



## **Urinary Organic Acid Analysis: A Powerful Clinical Tool, Potentially Muddled by Poor Testing Methods.**

Mark Newman, MS, Shalima Gordon, BSc., ND, and Raymond M. Suen, MT (ASCP)

Organic acids, also called carboxylic acids, comprise key intermediary compounds of many biochemical pathways as well as exogenous compounds. The tricarboxylic acid, TCA cycle (also called the citric acid cycle or Krebs cycle) in the mitochondrion, for example, comprised of nine organic acids and eight enzymes, is the central metabolic pathway for all fuel molecules; dietary carbohydrates, proteins, and fats. Deficiencies in any of the TCA cycle enzymes causes an inefficient cycling of the organic acid intermediates; any number of which consequently increase in concentration in the urine of the affected individual. Metabolic defects such as these are commonly known as organic acidurias, a well-established group of disorders classified under the term of Inborn Errors of Metabolism (IEM). In these patients, urinary organic acid profiling reveals greatly elevated excretion of organic acid intermediates due to a block in a metabolic pathway; a block which may arise from a defective or missing enzyme required in the pathway in question. Inborn Errors of Metabolism are generally rare, and potentially fatal abnormalities occurring in 1 in 5000 live births. Resulting life-threatening symptoms may arise from toxic accumulations of substrates upstream of the metabolic block, deficiencies of products downstream of the block, or from intermediates of alternative biochemical pathways.<sup>1</sup>

The key concept here is “defective enzyme activity”. And, it is important to note that this does not always result, or present as an aciduria from an IEM. Defective enzyme activity exists in varying degrees of severity encompassing a spectrum of effects from severe to mild. Less severe cases of variant enzyme activity may be due to a number of factors in which the net result is a decreased rate of reaction of the affected metabolic pathway with consequent overspill of organic acid intermediates, from the affected pathway, into the urine. Aging, oxidative damage, and nutritional inadequacy due to; malabsorption, effects of pharmacological drugs, and abnormalities of vitamin/nutrient metabolism/utilization, may account for sub-optimal functioning of enzyme pathways having a multitude of effects on cellular health.

In vitamin-dependent or vitamin-responsive disorders, many of which are inheritable conditions, use of pharmacological doses of a vitamin, sufficiently overcome the metabolic blockage for normal function to occur with symptom resolution, in many instances. Multiple carboxylase deficiency, megaloblastic anemia, methylmalonic aciduria, and B6-responsive anemia are just a few well-known disorders that respond favorably to pharmacological doses of vitamins. And, of new is the identification of single nucleotide polymorphisms (SNP's). With the advent of genomics and

identification of SNP's related to variant enzyme activity, it is estimated that as many as one-third of single mutations in a gene result in the corresponding enzyme having a decreased binding affinity for its coenzyme, resulting in a lower rate of reaction. Administration of relatively high doses of the vitamin component of the corresponding coenzyme serves to restore enzymatic activity for many inheritable defects due to SNP's.<sup>2</sup>

Methylmalonic acid (methylmalonate), for example, has long been known as a marker for vitamin B12 deficiency. As serum levels of cobalamin (B12) decrease, levels of urinary methylmalonate increase. Therefore, methylmalonate serves as an effective marker for B12 deficiency. The research of Miller et al., have recently shown that methylmalonate is a reliable index of defective enzyme activity, namely transcobalamin II (TCII), responsible for transporting B12 from the ileum to the tissues. Apparently, variant enzyme activity in TCII is due to a single nucleotide polymorphism (SNP) at base position 776 in the DNA sequence in which a G-to-C substitution results in the synthesis of this enzyme carrying an arginine residue in place of proline at codon position 259. This common polymorphism results in a decreased binding affinity of the enzyme for B12 with consequent compromised delivery of B12 to tissues, and overall decrease in B12 functional status.<sup>3</sup> Other SNP-organic acid relationships have not been elucidated to this degree, but it does add to the potential value in organic acid testing.

Dysfunction in any particular enzyme protein complex that involves the absorption, transport, activation, and/or utilization of a vitamin can result in an elevated urinary organic acid that is indistinguishable from one caused by dietary deficiency. With this in mind, measuring organic acids in urine can serve as a marker for depletion of nutrients at the cellular level, whether from a nutrient deficiency or a defective enzyme, thus making organic acid analysis the ultimate test of cellular need for a personally tailored nutritional approach. This is the real power of organic acid analysis.

Organic acid profiling in the realm of functional medicine typically encompasses a panel of 25 to 50 compounds under the classifications of; Glycolysis and Citric Acid Cycle Metabolites, Fatty Acid Oxidation, Ketone Metabolites, Cofactor and Neurotransmitter Markers, in addition to markers of detoxification. Clearly, a comprehensive analysis such as this holds the potential for a wealth of information on the physiological and pathophysiological status of different metabolic pathways and their interrelationships in the body, of which may provide a wealth of clinically relevant information for patient care.

Gas Chromatography/Mass Spectrometry (GC/MS) has been the gold standard for organic acid analysis over the last three decades, without equivocation. Its primary asset is that it allows for accurate and precise quantification of a myriad of compounds simultaneously. It has contributed greatly to the understanding of many disease states. For example, organic acid analysis via GC/MS has helped to identify diabetes mellitus as not only a defect in glucose metabolism but of amino acid and fatty acid metabolism as well.<sup>4</sup>

As strides are made to elucidate the full potential of organic acid analysis and its clinical applications, obtaining consistent, accurate and precise results are critical in defining the value, hence worth of this test. Recently, an alternative method of analysis, LC/MS/MS (Liquid Chromatography/Tandem Mass Spectrometry) has been applied for the functional assessment of urinary organic acids. In question now is which methodology, GC/MS or LC/MS/MS, is superior in yielding consistent, accurate, and precise results? Also in question is which method of urine collection is best suited for retaining organic acid stability during transport. Of new is the option of sending the testing laboratory a dried urine sample. Does drying the urine sample ensure superior analyte recovery? Answers to these questions were investigated in a thoroughly thought out experimental design using one urine sample; split, spiked and sent to another lab that utilizes LC/MS/MS technology for the functional assessment of organic acids.

The results of this study were conclusive. The most accurate and precise method of organic acid analysis proved to be GC/MS. LC/MS/MS analysis offered by the other lab, promoted as a means of “replacing less accurate GC/MS methods,” showed substantially higher variability than that of GC/MS. Coefficients of Variation (a measurement of an assays reproducibility) were more than four times higher with LC/MS/MS. Moreover, many of the organic acid values obtained from LC/MS/MS analysis exceeded generally accepted limits of random error . On the other hand, GC/MS analysis yielded analyte values with accuracy, and precision consistent with its reputation in clinical chemistry. Miller et al., showed a 30% increase in methylmalonate in individuals demonstrating the defective enzyme, TCII. GC/MS methodology was used in this metabolite assay. In order to monitor this SNP, CV values much less than 30% were required for a competent assay. LC/MS/MS results of methylmalonate, in our study, revealed a CV of 45.9% while GC/MS netted an 8.1% CV.

In addition, comparing the results from the sample collection technique, there was a decided advantage over utilizing the dried urine sample from collection strips vs. that from a liquid sample. Liquid urine analysis repeatedly showed dramatic degradation for certain chemically unstable compounds, despite the use of a preservative. The dried urine from collection strips, however, was accurately and precisely measured without degradation of key metabolites.

<b>STUDY COMPARISON SUMMARY</b>		
	<u>LC/MS/MS</u>	<u>GC/MS</u>
Literature Support	Supportive literature for only a limited number of organic acids. No known sources of simultaneous analysis of many organic acids	A wealth of papers supporting its use for as many as over 100 organic acids simultaneously
Reproducibility	Average Coefficients of Variation (CV) of 32.4% in this study, nearly 4 times greater than those generated from GC/MS analysis	Average CV of 8.3%, and all measured organic acids showed CV values of less than 20%
	24% of measurements of clearly elevated	No reported false-negatives from the seven

False-Negatives	organic acids were erroneously within the reference range	measurements of identically concentrated samples
False-Positives	7.9% of organic acids clearly within the reference range, reported as being abnormal	No reported false-positives from the seven measurements of identically concentrated samples
Specificity	Increases in fumarate caused by the addition of maleic acid, an isomer of fumarate	No known false-positives due to a lack of specificity
<u>Liquid Urine Analysis</u> vs. <u>Dried Urine Collection Strip Analysis</u>		
Reproducibility	Excellent reproducibility if using GC/MS	Excellent reproducibility if using GC/MS
Stability	Excellent stability for most organic acids Compromised stability observed for ketoacids, pyruvate, and succinate	Excellent stability for all organic acids tested and greatly improved over liquid analysis for ketoacids, pyruvate, and succinate

### **GC/MS vs. LC/MS/MS Organic Acids Analysis**

#### **A Brief Review of the Literature**

The first step in the development of an assay is to determine what is to be measured, and by what means it will be accomplished. When in search of the most effective means of measuring urinary organic acids, it is abundantly clear from the literature that the gold standard for doing so is by GC/MS. However, there is great appeal in the possibility of quantifying the same analytes via LC/MS/MS, as it generally allows for much higher throughput. While a review of the scientific literature will provide sources to justify the measurement of nearly every relevant organic acid by GC/MS,<sup>4-18</sup> only a handful of different organic acids have been successfully quantified by LC/MS/MS.<sup>19-24</sup>

Kushnir<sup>19</sup> and Magera<sup>20</sup> each measured methylmalonate (MMA) in urine samples by LC/MS/MS. Magera also investigated LC/MS/MS analysis of homovanillate (HVA), however, required a much different method to do so; the latter did not require chemical derivatization, via a butyl ester, prior to analysis.<sup>21</sup> This difference may suggest a difficulty in developing one method of analysis to quantify an entire panel of organic acids using LC/MS/MS methodology. There are a few literature sources, in addition to these, reviewing a very limited number of organic acids by LC/MS/MS, but lacking any clear success in quantifying a multitude of organic acids simultaneously. The most comprehensive panel, as found in chemistry literature, was accomplished by a group led by Dr. James Pitt generating an assay for 78 compounds of which only 30 comprised a variety of organic acids, and only a handful of these, representing metabolites commonly found on functional organic acid profiles. Pitt's analysis was performed using a tandem mass spectrometer (MS/MS), without employing the LC apparatus<sup>24</sup>. When personally asked about his success, Pitt readily admitted that he was not able to achieve successful quantitation for some organic acids due to a diversity of difficulties. The assay was an excellent screening method for some of the 30 organic acids, but he personally recommended GC/MS for quantitative analysis.

Specifically targeted, some organic acids can be analyzed via LC/MS/MS incredibly quickly without sacrificing the quality of the results. Unfortunately, the prospect of simultaneously analyzing the dozens of diverse organic acids needed for these

purposes would likely leave many in sub-optimal conditions. If this were the case, higher levels of variability and inaccuracy may result. Magera<sup>21</sup> was able to generate coefficients of variation (CV) for HVA of less than 12% when assaying solely for HVA, but our comparison study showed an LC/MS/MS CV of 43.0%, and a GC/MS CV of 4.8%. Likely, HVA can only be optimized for LC/MS/MS when the procedure's parameters are narrowed in on those optimal for HVA, at which point the successful analysis of other organic acids may be somewhat compromised.

The overriding advantage as reported in clinical chemistry in LC/MS/MS methodology is throughput, as opposed to the quality of the results obtained in the analysis. The results from GC/MS, on the other hand, are known to be of high quality, but the number of samples run per instrument is relatively low, as the run times are usually much longer ranging from 30-45 minutes. However, GC/MS is established for its ability to quantify a host of organic acids simultaneously, with some research groups reporting well over 100 metabolically important compounds via this method.<sup>4-18</sup> More importantly, these sources site success in quantifying many organic acids with great accuracy and precision. In addition, some of the most frequently published authors on both GC/MS and LC/MS/MS analysis of organic acids were asked in our survey, which method is preferable for quantitative organic acid profiles. None of the researchers asked deferred to LC/MS/MS, although many spoke of its future potential in many areas. Dr. Rashed, for example, Head of a Metabolic Screening Lab and author of many papers on GC and LC applications including that for MS/MS<sup>25</sup> replied in response to our survey, "if we're talking about organic acids, then GC/MS would beat LC/MS/MS hands down!"

## **Comparing the Performance Characteristics of LC/MS/MS and GC/MS**

### **A Reproducibility Experiment**

In our comparison study a replication experiment was performed to test, in parallel, the precision of GC/MS and LC/MS/MS. In methods validation, Standard Operating Procedures require a replication experiment as one of the first steps which must be successfully performed before an assay can be considered for clinical use. A reproducibility experiment is typically carried out by assaying the same specimen multiple times. Calculations of the mean, standard deviation, and coefficient of variation (CV) for each analyte allow for an examination of the method's precision, or reproducibility. CV values are often used to determine whether or not an assay is reliable enough to justify its use. For total cholesterol measurements, for example, acceptable performance criteria is +/- 10% (as set by Clinical Laboratory Improvements Amendment, CLIA). This is a very tight range, but is necessary if subtle changes in cholesterol are to be monitored. If CV values greatly exceed 10%, a change in cholesterol of 10% cannot be considered clinically significant since it could be due solely to lab variability. For organic acids, there is no standard level of precision that must be met, but generally less than 20% is considered acceptable<sup>26</sup>.

To test the precision of the LC/MS/MS and GC/MS analyses, a urine sample was split into three aliquots. LC/MS/MS measurements from the other lab were compared to

blinded in-house GC/MS measurements. For all three aliquots of urine, 27 organic acid concentrations were unaltered, all of which are reported by both labs. Other organic acids were altered to test for other assay characteristics such as stability (discussed later).

To ensure that at least some of the analytes were elevated, two of the three samples were spiked with 4.1ug/mg creatinine of the compounds, homovanillate and suberate. These analytes were chosen for this purpose because not only are they exceptionally stable, homovanillate has been successfully reported in the literature for both GC/MS and LC/MS/MS, and the two labs participating in this study report very similar reference ranges for these two compounds (see Table #1).

The three liquid samples were stored at -80°C and sent to the other lab for LC/MS/MS analysis. The samples were sent according to the lab's instructions in two-week intervals, each with different fictitious names. A request was made for all three samples to be repeated (samples are typically stored at -20°C such that repeated analyses may be performed). One sample in particular, was repeated twice, giving a total of seven measurements each of the 27 identically concentrated analytes, in addition to five measurements for the spiked homovanillate and suberate. This relatively large number of measurements provides an examination of the reproducibility of the LC/MS/MS assay for organic acids. After obtaining the results from the LC/MS/MS analysis, an attempt to repeat the experiment using in-house GC/MS testing was made, ensuring an equal number of duplicate runs per sample.

For our in-house analysis, the samples were collected onto our Dip 'N Dry urine collection strips in accordance to the instructions provided with each kit. The specimens were then stored at 6°C. One sample at a time was submitted approximately every two weeks for extraction and GC/MS analysis. Fictitious names were used on these specimens and combined into a lot of 10 other specimens to ensure a blinded run.

Table #1 represents the summary of the seven measurements, listed in order of the analytes' average concentrations as reported from the LC/MS/MS results. Although both labs report other analytes on their respective panels, only the ones shared between the two are reported here for ease of comparison. Also, a few analytes were not considered in this study because of their known instabilities. An accurate analysis through demonstrating repeatable results of the same specimen is not possible with these unstable analytes as one freeze-thaw cycle, would change the concentration of the analytes. This is true of all a-ketoacids, pyruvate, and succinate. All of these compounds will decrease in concentration with repeated analysis with the exception of succinate, which can increase by as much as 100% from one freeze-thaw cycle (this is due to the cleaving of succinate-glucuronide bonds thereby increasing the amount of 'free' succinate). This instability is a major issue in organic acid analysis, which often goes unchallenged. Additionally, fumarate levels were not included. This analyte was the subject for an interference study (to be discussed below - *An Additional LC/MS/MS Quagmire*).

The average CV reported by GC/MS was 8.3% for the 29 organic acids. While this demonstrates excellent reproducibility, it is not unexpected. The GC/MS procedure used in this study is based loosely on that of Duez<sup>7</sup> who reports Coefficients of Variation for 27 organic acids, with an average CV of 8.2%. In comparison, the average CV reported by LC/MS/MS in this study was 32.4%, greatly questioning the reproducibility of results through this methodology. Again, it is worth repeating here that no literature sources could be found using and substantiating LC/MS/MS for a comprehensive organic acid panel. This argues that at present, a CV as large as 32.4% makes LC/MS/MS methodology inappropriate for this use. Generally, a CV of less than 20% is desirable for a competent quantitative organic acid assay of this nature.<sup>26</sup> Higher CVs, on the other hand are more suited for a screening tool such as that for IEM.

Ideally, a replication experiment should have closer to 20 points of data, but seven being enough for reasonable comparisons with statistical significance. As the data from Table #1 demonstrate, the variability of the LC/MS/MS results exceeded that of the GC/MS assay for nearly every analyte (the exception was beta-hydroxyisovalerate, which showed a comparable CV of 9.0 and 10.9 for LC/MS/MS and GC/MS, respectively).

#### **A Significant Statistical Observation**

- A CV of 83% was observed for LC/MS/MS analysis of lactate (n=7).
- The reference range given for lactate is 2.0 – 11.0ug/mg creatinine.

Significance – For a CV of greater than 70%, a specimen containing an average quantity of lactate could be reported as ‘low’, within the reference range, or ‘high’ if repeatedly analyzed, due solely to the variability of the test results!

From seven measurements of the same sample, the lowest and the highest values reported should paint the same clinical picture. This is an indication of a test's reproducibility, and hence worth. For the GC/MS results of this study, this was indeed the case as the results from the low and high extremes were reasonably close to one another. The LC/MS/MS analysis, however, painted two very different pictures when considering the lowest and highest results. Homovanillate, for example, was said to be nearly ‘low’ in one measurement (1.7ug/mg creatinine) whereas it was significantly elevated in another (7.5). One disturbing example of irreproducible results is not shown on the table as it is for an analyte not reported on the GC/MS panel. Phenylpropionate is said to be a marker for bacterial dysbiosis. One LC/MS/MS result showed the sample as having an undetectable level of phenylpropionate (<0.03), while yet another placed the same patient in the elevated category significantly above the reference range (0 – 0.7). To report an organic acid as undetectable on one occasion and as extremely elevated in another, with five other results ranging in between, is rather disturbing, to say the least.

If a generous cut-off of acceptability at 30% CV is used (usually 20-25%), the LC/MS/MS assay would fail for 39% of the analytes (11/28) as seen in Table #2 (there are only 28, not 29 coefficients of variation because methylhippurate was reported as below the detection limit and a CV could not be calculated). Few of the organic acid CVs were less than 10% (3/28), and only 10/28 were less than 20% for LC/MS/MS analysis. Clinically, organic acids are monitored for slight aberrations. This is, which is impossible with results that are highly variable. GC/MS analysis, on the other hand, showed superb reproducibility. Every analyte in this study fell within the acceptable category of less than 20%, and over two-thirds (20/28), less than 10%.

Table #2 CV Summary LC/MS/MS vs. GC/MS		
CV Range	# of Organic Acids in CV Range	
	<u>LC/MS/MS</u>	<u>GC/MS</u>
<10	3	20
10 - 20	7	8
20 - 30	7	0
>30	11	0

### Comparing False-Positive and False-Negative Rates

For the sample in question here, there are three clear abnormalities. Suberate and homovanillate were intentionally spiked to be elevated, and citramalate was also elevated (see Table #1). From the repeat measurements of these three analytes, all values were reported as abnormal by GC/MS resulting in a 0% false-negative rate. Only 4 out of 5 results for suberate and homovanillate and 5 of 7 for citramalate were reported as abnormal by LC/MS/MS resulting in a 24% false-negative rate.

Of the 20 analytes whose average values were clearly within the reference range for both labs, 11 false-positives (7.9%) were reported by LC/MS/MS. Quinolinate, for example had an average value of only 2.13 with a reference range of 0 – 3.5. However, one of the seven measurements was listed as 5.6, generating a false-positive. GC/MS reported a 0% false-positive rate as all 20 analytes were reported within the reference range all seven times. Six organic acids (orotate, ethylmalonate, tartarate, hippurate, para-hydroxyphenylacetate, and tricarballylate) were near the reference range limit and not defined clearly as either ‘normal’ or ‘abnormal’. These analytes were therefore ignored for the purposes of this comparison.

The high rate of erroneous results obtained through LC/MS/MS creates a high level of uncertainty in the report with clinical ambiguity. GC/MS, on the other hand, showed incredibly consistent results. There was only one inconsistency in the state of the GC/MS results relative to the reference range for all seven reports. The value for orotate was reported within the reference range six times and once as being “high” at 1.02 (reference range given at 0 – 1.0). The average value (0.95) is a mere 5% from the high end of the reference range, at which point it is expected that a certain percentage of results will vary enough to creep out of the reference range. Unless a CV value of zero can be obtained (by returning exactly the same result in each repeated measurement), this is always possible if the result is very near the limit of the reference range. This “creeping” outside the reference range was seen for quinolinate from the LC/MS/MS results as only one out of the seven results was reported elevated.

However, the mean result, in this case, was nowhere near the limit of the reference range for either lab. The CV reported by LC/MS/MS was 79.3% (as compared to 4.4% for GC/MS analysis of orotate). At this level of variability, results that are in the middle of the reference range can vary to both extremes. The given reference range for quinolinate by LC/MS/MS is 0 – 3.5, and the results varied from 0.7 to 5.6. The two extremes (0.7 and 5.6) of the seven measurements were both generated from the third urine sample sent for analysis. While the initial analysis was 0.7, when the same urine sample was repeated for this compound, it increased by 800%. Unfortunately, physicians and patients only see one result, and assume it is accurate. It would seem that for LC/MS/MS analysis, this trust is not justified.

A replication study, followed by a comparison of coefficients of variation, can be an excellent means of determining how well respective labs perform a certain assay. However, directly comparing any one result (reported in ug/mg creatinine) between labs can be difficult and misleading. This is largely due to the fact that quantifying assays are often method dependent. As such, this can cause a shift in reference ranges with a respective shift in analyte results between labs. Compounding this issue are the differences due to different instrumentation. While the results and the reference ranges may be different between two labs, a result relative to its reference range should be comparable if both labs are yielding sound measurements. A more effective comparison can be made using the reported population ranking. Table #1 shows a different reference range for adipate from the two labs (0 – 8.4 and 0 – 3.0) with differing results (4.79 and 1.45); however, both labs report an identical patient population ranking as the average result was in approximately the 50<sup>th</sup> percentile (the submitting patient is about average in urinary output of adipate, higher than 50% of the population).

Patient population percentiles are determined by plotting a patient's result against the general population. If a result ranks higher than 80% of the population for a particular organic acid, the patient is said to be in the 80<sup>th</sup> percentile, for that analyte. These percentiles are broken up into 5 quintiles with the highest quintile representing the top 20 percent of the population. This top quintile can be used as the high limit of the reference range<sup>27</sup>.

This means of reporting can allow for a more accurate representation of the data. Alpha-hydroxybutyrate and tartarate, for example, have identical reference ranges (0 – 11.0) as reported by the other lab using LC/MS/MS, but the population distribution is very different between the two. An alpha-hydroxybutyrate reported as 6.2 is said to be higher than only 40% of the population, while an even lower tartarate of 5.7 is higher than 70% of the population, even though the reference ranges are identical. Alpha-hydroxybutyrate results are gently spread throughout the reference range, while tartarate tends to be low in most of the population and rise quite dramatically in a minority. This same trend was seen from in-house patient rankings for these two organic acids.

Depending on where one falls in the population distribution, what appears to be a significant change in a result may not alter the population percentile significantly. For the GC/MS assay, for example, a 100% increase in hippurate, from 40 to 80ug/mg creatinine would not be a significant difference. Both results would leave the patient's hippurate value at approximately the 10<sup>th</sup> percentile giving the same clinical picture for both values. On the other hand a 100% change from 350 to 700ug/mg creatinine is significant, as it would shift the hippurate result from the 70<sup>th</sup> to the 90<sup>th</sup> percentile.

The importance of obtaining repeatable data cannot be overstated. If the reproducibility or reliability of a test cannot be evaluated or confirmed, the test is worthless. That being said, variation between repeat analyses is only significant if the change alters the clinical interpretation. As such, a comparison between the two methods would be incomplete without comparing the reported population percentiles from the seven measurements.

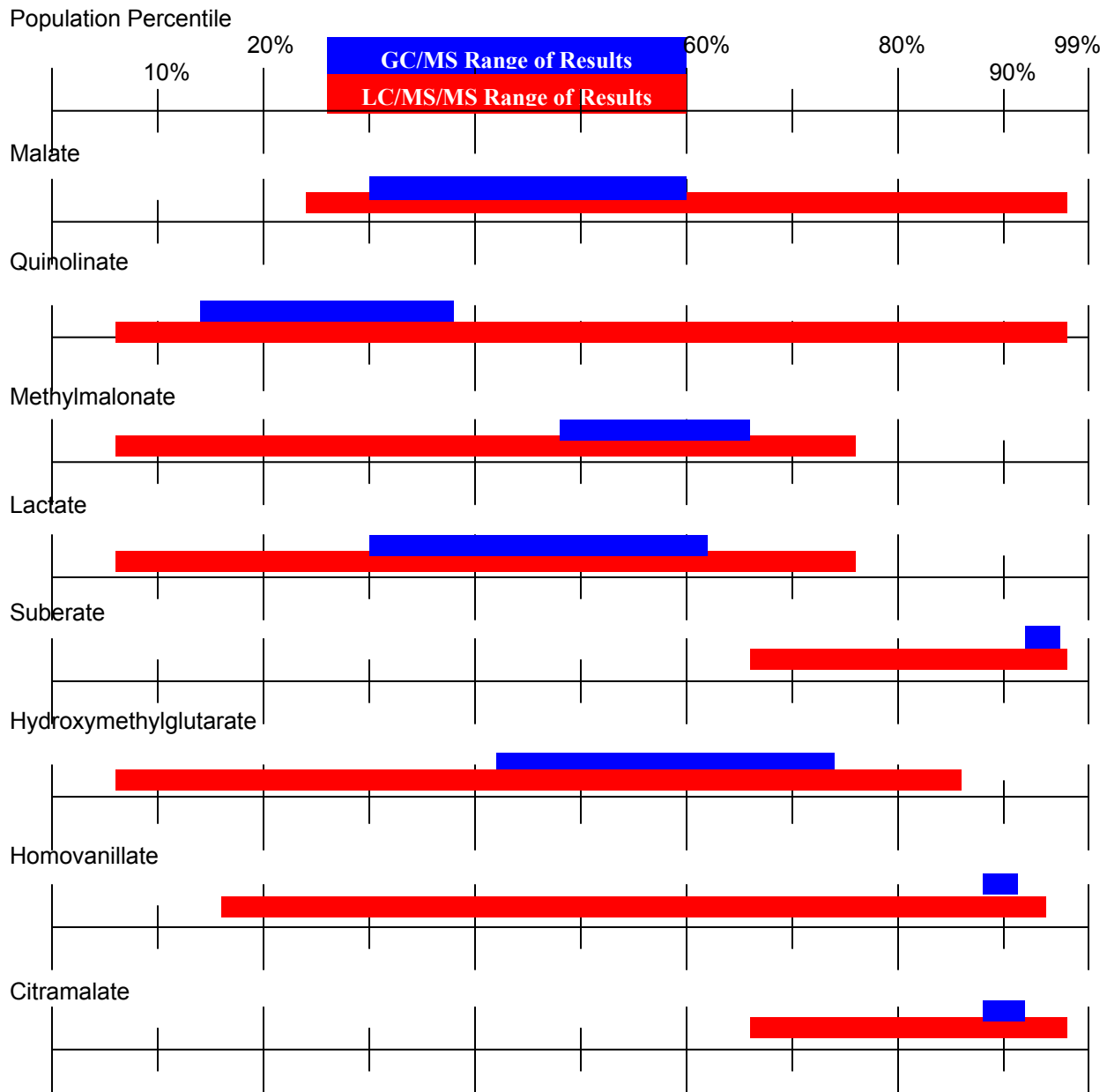
Population percentiles were compared for all 29 organic acids reported by both labs. Eight of the most dramatically variable results are shown in Figure #1. Juxtaposed to the LC/MS/MS results are the multiple measurements from the GC/MS assay. The bars on the chart represent the range of population percentiles reported from the lowest to the highest from the seven sets of results (five results for homovanillate and suberate). The bottom bar represents the LC/MS/MS reported percentiles, and the top bar represents that for GC/MS.

Recall that the concentrations for each analyte in the urine sample used in this study were identical. As figure 1 depicts, analytes with ideal precision, show results represented by a bar that covers 10 percentile or less. When considering all 29 organic acids, this ideal precision was observed for three analytes reported by LC/MS/MS and for 17 GC/MS reported values. Over half of the GC/MS acquired analytes (59%) show ideal precision, while only 10% of LC/MS/MS values match this. A general agreement is seen between the two results, but the LC/MS/MS data covers a much wider range from the identical samples.

Of the 29 analytes compared, eight (28%) from the LC/MS/MS data showed a population spread of at least 50 percentile between the measurements. GC/MS did not show this level of variability for any analytes. With this in mind, if follow-up testing is performed to track a patient's progress using GC/MS, the high reproducibility of this test, as confirmed in this study, makes for a very worthwhile profile. On the other hand, repeated profiles from LC/MS/MS yielded conflicting results. As seen from Fig. 1, quinolinate, hydroxymethylglutarate, and homovanillate were all reported in the lowest quintile (20 percentile) and also in the highest quintile, giving completely contradictory results. The seven LC/MS/MS results for quinolinate actually placed the patient in all five quintiles! This level of assurance from such varying results is only slightly higher than it would be if the numbers were randomly generated.

**FIGURE #1**

Population Percentiles of Selected Organic Acids:  
Range of Results from Seven Measurements of Identically Concentrated Samples



Recent publications have promoted the reporting of results against population quintiles, using the division between the fourth and fifth quintiles as the limit of the reference range. This allows for a more aggressive means of identifying a need for treatment or supplementation, as 20% of the population for any given marker will be classified as abnormal. Typically an abnormality corresponds to a recommendation for repletion in respective nutrient(s) for that marker. With regards to monitoring nutrient deficiencies with organic acid testing, more meaningful results can be obtained with this approach,

as opposed to the classic non-overlapping ranges for diseased and non-diseased populations<sup>27</sup>. However, the entire argument becomes superfluous if these population rankings are dependent on results that are excessively variable.

Comparing the population distributions reported by LC/MS/MS and GC/MS only serves to enhance the results of the reproducibility study. GC/MS results are simply more trustworthy than those provided currently by LC/MS/MS.

## Conclusions

Earlier we described how methylmalonate is higher in patients with a particular SNP. This abnormality creates compromised delivery of vitamin B12 to cells and tissues. A close look at the data from Miller's work shows that 1 in 5 of those studied, showed this DNA error, which caused about a 30% increase in output of the organic acid, methylmalonate. The average methylmalonate value as determined by LC/MS/MS in our comparison study was 62% of the reference range limit (2.99 average result, reference range 0 – 4.8). The average methylmalonate as determined by GC/MS was 67% of the reference range limit – incredible agreement! However, the problem lies in the reproducibility and precision of the results.

Hypothetically speaking, if one could afford to have the patient's sample analyzed 7 times, GC/MS and LC/MS/MS would prove equally valid, in this case, for methylmalonate quantification. Unfortunately, results obtained via LC/MS/MS deviated from this mean value by –57% and +57%. GC/MS values ranged from –8% of the mean to +13%. If a 30% increase in methylmalonate has high clinical significance (see paragraph above), results that vary from the true value by almost 60% are not useful. GC/MS results provide clinical significance; LC/MS/MS provides clinical ambiguity.

LC/MS/MS may one day provide a sound alternative to GC/MS analysis of organic acids. However, it appears that the scientific literature is scarce on this methodology for this purpose, and rightly so. LC/MS/MS results appear to be highly unreliable with dramatic variability when assaying for more than a few specifically targeted compounds. Additionally, there exists a documented lack of specificity, the extent of which is currently unknown (see enclosed article, *An Additional LC/MS/MS Quagmire*). Until these issues are resolved (at which point, LC/MS/MS would only meet, not exceed, the quality proven by GC/MS) GC/MS is the preferred method for quantitative organic acid analysis. With CVs exceeding 30%, LC/MS/MS may prove to be a more useful tool for screening for gross aberrations in organic acids as seen in Inborn Errors of Metabolism.

The one foreseeable patient/physician benefit to LC/MS/MS analysis is the ability to measure analytes that may not be available via GC/MS. Unfortunately, the compounds provided only via LC/MS/MS showed no improvement in reproducibility over the other analytes. Information gained from these added analytes can only be of benefit if the results are trustworthy. Unfortunately, this proved not to be the case in this study. Deoxyguanosine and glucarate, for example, are offered on the LC/MS/MS organic acid panel but not on any known GC/MS panel. However, the coefficients of variation for

these analytes were reported as 86% and 37% respectively, and deoxyguanosine showed a maximum population percentile shift of 50%. This leaves LC/MS/MS with a host of problems with little advantage over proven GC/MS methodology.

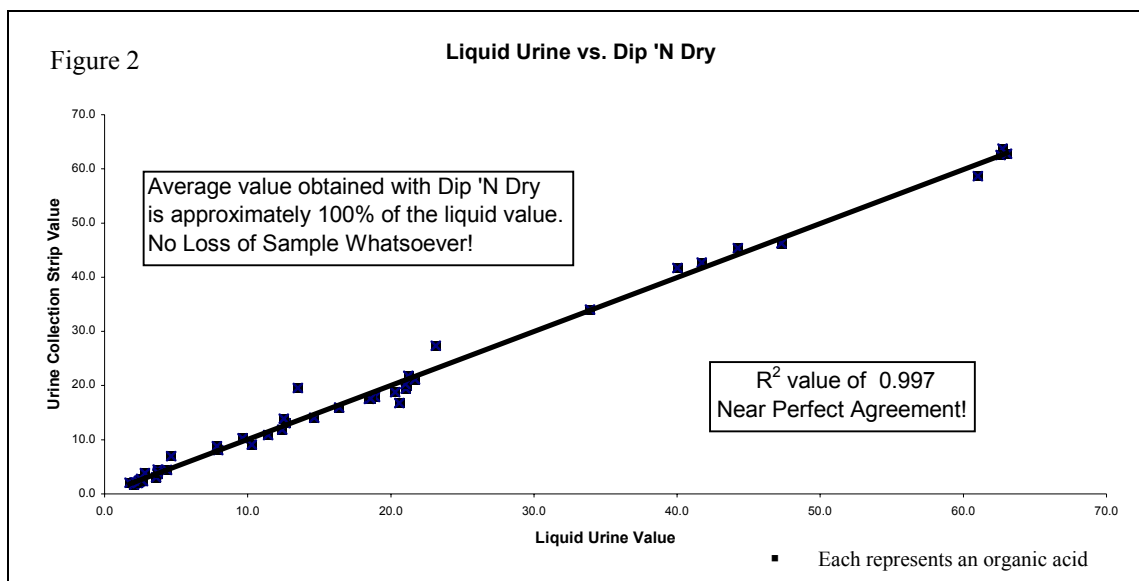
## Liquid Urine Analysis vs. The Dip 'N Dry Urine Collection Strip

### A Test of Organic Acid Stability

While innovations in technology have proven to be a step in the wrong direction with the implementation of LC/MS/MS for organic acid analysis, another innovation is erasing a chronic flaw in this highly promising field.

Initial investigations into the nature of organic acids showed insufficient stability of liquid urine samples. A more extensive study in which greatly elevated samples were sent to other labs revealed that the problem was worse than originally assessed. As such, an alternative means of sample collection has been developed. A simple urine collection strip (Dip 'N Dry) was developed by our laboratory along with a special kit to dry the sample quickly during transport. Upon arrival at the lab, the sample is rehydrated and quickly analyzed.

An extremely thorough validation of the Dip 'N Dry collection strip was performed. When compared to results from a fresh liquid urine sample, incredible agreement was observed (see Figure #2 with  $r^2$  of 0.997). These collection strips have two major advantages – stability and convenience. Absorbent materials used for the assessment of organic acids has been utilized for years in investigating Inborn Errors of Metabolism,<sup>12,14,17,19</sup> and, its full potential for functional analysis of organic acids are realized with the Dip 'N Dry collection strips.



As stated previously, the clinical use of organic acid analysis monitors subtle changes in urinary output. As such, an erroneous increase or decrease of any organic acid

concentration can alter the recommended treatment plan and clinical outlook. Unfortunately, in liquid urine, organic acid concentrations can change quite significantly between the time of collection and the time of analysis.

As mentioned in the replication study, alpha-ketoacids were not used for statistical analysis due to their lack of stability. To examine this instability, one sample was spiked with three alpha-ketoacids to greatly elevate levels. The results obtained from both collection methods, liquid and dried (Dip 'N Dry), were then compared. As can be seen from the Dip 'N Dry results in Table #3, all three compounds were verified to be well outside of the reference range in the spiked sample. The values from the baseline samples, on the other hand, were well within the reference ranges. However, the spiked sample analyzed by the other lab showed much less of the increased concentrations for alpha-ketoisocaproate and alpha-keto-beta-methylvalerate in addition to total loss of the spiked alpha-ketoisovalerate. It would appear that the liquid samples degraded during shipment or analysis. When the liquid extractions from the Dip 'N Dry samples were pulled from the freezer and analyzed a second time, all three alpha-ketoacids were substantially lower in concentration than when first analyzed. These unstable compounds are very difficult to analyze without degradation in liquid urine. The first step in GC/MS analysis is to stabilize these alpha-ketoacids by converting them to stable oxime derivatives. This can be done just minutes after the urine on the Dip 'N Dry is reconstituted with water, resulting in minimal degradation.

Table #3

Organic Acid (reference range):	* Baseline Concentration	# Spiked Sample Concentration	Observed Spike Concentration	% Increase
<b>Liquid Urine Results</b>				
alpha-ketoisovalerate (0-0.8)	0.2	0.2	0.0	0%
alpha-ketoisocaproate (0-0.5)	0.3	0.7 H	0.4	133%
alpha-keto-beta-methylvalerate (0-1.4)	0.8	3.6 H	2.8	350%
<b>Urine Collection Strip Results</b>				
alpha-ketoisovalerate (0-0.4)	0.14	1.87 H	1.73	1236%
alpha-ketoisocaproate (0-0.5)	0.25	2.17 H	1.92	768%
alpha-keto-beta-methylvalerate (0-0.9)	0.4	2.32 H	1.92	480%

\* Baseline Concentration is from the initial urine sample sent to both labs before spiking

# All concentrations are given in ug/mg creatinine

Most organic acids do not have a propensity for degradation. For those that do (the alpha-ketoacids listed above as well as pyruvate and succinate), the urine collection strip is a superior method of sample collection. Table #3 documents one small study showcasing the stability problems of liquid urine analysis. This experiment was by no means a fluke or aberration. Four different sets of samples have now been sent to two different labs and each and every one has shown dramatic degradation of these chemically unstable compounds. Despite the demonstrated lack of stability, liquid samples continue to be used because there has been no viable alternative – until now.

If subtle changes in organic acid profiles are to be monitored, clearly stability is a major concern, and the Dip 'N Dry provides a substantial advantage. The alpha-ketoacids are branched chain amino acid metabolites, important indicators for several nutrient

deficiencies, including conditions such as ketosis and lactic acidosis. Consistently erroneous results can be avoided by implementing the urine collection strip. However, reproducible values must be obtained from one strip to the next if a clinician is to effectively use the results. Fortunately, extensive research on this topic has been performed demonstrating incredible reproducibility for the Dip 'N Dry collection method. Since the Dip 'N Dry strips were utilized for the replication experiment (see results Table #1), this study provides an initial examination of the reproducibility in results obtained. CV values averaging 8.3% were obtained showing a reproducible assay. However, the question still exists, how much variation is added simply through the utilization of the Dip 'N Dry collection strip?

The baseline variability for the assay was observed by repeatedly measuring the organic acids in a liquid urine sample (without the implementation of the urine collection strip). This level of variability cannot be avoided, and it is fortunately very low (if using GC/MS). From this same urine sample, multiple urine collection strips were taken, dried, and analyzed according to the Standard Operating Procedure. Any increase in the CVs from the baseline would be caused solely by strip-to-strip inconsistency. If the increase in CVs is significant, the increased stability may be outweighed by the added variability.

When measuring the urine sample repeatedly, baseline CVs of 2.2-11.0% (average of 4.3%) were obtained for 38 organic acids (including the alpha-ketoacids, pyruvate, and succinate). This is the minimum variation, inherent to the assay. The strip-to-strip CV range was 3.7-12.3% (average of 6.1%). The average CV does not increase significantly with the introduction of the urine collection strip. Most importantly, the assay returns excellent CV values with or without the use of the urine collection strips. All coefficients of variation (CV) were quite acceptable (<20%) and proved that the Dip 'N Dry is not only dramatically superior in terms of its stability in analyte recovery, but also extremely reproducible.

The drawback to urine collection strips is the inability to monitor volatile acids, which largely evaporate with the liquid urine. Formic, acetic, and propionic acids, for example, are excessively volatile and must be analyzed via liquid urine samples. Fortunately, these analytes are not of interest, as these compounds are not known to be available on any organic acid panels. All organic acids of interest were reproducibly assayed with greatly improved stability using the Dip 'N Dry collection strip, making this the method of choice for organic acid analysis.

## **Discussion**

Organic acid analysis can be a powerful tool in the hands of a competent clinician, only if however, the results can be trusted. Organic acids have been measured and reported for more than two decades via GC/MS. And, now this accurate assay is combined with a sample collection technique in which the integrity of the organic acids is not compromised. This provides for a highly complex profile with trustworthy results.

Stability and reproducibility are essential components of successful laboratory assays. In this study, liquid urine analysis proved to be insufficiently stable for about one in seven organic acids. In addition, analysis via LC/MS/MS was insufficiently reproducible. Conversely, analysis by GC/MS, after collection with the Dip 'N Dry, optimized stability with incredibly accurate and precise results. Both stability and reproducibility were dramatically superior by this method, as proven in this study.

From the wealth of sound literature in support of GC/MS analysis, it was predicted that LC/MS/MS results may be outweighed by that acquired from GC/MS analysis. It was also expected that the Dip 'N Dry collection technique would outperform liquid urine analysis. This proved to be the case. What was not expected, though, was the extreme variability and instability encountered with LC/MS/MS and liquid urine collection, respectively. This poses some serious ramifications. Resulting false-positives and/or false-negatives provide for a questionable patient report with unnecessary, and perhaps inappropriate supplement recommendations. From a clinical perspective, this is disheartening. Organic acid analysis aims to monitor subtle changes in a variety of clinically relevant biological compounds, producing a test of high complexity. This complexity should not be burdened with results that are unreliable and irreproducible.

### **Acknowledgements**

The authors would like to offer a sincere thank you to Dr. Rashed and Dr. Pitt for their words of advice and technical expertise. Thanks to Robbyn Wolf for her help with the set-up and execution of the in-house experiment and to The Genesis Center for Integrative Medicine for facilitating data exchange.

### **Correspondence:**

Mark Newman, M.S.  
Senior Chemist  
US BioTek Laboratories  
13500 Linden Ave. N  
Seattle, WA 98133 USA  
Phone 206-365-1256  
Fax 206-363-8790  
mnewman@usbiotek.com  
www.usbiotek.com

### **References**

1. Weiner, Debra L. Pediatrics, Inborn Errors of Metabolism. *E Medicine* 2 Aug. 2001. 2 Jul. 2004
2. Ames, B., Elson-Schwab, I., Silver, E. High-dose vitamin therapy stimulates variant enzymes with decreased coenzyme binding affinity (increased Km): relevance to genetic disease and polymorphisms. *Am J Clin Nutr* 2002;75:616-58.
3. Miller, J., et al. Transcobalamin II 775G>C polymorphism and indices of vitamin B12 status in healthy older adults. *Blood* 2002;100:718-720.

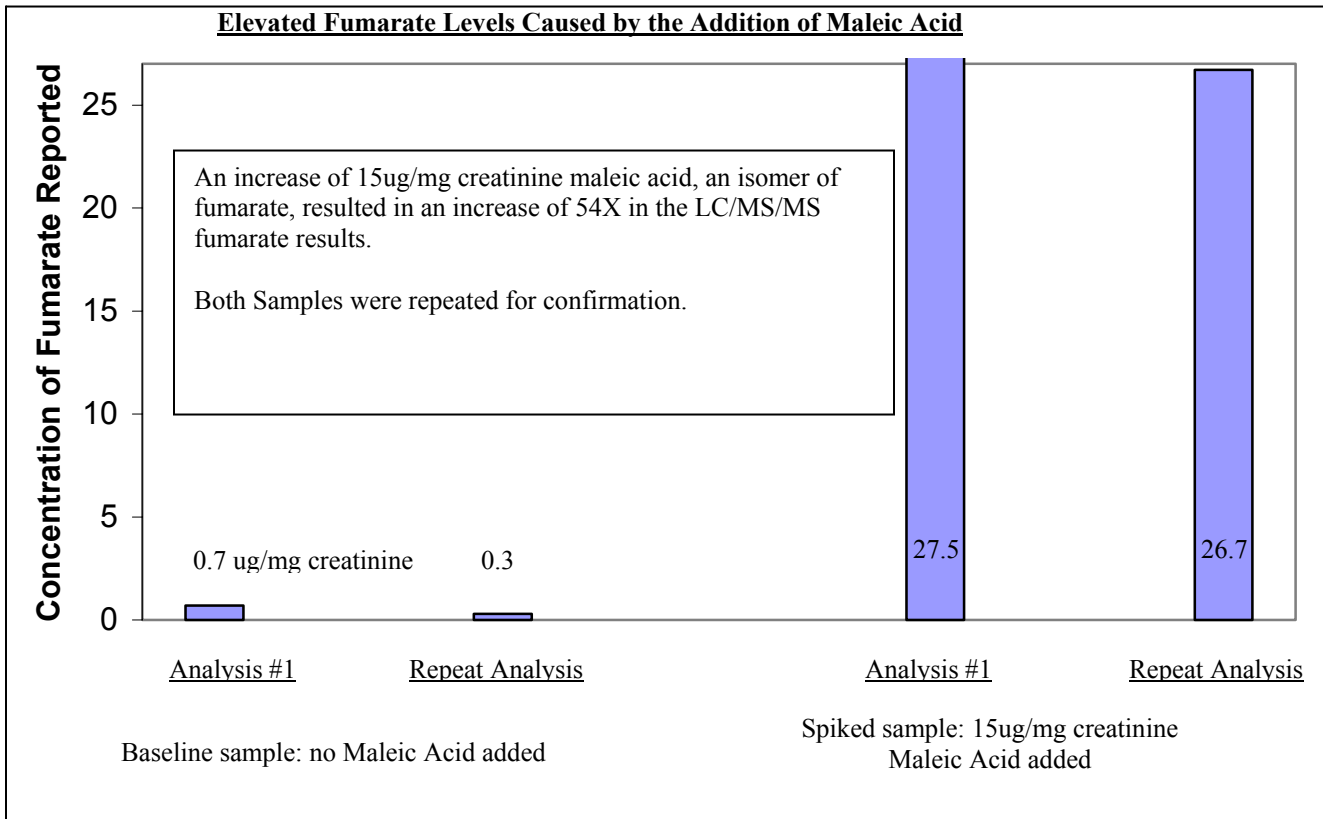
4. Liebich HM. Gas chromatographic profiling of ketone bodies and organic acids in diabetes. *J Chromatogr.* 20 (379): 347-66, 1986.
5. Chalmers R., Lawson, A., *Organic acids in man*. Chapman & Hall, London, 1982.
6. Cyr, D., et al. Stability of HVA and VMA on filter paper. *Early Human Development* 1997;49:149-152.
7. Duez, P., Kumps, A., Mardens, Y., GC/MS profiling of urinary organic acids evaluated as a quantitative method. *Clin. Chem.* 1996;42:1609-1615.
8. Fu, X., Iga, M., Kimura, M., Yamaguchi, S., Simplified screening for organic academia using GC/MS and dried urine filter paper: a study on neonatal mass screening. *Early Human Development* 2000;58:41-55.
9. Fu, X., Kimura, A., Iga, M., Yamaguchi, S., Gas chromatographic-mass spectrometric screening for organic acidemias using dried filter paper: determination of alpha-ketoacids. *J. Chromatography B Biomed Science Applications* 2001;758:87-94.
10. Greter, J., Jacobson, C., Urinary organic acids: isolation and quantification for routine metabolic screening. *Clin. Chem.* 1987;33:473-480.
11. Hoffman, G., et al. Quantitative analysis of organic acids in biological samples: batch isolation followed by gas chromatographic-mass spectrometric analysis. *Clin. Chem.* 1989;35:587-595.
12. Kuhara, T., Diagnosis of inborn errors of metabolism using filter paper urine, urease treatment, isotope dilution and gas chromatography-mass spectrometry. *J. Chromatography B Biomed Science Applications* 2001;758:3-25.
13. Shoemaker, J., Elliot, W., Automated screening of urine samples for carbohydrates, organic and amino acids after treatment with urease. *J. of Chromatography* 1991; 562:125-138.
14. Sweetman, L., *Organic acid analysis. Techniques in diagnostic human biochemical genetics. A laboratory manual*. Wiley-Liss, New York, 1991.
15. Tanaka, K., et al. Gas chromatographic method of analysis for urinary organic acids. I. retention indices of 155 metabolically important compounds. *Clin. Chem.* 1980;26:1839-1846.
16. Xu, K., et al. Screening for inborn errors of metabolism using gas chromatography-mass spectrometry. *J. of Chromatography B* 2001;758:75-80.
17. McCann, M., et al. Methylmalonic acid quantification by stable isotope dilution gas chromatography-mass spectrometry from filter paper urine samples. *Clin. Chem.* 1996;42:910-914.
18. Parnet, J., Divry, P., Vianey-Saban, C., Mathieu, M., Stable-isotope monitoring quantification of Methylmalonic acid in dried filter-paper urine samples. *J. Inherited Metabolic Diseases* 1996;19:635-637.
19. Kushnir, M., et al. Analysis of dicarboxylic acids by tandem mass spectrometry. High-throughput quantitative measurement of Methylmalonic acid in serum, plasma, and urine. *Clin. Chem.* 2001;47:1993-2002.

20. Magera, M., et al. Methylmalonic acid measured in plasma and urine by stable-isotope dilution and electrospray tandem mass spectrometry. *Clin. Chem.* 2000;46:1804-1810.
21. Magera, M., et al. Determination of homovanillic acid in urine by stable isotope dilution and electrospray tandem mass spectrometry. *Clinica Chimica Acta* 2001;306:35-41.
22. Marca, G., Casetta, B., Zammarchi, E., Rapid determination of orotic acid in urine by a fast liquid chromatography/tandem mass spectrometric method. *Rapid Communications in Mass Spectrometry* 2003;17:788-793.
23. Gonthier M., et al. Novel liquid chromatography-electrospray ionization mass spectrometry method for the quantification in human urine of microbial aromatic acid metabolites derived from dietary polyphenyls. *J Chromatography B Analyt Technol Biomed Life Sci* 2003;789:247-255.
24. Pitt, J., et al. Comprehensive screening of urine samples for inborn errors of metabolism by electrospray tandem mass spectrometry. *Clin. Chem.* 2002;48:1970-1979.
25. Rashed, M., Clinical applications of tandem mass spectrometry: ten years of diagnosis and screening for inherited metabolic diseases. *J. of Chromatography B* 2001;758:27-48.
26. Blau, N. et al. *Physicians Guide to the Laboratory Diagnosis of Metabolic Diseases*. Chapman and Hall, Germany, 1996.
27. Lord, R., Definitions of clinical laboratory reference limits. *Townsend Letter* 2004;246:81-85.
28. Christenson, W., David, M., Berndt, W., Alterations in the renal function of male and female rats exposed to maleic acid, dichloromaleic acid, and both compounds. *Toxicology* 1989;56:229-38.

### AN ADDITIONAL LC/MS/MS QUAGMIRE

In addition to the reported lack of reproducibility (see main paper), LC/MS/MS, as predicted, has shown a propensity for interference that is avoidable with GC/MS.

One specimen from a split urine sample was spiked with approximately 15ug/mg creatinine maleic acid. Both samples had identical concentrations of fumarate. Three different labs confirmed that the sample was well within given reference ranges for fumarate, a Citric Acid Cycle intermediate. Maleic acid is an isomer of fumarate sharing its molecular weight of 116.1. In-house studies have sufficiently proven that maleic acid does not interfere with fumarate quantitations with GC/MS analysis as the two are easily separated.



As can be seen, a gross error from the LC/MS/MS value was generated for the sample containing the fumarate isomer. Not only was the error quite large (the reference range for fumarate is 0 – 1ug/mg creatinine) but also consistent in repeated measurements of the samples. Either the assay is not able to decipher the difference between the two compounds (fumarate and maleic acid), or a consistent operator error exists in which the wrong chromatographic peak is being selected for quantitation. The results are quite startling and call into question alleged fumarate elevations reported by this method. Maleic acid is found in drinking water<sup>26</sup>, food additives, and pharmaceutical products. While it is not known if urinary levels of this particular isomer are found in appreciable levels, this demonstrates a principle of great importance. With dozens of organic acids having isomers existing in human metabolism and/or pharmaceutical

products, GC/MS provides a means of quantification that is less prone to erroneous numbers caused by a lack of instrumental specificity. This was precisely the warning given by experts in the field when asked about the feasibility of LC/MS/MS analysis of organic acids.