Factitious Diarrhea Induced by Stimulant Laxatives: Accuracy of Diagnosis by a Clinical Reference Laboratory Using Thin Layer Chromatography

JOSHUA H. SHELTON, CAROL A. SANTA ANA, DONALD R. THOMPSON, MICHAEL EMMETT,* and JOHN S. FORDTRAN

Background: Surreptitious ingestion of laxatives can lead to serious factitious diseases that are difficult to diagnose. Most cases involve ingestion of bisacodyl or senna. Thin layer chromatography (TLC) of urine or stool is the only commercially available test for these laxatives. Such testing is considered highly reliable, but its accuracy in clinical practice is unknown. Our aim was to evaluate the reliability of TLC laxative testing by a clinical reference laboratory (CRL) in the United States.

Methods: Diarrhea was induced in healthy volunteers by ingestion of bisacodyl, senna, or a control laxative (n = 11 for each laxative group). Samples of urine and diarrheal stool were sent in blinded fashion to CRL for bisacodyl and senna analysis.

Results: TLC testing for bisacodyl-induced diarrhea revealed a sensitivity of 73% and specificity of 91% when urine was tested and sensitivity and specificity of 91% and 96%, respectively, when stool was analyzed. When diarrhea was induced by senna, the TLC assay for senna failed to identify even a single urine or stool specimen as positive (zero% sensitivity).

Conclusions: Considering the expected prevalence of surreptitious laxative abuse in patients with chronic idiopathic diarrhea (2.4%–25%, depending on the clinical setting), TLC of urine or stool for bisacodyl by CRL would often produce misleading results, and testing for senna would have no clinical value. The major problems are false-positive tests for bisacodyl and false-negative tests for senna.

* Address correspondence to this author at: Baylor University Medical Center, 3500 Gaston Avenue, Dallas, TX 75246. Fax 214-820-4837; e-mail m.emmett@baylorhealth.edu.

Received August 7, 2006; accepted October 12, 2006.
Previously published online at DOI: 10.1373/clinchem.2006.077883

Drug Monitoring and Toxicology
laxatives ingested by the study participants, and it is possible that prior knowledge of laxative intake contributed to the perfect results. Moreover, a 2-year pilot study from the UK concluded that some laboratories approached 100% reliability in detection of stimulant laxatives in urine, whereas others were <50% reliable (26). The tested laboratories were presumably located within different hospitals, and presumably did not include external reference laboratories.

We know of no previous study on the reliability of aCRL, where good performance is usually taken for granted. Therefore we conducted an experiment to assess the accuracy of TLC for detection of bisacodyl and senna ingestion, as performed for clinical purposes in the US.

**Materials and Methods**

**STUDY DESIGN AND SAMPLE SIZE**

The purpose of the study was to test the hypothesis that the observed sensitivity and specificity of TLC laxative testing by CRL are substantially less than the 100% reported by the research laboratories (19, 23–25) that developed the TLC assay for laxatives. Data collection was planned before samples were collected and analyzed. Analysis of urine was the primary focus because TLC testing for senna is reported to be highly accurate when performed on urine but not when performed on stool (19), and because some clinical experts advocate testing of urine for both bisacodyl and senna (8, 26). CRL offers laxative testing on both urine and stool, however, and does not state a preference for either specimen. Therefore, a secondary goal of our study was to evaluate the accuracy of laxative TLC testing on stool.

The sample size requirements for this study were based on a normal approximation to the binomial distribution for comparing a single proportion to a known proportion (30). The known proportion is assumed to be 0.99, which corresponds to a sensitivity of 99% for TLC detection of bisacodyl and senna in urine (approximating the perfect results actually reported by research laboratories that developed and evaluated this method of laxative detection) (19, 23–25). A substantial difference from this known proportion is conjectured to be an absolute 20% decrease (i.e., 0.79). With a one-sided a-level of 0.05 and statistical power of ~90%, the required sample size was calculated to be 11. Therefore, each of the 3 study arms (bisacodyl, senna, and control laxatives) involved 11 participants.

**TLC ASSAY**

According to information provided by CRL, their assay for bisacodyl and senna is based on the method described by de Wolff et al. (19). The de Wolff procedure consists of high-performance TLC in 2 systems, after pretreatment of urine or stool with β-glucuronidase and subsequent column extraction. Reference standards for bisacodyl include bisacodyl itself as well as its metabolic products (mainly dihydroxy compounds). After ingestion of bisacodyl by healthy people, the stool contains unchanged bisacodyl, whereas the urine contains only bisacodyl metabolites (19).

Recommended reference standards for anthraquinone laxatives consist of rhein and danthron. Rhein is the metabolic product of senna and cascara, and its presence in urine indicates ingestion of one of these drugs. Rhein is not found in stool after senna ingestion, and this is presumably why the de Wolff method does not detect senna ingestion by analysis of stool (19). The danthron standard is used only for detection of danthron ingestion (21). Danthron was withdrawn from the US market in 1987 because of its association with development of hepatic and intestinal tumors in laboratory animals.

CRL reports results for bisacodyl and senna as either “positive” or “none detected”.

**EXPERIMENTAL PROTOCOL**

The studies were approved by the Institutional Board for Human Protection of Baylor University Medical Center and informed consent was obtained. Experiments were conducted in healthy adult volunteers, 12 men and 10 women, whose mean age was 32 years (range 18–57). Most participants were employees of our hospital or students at a nearby seminary; the remaining participants were spouses or friends of these 2 groups of people. All study participants denied use of laxatives, or recent use of any other medication. They were paid a stipend for their participation. Some individuals participated in more than one arm of this experiment. The interval between such experiments was always >1 week, however, much longer than the time required for complete elimination of bisacodyl metabolites (24 h) or senna metabolites (32 h) from the urine (19, 31).

On experimental test days, participants were admitted to a clinical research laboratory at 7 AM, after they had fasted for 10 h. Under the direct supervision of the authors, participants ingested bisacodyl, senna, or a control laxative (Milk of Magnesia, castor oil, or polyethylene glycol). The doses of bisacodyl and senna were higher than those used by previous investigators (Table 1). The order of laxative ingestion was random. Participants ate no food for 8 h after laxative ingestion, after which they were provided a standard meal and resumed a normal diet. Fluid intake was unrestricted, and no other medications were ingested. Each participant had a personal bathroom, which was used for no other purpose on experimental test days.

During the 24 h after laxative ingestion, each bowel movement was quantitatively collected in separate containers. Stool frequency and the weight of each bowel movement specimen were measured. Urine was collected quantitatively in a single container during the 24-h period. Stool and urine specimens were stored in a refrigerator (3–6 °C) or in a Styrofoam cooler with freezer packs (4–10 °C). At the end of the 24-h experimental period, aliquots of the specimens were frozen (−11 °C).
The assignment of fictitious names and labeling of samples were performed by 2 authors (JHS and CAS) to double-check their accuracy. Urine and stool samples from the same person were not identified as having a common origin. Submitted urine samples were aliquots of 24-h urine collections. Each submitted stool sample was taken from a single bowel movement after the onset of diarrhea. Frozen samples, packaged in dry ice, were mailed to CRL by overnight express.

**STATISTICAL ANALYSIS**

The reference standard for ingestion of bisacodyl, senna, or a control laxative was our direct observation of laxative ingestion. The sensitivity and specificity of TLC for detection of bisacodyl or senna ingestion, as performed by CRL, were calculated from laboratory reports from CRL, by use of standard methods (32). For calculations of statistical significance, one-sided testing was used and a P value of 0.05 or less was considered statistically significant. To test the hypothesis that the observed sensitivities and specificities were <99% and to determine confidence intervals, a normal approximation to the binomial distribution was used. In addition, Fisher’s exact test was used to compare sensitivity and specificity of urine vs stool.

**Results**

All experiments were carried out between June 30, 2005, and December 12, 2005. Participants taking bisacodyl and senna developed diarrhea with the characteristics shown in Table 1. Diarrhea also developed in all participants who ingested the 3 control laxatives (data not shown). There were no unexpected adverse events or complications associated with these experiments.

Results of bisacodyl and senna testing by CRL are shown in Fig. 1 and Table 2. For purposes of calculating specificity and predictive value, ingestion of senna was considered an additional control for the bisacodyl assay, and vice versa.

Eight of 11 urine specimens from participants with bisacodyl-induced diarrhea tested positive for bisacodyl. The observed sensitivity of urine testing for bisacodyl was therefore 73%, lower than the expected sensitivity of 99% (P <0.001). Two of 22 urine samples from participants who had diarrhea due to laxatives other than bisacodyl also tested positive for bisacodyl, yielding a specificity of 91% (P <0.001). One false-positive urine test for bisacodyl was from a participant ingesting senna; this person had received bisacodyl in a prior experiment 7 weeks earlier. The other false-positive urine result came from a partici-

| Table 1. Doses of bisacodyl and senna, and their effect on stool characteristics. |
|-----------------|-----------------|
|                  | Bisacodyl | Senna |
| Dose, mga        |          |      |
| deWolff et al (19) | 5        | 24    |
| Morton et al (25) | 5        | 8.6   |
| Bytzer et al (24) | 8–52     |       |
| This study       | 20       | 69 (n = 2) |
|                  |          | 103 (n = 9) |
| Stool characteristics (this study) |
| BM frequency     |          |      |
| mean (SE), no./day | 5 (0.7) | 3 (0.4) |
| Stool weightb     |          |      |
| mean (SE), g/day  | 757 (78) | 373 (47) |

a Laxative doses used in this study are compared with doses used in previous reports on TLC. In each case, laxatives were administered as a single dose. The usual starting doses of bisacodyl and senna are 5 and 17 mg, respectively.

b Daily stool weights in healthy volunteers in our laboratory are as follows: men (n = 28), 157 (13) g/day; women (n = 29), 87 (8) g/day.

| Table 2. TLC by a CRL for detection of bisacodyl or senna ingestion in urine and stool. |
|-----------------|-----------------|
|                  | Sensitivity, %  | Specificity, % |
| Bisacodyl        |          |      |
| Urine analysis   | 73 (8/11) | 91 (20/22) |
| Stool analysis   | 91 (10/11) | 95.5 (21/22) |
| Senna            |          |      |
| Urine analysis   | 0 (0/11)  | 100 (22/22) |
| Stool analysis   | 0 (0/11)  | 100 (22/22) |
pant ingesting polyethylene glycol, who had never taken bisacodyl.

As shown in Fig. 1 and Table 2, TLC testing for bisacodyl in stool had a higher sensitivity and specificity than TLC testing of urine, but the differences were not statistically significant. For stool and urine testing, there was no correlation between 24-h stool weight or urine volume and the likelihood of false-positive or false-negative tests for bisacodyl.

Of the 22 submitted urine and stool specimens from 11 participants with senna-induced diarrhea, none were reported as positive for senna, yielding a sensitivity of 0% (P < 0.001 compared with the perfect results reported by research laboratories). Similarly, all the urine and stool samples from participants who ingested nonsenna laxatives were negative for senna, yielding a specificity of 100%.

**Discussion**

Our results were obtained after participants ingested a single large dose of bisacodyl or senna. Even larger doses would presumably result in higher concentrations of laxatives or their metabolites in urine or stool, and this might improve the sensitivity of laxative testing. Smaller doses might have the opposite effect. By the same token, chronic ingestion over weeks or months may cause greater concentrations of the drugs and their metabolites in urine and stool and result in increased sensitivity of laxative detection tests. However, for several reasons it seems best to do performance testing of laxative detection methods in participants who have ingested a single dose of laxative. First, a number of previous studies (19, 23–25) reported excellent sensitivity and specificity when single doses (smaller than the doses used in the current study) were ingested by healthy volunteers. If a laxative detection test is reliable, it should be able to detect laxatives after a single dose, especially if the single dose is large enough to produce diarrhea. Second, some patients who surreptitiously abuse laxatives do so intermittently rather than on a daily basis. A reliable test for laxative ingestion should be able to detect intermittent as well as daily laxative abuse. Finally, giving laxatives on a long-term basis to volunteer participants for proficiency testing would be unsafe.

The results for bisacodyl and senna were strikingly different. Therefore they will be discussed separately.

**BISACODYL**

TLC analysis of urine for detection of bisacodyl ingestion was less sensitive and less specific than results reported by research laboratories. It is important to emphasize, however, that we submitted samples to CRL in a blinded fashion, whereas published studies from research laboratories made no claim that samples were analyzed in a blinded fashion. It is therefore possible that the differences between the present and previous results reflect observer bias in the historical data rather than technical inaccuracy on the part of CRL.

Stool specimen results obtained by CRL were somewhat more accurate than those achieved with urine specimens, possibly reflecting high concentrations of bisacodyl in stool and relatively low concentrations of bisacodyl metabolites in urine (19). Although most current recommendations advise testing of urine rather than stool specimens for detection of laxatives, our results suggest that stool analysis is just as good, and possibly better, for detection of bisacodyl.

Based on the observed sensitivity and specificity obtained with stool samples analyzed by CRL (Table 2), we calculated the predictive value of a positive test for bisacodyl in stool as a function of the prevalence of bisacodyl ingestion in various populations. Their combined findings revealed surreptitious ingestion of bisacodyl in 8 of 330 patients (2.4%). At this prevalence, only 29% of positive stool tests for bisacodyl would correctly identify that a patient had ingested bisacodyl, and 71% would represent false-positive results. On the other hand, bisacodyl testing would be more accurate later in the diagnostic evaluation of patients with idiopathic diarrhea, after extensive testing had failed to reveal a cause of diarrhea, when the prevalence of factitious diarrhea is ~25% (4). Even at this higher prevalence, however, ~15% of positive stool tests for bisacodyl would be false positives (Fig. 2), and 9% of negative tests would be false negatives. If urine results were used, the inaccuracy of interpretation of positive and negative tests would be higher than depicted in Fig. 2 for stool.

If a positive test from either a urine or a stool sample from an individual was considered a positive result for that person, the sensitivity of TLC bisacodyl testing by CRL would be 100%, but the specificity would fall to 86%.
If the prevalence of bisacodyl ingestion in a population were 10%, then more than half of positive tests (from either urine or stool) would represent false positives.

SENNAA
Our participants ingested 69–103 mg of senna (8–12 tablets) and developed diarrhea, yet TLC testing by CRL did not detect senna in any of the 11 urine or 11 stool samples that were submitted. These findings are remarkable, given published reports from research laboratories that TLC of urine detects senna with 100% sensitivity (19, 23–25). The discrepancy cannot be explained by the amount of ingested senna, because the doses ingested by our participants were higher than the doses used in the published studies. Because CRLs receive mailed samples, whereas the published data may have been derived from samples collected on site, it is possible that delays and differences related to packaging and mailing might contribute to poor performance by a CRL. Laxatives and their metabolic products are stable in urine over time, however, and special collection or storage conditions are believed to be unnecessary (26). The cause(s) of the discrepancy are therefore unknown, although the failure to detect even a single instance of senna-induced diarrhea suggests a technical problem in one of the essential steps of the TLC assay, such as the failure to use an appropriate standard (rhein) or failure to pretreat samples with β-glucuronic acid (19).

Anthraquinone drugs are used in ~50% of patients with factitious diarrhea (3, 4, 8). In the US, all anthraquinone-based laxatives sold in pharmacies contain senna, although products containing cascara can be purchased in health food stores. Because the TLC urine assay for senna and cascara depends on detection of the same metabolic product (rhein) (19–21, 23), the low sensitivity of the TLC assay for senna may portend a similarly low sensitivity of TLC for detection of cascara.

Conclusions
The sensitivities and specificities observed in this study must generate major concern about the validity of TLC testing for the diagnosis of surreptitious ingestion of bisacodyl and senna. False-negative tests, particularly for senna, may delay the diagnosis of factitious disease for years and contribute to complications from unnecessary or inappropriate medical or surgical therapy and overutilization of healthcare resources. Conversely, false-positive tests for bisacodyl could prematurely stop the diagnostic evaluation for an authentic disease. False-positive tests might also result in a patient being incorrectly accused of self-inducing his or her illness, which would likely destroy the physician-patient relationship and could cause severe psychological harm. In the case of Munchausen syndrome by proxy, false-negative tests for bisacodyl and senna could exonerate a guilty mother and endanger the child, whereas false-positive results for bisacodyl could lead to separation of an innocent mother from her child.

It is therefore important to improve the performance of TLC, or to encourage the commercial development of other methods (21, 31, 33, 34) that may be more accurate than TLC. The goal should be high sensitivity and near-perfect specificity.

It is important for clinicians to know that a positive test for bisacodyl increases the likelihood of surreptitious bisacodyl ingestion but that a positive test might also represent a false-positive result, and that a negative test for bisacodyl does not rule out bisacodyl-induced disease. In our opinion, TLC testing for anthraquinone laxatives such as senna should not be performed until a reliable test has been established. Therefore, as with many other factitious diseases, physicians must use clinical evidence from multiple sources to establish a diagnosis of surreptitious laxative ingestion, rather than rely primarily on laboratory tests.

Michelle Secic, of Secic Statistical Consulting, Inc., Chardon, Ohio, provided expert statistical advice and analysis, as well as help in preparing the manuscript. Drs. Frank Hall, Michael Nicar, and Alan Hofmann provided excellent suggestions on the design of the study and the preparation of the manuscript. This work was supported by the Southwest Digestive Disease Foundation.

References
7. Savino AC, Fordtran JS. Factitious disease: clinical lessons from case studies at Baylor University Medical Center. Proceedings (Baylor University Medical Center) 2006;19(3):195–208.