



## Troubleshooting

**Laurie A. Cooley and  
John W. Dolan**

**But it worked for  
someone else . . .**

# Reproducibility and Carryover — A Case Study

**A**s part of our training program for new employees, our laboratory uses several integrated tests to verify that chemists can perform certain routine tasks with the acceptable level of expertise. One of these tests is a bioanalytical proficiency test that combines preparation of spiked plasma samples, solid-phase extraction (SPE), liquid chromatography (LC) analysis, and data work-up. During training, new analysts typically are taught how to use pipettes and balances, prepare standards and mobile phases, and operate the LC system. However, it is not until all these techniques are used together that supervisors have confidence that analysts have a good grasp of the techniques. This test is intended to simulate a typical process that workers might encounter with client samples. The accompanying sidebar, "Skills Included in the Proficiency Test," lists some of the analytical techniques involved.

This month's "LC Troubleshooting" column deals with reproducibility and carryover problems that were encountered during the development of a method used to verify analysts' proficiency. The problems encountered and the approach to isolating and correcting the problem sources will serve as examples for readers who may experience similar problems with methods for analyzing drugs in plasma.

### The Method

The method under development for bioanalytical proficiency training is an LC method for analyzing a nonsteroidal anti-inflammatory drug (naproxen) in dog plasma. Plasma samples are spiked with the drug and an internal standard (indomethacin). Analysts use these spiked samples to generate a standard curve for calibration and to simulate samples that are quantified using the standard curve. (It should be noted that this discussion concerns a method that is still in the development stages. It has not been validated and is not used for measuring naproxen levels in real

samples; rather, it is used only as a proficiency test.)

Sample pretreatment is by SPE. The C18 SPE cartridge is conditioned with methanol and again with 0.1 M phosphate buffer (pH 3). Analysts load an aliquot of the sample on a cartridge and wash it with buffer, buffer with 5% methanol, and finally buffer with 25% methanol to remove hydrophilic interferences. The drug then is eluted with methanol and blown to near dryness in a water bath under a stream of dry nitrogen. The sample is reconstituted in 25 mM phosphate buffer (pH 3).

A 150 mm  $\times$  4.6 mm, 5- $\mu$ m  $d_p$  C18 column is operated at 35 °C isocratically at 1.5 mL/min using a mobile phase of 65:25 (v/v) methanol–25 mM phosphate buffer (pH 3). UV detection is performed at 254

### *Skills Included in the Proficiency Test*

- Following a written procedure
- Using balances, pipettes, and pH meters
- Preparing buffers and mobile phases
- Handling biological samples
- Preparing standard curve samples in plasma
- Using internal standards
- Using solid-phase extraction techniques
- Using general laboratory equipment, including vortex mixers, centrifuges, and nitrogen evaporators
- Setting up and operating LC systems
- Determining system suitability
- Demonstrating data system proficiency (sequence files, data acquisition, calibration files, integration, batch processing, and report formats)
- Calculating recovery, linearity, and assay content
- Entering data in laboratory notebook
- Preparing reports
- Filing and records disposition

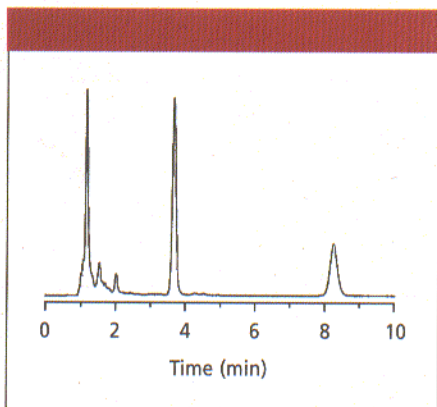
**John W. Dolan**  
LC Troubleshooting Editor

nm. The sample injection size is 100  $\mu$ L. Workers use an autosampler flush solvent of 50:50 (v/v) methanol–water.

The analysts determine system suitability by injecting six replicate samples of an aqueous solution of a drug and an internal standard. Performance is acceptable if the relative standard deviation (RSD) is less than 2% for peak area and retention time. Carryover between the last system-suitability sample and a blank injection should be less than 1% of the area of the naproxen peak.

Following the system-suitability samples, the standard curve calibration samples are analyzed, and then the spiked samples are injected and quantified against the standard curve using internal standardization. Figure 1 shows a typical chromatogram (30 ng/mL). The standard curve fit should be linear with  $1/x^2$  weighting of peak areas with the regression coefficient ( $r^2$ ) greater than 0.95. Performance is acceptable if all points on the standard curve are within 15% of the curve, and carryover after the highest standard curve level (300 ng/mL) is less than 1%. Recovery for all of the assayed samples must be within 15% of the expected assay value.

Readers who do not regularly perform bioanalytical work may be shocked by the seemingly broad limits for precision and accuracy of this method. These levels are typical of bioanalytical methods that involve extensive sample preparation, single injection, and nanograms-per-milliliter drug levels. Most of these methods specify recovery values within 15% of the expected amount at all levels except the limit of quantification, at which  $\pm 20\%$  is acceptable.



**Figure 1:** Typical chromatogram for a 100- $\mu$ L injection of a 30-ng/mL sample of naproxen (3.8 min) and indomethacin internal standard (8.3 min).

## The Results: Runs 1 and 2

An analyst performed the method as described above. The system-suitability test showed retention and peak area variability significantly less than the 2% limits. However, carryover was greater than 5%. When analyzing the spiked samples, the analyst observed carryover of more than 6% for a blank injected after the 300-ng/mL sample. Furthermore, assay values for the samples (ranging from 30 to 300 ng/mL) were  $\pm 5$ –70% of the expected amount. The analyst discussed these clearly unacceptable results with the supervisor. Following a review of the data, the supervisor felt that the problem may have originated in the sample preparation steps, so the analyst repeated the entire procedure.

The analyst obtained similar results on the second execution of the procedure. The standard curve samples had deviations of as much as 19%, error for some of the samples exceeded 40%, and more than 6% carryover was observed in a blank sample.

## Potential Problem Sources

At this point, we did not know if the observed problems were related to the instrumentation or to sample preparation. When analysts encounter problems that may come from one of several different sources, they should divide the suspected causes into several broad categories. The possibilities were broken down into four potential problem areas: system suitability, injection reproducibility, carryover, and sample preparation. We decided to look at each of these areas in a series of experiments, which we will discuss in the following sections.

In addition, we made two changes to the original method. First, we changed the reconstitution solvent from aqueous buffer to 50:50 (v/v) methanol–buffer. During standard preparation we had observed that the analytes were poorly soluble in water. We reasoned, therefore, that if no organic solvent was included in the reconstitution solvent, sample recovery at this step could be variable. As long as the injection solvent is no stronger than the mobile phase, we expect that it should cause no degradation of the separation. With a 65:35 (v/v) methanol–buffer mobile phase, the 50:50 (v/v) methanol–buffer injection solvent should be satisfactory.

Second, the strength of the autosampler wash solvent was increased from 50:50 (v/v) methanol–water to 75:25 (v/v) methanol–water. We wanted to ensure that the wash solvent was completely removing

any sample residue from previous injections. In general, we would like the autosampler wash to be at least as strong as the mobile phase (65:35 [v/v] methanol–buffer), so we chose 75:25 (v/v) methanol–water. We hoped that this change would eliminate the carryover we observed. Typically we do not like to use buffers in the autosampler wash because of the potential for microbial growth and residues during evaporation, so the wash solvent was unbuffered.

## System-Suitability Tests

System-suitability tests should be designed to show that the LC system hardware and basic chromatographic separation are performing in a manner acceptable for routine analysis. A typical LC system with an autosampler and a UV detector should be able to generate peak areas varying less than 1% in area and less than 0.1 min in retention time under conditions similar to those used in this method. Thus, 2% limits in retention time and area are quite generous.

To examine the system-suitability test more thoroughly, we injected a series of six 100- $\mu$ L samples of the analyte at 300 ng/mL (in 50:50 [v/v] methanol–buffer) followed by two blank injections of only injection solvent. This process was repeated for a total of 16 injections. The RSD for peak area for naproxen was 0.5% for the first set of injections and 0.4% for the second set. We obtained similar results for the internal standard areas (0.3% and 0.4%). The RSD for retention times for all peaks in both runs was less than 0.03%. Carryover was 0.4% and 0.5% for the two series. All of these values passed the system-suitability requirements, which indicates that the injection, pumping precision, and carryover in the LC hardware were adequate for the method. We felt that the change in injection and wash solvents was responsible for eliminating the carryover problem, at least for the system-suitability samples.

## Injection Precision of Spiked Samples

The successful system-suitability tests showed us that the hardware was working properly, but we weren't sure if the system would perform in a similar manner with real samples. The high variability in assay values obtained initially could be attributed to sample-preparation variability or injection variability. When we made one injection of each sample, it was impossible for us to determine which of these factors was



the problem source. To help determine which factor was responsible, we prepared four plasma samples at 300 ng/mL and four at 30 ng/mL. The samples at each level were pooled to generate a sufficient amount of sample at each level for multiple injections of identical solutions. We made four 100- $\mu$ L injections of the 300-ng/mL pool, followed by two plasma blanks, then four 30-ng/mL samples, and finally two more blanks.

These naproxen experiments yielded area RSD values of 0.4% at the 300-ng/mL level and 0.1% at the 30-ng/mL level. The internal standard RSDs were less than 0.3%. In addition, retention time deviation for all peaks was less than 0.3%. Carryover following the high-level sample was 0.1% and less than 2% for the low level, both of which are within specifications. These results convinced us that the LC system was working properly. We concluded that the high variability and carryover observed initially was not caused by autosampler performance or sample matrix effects.

### Carryover

The system-suitability and injection-precision tests included measurements of carryover. In addition, we set up three series of injections of a high-concentration plasma sample (300 ng/mL) followed by two blank injections. These samples yielded results consistent with those of the other tests. Carryover was less than 0.2% for all three carryover series.

In each case, blank injections yielded peaks with approximately 4000 area counts, which were consistent with the

results of the injection-precision tests at both high and low levels. For data system and methods such as this one that generate fairly quiet baselines, an area of approximately 400 units is about as small an area as we can confidently call a peak. The examples of Figure 2 show this situation. It is clear that a real peak is present for the plot of Figure 2a, which has an area for naproxen of roughly 4000 units. However, in the plot shown in Figure 2b, the peak has an area of just more than 100 units and is difficult to distinguish from long-term baseline noise (note that Figure 2b is plotted on a scale that is 10 times more sensitive than that in Figure 2a). So, although all the carryover tests yielded results less than the 2% limit, the carryover as shown in Figure 2a certainly is not zero.

### Sample Preparation

The remaining portion of the method to test was sample preparation. When the method was run initially, it was unclear whether the variability was due to sample preparation, injector performance, or some other problem. With successful results from the tests performed to this point, we suspected that sample preparation was the central cause of problems. We prepared two plasma pools: one spiked at 300 ng/mL and one at 30 ng/mL. From these pools we ran four separate samples at each level through the sample preparation process (eight samples total) and injected each sample twice. We felt that the increased number of samples and replicate injections should help to isolate problems in the sample preparation process. All samples yielded results that were within 10% of the expected levels and well within the  $\pm 15\%$  method limits.

At this point, we concluded that the changes to the autosampler wash solvent and the sample reconstitution solvent corrected the earlier problems with the method. The next step was to repeat the entire protocol to ensure that it worked as required.

### Results for Run 3

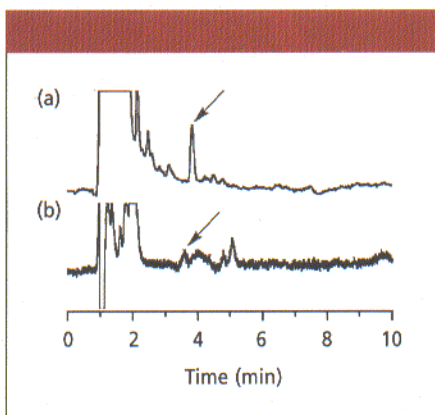
We repeated the entire protocol and expected the results to meet the acceptance criteria with ease. The standard curve was acceptable only if the 50-ng/mL standard was dropped. The deviations in assay values for all the samples were within the acceptance limits, but blanks continued to exceed 5% of the area of the preceding 300-ng/mL injection. This result was not what we expected to see.

### Equipment Problem

During the work-up of the last set of SPE samples, an analyst noticed that one of the Luer fittings on the SPE manifold was cracked, so a vacuum leak occurred for that SPE tube. When we disassembled the manifold lid to replace this cracked fitting, we noticed that the manifold tips were cracked and had crusty deposits. The manifold normally is fitted with small plastic tips, similar to disposable tips used for automatic pipettes, that act as small funnels to direct the SPE effluent into collection tubes. During normal operation, these tips are rinsed with SPE eluent during processing and rinsed with the cartridge activation solvents at the beginning of the next cycle. By our experience, these washes are sufficient to clean the tips so that no separate cleaning is necessary. In this case, however, it appeared that the tips had become damaged and leaked sample down the outside, as well as inside the tips. The crusty residue made us suspicious that we had located the major source of carryover.

As a check of our theory, we removed two sets of eight tips each from the manifold and extracted their deposits by sonication in methanol. The extract was blown to dryness and reconstituted in 50:50 (v/v) methanol-buffer injection solvent. We observed large peaks (about 100,000 area units) in each sample, which semiquantitatively correlated with approximately 15 ng of naproxen extracted from each tip.

An additional test was made by preparing two blank plasma samples. One sample was prepared using a dirty, unsonicated manifold tip, and the other sample was prepared using one of the sonicated tips. The dirty tip produced a peak about twice the size of the peak observed for the sonicated tip, as Figure 3 shows. It is interesting to note that the peak in the dirty tip sample had an area of approximately 4000, which was typical for the carryover peaks observed in previous experiments. In this case, however, we injected a reagent blank rather than a high-level sample before the plasma blank, which confirmed that the peak was picked up by the plasma blank during processing. Because the plasma had been tested and found to be free of naproxen and any other coeluted interferences, the naproxen must have originated from a dirty tip in the sample processing step.



**Figure 2:** Chromatograms showing (a) a typical carryover peak for naproxen (arrow) with an area of 3799 units and (b) a peak with an area (118 units) near the limits of detection. The intensity scale for run (b) is 10-fold more sensitive than for (a); see text for details.

## Final Run

We replaced all the manifold tips and repeated the entire procedure. The results easily passed the acceptance criteria in all cases.

## Conclusions

This case study can teach us a number of things that may help us solve future problems. One of our conclusions is the more eyes that review the method, the better. Although the method had been developed and had worked successfully in one person's hands, it did not work well for someone else. Two aspects of the method — the injection solvent and the autosampler wash — may not have been a problem, but during the review of the method by a more experienced chromatographer, these two items stood out as undesirable selections. It is a good idea to have several people examine a method before finalizing it. Of course, normal exercise of the method in the prevalidation stage of method development should identify the aspects of a method that affect method ruggedness.

When we obtained the initial results, we could not determine whether the problem was related to the analyst or the method.

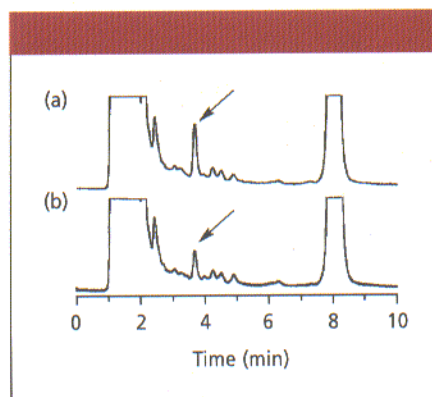
Usually the easiest approach at that point is to repeat the entire analysis. In troubleshooting, sometimes we refer to this as the *Rule of Two* — make sure that a problem occurs at least two times, otherwise it will be difficult, if not impossible, to isolate the problem source. Again, these tests should be performed during prevalidation or validation when the method is tested by more than one operator.

It is a good assumption that a problem with a method is due to one cause or one primary cause. In the present case, it is unclear whether the method adjustments really made a difference or not, but certainly the cracked fitting and dirty tips on the SPE manifold were the primary source of problems.

To isolate a problem, the divide-and-conquer technique often is the most expedient. In the present case, we were unsure if the problem was related to the LC equipment, the SPE method, or other aspects of the protocol. We selected four areas to review. Although tests of these areas were discussed as separate experiments, they were performed in one series of runs to minimize the total investment in troubleshooting time. The results from

each test increased our confidence in the performance of different aspects of the method and helped to isolate the problem source.

It is a good idea to run quality-control samples randomly among samples of unknown assay levels. Quality-control samples are samples that are spiked at known levels, then interspersed with real samples and run throughout the entire



**Figure 3:** Carryover peaks in chromatograms generated using (a) a dirty SPE manifold tip and (b) a sonicated SPE tip for sample preparation. Scale same as in Figure 2a. See text for details.



procedure. The percentage recovery of a known level of analyte helps ensure that the method is performing as expected. For the present test, all the samples fall in the quality-control category because their content was known in advance.

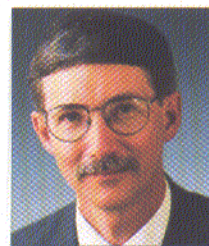
Finally, this example illustrates the value of an all-up proficiency test — our terminology for a test that combines all the different factors and examines the whole system as it would be used normally, rather

than a test that evaluates the individual parts. Just as the chemist had been trained to perform each of the steps in isolation, each of our tests showed that the different parts of the protocol were working properly. However, when the entire protocol was run as a unit, it failed. We feel that these types of tests increase the quality of the training experience and the confidence of new operators before they begin working with client samples.

*Laurie Cooley is a chemist for LC Resources Inc. in McMinnville, Oregon. Her research interests include development and application of LC methods for drugs in plasma and other tissues.*

#### **John W. Dolan**

*"LC Troubleshooting" editor John W. Dolan is president of LC Resources Inc. of Walnut Creek, California, and a member of LCGC's editorial advisory board.*



*Direct correspondence about this column to "LC Troubleshooting," LCGC, 859 Willamette Street, Eugene, OR 97401, e-mail [John.Dolan@LCResources.com](mailto:John.Dolan@LCResources.com).*

*For an ongoing discussion of LC troubleshooting with John Dolan and other chromatographers, visit the Chromatography Forum discussion group at <http://www.chromforum.com>.*

#### **Errata**

The plot of Figure 3c in December's "LC Troubleshooting" ("Reducing Column Diameter in Gradient Elution — A Case Study," *LCGC* **18**[12], 1230 [2000]) should show an isocratic hold (dwell) of three column volumes, not five; the gradient volume is correct as given.

The last line of Table III in February's "LC Troubleshooting" ("Autosampler Carryover," *LCGC* **19**[2], 164 [2001]) is incorrect. The concentration should be 10, the response should be 25.1, and the error should be 150.9%. The accompanying text should read "Errors range from less than 1% to more than 150% for this particular sequence of functions."