Outcome Oriented

The Online Newsletter of the
Center for Outcome Measurement in Brain Injury **(COMBI)**

Winter 2001

Measuring Up!

The COMBI continues to add more important scales to its resource center. As of December 2001 there are currently twenty measures featured and detailed in the COMBI.

- Agitated Behavior Scale (ABS)
- Awareness Questionnaire (AQ)
- Coma/Near Coma Scale (CNC)
- Community Integration Questionnaire (CIQ)
- The Craig Handicap Assessment and Reporting Technique (CHART)
- The CHART Short Form (CHART-SF)
- The Craig Hospital Inventory of Environmental Factors (CHIEF)
- Disability Rating Scale (DRS)
- The Family Needs Questionnaire (FNQ)
- Functional Assessment Measure (FAM)
- Functional Independence Measure (FIM)
- Glasgow Outcome Scale (GOS)
- Level of Cognitive Functioning Scale (LCFS)
- Mayo Portland Adaptability Inventory (MPAI)
- Neurobehavioral Functioning Inventory (NFI)
- The Orientation Log (O-Log)
- The Patient Competency Rating Scale (PCRS)
- Satisfaction With Life Scale (SWLS)
- Service Obstacle Scale (SOS)
- Supervision Rating Scale (SRS)

What's Inside

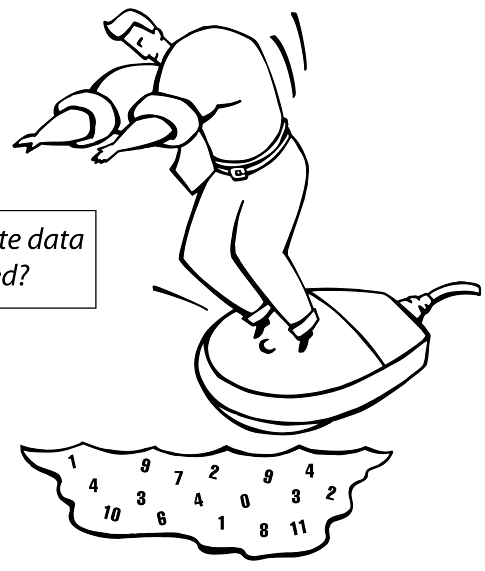
Dealing With Missing Data

Scott R. Millis, PhD
Traumatic Brain Injury National Data Center
Kessler Medical Rehabilitation Research & Education Corporation

Scott R. Millis, PhD is co-director of the Traumatic Brain Injury National Data Center and a Senior Research Scientist at Kessler Medical Rehabilitation Research and Education Corporation. Dr. Millis will be writing for Outcome Oriented on issues relating to statistics and interpretation.

Dealing with missing data is a problem commonly encountered by investigators. The problem may be exacerbated in longitudinal studies because research participants need to be followed over extended periods. Data may be missing for a variety of reasons:

Can an incomplete data set be rescued?



- Research participants may move and the investigator is unable to find new addresses or phone numbers;
- Research participants may no longer wish to participate in the study;
- A research assistant may forget to collect data on certain items in the protocol;
- The research participant may be too ill or impaired to complete specific cognitive tests in the protocol;
- A case report form containing data is lost or misplaced;
- A computer data file becomes corrupt;
- A research participant does not have transportation to the clinic on the day of data collection;
- The research protocol may change and variables are dropped;
- The research participant is discharged from the hospital before all the test data could be collected.

When faced with missing data, many investigators will simply analyze only those cases with complete data – also known as listwise deletion or complete case analysis. However, there are many problems with this approach. The sample size may shrink dramatically which may preclude performing some types of statistical analyses. There is a loss of statistical power to detect treatment effects or predictors of outcome. In addition, the results obtained may be quite inaccurate or biased. In some cases, entire research projects can be jeopardized by missing data when complete case analysis is used.

Fortunately, there have been recent developments for handling missing data. The first step is to recognize that there are different patterns of missing data, which will help determine analytic strategies:

Missing Completely at Random (MCAR):
This means that the missing data for a variable, for example, age, is unrelated to the value of age itself or to the values of any variables whether missing or observed (Allison, 2001). An example of a process leading to MCAR would be one of a research assistant randomly losing a sample of the

Continued on Page 2

Continued from Page 1

case report forms. When data are MCAR, complete case analysis (i.e., analyzing only the cases having complete data) does not yield biased results. However, MCAR is a very stringent assumption and one that is probably rarely encountered in the real world of research.

Missing at Random (MAR): MAR is a less stringent assumption than MCAR. MAR implies that participants with incomplete data may differ from participants with complete data but the pattern of missingness is predictable from other variables in the data set (Allison, 2001). For example, if persons with severe traumatic brain injury (TBI) are less likely to be able to complete a memory test, then the data are MAR as long as some persons with severe TBI were able to complete the test. In other words, the missing data memory test can be explained by other variables in the study, i.e., injury severity. When data are MAR, there are some powerful statistical techniques to deal with the missing data – which will be illustrated in the following section.

Nonignorable: When data are not missing at random and they are not predictable from other variables in your database, the missing data are said to be nonignorable. This situation is particularly challenging from a data analytic perspective. It requires a sophisticated approach on which we will touch in the next section of this guide.

Strategies for Handling Missing Data

Recognizing the limitations of complete case analysis, investigators have tried other techniques for handling missing data:

Pairwise Data Deletion: A correlation or covariance matrix can be calculated from available pairwise data. That is, a participant with missing data on one variable will be used only in calculations that do not involve that variable. In this manner, the sample size is often larger than when using complete case analysis. Linear regression can then be performed with the covariance matrix. However, unless the data are MCAR, pairwise deletion produces biased estimates and is not recommended for use (Allison, 2001; Roth, 1994).

Mean Substitution: With this method, the variable's mean value is calculated from the available cases and is used as the imputed value for the missing cases. As with the pairwise deletion method, mean substitution has a high likelihood of producing biased estimates and is not recommended for use.

Full Information Maximum Likelihood (FIML): When data are MAR, FIML should be considered. It uses all available data to calculate a vector of means and covariance matrix (i.e., maximum likelihood-based sufficient statistics) in a way that is superior to other methods (Wothke, 1998). FIML can be used to handle missing data in a variety of models, such as regression, ANOVA, ANCOVA, and structural equation models. It can be easily implemented with AMOS (Analysis of Moment Structures) software.

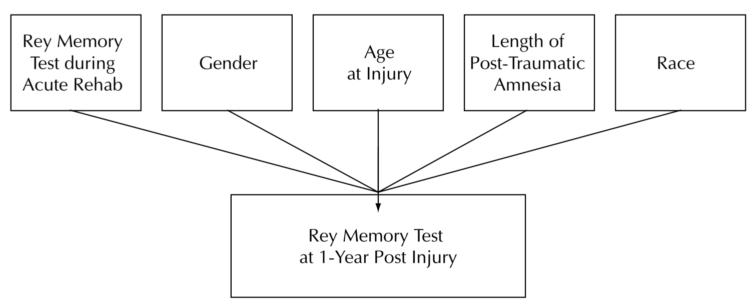
Multiple Imputation (MI): MI is similar to FIML but it creates five to 10 data sets in which raw data are generated that can be used to fill in the missing data. The data from the imputed data set are then pooled and parameters are estimated. MI can be performed with a variety of software packages, including SAS, S-Plus, and Solas.

Currently, FIML and MI methods appear to be the methods of choice for handling missing data. There are other methods such as expectation maximization, regression, and hot deck imputation, but they do not have any notable advantages over FIML or MI. MI methods are particularly flexible for a wide variety of linear and nonlinear models. Even when missing data are nonignorable, Wothke (1998) has shown that FIML outperforms pairwise deletion and complete case analysis methods. Methods to handle nonignorable missing data are still being developed. Hedeker and Gibbons (1997) applied pattern-mixture models to this pattern of missing data.

An Example

Using data from the TBI Model Systems, I wanted to predict performance on a memory measure (total score from the Rey Auditory Verbal Learning Test) at one-year post-injury based on the participant's score on the same test during acute rehabilitation plus the participant's gender, age at time of injury, length of post-traumatic injury, and race (Figure 1). Memory deficits following TBI can interfere with a person's capacity to function independently in the community. Developing a model to predict long-term outcome can be very useful for treatment planning and family counseling, especially if it is based on information known soon after the injury. Typically, standard multiple linear regression is used to estimate models like this one. There were 532 cases in this dataset but only 69 participants (13%) had complete data! Table 1 shows the missing data patterns in this data set. Virtually all regression routines in the major statistical packages use complete case analysis as the default analytic option. Hence, standard analysis of this model would have wasted 87% of the data. Pairwise deletion would increase the sample size to 88 participants – still using only a fraction of the total sample.

Figure 1. Model to Predict Performance on Rey Auditory Verbal Learning Test at One-Year Post-Injury



Next, I estimated models using mean substitution, FIML (using AMOS) and multiple imputation (using SAS Proc MI). The parameter estimates (regression coefficients and standard errors) for all methods including listwise deletion (complete case analysis) and pairwise deletion appear in Table 2. Although all methods identified the Rey memory test (given during acute rehabilitation) as a statistically significant predictor of memory performance at one-year post-injury, the magnitude of Rey's regression coefficient varied widely across methods, from .08 (mean substitution) to .78 (multiple imputation). The size of the standard errors also showed variability. Moreover, the variables found to be significant predictors varied across methods. Pairwise deletion identified length of post-traumatic amnesia as an important predictor whereas mean substitution found race to be significant. FIML and MI showed the greatest congruence in terms of size of coefficients and standard errors.

Table 1. Missing Data Patterns (X denotes data present)

REY1	REY2	SEX	AGE	PTA	RACE	Freq	Percent
X	X	X	X	X	X	69	12.97
X	X	X	X	.	X	19	3.57
X	.	X	X	X	X	217	40.79
X	.	X	X	.	X	93	17.48
X	.	.	.	.		1	0.19
.	X	X	X	X	X	57	10.71
.	X	X	X	.	X	32	6.02
.	.	X	X	X	X	24	4.51
.	.	X	X	.	X	20	3.76

Table 2. Regression Coefficients and Standard Errors of Different Methods for Handling Missing Data (* p < .05)

	Listwise	Pairwise	Mean S	ML	MI
Rey1	**.66***	**.66***	**.08***	**.72***	**.78***
	(.10)	(.08)	(.03)	(.06)	(.08)
Sex	-1.01	-.78	-1.00	-2.73	-3.37
	(2.42)	(2.42)	(.78)	(1.82)	(1.86)
Age	**-.16***	-.09	**-.08***	**-.12***	-.13*
	(.08)	(.08)	(.02)	(.06)	(.07)
PTA	.03	**.06***	.01	.03	.04
	(.04)	(.03)	(.01)	(.03)	(.03)
Race	-2.07	1.50	**1.52***	1.58	1.11
	(2.24)	(2.13)	(.68)	(1.61)	(2.97)

Although the listwise deletion method also identified the baseline Rey memory test score and age as significant predictors of memory performance at follow-up, the size of its standard errors tended to be larger than those produced by FIML and MI. In addition, listwise deletion's estimation of the impact of race was opposite of the association found by the other methods. Based on only 13% of the data, the listwise deletion method was also hampered by a substantial loss of statistical power and an inadequate subject-to-variable ratio. Consequently, parameter estimates from the listwise deletion method may fail to replicate on a new sample of patients.

Which method is correct? In all likelihood, the TBI Model Systems data are not MCAR. Participants who return at year one for follow-up to take the Rey memory test do not make up a random sample. At best, the data may be MAR. Hence, I would choose the findings from the FIML and MI methods. It certainly makes sense that the level of performance on the Rey memory test during rehabilitation should be highly predictive of performance at year one. In addition, we know that increased age is associated with declining performance on the Rey in the normal population. The other variables may be less important because their influence might be captured indirectly by the initial performance on the Rey.

For more information or questions, please contact Dr. Millis at **<smillis@kmrrec.org>**

References

Allison, P. D. (2001) *Missing data.* Thousand Oaks, CA: Sage.

Hedeker, D., Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods,* 2, 64-78.

Roth, P. (1994). Missing data: A conceptual overview for applied psychologists. *Personnel Psychology,* 47, 537-560.

Wothke, W. (1998). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data.* Mahwah, NJ: Lawrence Erlbaum Associates. ☑

Assessing The COMBI

Log Files 101

Did you know that every time you access a web page, a record of what you did is created? These records, called log files, give webmasters a lot of information about you and what you looked at on the site. Programs that interpret log files can tell you what countries your users come from, what pages they looked at, what files they downloaded, what site referred them, even what operating system they use.

The Stats

In the last 5 months (July 01–November 01) the COMBI has logged in 25,265 visitors. That's almost 165 users a day! During this period 77,106 pages of information were reviewed (that's 868 megabytes of data).

The COMBI logs show that 88% of our users are within the United States and 12% are from 60 other countries. The COMBI is especially popular in Canada, the United Kingdom, Australia, New Zealand, Italy, and Japan. Our biggest referrals come from Google, the Brain Attack Coalition (www.stroke-site.org), MSN.com, and Yahoo.

The COMBI newsletter, *Outcome Oriented*, is primarily disseminated in Portable Document Format (PDF) from the website. Over the last five months, 1616 newsletters were downloaded by COMBI users.

The COMBI continues to be very successful as a dissemination effort. In the past five months over 7,671 rating forms were downloaded. Itemized scale activity is summarized in the table below .
But please, no wagering. ☑

Scale Activity (Number of Visitors & Downloads)
July 2001–November 2001

Scale	Visitors	Downloads
ABS	1291	311
AQ	604	1153
CHART	459	616
CHART-SF	429	623
CHIEF	328	360
CIQ	523	431
CNC	558	468
DRS	788	168
FAM	754	739
FIM	1768	na
FNQ	359	na
GOS	2019	na
LCFS	596	153
MPAI	515	785
NFI	454	na
O-LOG	315	338
PCRS	471	1127
SOS	235	179
SRS	331	221
SWLS	743	na

Future Directions

The COMBI will continue to add new measures and act as a resource for the rehabilitation community. Planned additional instruments include the Extended Glasgow Outcome Scale (GOS-E), and the Expanded Rancho (LCFS) Scale.

We are also seeking scales that focus on employment, vocational, and family dynamics issues.

We are looking to add more training and testing materials for COMBI measures, and to make the existing materials more interactive (automatic email of results from testing exercises).

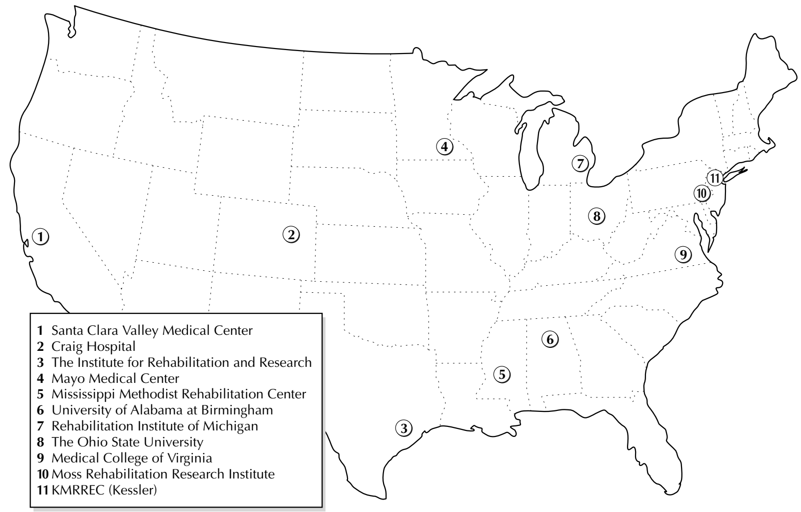
Please email us at <combi@tbi-sci.org> with your thoughts and suggestions. Let us know how we measure up! ☑

Outcome Oriented *is a project of the Center for Outcome Measurement in Brain Injury (COMBI) which is funded by the U.S. Department of Education, Office of Special Education and Rehabilitative Services, National Institute on Disability and Rehabilitation Research.*

Address inquiries to
***Jerry Wright**, Editor. PHONE (408) 295-9896 ext 20; FAX (408) 295-9913; EMAIL combi@tbi-sci.org*

This document is available online at:
<www.tbims.org/combi/combinews.html>

Credit to Our Collaborators



The COMBI is a collaborative project of eleven brain injury centers located across the US. Without the expertise of these centers this project would not be possible. We would like to offer special recognition to the individuals at these facilities who have taken the time to prepare materials for the COMBI and act as contacts:

Tamara Bushnik, PhD, Jerry Wright, BA, Maurice Rappaport, MD, PhD, & Mary Lou Kohlmiller, RN, BSN at Santa Clara Valley Medical Center (Lead Center)

Dave Mellick, MA and Cindy Harrison-Felix, MS at Craig Hospital

Corwin Boake, PhD at The Institute for Rehabilitation Research

James F. Malec, PhD, LP at the Mayo Medical Center

Mark Sherer, PhD, ABPP-Cn at the Mississippi Methodist Rehabilitation Center

Tom Novack, PhD at University of Alabama at Birmingham

Marcel Dijkers, PhD at Mount Sinai School of Medicine (Formerly at the Rehabilitation Institute of Michigan)

Jennifer Bogner, PhD & John D. Corrigan, PhD at the Ohio State University

Jeffrey Kreutzer, PhD and Jenny Marwitz, MA at Medical College of Virginia

Tessa Hart, PhD at Moss Rehabilitation Research Institute

Scott Millis, PhD at Kessler Medical Rehabilitation Research and Education Corporation ☑

SANTA CLARA VALLEY MEDICAL CENTER

Rehabilitation Research Center for TBI & SCI
Santa Clara Valley Medical Center
950 South Bascom Avenue, #2011
San Jose, CA 95128

UPDATE
Center for Outcome Measurement in Brain Injury (COMBI)
<www.tbims.org/combi>

Winter 2001