ABSTRACT
With any manual data entry process, it is expected that data entry errors exist within the database. To assess the integrity of the data, it is often of particular interest to measure the extent of these errors. This paper will demonstrate a method that can be used to calculate an error rate.

The process of determining the error rate involves the following: a visual comparison of a random sample of entered pages to the database; the recording of any errors in a separate database; and the conversion of this "error" database to a SAS dataset using PROC IMPORT. In addition, the Data Step is used to perform minor data management. Ultimately, PROC MEANS in conjunction with the Output Delivery System will produce summary statistics that can be used to calculate an overall error rate as well error rates of specific interest.

INTRODUCTION
The Lewin Group, Inc. is a health care consulting firm which serves the spectrum of the US and international health care services and products. Our Medical Technology Practice specializes providing expertise to the pharmaceutical industry by conducting with late phase clinical trials, effectiveness trials, and health outcomes research, and providing medical product reimbursement strategy and support.

Data entry is an integral part of the provision of our services. Data provided by study participants is generally received via mail or fax on paper case report forms (CRFs) and manually entered into a database. Data can also be faxed directly into a database, or entered electronically over a secure internet site. Other forms of data collection include touch-tone phone and computer-assisted telephone interview (CATI).

All data entry processes are subject error. Measuring the integrity of data can involve assessing the level of error present. This paper will demonstrate one method that can be used to calculate an error rate.

Basic Study Design
Physicians are contacted to enroll in disease-specific studies. The physicians, in turn, enroll eligible patients affected with the disease of interest. As a part of the study, patients are asked to complete a questionnaire which often measures treatment satisfaction, quality-of-life, or health resource use. Patients are common asked such questions as:

In general, would you say your health is (circle one number):

- Excellent 1
- Very good 2
- Good 3
- Fair 4
- Poor 5

The questionnaires are then sent to us, and are manually entered into a database. These data are finally imported into SAS.

METHODS
A quality assurance audit can be conducted while a study is in progress, or at the completion of a study. In either case, assessing the level of error present due to data entry is essentially the same. For the most objective results, a random sample of data must be obtained and compared to the source documentation, the questionnaire completed by the patient. Errors must be tabulated, as well as the number of items on the questionnaire. An error rate can then be calculated.

Obtaining a Random Sample
A random sample of patient data can be gathered in several ways. Samples can be obtained by randomly selecting questionnaires immediately after they are manually entered. Alternatively, a random sample of patients may be obtained using the following code:

```sas
data library.random (keep = subject invid);
retain k 10 n;
set library.demog nobs = total;
if _n_ = 1 then n = total;
if ranuni(747088789) <= k/n then do;
output;
k = k-1;
end;
n = n-1;
if k = 0 then stop;
run;
```

Which produces the following output:
To random select a different number of patients, simply change the number in the line:

\[ \text{retain } k \text{ 10 } n; \]

**Computation of the Error Rate**

We are now ready to compute the error rate. PROC MEANS, along with the Output Delivery System, is used to determine the total number of item and errors in the dataset.

...
CONCLUSION
Periodically assessing the error rate of a manually entered data is helpful in assessing and maintaining the integrity of that data. Such assessments are simple to implement, and provide valuable information about the accuracy of manual data entry.

REFERENCES

ACKNOWLEDGMENTS
I would like to thank to Stefanie J. Silva, MS, for her guidance and assistance.

CONTACT INFORMATION
If you have further questions about this paper, please contact the author at:
   Edem Kuevor
   The Lewin Group
   490 Second Street, Suite 201
   San Francisco, CA 94107
   415.633.3100
   415.633.3133
   edem.kuevor@lewin.com