

Gross TG, Hale GA, He W, Camitta BM, Sanders JE, Cairo MS, Hayashi RJ, Termuhlen AM, **Zhang MJ**, Davies SM, and Eapen M, Hematopoietic stem cell transplantation for refractory or recurrent non-hodgkin lymphoma in children and adolescents: A report from the Center for International Blood and Marrow Transplant Research, *Biol Blood Marrow Transplant*, 2009 Sep 29. [Epub ahead of print]

Ananthakrishnan AN, **McGinley EL**, Saeian K, Binion DG. Laproscopic Resection for Inflammatory Bowel Disease: Outcomes from a Nationwide Sample. *Journal of Gastrointestinal Surgery*, 2009 Sept 17. [Epub ahead of print].

Trivedi H, Daram S, **Szabo A**, Bartorell AL and Marenzi G, High-Dose N-acetylcysteine for the Prevention of Contrast-induced Nephropathy, *American Journal of Medicine*; 122(9):874.e9-15, 2009.

Wang T and Zeng ZB, Contribution of genetic effects to genetic variance components with epistasis and linkage dis-equilibrium, *BMC Genet*; 10(1):52, 2009.

For details on other publications and presentations, visit our websites.

Division of Biostatistics: **www.mcw.edu/biostatistics/overview/publicationspresentations.htm**

Epidemiology Data Service Center: **www.mcw.edu/edsc/publications.htm**

About the

BIOSTATISTICS & EPIDEMIOLOGY KEY FUNCTION

Clinical and Translational Science Institute:
The Clinical and Translational Science Institute (CTSI) of Southeastern Wisconsin represents a unique and transformative collaboration among the Medical College of Wisconsin, its campus research partners: Froedtert Hospital, Children's Hospital of Wisconsin, Zablocki Veterans Affairs Medical Center, and the Blood Research Institute, and the major academic institutions in southeastern Wisconsin: Marquette University, University of Wisconsin-Milwaukee, and Milwaukee School of Engineering. The CTSI is a new and innovative infrastructure to support and advance education, collaboration, and research in clinical and translational science.

Epidemiology Data Service Center:
The Epidemiology Data Service Center (EDSC) is a centralized secondary data resource for researchers in epidemiology, health services, and other related disciplines. The services range from providing simple summary statistics, to preparing data set extractions, to lending data management and preparation expertise, to long-term research projects, to mapping and other spatial analysis. The EDSC can also provide answers to your general questions about data resources, and offers guidance to users of geographic information systems (GIS). To learn more about the EDSC, view our online database catalog, or browse newsletter archives, visit our website at **www.mcw.edu/edsc.htm**.

Biostatistics Consulting Service:
The Biostatistics Consulting Service provides statistical support to investigators at the Medical College, its affiliates and the general public. This support includes assistance with design and analysis of clinical trials, observational studies, and surveys, assistance with public databases, sample size and power calculations, and data analysis and interpretation. All investigators preparing clinical or translational research are provided free statistical support. The Biostatistics Consulting Service also offers free drop-in assistance at different locations. For more information about the consulting service, or to find times and locations for drop-in sessions, please visit our website at **www.mcw.edu/biostatsconsult.htm.**

*Visit **www.mcw.edu/biostatistics/CTSIKeyFunction/Newsletter.htm** to subscribe to Datum.*

Biostatistics Consulting Drop In Service
Schedule & Locations

FROEDTERT HOSPITAL
Monday, Wednesday and Friday
Building: Pavilion
Room: #L772A—TRU Offices
Time: 1:00 PM—3:00 PM

MEDICAL COLLEGE OF WISCONSIN
Tuesday and Thursday
Building: Health Research Center
Room: H2400 Biostatistics
Time: 1:00 PM—3:00 PM

VETERANS AFFAIRS MEDICAL CENTER
1st and 3rd Monday of the month
Building 111
Room B-5423
Time: 8:30—11:30 AM

MARQUETTE UNIVERSITY
Tuesday
Office of Research and Scholarship
Clark Hall
Room: 112D
Time: 9:00 AM—5:00 PM

NEW