Clinical Chemistry 59:11
1657–1667 (2013)



# Reproducibility of Metabolomic Profiles among Men and Women in 2 Large Cohort Studies

Mary K. Townsend,[1*] Clary B. Clish,[2] Peter Kraft,[3] Chen Wu,[3] Amanda L. Souza,[2] Amy A. Deik,[2] Shelley S. Tworoger,[1,3†] and Brian M. Wolpin[4,5†]

**BACKGROUND:** Rigorous studies are necessary to demonstrate suitability of metabolomics platforms to profile metabolites in archived plasma within epidemiologic studies of human disease, for which attenuation of effect estimates due to measurement error is a key concern.

**METHODS:** Using a liquid chromatography–tandem mass spectrometry platform, we quantified 257 metabolites from archived plasma to evaluate metabolite interassay reproducibility, reproducibility with delayed processing, and within-person reproducibility over time. Interassay reproducibility was assessed with CVs from 60 duplicate plasma samples donated by participants in the Nurses' Health Study and Health Professionals Follow-up Study, and 20 QC pool plasma replicates. Metabolite reproducibility over a 24- to 48-h processing delay (n = 48 samples) and within-person reproducibility over 1–2 years (n = 80 samples) were assessed using Spearman and intraclass correlation coefficients (ICCs).

**RESULTS:** CVs were <20% for 92% of metabolites and generally were similar by plasma anticoagulant type (heparin or EDTA) and fasting time. Approximately 75% of metabolites were reproducible over delays in processing of blood samples (Spearman correlation or ICC ≥0.75, comparing immediate and 24-h delayed processing). Carbohydrates and purine/pyrimidine derivatives were most adversely affected by the processing delay. Ninety percent of metabolites were reproducible over 1–2 years within individuals (Spearman correlation or ICC ≥0.4).

**CONCLUSIONS:** For potential use in epidemiologic studies, the majority of plasma metabolites had low CVs and were reproducible over a 24-h processing delay and within individuals over 1–2 years. Certain metabolites, such as carbohydrates and purine/pyrimidine derivatives, may be challenging to evaluate if samples have delayed processing.



High-throughput identification and quantification of small molecule metabolites produced by metabolism, or metabolomics, is increasingly used to evaluate alterations in metabolism and their relationship to human disease *(1–5)*. In epidemiologic studies, in which participants have provided information on a variety of exposures and diseases in addition to biological samples, metabolomics has particular potential to advance understanding of disease. Understanding sources of error in metabolite measurement is critical for epidemiologic studies because measurement error may lead to attenuation of metabolic profile–disease effect estimates and inhibit ability to detect true associations.

In addition to laboratory variability, other sources of error that may be unique to epidemiologic studies need to be considered *(6, 7)*. For example, when participants reside across a wide geographic area, peripheral blood can be drawn locally and mailed to a central laboratory for processing. Some circulating analytes may be unstable during delays in plasma processing, contributing to measurement error. Further, because of limited resources, many studies collect only a single biological sample from participants. Yet, if a single metabolite measurement is not representative of longer-term levels, a prospective epidemiologic investigation may miss an association between a metabolite profile and disease, particularly for diseases with long latency periods *(6, 7)*.

However, only limited data on reproducibility over processing delays and over time in archived plasma samples typical of epidemiologic studies are available for many of the metabolites measured by

[1] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; [2] Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA; [3] Department of Epidemiology, Harvard School of Public Health, Boston, MA; [4] Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; [5] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA.

* Address correspondence to this author at: Channing Division of Network Medicine, 181 Longwood Ave., Boston, MA 02115. Fax 617-525-2008; e-mail nhmkt@channing.harvard.edu.

† Shelley S. Tworoger and Brian M. Wolpin contributed equally and should be considered co–senior authors.

Received November 8, 2012; accepted July 8, 2013.
Previously published online at DOI: 10.1373/clinchem.2012.199133

**Table 1. Summary of pilot studies.**

| Pilot study | Purpose | Description of pilot samples |
|---|---|---|
| Blinded duplicates | To assess interassay reproducibility, overall, by anticoagulant type, and by fasting status | Two aliquots of NHS (sodium heparin) and HPFS (EDTA) participant plasma samples, plus replicate samples from QC pools of noncohort participant plasma interspersed randomly across the 3 pilot studies |
| Delayed processing | To assess reproducibility of metabolites in whole blood samples over processing delays of 24–48 h | Noncohort participant sodium heparin and EDTA whole blood samples divided into 3 aliquots: aliquot #1 processed into plasma, RBC,[a] and WBC and frozen at less than −130 °C immediately after blood draw; aliquot #2 stored with an icepack in a Styrofoam container for 24 h after blood draw and then processed and frozen; aliquot #3 stored with an icepack in a Styrofoam container for 48 h after blood draw and then processed and frozen |
| Within-person reproducibility over time | To assess reproducibility of metabolites within individuals over 1–2 years | Two sodium heparin plasma samples, both ≥8 h fasting, from NHS participants donated a mean of 1.4 (range, 0.8–2.3) years apart |

[a] RBC, red blood cell; WBC, white blood cell.

metabolomics platforms *(8–12)*. We conducted 3 pilot studies to assess interassay reproducibility, reproducibility over delays in processing, and within-person reproducibility over time of over 250 plasma metabolites measured by a liquid chromatography–tandem mass spectrometry (LC-MS/MS)[6] metabolomics platform. These studies used archived biological samples donated by men and women enrolled in 2 large, prospective cohort studies as well as noncohort participant volunteers.

## Materials and Methods

### NURSES' HEALTH STUDY

The Nurses' Health Study (NHS) began in 1976 when 121 700 female registered nurses, age 30–55 years, responded to a mailed questionnaire. In 1989–1990, 32 826 participants arranged to have their blood collected in two 15-mL sodium heparin tubes and shipped the samples with an icepack in a Styrofoam container to our laboratory via overnight courier *(13)*. On arrival, blood samples were processed (i.e., centrifuged and divided into aliquots of plasma, white blood cell, and red blood cell components), and stored in the vapor phase of liquid nitrogen freezers at less than −130 °C.

To assess within-person reproducibility of biological markers, a subset of 390 participants who returned a blood sample in 1989–1990, were postmenopausal, had not used postmenopausal hormones for at least 3 months, and had no previous diagnosis of cancer (except nonmelanoma skin cancer), provided 2 additional blood samples in 1991–1992 using the same collection protocol and materials as in the initial collection.

### HEALTH PROFESSIONALS FOLLOW-UP STUDY

The Health Professionals Follow-Up Study (HPFS) was initiated in 1986 when 51 529 male US dentists, pharmacists, veterinarians, podiatrists, and osteopathic physicians, age 40–75 years, completed a mailed questionnaire. In 1993–1995, EDTA blood samples (three 10-mL tubes) were obtained from 18 225 participants *(14)*. Procedures for collecting, processing, and storing the samples were identical to those for the NHS samples.

### PILOT STUDIES

Table 1 summarizes 3 pilot studies performed to evaluate an LC-MS/MS metabolomics platform based at the Broad Institute of Massachusetts Institute of Technology and Harvard University (Cambridge, MA) for use in epidemiologic studies. Detailed information about the numbers of plasma samples included in each pilot is shown in Table 2. Briefly, the blinded duplicates pilot evaluated interassay reproducibility—overall, by anticoagulant type, and by fasting status—in duplicate samples from 30 NHS and HPFS participants and 20 replicate samples from 6 large QC pools (5 sodium heparin pools and 1 EDTA pool). The QC pools were created using discarded plasma from blood donation centers and interspersed randomly across the 3 pilot studies. The delayed processing pilot included 48 sam-

[6] Nonstandard abbreviations: LC-MS/MS, liquid chromatography–tandem mass spectrometry; NHS, Nurses' Health Study; HPFS, Health Professionals Follow-up Study; HILIC, hydrophilic interaction chromatography; BMI, body mass index; ICC, intraclass correlation coefficient.

**Table 2. Description of samples included in metabolomics platform pilot studies.**

| Variables | Blinded duplicates | Delayed processing | Within-person reproducibility over time |
|---|---|---|---|
| Total samples (QC and participant) | 66 | 54 | 88 |
| Standard QC samples,[a] n | 6 | 6 | 8 |
| Plasma QC pools | 3 | 3 | 2 |
| Samples per QC pool | 2 | 2 | 4 |
| Study participants, n | 30 | 12 | 40 |
| Plasma samples per participant | 2 | 3 or 6[b] | 2 |
| Plasma sample anticoagulant | | | |
| EDTA | 15 | 8[b] | 0 |
| Heparin | 15 | 8[b] | 40 |
| Time since last meal, h, n (%) | | | |
| ≥8 | 10 (33) | 6 (50) | 40 (100) |
| 6–7 | 10 (33) | 2 (17) | 0 |
| 4–5 | 10 (33) | 4 (33) | 0 |
| Mean age, years (range) | 58 (43–77) | 33 (26–45) | 63 (54–70) |
| Female, n (%) | 15 (50) | 5 (42) | 40 (100) |
| White race/ethnicity, n (%) | 30 (100) | 7 (58) | 40 (100) |

[a] Data from all QC samples were analyzed as part of the blinded duplicates pilot study.
[b] Each participant's initial blood collection was divided into 3 aliquots. Four participants provided 3 heparin aliquots, 4 participants provided 3 EDTA aliquots, and 4 participants provided 3 heparin and 3 EDTA aliquots.

ples donated by 12 individuals and evaluated reproducibility of metabolites in EDTA and sodium heparin whole blood samples that were processed and then frozen at less than −130 °C immediately or after a 24- or 48-h delay. The within-person reproducibility over time pilot study evaluated whether a single metabolite measurement reasonably reflected a measurement obtained 1–2 years later. Previous studies have demonstrated that, within the context of epidemiologic studies, a small number of replicates obtained several years apart is useful to estimate the extent of attenuation in observed effect estimates due to within-person variability over time *(7, 15–19)*. For this pilot, we randomly selected 40 NHS participants who returned at least 2 separate blood samples, each collected at least 8 h after the last meal, over 1–2 years. To avoid bias in metabolite profiling, all analyzed samples, including QCs, were given alias identification numbers and randomly arranged. All study participants provided implied consent by returning the questionnaires and blood samples. The Institutional Review Board of Brigham and Women's Hospital approved these studies.

### METABOLITE PROFILING

Profiles of endogenous polar metabolites and lipids were obtained using LC-MS/MS. The polar metabolite profiling methods were developed using reference standards of each metabolite to determine chromatographic retention times and MS multiple reaction–monitoring transitions, declustering potentials, and collision energies (see Tables 1–3 in the Data Supplement that accompanies the online version of this report at http://www.clinchem.org/content/vol59/issue11 for LC-MS/MS parameters for each method).

Negative ionization mode data were acquired using an ACQUITY UPLC (Waters) coupled to a 5500 QTRAP triple quadrupole mass spectrometer (AB SCIEX) running a modified version of the hydrophilic interaction chromatography (HILIC) method described by Bajad et al. *(20)*. Plasma samples (30 μL) were extracted using 120 μL of 80% methanol (VWR) containing 0.05 ng/μL inosine-$^{15}N_4$, 0.05 ng/μL thymine-$d_4$, and 0.1 ng/μL glycocholate-$d_4$ as internal standards (Cambridge Isotope Laboratories). The samples were centrifuged (10 min, 9000*g*, 4 °C) and the supernatants (10 μL) were injected directly onto a 150 × 2.0–mm Luna NH2 column (Phenomenex) that was eluted at a flow rate of 400 μL/min with initial conditions of 10% mobile phase A [20 mmol/L ammonium acetate and 20 mmol/L ammonium hydroxide (Sigma-Aldrich) in water (VWR)] and 90% mobile phase B [10 mmol/L ammonium hydroxide in 75:25 v/v acetonitrile/methanol (VWR)] followed by a 10-min linear gradient to 100% mobile phase A. The ion

spray voltage was −4.5 kV and the source temperature was 500 °C.

Positive ionization mode data were acquired as in Wang et al. with a modification to the MS acquisition in which all multiple reaction–monitoring transitions were scheduled in a single method file *(3)*. Briefly, the LC-MS/MS system consisted of a 4000 QTRAP triple quadrupole mass spectrometer (AB SCIEX) coupled to an 1100 Series pump (Agilent) and an HTS PAL autosampler (Leap Technologies). Plasma samples (10 μL) were extracted using 9 volumes of 74.9:24.9:0.2 (v/v/v) acetonitrile/methanol/formic acid containing stable isotope-labeled internal standards [0.2 ng/μL valine-d8 (Isotec) and 0.2 ng/μL phenylalanine-d8 (Cambridge Isotope Laboratories)]. The samples were centrifuged (10 min, 9000*g*, 4 °C) and the supernatants (10 μL) were injected onto a 150 × 2.1–mm Atlantis HILIC column (Waters). The column was eluted isocratically at a flow rate of 250 μL/min with 5% mobile phase A (10 mmol/L ammonium formate and 0.1% formic acid in water) for 1 min followed by a linear gradient to 40% mobile phase B (acetonitrile with 0.1% formic acid) over 10 min. The ion spray voltage was 4.5 kV and the source temperature was 450 °C.

Lipids were profiled as described in Rhee et al. *(21)*. Briefly, plasma samples (10 μL) were extracted for lipid analyses with 190 μL of isopropanol containing 0.25 ng/μL 1-dodecanoyl-2-tridecanoyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids). After centrifugation, supernatants (10 μL) were injected directly onto a 150 × 3.0–mm Prosphere HP C4 column (Grace). The column was eluted isocratically with an 80% mobile phase A (95:5:0.1 v/v/v 10 mmol/L ammonium acetate/methanol/acetic acid) for 2 min followed by a linear gradient to 80% mobile-phase B (99.9:0.1 v/v methanol/acetic acid) over 1 min, a linear gradient to 100% mobile phase B over 12 min, then 10 min at 100% mobile-phase B. MS analyses were carried out using electrospray ionization and Q1 scans in the positive ion mode. Ion spray voltage was 5.0 kV and source temperature was 400 °C.

Before each set of analyses, LC-MS/MS system sensitivity and chromatography quality were checked by analyzing reference samples: synthetic mixtures of reference metabolites (Sigma) and a lipid extract prepared from a pooled human plasma stock (Bioreclamation). During the application of each method, internal standard peak areas were monitored for QC. Moreover, reference pooled plasma samples (which were not blinded to the laboratory, unlike the QC pooled plasma samples in the blinded duplicates pilot) were included in each set of analyses, with samples inserted at the beginning and after sets of approximately 20 study samples. The reproducibility of each metabolite in the pooled plasma samples was determined to confirm the overall quality of the analyses (see online Supplemental Tables 4–6). MultiQuant 1.2 software (AB SCIEX) was used for automated peak integration and metabolite peaks were manually reviewed for quality of integration and compared against a known standard to confirm identity. Metabolites with a signal-to-noise ratio <10 were considered unquantifiable. For these analyses, metabolite signals were retained as measured LC-MS/MS peak areas, which are proportional to metabolite concentration and are appropriate for metabolite clustering and correlative analyses.

### STATISTICAL ANALYSIS

The 3 pilot studies were completed sequentially. For these analyses, we considered 257 metabolites that were measured in both the delayed processing and within-person reproducibility over time pilots. For presentation purposes, metabolites were classified into 11 categories (see online Supplemental Tables 7 and 8): lipids and lipid metabolites (n = 118); bile acids (n = 5); amino acids (n = 22); amino acid derivatives (n = 28); amines (n = 7); alcohols (n = 3); carbohydrates and intermediates (n = 13); organic acids (n = 17); purines, pyrimidines, and derivatives (n = 28); vitamins (n = 7); and other (n = 9). In each pilot study, we excluded metabolites that were unquantifiable in >50% of samples or had CVs >20% in the analyzing laboratory's reference pooled plasma samples (see online Supplemental Tables 4–6) from statistical analyses; thus, the number of metabolites included in each analysis varies slightly. Of note, 16 metabolites were either unquantifiable in >50% of samples or had CVs >20% in the reference pooled plasma samples across all 3 pilot studies and therefore do not have available data for any of the 3 pilots.

To quantify interassay reproducibility, we used data from the blinded duplicate participant samples and QC replicates in the 3 pilot studies. For each metabolite, we calculated the CV for each participant or QC pool by dividing the SD by the mean peak area and multiplying by 100. To calculate the mean interassay CV, we averaged the CVs across participants and QC pools. Within each metabolite category, we averaged mean CVs across metabolites and calculated the percentage of metabolites with mean CVs <20%, a level generally considered acceptable with respect to laboratory error *(22)*. To evaluate whether assay variability differed by anticoagulant type or fasting status, CVs were assessed separately for EDTA and heparin samples and for blood samples donated by participants fasting ≥8 and <8 h. Regarding fasting status, we hypothesized that CVs might be higher in samples donated after ≥8 vs <8 h fasting because of lower concentrations of some metabolites with longer fasting time. To assess whether platform reproducibility was

associated with metabolite peak area, we calculated Spearman correlation coefficients of metabolites' mean peak area with mean CV.

To assess metabolite reproducibility in blood samples processed after a delay, we calculated Spearman correlation coefficients to compare rankings of metabolite peak areas in samples processed immediately vs 24 or 48 h after collection. Additionally, to provide a summary measure accounting for the impact of within-person relative to total variation in metabolite peak areas, as well as the absolute values of metabolite peak areas across processing times, we calculated intraclass correlation coefficients (ICCs), defined as the between-person variance divided by the sum of the within- and between-person variances; between- and within-person variances were estimated using a mixed model, with participant as the random variable. Metabolite peak areas were natural log transformed for ICC calculations to improve the normality of peak area distributions. ICCs in samples processed immediately vs 24 h after collection as well as across all 3 processing times were calculated. An ICC ≥0.75 indicates excellent reproducibility, 0.4–0.75 indicates fair to good reproducibility, and <0.4 indicates poor reproducibility *(22)*. Spearman correlation coefficients and ICCs were calculated separately for EDTA and heparin samples.

To assess metabolite peak area reproducibility over 1–2 years, Spearman correlation coefficients and ICCs were calculated as described above. In additional analyses, we calculated ICCs adjusted for age at initial blood draw (continuous), body mass index (BMI) at initial blood draw (continuous, $kg/m^2$), and time between collections (continuous) by including these variables in the mixed model as fixed effects. For these analyses, 0.4 was considered the lower bound of acceptability; for context, an ICC of 0.4 would attenuate a true odds ratio of 2.0 to an observed odds ratio of 1.3 in epidemiologic studies *(7, 22)*. Also, for comparison, ICCs for serum cholesterol measured 1 year apart and plasma prolactin measured 2–3 years apart are 0.65 and 0.45, respectively; both have been associated with disease in epidemiologic studies *(23, 24)*.

We conducted secondary analyses to examine the influence of fasting status and anticoagulant type on metabolite peak areas. For each metabolite, we averaged metabolite peak areas separately among participants in the blinded duplicates pilot who fasted ≥8, 6–7, and 4–5 h before blood collection. Then, we calculated the percentage difference in mean metabolite peak areas between the group fasting 6–7 or 4–5 h vs ≥8 h. We also compared metabolite peak areas by anticoagulant type in samples processed immediately after collection among 4 individuals in the delayed processing pilot study who donated both heparin and EDTA blood samples. Percentage differences in metabolite peak areas in heparin vs EDTA were calculated by subtracting the EDTA peak area from the heparin peak area and dividing by the EDTA peak area.

## Results

Characteristics of participants included in each pilot study are shown in Table 2. Pilot study results for individual metabolites are provided in online Supplemental Tables 7 and 8.

### BLINDED DUPLICATES PILOT

The majority of metabolites (92%) had CVs <20%, indicating acceptable interassay reproducibility (Table 3). Purines, pyrimidines, and derivatives were measured with the most variability, with 55% of CVs ≥20%. CVs generally were comparable between EDTA and heparin samples. However, variability appeared greater in heparin vs EDTA samples for bile acids and carbohydrates and intermediates. Fasting status did not appear to affect laboratory variability for most metabolites, although CVs tended to be higher in samples donated after ≥8 vs <8 h fasting, possibly because of lower concentrations of some metabolites in the fasting condition. Larger metabolite peak area tended to be associated with lower CVs, with an overall correlation of −0.54.

When comparing metabolite peak areas across fasting status categories, we noted small differences in measured peak areas for most metabolites (Table 4). Among all metabolites, median differences in metabolite peak area were −1% and 6% comparing participants fasting for 6–7 and 4–5 h vs ≥8 h, respectively, although ranges were wide (10th to 90th percentile, −17% to 30% and −13% to 27%, respectively).

### DELAYED PROCESSING PILOT

Among all metabolites, >75% had Spearman correlations or ICCs ≥0.75 when we compared samples processed immediately vs 24 h after collection, indicating good to high reproducibility of most metabolites in whole blood samples after a 1-day processing delay (Table 5). However, <15% of purines, pyrimidines, and derivatives and <30% of carbohydrates and intermediates had Spearman correlations or ICCs ≥0.75. Reproducibility over processing delays was similar, although slightly reduced, when we compared metabolite peak areas in samples processed immediately vs after 48 h (data not shown).

We compared metabolite peak areas by anticoagulant type among 4 noncohort participants, using samples processed immediately after collection (see online Supplemental Table 9). Among all metabolites, the median difference in peak area in heparin vs EDTA samples was 4% (10th to 90th percentile, −18% to 50%). In

**Table 3. Blinded duplicates pilot: interassay CVs for metabolite LC-MS/MS peak areas overall, by anticoagulant type, and by fasting status.**

| | | CV, % | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Overall | | EDTA | | Heparin | | Fasting ≥8 h | | Fasting <8 h | |
| Metabolite category | n[a] | Mean (range) | n (%) <20% | Mean (range) | n (%) <20% | Mean (range) | n (%) <20% | Mean (range) | n (%) <20% | Mean (range) | n (%) <20% |
| Lipids and lipid metabolites | 116 | 8 (2–17) | 116 (100) | 8 (2–21) | 115 (99) | 8 (2–19) | 116 (100) | 10 (2–25) | 112 (97) | 8 (2–17) | 116 (100) |
| Bile acids | 4 | 16 (8–22) | 3 (75) | 6 (2–14) | 4 (100) | 18 (9–24) | 1 (25) | 32 (16–39) | 1 (25) | 6 (2–14) | 4 (100) |
| Amino acids | 22 | 7 (3–13) | 22 (100) | 6 (3–12) | 22 (100) | 7 (3–13) | 22 (100) | 6 (3–10) | 22 (100) | 7 (3–15) | 22 (100) |
| Amino acid derivatives | 13 | 15 (5–40) | 10 (77) | 16 (5–49) | 9 (69) | 16 (4–40) | 10 (77) | 16 (5–42) | 9 (69) | 13 (5–32) | 11 (85) |
| Amines | 6 | 7 (4–14) | 6 (100) | 6 (3–14) | 6 (100) | 7 (4–13) | 6 (100) | 6 (3–13) | 6 (100) | 7 (3–14) | 6 (100) |
| Alcohols | 3 | 11 (10–12) | 3 (100) | 9 (3–13) | 3 (100) | 11 (9–14) | 3 (100) | 17 (12–26) | 2 (67) | 7 (2–10) | 3 (100) |
| Carbohydrates and intermediates | 8 | 20 (6–44) | 5 (63) | 11 (2–33) | 7 (88) | 25 (10–61) | 4 (50) | 19 (9–36) | 3 (38) | 19 (6–44) | 5 (63) |
| Organic acids | 11 | 12 (6–22) | 10 (91) | 8 (0–30) | 9 (82) | 12 (6–22) | 10 (91) | 13 (5–26) | 8 (73) | 9 (2–17) | 10 (91) |
| Purines, pyrimidines, and derivatives | 11 | 25 (1–55) | 5 (45) | 16 (1–50) | 5 (45) | 27 (5–55) | 4 (36) | 30 (7–64) | 3 (27) | 13 (1–42) | 6 (55) |
| Vitamins | 5 | 16 (7–24) | 3 (60) | 14 (7–19) | 4 (80) | 16 (5–26) | 3 (60) | 23 (10–45) | 2 (40) | 13 (5–23) | 4 (80) |
| Other | 6 | 12 (5–16) | 6 (100) | 8 (1–16) | 6 (100) | 12 (4–20) | 5 (83) | 14 (4–39) | 5 (83) | 9 (1–17) | 6 (100) |
| All metabolites | 205 | 11 (1–55) | 189 (92) | 9 (0–50) | 190 (93) | 11 (2–61) | 184 (90) | 12 (2–64) | 173 (84) | 9 (1–44) | 193 (94) |

[a] Number of metabolites included in analyses (45 metabolites excluded because the metabolite CV in the analyzing laboratory's reference pooled plasma samples was >20%; 7 metabolites excluded because there was no quantifiable signal in >50% of samples).

**Table 4. Blinded duplicates pilot: comparison of metabolite LC-MS/MS peak areas by fasting time.**

| Metabolite category | n[a] | Percentage difference in mean metabolite peak areas, median (10th, 90th percentile) | |
|---|---|---|---|
| | | Fasting 6–7 vs ≥8 h[b] | Fasting 4–5 vs ≥8 h[b] |
| Lipids and lipid metabolites | 106 | 3 (−9, 44) | 7 (−8, 33) |
| Bile acids | 2 | −12 (−13, −10) | 3 (−13, 19) |
| Amino acids | 22 | −5 (−18, 1) | 6 (−6, 15) |
| Amino acid derivatives | 8 | −9 (−21, 23) | −4 (−27, 23) |
| Amines | 6 | −7 (−49, 7) | −14 (−67, 11) |
| Alcohols | 2 | −48 (−61, −35) | −19 (−34, −3) |
| Carbohydrates and intermediates | 3 | −8 (−19, 85) | 4 (−12, 23) |
| Organic acids | 3 | −21 (−61, −16) | 0 (−22, 6) |
| Purines, pyrimidines, and derivatives | 2 | −7 (−23, 10) | −23 (−29, −16) |
| Vitamins (pantothenate) | 1 | 24 (24, 24) | 43 (43, 43) |
| Other | 3 | −7 (−14, 3) | 9 (−6, 11) |
| All metabolites | 158 | −1 (−17, 30) | 6 (−13, 27) |

[a] Number of metabolites included in analyses (53 metabolites excluded because they were not measured in the blinded duplicates pilot; 39 metabolites excluded because the metabolite CV in the analyzing laboratory's reference pooled plasma samples was >20%; 7 metabolites excluded because there was no quantifiable signal in >50% of samples).
[b] Comparisons of metabolite concentrations among 30 participants, of which 10 each provided plasma 4–5 h, 6–7 h, and ≥8 h after their last meal.

8 of 11 metabolite categories, median percentage differences in peak areas were <15%, although percentage differences were variable in certain categories, most notably, purines, pyrimidines, and derivatives.

#### WITHIN-PERSON REPRODUCIBILITY OVER TIME PILOT

Among participants who donated 2 blood samples 0.8–2.3 years apart, women were on average 1.4 years older and their mean (range) BMI was 0.17 kg/m$^2$ (−1.8, +2.7 kg/m$^2$) higher at the second vs first blood collection. Comparing measurements in the 2 samples, 90% of metabolites had Spearman correlations or ICCs ≥0.40, indicating that a single measurement may reasonably represent longer-term levels in epidemiologic studies (Table 6). However, we observed less robust within-person reproducibility for carbohydrates and intermediates; amines; and purines, pyrimidines, and derivatives (Spearman correlations or ICCs <0.40 for 29%–60% of metabolites in the category). Nevertheless, samples in this pilot were subject to delayed processing, which may have added variability to metabolite measurements for these categories, and impacted estimates of within-person reproducibility over time.

In additional analyses, we adjusted the ICCs for factors that could potentially affect reproducibility (age, BMI, time between blood collections). ICCs tended to attenuate; 74% of metabolites had ICCs ≥0.40 compared with 83% in unadjusted analyses (Table 6; also see online Supplemental Table 8). Between-person variability in metabolites explained by adjustment factors may explain the attenuation in ICCs.

## Discussion

To determine the suitability of an LC-MS/MS metabolomics platform for use in large-scale epidemiologic studies, we performed 3 pilot studies of over 250 metabolites measured in archived human plasma. These pilots were designed to evaluate sources of measurement error of particular concern in epidemiologic studies: interassay reproducibility, reproducibility over processing delays, and within-person reproducibility over time. Approximately 90% of metabolites were measured with CVs <20%; CVs tended to be lower for metabolites with larger peak areas. CVs generally were comparable by anticoagulant type or fasting status, although specific metabolites demonstrated differences. This finding suggests systematic differences between referent and comparison groups in anticoagulant type or fasting time have the potential to introduce bias in epidemiologic studies.

Our delayed processing pilot, in which approximately 75% of metabolites had Spearman correlations or ICCs ≥0.75 when we compared immediately processed samples to those processed after a 24-h delay, highlighted a set of metabolites (e.g., carbohydrates and purines, pyrimidines, and derivatives) that are un-

**Table 5. Delayed processing pilot: Spearman correlation coefficients and ICCs for metabolite LC-MS/MS peak areas in EDTA and heparin samples comparing samples processed immediately vs 24 h after collection.**

| | | | Spearman *r* | | | | ICC[b] | | | | Spearman *r* or ICC | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | n[a] | | Median (10th to 90th percentile) | | n (%) ≥0.75 | | Median (10th to 90th percentile) | | n (%) ≥0.75 | | n (%) ≥0.75 | |
| Metabolite category | E | H | E | H | E | H | E | H | E | H | E | H |
| Lipids and lipid metabolites | 116 | 116 | 0.90 (0.57 to 1.00) | 0.93 (0.60 to 1.00) | 93 (80) | 100 (86) | 0.93 (0.62 to 0.99) | 0.95 (0.69 to 0.99) | 95 (82) | 103 (89) | 101 (87) | 105 (91) |
| Bile acids | 3 | 4 | 0.95 (0.83 to 1.00) | 0.99 (0.90 to 1.00) | 3 (100) | 4 (100) | 0.99 (0.37 to 0.99) | 0.99 (0.99 to 0.99) | 2 (67) | 4 (100) | 3 (100) | 4 (100) |
| Amino acids | 22 | 22 | 0.93 (0.81 to 0.98) | 0.93 (0.76 to 0.98) | 20 (91) | 20 (91) | 0.94 (0.00 to 0.98) | 0.89 (0.00 to 0.97) | 18 (82) | 15 (68) | 21 (95) | 20 (91) |
| Amino acid derivatives | 17 | 17 | 0.71 (0.43 to 0.98) | 0.83 (0.24 to 0.98) | 8 (47) | 10 (59) | 0.75 (0.20 to 0.97) | 0.85 (0.31 to 0.98) | 9 (53) | 11 (65) | 10 (59) | 12 (71) |
| Amines | 7 | 7 | 0.69 (0.55 to 1.00) | 0.62 (0.26 to 0.98) | 3 (43) | 3 (43) | 0.79 (0.00 to 0.99) | 0.50 (0.00 to 0.99) | 4 (57) | 3 (43) | 4 (57) | 3 (43) |
| Alcohols | 3 | 3 | 0.81 (0.62 to 0.98) | 0.90 (0.31 to 0.98) | 2 (67) | 2 (67) | 0.71 (0.46 to 0.96) | 0.57 (0.11 to 0.92) | 1 (33) | 1 (33) | 2 (67) | 2 (67) |
| Carbohydrates and intermediates | 7 | 5 | 0.69 (−0.29 to 0.95) | 0.12 (−0.19 to 0.95) | 2 (29) | 1 (20) | 0.00 (0.00 to 0.97) | 0.00 (0.00 to 0.04) | 1 (14) | 0 (0) | 2 (29) | 1 (20) |
| Organic acids | 14 | 14 | 0.82 (−0.05 to 0.98) | 0.83 (0.07 to 0.98) | 8 (57) | 8 (57) | 0.29 (0.00 to 0.96) | 0.57 (0.00 to 0.96) | 3 (21) | 4 (29) | 8 (57) | 9 (64) |
| Purines, pyrimidines, and derivatives | 14 | 15 | 0.58 (0.02 to 0.71) | 0.50 (0.21 to 0.86) | 1 (7) | 2 (13) | 0.00 (0.00 to 0.53) | 0.17 (0.00 to 0.56) | 1 (7) | 0 (0) | 2 (14) | 2 (13) |
| Vitamins | 5 | 5 | 0.76 (0.26 to 0.98) | 0.83 (−0.70 to 0.95) | 3 (60) | 3 (60) | 0.58 (0.04 to 0.98) | 0.80 (0.00 to 0.98) | 2 (40) | 4 (80) | 3 (60) | 4 (80) |
| Other | 8 | 8 | 0.94 (0.74 to 0.98) | 0.87 (0.28 to 1.00) | 7 (88) | 6 (75) | 0.95 (0.78 to 0.98) | 0.94 (0.00 to 0.98) | 8 (100) | 7 (88) | 8 (100) | 7 (88) |
| All metabolites | 216 | 216 | 0.88 (0.43 to 0.98) | 0.90 (0.36 to 1.00) | 150 (69) | 159 (74) | 0.90 (0.00 to 0.99) | 0.92 (0.04 to 0.99) | 144 (67) | 152 (70) | 164 (76) | 169 (78) |

[a] Number of metabolites included in analyses. For EDTA (E) and heparin (H), 29 metabolites were excluded because the metabolite CV in the analyzing laboratory's reference pooled plasma samples was >20%; 12 metabolites were excluded because there was no quantifiable signal in >50% of samples.

[b] Intraclass correlations, defined as the between-person variance divided by the sum of the between- and within-person variance, were calculated using natural log transformed metabolite LC-MS/MS peak areas.

**Table 6. Within-person reproducibility over time pilot: Spearman correlation coefficients and ICCs for pilot assessing reproducibility of metabolite LC-MS/MS peak areas within individuals over time.**

| | | Spearman *r* | | ICC[a] | | Spearman *r* or ICC |
|---|---|---|---|---|---|---|
| Metabolite category | N[b] | Median (10th to 90th percentile) | n (%) ≥0.40 | Median (10th to 90th percentile) | n (%) ≥0.40 | n (%) ≥0.40 |
| Lipids and lipid metabolites | 116 | 0.64 (0.44–0.79) | 111 (96) | 0.68 (0.47–0.82) | 113 (97) | 114 (98) |
| Bile acids | 5 | 0.53 (0.51–0.63) | 5 (100) | 0.55 (0.51–0.68) | 5 (100) | 5 (100) |
| Amino acids | 22 | 0.57 (0.37–0.71) | 18 (82) | 0.44 (0.36–0.58) | 17 (77) | 20 (91) |
| Amino acid derivatives | 17 | 0.62 (0.41–0.79) | 16 (94) | 0.56 (0.31–0.78) | 13 (76) | 16 (94) |
| Amines | 7 | 0.39 (0.13–0.65) | 3 (43) | 0.39 (0.07–0.67) | 3 (43) | 3 (43) |
| Alcohols | 2 | 0.42 (0.37–0.47) | 1 (50) | 0.61 (0.48–0.74) | 2 (100) | 2 (100) |
| Carbohydrates and intermediates | 7 | 0.25 (0.14–0.70) | 2 (29) | 0.31 (0.22–0.43) | 1 (14) | 2 (29) |
| Organic acids | 16 | 0.48 (0.23–0.74) | 10 (63) | 0.46 (0.25–0.75) | 11 (69) | 12 (75) |
| Purines, pyrimidines, and derivatives | 10 | 0.48 (0.17–0.78) | 6 (60) | 0.34 (0.06–0.78) | 4 (40) | 6 (60) |
| Vitamins | 2 | 0.81 (0.81–0.81) | 2 (100) | 0.80 (0.73–87) | 2 (100) | 2 (100) |
| Other | 6 | 0.49 (0.44–0.80) | 6 (100) | 0.44 (0.00–0.82) | 4 (67) | 6 (100) |
| All metabolites | 210 | 0.60 (0.37–0.79) | 180 (86) | 0.58 (0.33–0.80) | 175 (83) | 188 (90) |

[a] Intraclass correlations, defined as the between-person variance divided by the sum of the between- and within-person variance, were calculated using natural log transformed metabolite LC-MS/MS peak areas.
[b] Number of metabolites included in analyses (31 metabolites excluded because the metabolite CV in the analyzing laboratory's reference pooled plasma samples was >20%; 16 metabolites excluded because there was no quantifiable signal in >50% of samples).

likely to be accurately measured in studies with delayed processing. Poor results for carbohydrates are expected given the glycolytic processes in whole blood. Also, ongoing enzymatic reactions in whole blood likely contribute to instability of purines and pyrimidines *(25)*. Inclusion of these metabolites may increase the likelihood of false-negative results and, in studies of hundreds of analytes, may increase the penalty for multiple testing without contributing meaningful data.

Ninety percent of metabolites demonstrated Spearman correlations or ICCs ≥0.40 when we compared measurements in samples donated 0.8–2.3 years apart. However, ICCs may vary between individuals who did and those who did not experience changes in health between blood collections *(26)*, and our findings represent average results among all women in the study population. Nonetheless, our results suggest that, within the context of an epidemiologic study in which attenuation of effect estimates is a key concern, a single plasma measurement likely is reasonably representative of longer-term exposure for the vast majority of metabolites on the platform. This is of particular importance for prospective epidemiologic studies of chronic diseases with long latency periods.

Findings from previous studies support the results we observed for the best-performing metabolite categories, including lipids *(23, 27–31)* and amino acids *(32)*. Of note, several individual metabolites [e.g., ADMA (asymmetric dimethylarginine), homocysteine, urate, and folate] performed poorly in these metabolite profile pilot studies but have had acceptable results when measured using single assays. For example, folate concentrations were largely undetectable with the metabolomics platform but were measurable in NHS samples using a radioassay kit *(33)*. Such differences in performance likely reflect differences in assay optimization between the metabolomics platform and a single assay.

Several limitations of our study should be considered. First, although we assessed interassay reproducibility, CVs may vary by laboratory and project size. Investigators should evaluate CVs in replicate QC samples in their own studies; including 10% of the total sample size as QCs is one standard approach *(6)*. Our analysis of differences in LC-MS/MS peak area by fasting status evaluated separate groups of individuals, rather than changes within individuals; additional studies are needed to better understand this issue *(34)*. Also, because our examination of long-term within-person reproducibility of metabolite measurements used actual NHS samples, variability in metabolite measurements due to delayed processing was incorporated into the within-person variance estimates, possibly causing underestimation of reproducibility over

time for metabolites during the processing delay. Finally, we were not able to assess certain metabolites because of undetectable concentrations in our participants. However, when measuring over 250 metabolites, it is expected that a small proportion will not be detectable in a particular population.

In summary, our pilot studies indicated that the large majority of metabolites included on a high-throughput metabolomics platform could be accurately measured in conditions mimicking those of prospective epidemiologic studies. Most metabolites, particularly lipids and lipid metabolites, amino acids, and bile acids, were measured with acceptable laboratory variability and were reproducible over processing delays and within individuals over 1–2 years. However, our results also suggested that metabolite profiles involving carbohydrates and intermediates; purines, pyrimidines, and derivatives; and amines may be more difficult to evaluate in epidemiologic studies in which blood samples were processed after a 24-h or longer delay. Future studies are needed to obtain estimates of long-term within-person reproducibility of metabolites in samples processed immediately after collection.

**Author Contributions:** *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

**Authors' Disclosures or Potential Conflicts of Interest:** *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

**Employment or Leadership:** A.A. Deik, the Broad Institute.
**Consultant or Advisory Role:** None declared.
**Stock Ownership:** None declared.
**Honoraria:** None declared.
**Research Funding:** NHS funded by NIH grants P01-CA87969 and R01-CA49449, HPFS funded by NIH grant P01-CA055075, Metabolite profiling supported by a gift from the Nestle Research Center to the Broad Institute; B.M. Wolpin, grant K07-CA140790 from the National Cancer Institute, American Society of Clinical Oncology Career Development Award, Howard Hughes Medical Institute Early Career Physician-Scientist Award, and the Lustgarten Foundation for Pancreatic Cancer Research.
**Expert Testimony:** None declared.
**Patents:** None declared.

**Role of Sponsor:** The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

## References

1. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature 2009;457:910–4.
2. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature 2011;477:54–60.
3. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nat Med 2011;17:448–53.
4. Shaham O, Slate NG, Goldberger O, Xu Q, Ramanathan A, Souza AL, et al. A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent media from cultured muscle cells. Proc Natl Acad Sci U S A 2010;107:1571–5.
5. Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation 2012;125:2222–31.
6. Tworoger SS, Hankinson SE. Use of biomarkers in epidemiologic studies: minimizing the influence of measurement error in the study design and analysis. Cancer Causes Control 2006;17:889–99.
7. White E. Measurement error in biomarkers: sources, assessment, and impact on studies. In: Rothman N, Hainaut P, Schulte P, Smith M, Boffetta P, Perera F, eds. Molecular epidemiology: principles and practices. Lyon: International Agency for Research on Cancer; 2011. p 143–61.
8. Yin P, Peter A, Franken H, Zhao X, Neukamm SS, Rosenbaum L, et al. Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. Clin Chem 2013;59:833–45.
9. Dunn WB, Broadhurst D, Ellis DI, Brown M, Halsall A, O'Hagan S, et al. A GC-TOF-MS study of the stability of serum and urine metabolomes during the UK Biobank sample collection and preparation protocols. Int J Epidemiol 2008;37(Suppl 1):i23–30.
10. Shurubor YI, Matson WR, Willett WC, Hankinson SE, Kristal BS. Biological variability dominates and influences analytical variance in HPLC-ECD studies of the human plasma metabolome. BMC Clin Pathol 2007;7:9.
11. Wood JT, Williams JS, Pandarinathan L, Courville A, Keplinger MR, Janero DR, et al. Comprehensive profiling of the human circulating endocannabinoid metabolome: clinical sampling and sample storage parameters. Clin Chem Lab Med 2008;46:1289–95.
12. Yu Z, Kastenmuller G, He Y, Belcredi P, Moller G, Prehn C, et al. Differences between human plasma and serum metabolite profiles. PLoS One 2011;6:e21230.
13. Hankinson SE, Willett WC, Manson JE, Colditz GA, Hunter DJ, Spiegelman D, et al. Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women. J Natl Cancer Inst 1998;90:1292–9.
14. Chu NF, Spiegelman D, Yu J, Rifai N, Hotamisligil GS, Rimm EB. Plasma leptin concentrations and four-year weight gain among US men. Int J Obes Relat Metab Disord 2001;25:346–53.
15. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. Am J Epidemiol 1992;136:1400–13.
16. Al-Delaimy WK, Natarajan L, Sun X, Rock CL, Pierce JP, Women's Healthy Eating and Living (WHEL) Study Group. Reliability of plasma carotenoid biomarkers and its relation to study power. Epidemiology 2008;19:338–44.
17. Comstock GW, Burke AE, Hoffman SC, Norkus EP, Gross M, Helzlsouer KJ. The repeatability of serum carotenoid, retinoid, and tocopherol concentrations in specimens of blood collected 15 years apart. Cancer Epidemiol Biomarkers Prev 2001;10:65–8.
18. Gu Y, Zeleniuch-Jacquotte A, Linkov F, Koenig KL, Liu M, Velikokhatnaya L, et al. Reproducibility of serum cytokines and growth factors. Cytokine 2009;45:44–9.
19. Zeleniuch-Jacquotte A, Adlercreutz H, Akhmedkhanov A, Toniolo P. Reliability of serum measurements of lignans and isoflavonoid phytoestrogens over a two-year period. Cancer Epidemiol Biomarkers Prev 1998;7:885–9.
20. Bajad SU, Lu W, Kimball EH, Yuan J, Peterson C, Rabinowitz JD. Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. J Chromatogr A 2006;1125:76–88.
21. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest 2011;121:1402–11.
22. Rosner B. Fundamentals of biostatistics. 6th ed. Belmont (CA): Duxbury Press; 2005.
23. Shekelle RB, Shryock AM, Paul O, Lepper M,

Stamler J, Liu S, Raynor WJ Jr. Diet, serum cholesterol, and death from coronary heart disease. The Western Electric study. N Engl J Med 1981; 304:65–70.

24. Tworoger SS, Eliassen AH, Sluss P, Hankinson SE. A prospective study of plasma prolactin concentrations and risk of premenopausal and postmenopausal breast cancer. J Clin Oncol 2007;25: 1482–8.
25. Jabs CM, Neglen P, Eklof B, Thomas EJ. Adenosine, inosine, and hypoxanthine/xanthine measured in tissue and plasma by a luminescence method. Clin Chem 1990;36:81–7.
26. Hankinson SE, Manson JE, Spiegelman D, Willett WC, Longcope C, Speizer FE. Reproducibility of plasma hormone levels in postmenopausal women over a 2–3-year period. Cancer Epidemiol Biomarkers Prev 1995;4:649–54.
27. Giltay EJ, Geleijnse JM, Schouten EG, Katan MB, Kromhout D. High stability of markers of cardiovascular risk in blood samples. Clin Chem 2003; 49:652–5.
28. Hankinson SE, London SJ, Chute CG, Barbieri RL, Jones L, Kaplan LA, et al. Effect of transport conditions on the stability of biochemical markers in blood. Clin Chem 1989;35:2313–6.
29. Kotsopoulos J, Tworoger SS, Campos H, Chung FL, Clevenger CV, Franke AA, et al. Reproducibility of plasma and urine biomarkers among premenopausal and postmenopausal women from the Nurses' Health Studies. Cancer Epidemiol Biomarkers Prev 2010;19:938–46.
30. van Eijsden M, van der Wal MF, Hornstra G, Bonsel GJ. Can whole-blood samples be stored over 24 hours without compromising stability of C-reactive protein, retinol, ferritin, folic acid, and fatty acids in epidemiologic research? Clin Chem 2005;51:230–2.
31. Zivkovic AM, Wiest MM, Nguyen UT, Davis R, Watkins SM, German JB. Effects of sample handling and storage on quantitative lipid analysis in human serum. Metabolomics 2009;5:507–16.
32. Davis JS, Darcy CJ, Piera K, McNeil YR, Woodberry T, Anstey NM. Ex-vivo changes in amino acid concentrations from blood stored at room temperature or on ice: implications for arginine and taurine measurements. BMC Clin Pathol 2009;9:10.
33. Zhang SM, Willett WC, Selhub J, Hunter DJ, Giovannucci EL, Holmes MD, et al. Plasma folate, vitamin B6, vitamin B12, homocysteine, and risk of breast cancer. J Natl Cancer Inst 2003;95:373–80.
34. Ho JE, Larson MG, Vasan RS, Ghorbani A, Cheng S, Rhee EP, et al. Metabolite profiles during oral glucose challenge. Diabetes 2013; 62:2689–98.