The Problem: We’re Vulnerable

What do sociologist Lenore Weitzman, the state of Arizona’s Independent Redistricting Commission, and the National Aeronautics and Space Administration (NASA) Mars Polar Explorer have in common? All were the subject of considerable media attention in their day. All were carrying out important work. All employed bright and talented people dedicated to getting the job done, yet each fell victim to the same problem: work delegated to others was either inadequately communicated or mismanaged, with disastrous consequences.

Published in 1985, Weitzman’s study of no-fault divorce in California—“The Divorce Revolution”—was heralded as a groundbreaking indictment of the legal system’s victimization of women.1 Following a divorce, Weitzman reported, females’ standard of living dropped 73%, while males’ increased by 42%.2,3 Weitzman’s finding captured remarkable public attention despite its inconsistency with other published work. In the decade that followed, Weitzman’s book was cited by 175 newspaper and magazine stories, 348 social science articles, 250 law review articles, 24 legal appeals, Supreme Court cases, and President Clinton’s 1996 budget. “There’s only one problem,” reported an Associated Press news story in 1996. “Her figures are wrong.”4

Investigation revealed that Weitzman had turned her calculations over to a research assistant, who had apparently made 1 or more error(s) in data analysis or processing. Another researcher, given access to the data in 1996, found not only discrepancies in the calculations themselves but also paper data collection records that did not match to Weitzman’s computer files.5

Charged with the controversial task of redrawing Arizona’s legislative boundaries, the state’s Independent Redistricting Commission encountered serious trouble in April of 2002. Testimony given in a legal deposition revealed that a serious miscommunication between the commission and a consulting firm had taken place months before. Active and inactive voters had inappropriately been combined in tallying district populations. “The mistake means further chaos for the state’s election system,” a local newspaper reported. “As filing deadlines approach, candidates don’t know where to collect the signatures and donations they need to run for office.”6

The Mars Polar Explorer was launched in January 1999 to find evidence of precipitation on Mars, but it crashed into the planet’s surface upon arriving on December 3, 1999. Investigation revealed that the 2 teams working on the project had used different measurement units, one metric (e.g., kilometers, kilograms), the other English (e.g., miles, pounds). Neither team knew how the other was carrying out its work, resulting in the failure to place the spacecraft into proper orbit. A NASA administrator assessed the situation: “People sometimes make errors. The problem here was not the error; it was the failure of NASA’s systems engineering and the checks and balances in our processes to detect the error. That’s why we lost the spacecraft.”7

While few of us working in managed care pharmacy are likely ever to be responsible for writing Supreme Court legal opinions, planning an election, or navigating a spacecraft, many of us routinely use quantitative analyses to inform decisions that affect the lives and health care of thousands and even millions of members of health plans. This makes us vulnerable to errors that can diminish the quality of the information we provide to others, affecting benefit management or possibly even patient care. And the more emotionally or financially attached we are to the results of our analyses, the more vulnerable we are to mistakes. Personal investment in our own results, a habit that most of us are guilty of at one time or another, tempts us to ignore the only reliable method to prevent small mistakes from becoming big problems: data quality control (DQC).

Data Quality Control Principles

Most of us confront DQC with little or no inherent interest. Results are much more intriguing than the “nuts and bolts” methodology that generated them. However, DQC can and should be incorporated into organizational culture as an essential precursor to information release and utilization. The idea is not to prevent every error, which is impossible given human fallibility; instead, the point is to keep errors from causing avoidable damage. Implemented properly, good DQC provides the assurance that all disseminated work has been appropriately verified for accuracy. DQC also facilitates easy responses to information requests and produces a net time savings from a relatively small time investment. Built on the fundamental principle that everyone makes mistakes, DQC relies on procedures instead of on individuals. People have bad days; procedures do not.

A good DQC program is composed of 3 components: (1) verification of what is received from others; (2) verification of what is provided to others, including documentation standards; and (3) internal quality control peer review procedures that are never overlooked, no matter how urgent the need for information (Table 1).

Verification of What Is Received From Others

In performing quantitative analyses, it is common to receive data (e.g., claims files, eligibility records, questionnaires) from others. Data received according to request should always be verified against stated criteria. For example, assume a researcher has requested a file of all pharmacy claims for the year 2005 for health plan members who were continuously enrolled through the year 2005 and had at least 1 diagnosis of rheumatoid arthritis (International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM] 714.xx) during the year’s first quarter. Upon receipt of the file, each requirement of the request (all pharmacy claims, continuously enrolled, diagnosis of rheuma-
Primary care providers reported that they had deliberately linked to payment. For example, in a 1994 survey, 50% of the patient’s condition is socially stigmatized or in any way affects patients’ out-of-pocket cost (e.g., when a pharmacy pharmacy technician at the point of service and potentially payment to take place, and (c) often edited real time in claims field on a drug claim is usually reliable because it is (a) typically tied to a pharmacy’s billing and inventory system, (b) essential for payment to take place, and (c) often edited real time in claims transaction processing systems. The “days supply” field, which is entered in the claim transaction by the pharmacist or pharmacy technician at the point of service and potentially affects patients’ out-of-pocket cost (e.g., when a pharmacy benefit plan calls for a 30-day supply maximum but the physician has prescribed a 35-day supply), is less reliable. Diagnosis codes pose a particular challenge, especially when the patient’s condition is socially stigmatized or in any way linked to payment. For example, in a 1994 survey, 50% of primary care providers reported that they had deliberately miscoded major depression as a different diagnosis at least once during the prior 2 weeks, most commonly because of uncertainty about the diagnosis or concerns about obtaining reimbursement for a mental health diagnosis. In a 2004 letter to JMCP, Dr. John Barbuto told a similar story: because “tension headache” was once considered a psychiatric diagnosis and therefore payable at a lower rate, physicians would avoid using that diagnosis and “miraculously, everyone seemed to have migraine.”

One type of analysis common to managed care pharmacy, assessment of the effect of a benefit design change, requires special verification procedures. Copayments, deductibles, and other benefit design features (e.g., coinsurance or mandatory payment of the brand-generic cost differential for multiple-source brand drugs) should be verified before analysis. Typically, it is necessary to remove the lowest-cost medications from these verifications because of policies that limit out-of-pocket outlays to the lesser of the expected copayment or the medication cost. For example, for evaluation of a 3-tier structure of $10/$20/$30, the analyst should select generic drugs, with an ingredient cost >$10; preferred brand drugs, with an ingredient cost >$20; and nonpreferred brand drugs with an ingredient cost >$30 and verify that copayment amounts are as expected for each copayment tier. If the copayment amounts on the claims do not fall into the expected pattern ($10/$20/$30 in this example), additional analyses will be necessary to detect the source of the discrepancy. The analyst should examine copayments by month (did the copayment distribution inexplicably change over time?) and by therapy class or drugs (do the copayment distributions look accurate for some classes or drugs but not others?). Common sources of copayment discrepancies are: (1) “grandfathering” (i.e., charging the tier-2 copayment for the first month after a medication’s status changes to non-formulary [tier-3 copayment]), (2) mid-year changes in tier status, and (3) charging the generic copayment amount for certain drugs, e.g., maintenance drugs. Decision rules are needed to handle each of these situations. This verification process, though somewhat tedious, helps the analyst to make conscious a priori choices about how to handle special, potentially important factors affecting the integrity of the work.

Claims processing time and payment cycles should be considered as well. So that results are not affected by possibly irregular payment cycles, date of service, not paid date, should be used to define time periods whenever possible. Analysts should also be aware of the sources of the claims data being used for a project. Paper claims, which must be submitted and manually keyed before processing into a data file, introduce both delay and possible inaccuracy into the claims data. Analysts working on longitudinal studies should be especially aware that claims with dates of service 10 or 15 years ago might have been processed very differently than today, when much processing is carried out with “real time” verification. Claims submitted by members are subject to the so-called “shoebox
**TABLE 1** Checklist for Data Quality Control

<table>
<thead>
<tr>
<th>1. Documentation</th>
<th>2. Work performed matches specifications (calculations in printout match printout)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Written specifications indicate work to be performed and clear definitions of terms (e.g., washout, termination, cost)</td>
<td>4. Computer printouts include annotation at the start of the printout indicating the program name, date created, where stored, and identity of analyst</td>
</tr>
<tr>
<td>5. Computer printouts include annotations throughout explaining each major step</td>
<td>6. Each step in analysis includes a verification of accuracy</td>
</tr>
<tr>
<td>7. Sample tracking document shows N at each stage of sampling process</td>
<td>8. Sample tracking document shows N at each stage of sampling process</td>
</tr>
<tr>
<td>9. File names match what the files are (e.g., claims, eligibility)</td>
<td>10. File is as specified (e.g., patient demographics and utilization, dates, N)</td>
</tr>
<tr>
<td>11. Previously published or known results are replicated before using file</td>
<td>12. Benefit design features (e.g., copayments, deductibles) are verified before beginning analysis</td>
</tr>
<tr>
<td>13. “Claims completion” is verified for entire analysis period</td>
<td>14. Data from different files (e.g., claims and eligibility data) are consistent</td>
</tr>
<tr>
<td>15. Results (e.g., prevalence rates) are reasonable</td>
<td>16. Results (e.g., prevalence rates) are reasonable</td>
</tr>
<tr>
<td>17. Documentation is clear even to persons unfamiliar with the work</td>
<td>18. Counts of patients or cases track throughout document (N’s consistent from one stage to the next and in sample tracking document versus printouts)</td>
</tr>
</tbody>
</table>

**TABLE 2** Suggested Principal Investigator (PI)-Analyst Responsibilities for Data Quality Control

**Principal Investigator**

1. Specifies clearly project objectives and methods, in writing whenever possible
2. Obtains Institutional Review Board approval or waiver before beginning work
3. Documents any changes necessary, the date made, and the reason for the decision
4. Verifies that the analysts’ work products meet specification(s)
5. Verifies that written reports (e.g., papers, presentations) match the procedures actually performed by the analyst
6. Permits information release only after internal peer review verification is performed

**Analyst**

1. Ensures that work performed matches project specifications provided by PI
2. Ensures accuracy of quantitative analysis before providing PI with results
3. Ensures that project documentation contains sufficient information for PI or other objective third party to verify quality of work
4. Maintains easily accessible records
5. Safeguards any protected health information, per HIPAA guidelines
6. Discloses results of quantitative analysis only after PI has completed review and authorized release

**HIPAA = Health Insurance Portability and Accountability Act.**

---

**Editorial**

Effect,” in which claims can be stored for months before being submitted to the payer for reimbursement. “Claims completion,” i.e., whether all claims for a given time period have been received and processed through to the administrative data files, should be verified by examining claim counts for each month of the time period of interest.

**Verification of What Is Provided to Others**

A fundamental factor in providing accurate information to others is to have well-defined roles for data analysts and study decision makers, particularly principal investigators (PIs). This does not imply that rigid roles are necessary. The appropriate roles might vary from one organization to another; a reasonable structure is shown in Table 2. The goal is to ensure a comprehensive approach to quality, which prevents gaps caused by each party assuming that the other was responsible for a piece of the process. In implementing policies of this type, it should be emphasized that roles are minimum guidelines, not excuses to shift responsibility for accuracy to someone else.

Documentation standards for the PI are also key to a successful quality control process. Before work begins, PIs should specify in writing:

- **Who** – sample selection criteria; inclusion, exclusion
- **When** – time periods for sample selection and analysis
- **What** – procedures to be performed
- **Why** – briefly give sense of context, what it is hoped will be learned (helps analyst watch for possible problems or opportunities)
- **How** – describe any special techniques to be used, along with resources (textbooks, manuals, etc.) to be consulted

One important responsibility of a PI, verification of the analyses performed, requires that work output (e.g., study reports, data files, computer printouts, spreadsheets) be stored in an accessible location, using intuitively obvious file names. Files of claims data, for example, should have the word “claims” somewhere in their name so that even when the primary analyst is not available, the decision maker or another analyst can locate needed data quickly. Whether stored in electronic or paper format, work output should be clearly verifiable in ways that are understandable and accessible to the PI.

Annotation of printouts is a key component of that accessibility. When writing analytic code, analysts should begin each program with a comment line explaining what the program does, where it is located, the date that it was last modified, and the analyst’s name. At each step in the program, a comment line should explain in simple terms what the statistical code does so that someone reading it at a later date can easily understand the procedures and how they were applied. A sample annotation at the start of the program should provide a brief explanation of what the program does (e.g., “This job calculates treatment termination rates for each age group.”) along with the name and location of the program file, the identity of the programmer,
and the date that the program was created. A sample annotation at the start of a section of code should explain what steps the code is taking and why (e.g., ‘Identify claim date; this will be used to identify the earliest and latest use dates for the target medication.’).

It is also strongly recommended that analysts complete each major step of a computer program with a visible verification that the step was successful. This verification is a check not only on the coding process used but also on the source data. For example, an analysis of cost data should include basic descriptive measures, such as minimum, maximum, median, mean, and interquartile range, so that the analyst and PI can verify that the calculations used (e.g., mean cost) are appropriate for the data. A classification of patients by age group should be verified using a check of descriptive measures on age for each group (e.g., verifying that patients in an “18-to-24-year-old” age group have minimum and maximum ages of 18 and 24 years, respectively). A matching of claims to eligibility data should include verification that every person with a claim has an eligibility record and that a reasonable proportion of eligible members has at least 1 claim.

The appropriate form of verification varies depending on the situation. For example, a reclassification of one variable’s discrete categories into another variable can be verified with a simple cross tabulation of the 2 variables. A “list cases” command can be used to verify some types of code. A good “rule of thumb” for the analyst to follow is that output should include all the information necessary for a reasonable person to verify, based on the job output alone and with no other knowledge of the project procedures, that the work performed matches specifications.

The tracking of sample size from the beginning to end of the sampling process is both a critical component of verification of study procedures and very strongly recommended for articles published in JMCP. A summary table (see Figure 1 for an example) should track the effect of each inclusion or exclusion criterion on sample size. For each step in the process, numbers in the table should match exactly to the job output available to the PI. Counts that do not match are a sign of a coding error, methodology that mistakenly excluded certain subgroups, or an incompatibility between code and data; these occurrences require further investigation.

One particularly helpful step for work intended for publication or presentation is to document the source of each finding in the internal study files. Analysts should keep copies of each data table included in a publication or presentation, along with an annotation indicating the name of the computer program that produced the finding. (Example annotation: “Source: Job name is ‘demographics.sps’ in the XYZ subdirectory on the ABC computer, run date 9-22-06”.) The annotation enables the PI to locate and respond to questions easily even when the analyst is not available because the findings table can quickly be tracked back to the source code. Following receipt of peer-review comments, returning to the initial study files to perform any necessary reanalysis becomes a relatively easy task. Reanalysis is not uncommon, given the need by journal editors to ensure that results provide useful information for readers.

**Internal Quality Control Peer Review**

No quantitative results should be released until they are verified by a second person, an “internal peer reviewer” (IPR) who compares the work performed with the project specifications, checks the integrity of the figures produced, and looks for problems affecting interpretation. The IPR process should not be seen as an excuse to transfer all responsibility for quality to a second person; instead, both the original producers of the findings and the IPR share in responsibility for the final work product.

Having a relatively standard package of materials to provide to an IPR facilitates his/her work. The package should include: (1) written specifications for the project, such as a study plan or proposal document; (2) printouts (or links to electronic versions) of all programs relevant to the final study work product, such as computer jobs that create new data files from source files, pull a sample, or perform calculations; (3) a guide to the order in which steps were undertaken, such as the sequence of computer jobs; (4) the sample tracking document; and (5) any other information necessary for the IPR to understand how the work was performed and why.

The IPR should compare all steps in sampling, processing, and calculation with the initial written specifications for the project. Deviations from the plan should be reasonable and, preferably, documented. For example, if the study’s original plan was to analyze results separately for different clinics but one
The clinic was dropped because it had only a few patients, this change should be noted somewhere in writing, preferably with the date of the decision and the person responsible.

The IPR should also ensure that counts of cases are consistent throughout all calculations and match the sample tracking document. For example, if there were 1,000 cases in the dataset at the end of the first step in sampling, both the tracking document and all computer printouts should indicate 1,000 cases not only at the end of the first step but also at the start of the second step. Cases should never just “disappear;” a change in case counts should be both explainable and visible in the documentation. Finally, calculations and processing steps should be verified. For example, if a grouped-age variable was to be created from date of birth, the IPR should verify that the grouping was accurate and consistent with the data. Was the coding performed correctly? Did any cases have missing or out-of-range values for age, and, if so, how did the grouping handle them? The IPR should note where the documentation does not contain sufficient information to answer these questions.

A key element of success is the authority of the IPR to request additional information when necessary. An IPR who is charged with “signing off” on a project but not given access to all the necessary information will become, at best, frustrated and, at worst, a “rubber stamp.” It should be clearly understood by all analysts and PIs that internal peer review is in their best interest and that they are obligated to cooperate fully with the process. An organizational culture that embraces peer review is ultimately necessary to ensure high-quality administrative claims research.

Conclusion
As the cases of Lenore Weitzman, the Arizona redistricting project, and the Mars Polar Explorer demonstrate, catastrophic errors are often the result of simple and avoidable mistakes. DQC procedures greatly reduce the possibility of these unfortunate events. When DQC is consistently applied, the ubiquitous problems of human error and miscommunication can be reduced to measurable and largely controllable impediments to producing high-quality, reliable quantitative information. The lives of health plan members will be affected by our attention to DQC.

Kathleen A. Fairman, MA
Outcomes Research Consultant
JMCP Associate Editor and Senior Methodology Reviewer
kathleenfairman@qwest.net

DISCLOSURES
The author discloses no potential bias or conflict of interest relating to this article.

REFERENCES