Anal. Chem. **2009,** *81,* 1357–1364

Development of a Robust and Repeatable UPLC–MS Method for the Long-Term Metabolomic Study of Human Serum

Eva Zelena,[†] Warwick B. Dunn,*[,‡] David Broadhurst,[†] Sue Francis-McIntyre,[†] Kathleen M. Carroll,[‡] Paul Begley,[†] Steve O'Hagan,[†] Joshua D. Knowles,[§] Antony Halsall,[†] HUSERMET Consortium, Ian D. Wilson,[||] and Douglas B. Kell[†,‡]

Bioanalytical Sciences Group and Manchester Centre for Integrative Systems Biology, School of Chemistry, Manchester Interdisciplinary Biocentre, University of Manchester, M1 7DN, U.K., School of Computer Science, Manchester Interdisciplinary Biocentre, University of Manchester, M1 7DN, U.K., and Department of Clinical Pharmacology, Drug Metabolism and Pharmacokinetics, AstraZeneca, Mereside, Alderley Park, Macclesfield, Cheshire, SK10 4TG, U.K.

A method for performing untargeted metabolomic analysis of human serum has been developed based on protein precipitation followed by Ultra Performance Liquid Chromatography and Time-of-Flight mass spectrometry (UPLC–TOF-MS). This method was specifically designed to fulfill the requirements of a long-term metabolomic study, spanning more than 3 years, and it was subsequently thoroughly evaluated for robustness and repeatability. We describe here the observed drift in instrumental performance over time and its improvement with adjustment of the length of analytical block. The optimal setup for our purpose was further validated against a set of serum samples from 30 healthy individuals. We also assessed the reproducibility of chromatographic columns with the same chemistry of stationary phase from the same manufacturer but from different production batches. The results have allowed the authors to prepare SOPs for "fit for purpose" long-term UPLC–MS metabolomic studies, such as are being employed in the HUSERMET project. This method allows the acquisition of data and subsequent comparison of data collected across many months or years.

Metabolomics, and the related field of metabonomics, are increasingly being applied in post-genomic research and systems biology. A number of advantages are observed when compared to the study of the genome, transcriptome, and proteome including discriminatory power, throughput, and cost.[1,2] The objective of metabolomics is to characterize the quantitative composition of low molecular weight chemicals in biological systems (defined as the metabolome) and apply the data to real world problems. Studies include defining phenotypic changes of microbial, plant, and mammalian systems related to genetic or environmental modifications and stress.[3–8] Metabolomics is also an integral component of systems biology investigations.[9–13] Of the biological systems studied, mammals are one of the most complex in terms of the composition and size of the metabolome and observed changes in the phenotype are dependent on small differences in gender, age, and diurnal variation,[14,15] diet and lifestyle,[16] and disease.[17–20]

Numerous experimental strategies and technologies are used in the metabolomics pipeline[21] to study intracellular and extra-

* Corresponding author. Warwick Dunn, e-mail warwick.dunn@manchester.ac.uk; phone +44 (0) 161 306 5146; fax +44 (0) 161 306 4556.

[†] Bioanalytical Sciences Group, University of Manchester.

[‡] Manchester Centre for Integrative Systems Biology, University of Manchester.

[§] School of Computer Science, University of Manchester.

[||] AstraZeneca.

(1) Dunn, W. B. *Phys. Biol.* **2008**, *5*.

(2) Kell, D. B. *Drug Discovery Today* **2006**, *11*, 1085–1092.

(3) Hall, R. D. *New Phytologist* **2006**, *169*, 453–468.

(4) Kaddurah-Daouk, R.; Kristal, B. S.; Weinshilboum, R. M. *Annu. Rev. Pharmacol. Toxicol.* **2008**, *48*, 653–683.

(5) Kenny, L. C.; Broadhurst, D.; Brown, M.; Dunn, W. B.; Redman, C. W. G.; Kell, D. B.; Baker, P. N. *Reprod. Sci.* **2008**, *15*, 591–597.

(6) Lin, C. Y.; Viant, M. R.; Tjeerdema, R. S. *J. Pest. Sci.* **2006**, *31*, 245–251.

(7) Mashego, M. R.; Rumbold, K.; De Mey, M.; Vandamme, E.; Soetaert, W.; Heijnen, J. J. *Biotechnol. Lett.* **2007**, *29*, 1–16.

(8) Schnackenberg, L. K.; Beger, R. D. *Toxicol. Mech. Methods* **2008**, *18*, 301–311.

(9) Castrillo, J. I.; Zeef, L. A.; Hoyle, D. C.; Zhang, N.; Hayes, A.; Gardner, D. C. J.; Cornell, M. J.; Petty, J.; Hakes, L.; Wardleworth, L.; Rash, B.; Brown, M.; Dunn, W. B.; Broadhurst, D.; O'Donoghue, K.; Hester, S. S.; Dunkley, T. P. J.; Hart, S. R.; Swainston, N.; Li, P.; Gaskell, S. J.; Paton, N. W.; Lilley, K. S.; Kell, D. B.; Oliver, S. G. *J. Biol.* **2007**, *6*.

(10) Daran-Lapujade, P.; Rossell, S.; van Gulik, W. M.; Luttik, M. A. H.; de Groot, M. J. L.; Slijper, M.; Heck, A. J. R.; Daran, J. M.; de Winde, J. H.; Westerhoff, H. V.; Pronk, J. T.; Bakker, B. M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15753–15758.

(11) Kell, D. B. *IUBMB Life* **2007**, *59*, 689–695.

(12) Martin, F. P. J.; Dumas, M. E.; Wang, Y.; Legido-Quigley, C.; Yap, I. K. S.; Tang, H.; Zirah, S.; Murphy, G. M.; Cloarec, O.; Lindon, J. C.; Sprenger, N.; Fay, L. B.; Kochhar, S.; Van Bladeren, P.; Holmes, E.; Nicholson, J. K. *Mol. Syst. Biol.* **2007**, *3*.

(13) Weckwerth, W. *Annu. Rev. Plant Biol.* **2003**, *54*, 669–689.

(14) Plumb, R.; Granger, J.; Stumpf, C.; Wilson, I. D.; Evans, J. A.; Lenz, E. M. *Analyst* **2003**, *128*, 819–823.

(15) Williams, R. E.; Lenz, E. M.; Lowden, J. S.; Rantalainen, M.; Wilson, I. D. *Mol. Biosyst.* **2005**, *1*, 166–175.

(16) Lenz, E. M.; Bright, J.; Wilson, I. D.; Hughes, A.; Morrisson, J.; Lindberg, H.; Lockton, A. *J. Pharm. Biomed. Anal.* **2004**, *36*, 841–849.

(17) Atherton, H. J.; Bailey, N. J.; Zhang, W.; Taylor, J.; Major, H.; Shockcor, J.; Clarke, K.; Griffin, J. L. *Physiol. Genomics* **2006**, *27*, 178–186.

(18) Dunn, W. B.; Broadhurst, D. I.; Deepak, S. M.; Buch, M. H.; McDowell, G.; Spasic, I.; Ellis, D. I.; Brooks, N.; Kell, D. B.; Neyses, L. *Metabolomics* **2007**, *3*, 413–426.

(19) Kind, T.; Tolstikov, V.; Fiehn, O.; Weiss, R. H. *Anal. Biochem.* **2007**, *363*, 185–195.

10.1021/ac8019366 CCC: $40.75 © 2009 American Chemical Society
Published on Web 01/26/2009

cellular metabolomes.[22−25] Two generalized strategies are commonly employed; targeted analysis and metabolic profiling. Metabolic profiling (metabolite profiling or untargeted analysis) provides a holistic and global picture of the metabolome in a hypothesis-generation strategy.[26] Thousands of metabolites can be detected with limited *a priori* knowledge of which metabolites are present. Detection aims to be relatively unbiased, though a degree of bias may be introduced through the methods of sample preparation and the analytical platform chosen. A range of analytical platforms are employed in metabolic profiling,[22] chromatography−mass spectrometry and NMR being the most widely applied. The global metabolic profiles produced in this type of work are providing new insights into the phenotype of biological systems and in mammalian systems the physiological and pathophysiological processes related to diet, lifestyle, diseases, and pharmaceutical intervention.[2,4,13,27−31]

To enable a greater understanding of the metabolic status of humans, large-scale epidemiological studies are required in order to take account of the substantial diversity observed in physiology, metabolic status, and lifestyle in the general human population. Large-scale studies are required also to boost the power of any subsequent statistical analysis, so that subtle differences within the subject cohort can be detected.[32] Most reported studies using LC−MS-based techniques have involved relatively small sample numbers and single blocks, with fewer than 150 injections, (e.g., see refs 19, 33−35). This vast increase in sample throughput requires improvements in the analytical technologies and methodologies to provide repeatable, robust, and high-quality data. The availability of high-throughput analytical methodologies for metabolomic investigations allows these large-scale studies to be performed. The HUSERMET project (Human Serum Metabolome)[36] is investigating the serum metabolomes of "healthy" subjects and those from two diseased populations (ovarian cancer and Alzheimer's disease). Greater than 7000 samples will be collected during this project and subsequently analyzed. Ultra Performance Liquid Chromatography−mass spectrometry (UPLC−MS) is one analytical platform being applied in the HUSERMET study and requires methods and procedures to be in place that ensure analytical stability and comparability of data that may be acquired several years apart. A number of publications have highlighted the requirements for long-term metabolomic studies,[1,32,37,38] and a range of issues should be considered in these long-term studies including machine drift, analysis order, and choice of technology. All these issues provide significant challenges which have not been comprehensively addressed. The issue of "machine drift" may be compensated by various modeling/statistical techniques; however, due to the highly multivariate nature of the data this cannot be assumed (i.e., not every measured metabolite will be affected in a predictable way). With this in mind, together with general instrument maintenance issues, the HUSERMET study has been organized as a series of "blocked" analytical experiments. That is, sets of N subject samples are analyzed in a single sub-experiment. Therefore the final data analysis will involve comparing profiles within each block as well as between the many block experiments run over the course of the epidemiological study. In this context the definition of a block originates from epidemiological studies. In analytical terminology this is usually defined as a batch, i.e., the continuous analysis of samples without instrument maintenance. In this paper we shall employ the term block.

The performance of analytical instrumentation has to be assessed robustly to ensure that data are of comparable high quality within and between blocks. An approach based on the periodic analysis of a standard biological Quality Control sample (QC sample) together with the subject samples[39] is now starting to be accepted as a quality assurance strategy in metabolic profiling. This approach has been applied to metabolomic studies as a pragmatic solution to fulfill the requirement to simultaneously assess repeatability for hundreds of endogenous metabolites. Pooled QC samples which were prepared by pooling equal aliquots from all the samples of interest have been shown as appropriate for urine in short-term studies (maximum of 114 injections).[40,41]

The development and assessment of UPLC−MS instrumentation and methodologies to allow the preparation of a standard operating procedure (SOP) for the analysis of serum samples using Ultra Performance Liquid Chromatography−time-of-flight mass spectrometry (UPLC−TOF-MS) will be discussed here in detail. Specific areas developed include appropriate UPLC−MS analytical methods for high-throughput metabolomics studies and the applicability of the developed SOP for the investigation of the serum metabolomes of healthy subjects. Other sources of variation will also be described including the comparability of data acquired on UPLC columns of the same stationary phase but different manufacturing batches and the length of an analytical block.

(20) Sabatine, M. S.; Liu, E.; Morrow, D. A.; Heller, E.; McCarroll, R.; Wiegand, R.; Berriz, G. F.; Roth, F. P.; Gerszten, R. E. *Circulation* **2005**, *112*, 3868–3875.
(21) Brown, M.; Dunn, W. B.; Ellis, D. I.; Goodacre, R.; Handl, J.; Knowles, J. D.; O'Hagan, S.; Spasic, I.; Kell, D. B. *Metabolomics* **2005**, *1*, 39–51.
(22) Dunn, W. B.; Bailey, N. J. C.; Johnson, H. E. *Analyst* **2005**, *130*, 606–625.
(23) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155–171.
(24) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends Biotechnol.* **2004**, *22*, 245–252.
(25) Hollywood, K.; Brison, D. R.; Goodacre, R. *Proteomics* **2006**, *6*, 4716–4723.
(26) Kell, D. B.; Oliver, S. G. *BioEssays* **2004**, *26*, 99–105.
(27) Fernie, A. R.; Trethewey, R. N.; Krotzky, A. J.; Willmitzer, L. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 763–769.
(28) Kell, D. B. *FEBS J.* **2006**, *273*, 873–894.
(29) Lindon, J. C.; Holmes, E.; Nicholson, J. K. *Curr. Opin. Mol. Ther.* **2004**, *6*, 265–272.
(30) Thomas, C. E.; Ganji, G. *Curr. Opin. Drug Discovery Dev.* **2006**, *9*, 92–100.
(31) van der Greef, J.; Hankemeier, T.; McBurney, R. N. *Pharmacogenomics* **2006**, *7*, 1087–1094.
(32) Broadhurst, D. I.; Kell, D. B. *Metabolomics* **2006**, *2*, 171–196.
(33) Hodson, M. P.; Dear, G. J.; Roberts, A. D.; Haylock, C. L.; Ball, R. J.; Plumb, R. S.; Stumpf, C. L.; Griffin, J. L.; Haselden, J. N. *Anal. Biochem.* **2007**, *362*, 182–192.
(34) van Ginneken, V.; Verhey, E.; Poelmann, R.; Ramakers, R.; van Dijk, K. W.; Ham, L.; Voshol, P.; Havekes, L.; Van Eck, M.; van der Greef, J. *Biochim. Biophys. Acta* **2007**, *1771*, 1263–1270.
(35) Verhoeckx, K. C. M.; Bijlsma, S.; Jespersen, S.; Ramaker, R.; Verheij, E. R.; Witkamp, R. F.; Van Der Greef, J.; Rodenburg, R. J. T. *Int. Immunopharmacol.* **2004**, *4*, 1499–1514.
(36) http://www.husermet.org/.
(37) Bijlsma, S.; Bobeldijk, I.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; Van Ommen, B.; Smilde, A. K. *Anal. Chem.* **2006**, *78*, 567–574.
(38) De Vos, R. C. H.; Moco, S.; Lommen, A.; Keurentjes, J. J. B.; Bino, R. J.; Hall, R. D. *Nat. Protoc.* **2007**, *2*, 778–791.
(39) Sangster, T.; Major, H.; Plumb, R.; Wilson, A. J.; Wilson, I. D. *Analyst* **2006**, *131*, 1075–1078.
(40) Gika, H. G.; Macpherson, E.; Theodoridis, G. A.; Wilson, I. D. *J. Chromatogr., B* **2008**, *871*, 299–305.
(41) Gika, H. G.; Theodoridis, G. A.; Wingate, J. E.; Wilson, I. D. *J. Proteome Res.* **2007**, *6*, 3291–3303.

MATERIALS AND METHODS

Chemicals. All chemicals (uridine, nicotinic acid, tryptophan, hippuric acid, raffinose, alanine, glycine, leucine, phenylalanine, citric acid, glutamic acid, caffeine, nicotine, formic acid, and glycocholic acid) and solvents (methanol, water, and acetonitrile) were of analytical grade or higher purity and were obtained from Sigma-Aldrich (Gillingham, U.K.).

Samples. Four different types of samples were used in the study. These are (1) Sigma serum QC (S7023, Sigma-Aldrich, Gillingham, U.K.); (2) pooled serum QC samples, 100 μL aliquots pooled from 30 samples in the study set; and (3) pooled spiked serum QC samples; pooled serum QC samples were spiked with a selection of metabolites to produce a sample with elevated metabolite concentrations to mimic inborn errors of metabolism or dietary intake. Serum was spiked with a standard mixture (3:1 ratio of serum to spiking solution) containing 100 μg mL^{-1} of alanine, glycine, leucine, phenylalanine, citric acid, glutamic acid, caffeine, and nicotine dissolved in 1:1 methanol/water. Final concentrations (which include the initial background metabolite concentration) in serum ranged from 200 to 400 μmol·L^{-1} and relate to typical concentrations as described in the Human Metabolome Database.[42] (4) Serum samples from 30 healthy individuals. These were randomly selected samples from the HUSERMET sample database. Further descriptions of the application of QC samples can be seen elsewhere.

Serum Sample Preparation. All serum samples were thawed on ice at 4 °C followed by deproteinization by the addition of methanol (1:3 ratio of serum to methanol, room temperature), vortex mixed for 15 s and centrifuged for 15 min at 15 871*g*. The supernatants were transferred to new Eppendorf tubes, lyophilized at 45 °C for 16 h (HETO VR MAXI vacuum centrifuge attached to a Thermo Svart RVT 4104 refrigerated vapor trap; Thermo Life Sciences, Basingstoke, U.K.) and stored at 4 °C prior to analysis. The samples were reconstituted in HPLC grade water (1:1 ratio of original serum volume to water), vortex mixed and centrifuged for 15 min at 15 871*g*. The supernatants were transferred to analytical vials, stored in the autosampler at 4 °C and analyzed within 48 h of reconstitution.

UPLC Column Test Mix Samples. For assessment of the performance of UPLC columns, a simple mixture of seven metabolites (100 μg·mL^{-1} of uridine, nicotinic acid, tryptophan, hippuric acid, raffinose, phenylalanine, and glycocholic acid) in water was used. Aliquots (200 μL) of this solution were lyophilized and handled in the same way as other samples. This UPLC column test mix solution was analyzed at the beginning and end of each analytical block with retention times and peak heights being compared for all detected compounds.

Instrumentation. All analyses were performed with a UPLC system (UPLC Acquity, Waters Ltd. Elstree, U.K.) coupled online to a TOF mass spectrometer (LCT Premier, Waters MS Technologies, Ltd., Manchester, U.K.). For high sensitivity, the instrument was operated in V mode, with DRE and lock mass correction. Optimized UPLC and MS settings for analysis in the ESI+ and ESI− modes are shown in Supplementary Tables 1 and 2 in the Supporting Information.

Routine instrument maintenance consisted of pumping 100% acetonitrile through the pump heads and UPLC column for a minimum of 30 min at a flow rate of 0.4 mL·min^{-1} and a column temperature of 50 °C. The MS source (sample cone and baffle components) was cleaned by sonication in 50/40/10 methanol/water/formic acid for 15 min. The frequency of cleaning varied and is detailed in the Results and Discussion section.

Data Preprocessing. All raw data files were converted to NetCDF format using the Waters DataBridge software on a Windows PC. The freely available XCMS software[43,44] was employed to convert (or deconvolve) each 3-D data matrix (intensity × m/z × time − one per sample) into a matrix of detected peaks vs sample identification (ID) with peak response for detected metabolites reported, where a peak response is defined as the sum of intensities over a window of specified mass and time ranges. Default settings were employed in XCMS with the exception of S/N threshold (3), mass limit (0.1 amu), time limit (15 s), and sample limit (3).[45] Subsequent statistical analysis was performed in Matlab (version 7.4).

RESULTS AND DISCUSSION

There are a number of experimental variables and sources of variation that can impact the repeatability and validity of metabolic profile data, and this can limit the data's biological relevance and applicability. The importance of such parameters is amplified in long-term studies. Here, we show the development of methodologies and application of UPLC coupled to an electrospray time-of-flight mass spectrometer used for the long-term metabolomic study of human serum. We also demonstrate how a quality control (QC) sample can be used to assess repeatability for such a method.

Closed-Loop Optimization of UPLC−MS Methods. A set of optimal UPLC (reversed phase with water/methanol gradients) and mass spectrometer settings for investigation of the serum metabolome were determined by applying a closed-loop multiobjective optimization process, as previously applied for GC/MS and comprehensive GC × GC/MS.[46,47] The three objectives were number of detected peaks (optimal is high), chromatographic resolution (optimal is high), and analysis time (optimal is low). These objectives were chosen as distinct features of the separation and to provide a wide coverage of the metabolome while ensuring adequate sample throughput. Two separate methods were obtained from 130 and 150 injections for positive (ESI+) and negative (ESI−) ion modes, respectively. These methods are described in Supplementary Tables 1 and 2 in the Supporting Information. For QC serum samples, this methodology resulted in detection on average of greater than 1600 and 1100 features in ESI+ and ESI−, respectively, with fewer than 20% missing values. Representative chromatograms for analysis of QC serum samples in ESI+ and ESI− are shown in Figure 1.

A greater number of detected features were observed as the analysis time increased, as has been reported previously.[46−48] There is thus a relationship between throughput and acquired biological information which is of interest for metabolomics, where hundreds of samples are analyzed compared to the typically considerably fewer samples in proteomic and transcriptomic

(42) http://www.hmdb.ca/.

(43) http://masspec.scripps.edu/xcms/xcms.php.

(44) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.

(45) Dunn, W. B.; Broadhurst, D.; Brown, M.; Baker, P. N.; Redman, C. W.; Kenny, L. C.; Kell, D. B. *J. Chromatogr., B* **2008**, *871*, 288–298.

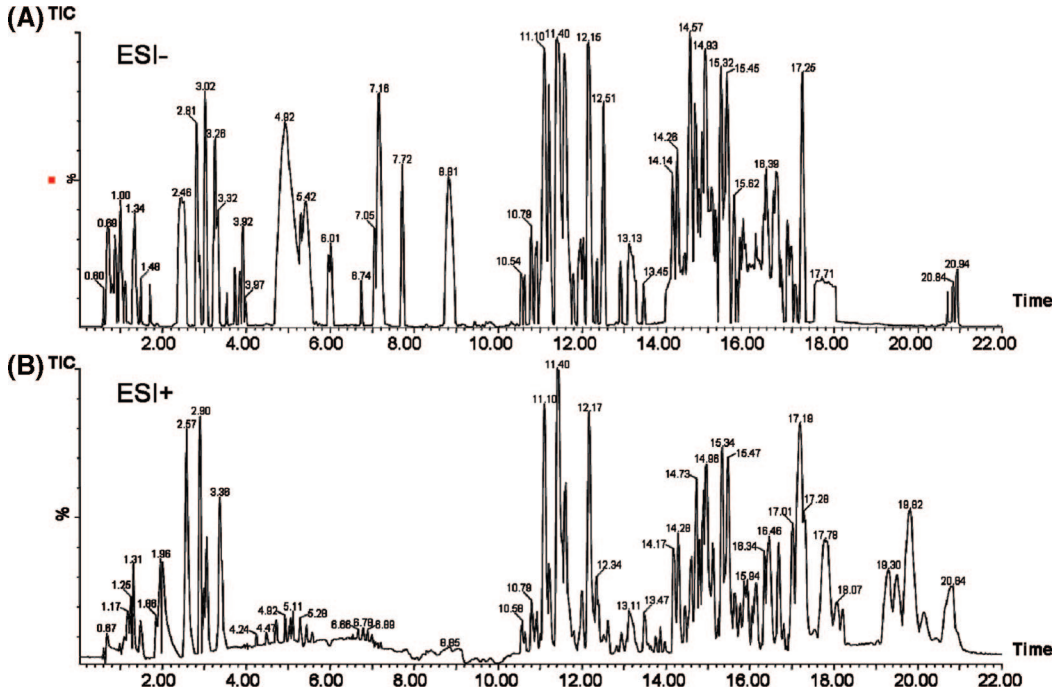

Figure 1. Base peak ion (BPI) chromatograms of pooled serum QC acquired using optimized UPLC−TOF-MS methods in (A) ESI− and (B) ESI+.

studies. Appropriate mass spectrometer tuning to allow detection of ions across wide mass ranges (50−1000 Da) was confirmed to be important. Radio frequency lens and ion guide voltages significantly affected both the detectable mass ranges and responses. Tuning on a single compound (as is typically applied) is not always appropriate to optimize these settings. Of interest in this method is that sub-optimal mobile phase velocities/flow rates are being applied, and typically peak widths of 3−10 s are observed as shown previously.[45] Although higher linear velocities are generally required for high chromatographic resolution on UPLC instruments, there is a tradeoff between scan speed and sensitivity, and this was considered, so a compromise was to use acquisition times in the range 0.2−0.5 s. In fact, for the TOF mass analyzer applied here, scan times greater than 0.40 s were found to provide a higher sensitivity.

Sample preparation consisted of methanol-based protein precipitation, followed by lyophilization and reconstitution in HPLC grade water prior to analysis. Protein precipitation in methanol was chosen as previous research has shown this as "fit for purpose" for efficient precipitation of proteins.[49,50] However, it was observed that differences in the reconstitution volume of samples was necessary for ES+ (no dilution) and ES− (2:1 dilution) to ensure no detector saturation for high concentration metabolites. This shows that analytical workflows are instrument-dependent for a given biofluid and tuning of ion production and transfer processes, which can be biased to certain sections of the mass range, can influence the sample concentration required.

UPLC Column Comparison. In long-term studies, many chromatographic columns will be used and these will be sourced from more than one production batch from a single manufacturer. There is the requirement for columns packed with different production batches of chromatographic stationary phase not only to provide reproducible performance over their operational lifetimes but also for data generated from columns originating from different batches to be comparable. To assess inter-batch column reproducibility, an experiment was performed in ESI+ where two identical blocks of 60 serum QC samples were prepared and consecutively injected on two UPLC columns from different production batches. Each column had not previously been used and the mass spectrometer source was maintained to remove any debris, or sample build-up, that may influence the sensitivity of ionization or detection.

The distribution of peak-wise relative standard deviation (RSD) for each column was similar, with more than 1000 reproducible peaks having RSD < 20% for each column. Principal Component Analysis (PCA) shows that no separation was observed in the PCA space of data obtained on both of these columns as shown in Figure 2. These results indicate that data obtained on the two assessed UPLC columns were comparable with "fit for purpose" reproducibility. Data describing reproducibility for analyses performed over 3 years is currently being acquired. Confidence in the reproducibility of these columns through QA manufacturing processes is essential in long-term studies. In the authors' experience up to 400 samples can be injected onto UPLC columns, with intermittent cleaning, before degradation of data quality.

Time-Related Drift in Instrumental Performance. The quality of analytical data in metabolomic studies acquired using chromatography−mass spectrometry platforms may in part be

(46) O'Hagan, S.; Dunn, W. B.; Brown, M.; Knowles, J. D.; Kell, D. B. *Anal. Chem.* **2005**, *77*, 290–303.
(47) O'Hagan, S.; Dunn, W. B.; Knowles, J. D.; Broadhurst, D.; Williams, R.; Ashworth, J. J.; Cameron, M.; Kell, D. B. *Anal. Chem.* **2007**, *79*, 464–476.
(48) Guy, P. A.; Tavazzi, I.; Bruce, S. J.; Ramadan, Z.; Kochhar, S. *J. Chromatogr., B* **2008**, *871*, 253–260.
(49) Jiye, A.; Trygg, J.; Gullberg, J.; Johansson, A. I.; Jonsson, P.; Antti, H.; Marklund, S. L.; Moritz, T. *Anal. Chem.* **2005**, *77*, 8086–8094.
(50) Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 743–752.

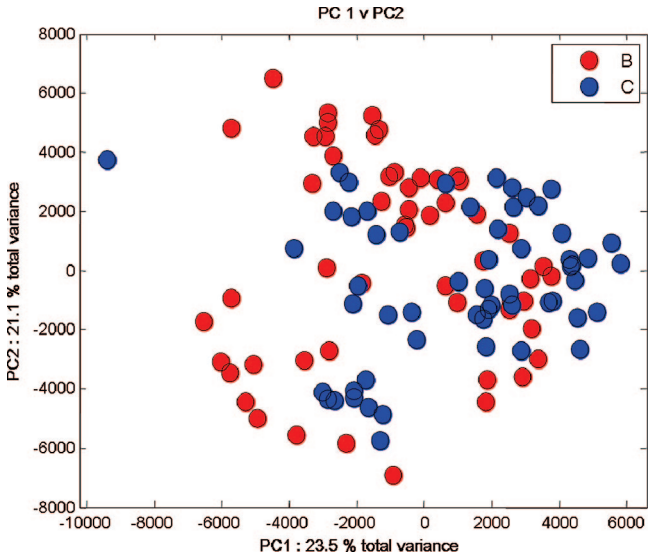

Figure 2. PCA scores plot for data acquired using two UPLC columns from different manufacturing batches. The scores for all the samples analyzed on both columns are overlaid and there is no visual separation between the results from both columns.

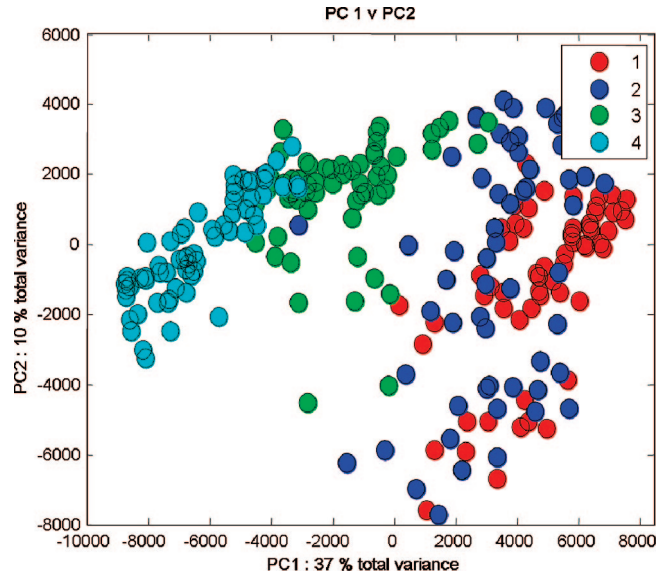

Figure 3. PCA scores plot for four identical blocks (each consisting of 60 Sigma QC samples) which were analyzed over a period of 5 days. The scores for block one and two are overlaid, but there is a significant drift observed for blocks three and four. As principal component 1 contributes to about 37% of the total observed variance, it is obvious that the data from the third and fourth blocks are not identical.

influenced by the length of the analytical run. The appropriate length of the analytical block was assessed to provide the number of injections acceptable before data quality (sensitivity and retention time) is irretrievably affected. As a first experiment, 240 identical quality control (QC) samples were analyzed in a continuous run of four blocks (60 QC serum samples in each block), requiring analysis over a 5 day period (without any stop during the analysis, nor preventative maintenance). Samples from each block were freshly reconstituted at the start of each of the 4 days to ensure representative stability of all four sets of 60 QC samples.

Figure 3 shows a PCA scores plot (PC1 vs PC2) for all the samples analyzed over the 5-day period (color coded into 4 blocks of 60 samples). The scores for block one and two show no separation implying that these two blocks are statistically similar when compared to blocks three and four where considerable drift is observed in PC1, which contributes to 36.9% of the variance.

Shewhart's charts[51] enable the visual assessment of the changes in responses over time for individual peaks. It was observed that the change in response for any peak was generally nonlinear over time and also that the change in response over time was not comparable between different peaks (the peak area of some peaks decreased over the period of 5 days, while the peak area of other peaks increased). This variability is most likely caused by contamination of the source or column.

From these results we can conclude that there is significant observable drift in response over time. Generally, this drift is observed to be nonlinear, and not all the metabolites in the sample follow the same trends. These time-related (or longitudinal) artifacts can not be easily compensated for, with, for example, the application of internal standards, where an internal standard for each metabolite would be appropriate because of the wide chemical and physical diversity of metabolites.

While there are no generally accepted criteria for the assessment of repeatability in metabolomic data sets, the Food and Drug Administration (FDA) suggests a range of criteria that should be applied. In the guidance for bioanalytical method validation in industry,[52] the FDA recommends for single analyte tests that tolerance limits are set such that the measured response detected in two-thirds of QC samples is within 15% of the QC mean, except for compounds with concentrations at or near the limit of quantification (LOQ), in these cases a tolerance of 20% is acceptable. In our case, the methods are not specific for one analyte of interest, but instead we aim to detect thousands of analytes, therefore an acceptance tolerance of 20% would seem to be appropriate. For the above experiment, the acceptance criteria for repeatability were generally achieved for injections 1 to 100, after which unacceptable drift was observed (data not shown). Supplementary Figure 1 in the Supporting Information shows the distribution of RSDs for injections 1−240. Only about 450 peaks, out of over 1600 peaks detected, fall within 20%.

Optimal Length of the Analytical Block. The results described above show that the instrumental drift becomes intolerable after about 100 injections. In view of the analysis time for UPLC−MS and length of the HUSERMET study, a maximum number of injections per week of 180 was deemed appropriate for UPLC−MS so not to require overweekend runs. This set of injections can be split into 120 subject samples and 60 QC samples for UPLC−MS and requires a 4 day period of sample preparation. Applying these sample numbers allows samples to be prepared for both GC/MS and UPLC−MS from a single aliquot of the serum sample, therefore using precious samples in an efficient manner. Two additional experiments were performed to assess more closely the optimal length of an analytical block and the influence of preventive maintenance on the quality of acquired data.

(51) Wheeler, D. J. *45th Annual Quality Congress Transactions*, Milwaukee, WI, 1991; pp 533−538 (published by ASQC).

(52) FDA Guidance for Industry, Bioanalytical Method Validation, Food and Drug Administration, Centre for Drug Valuation and Research (CDER), May 2001.

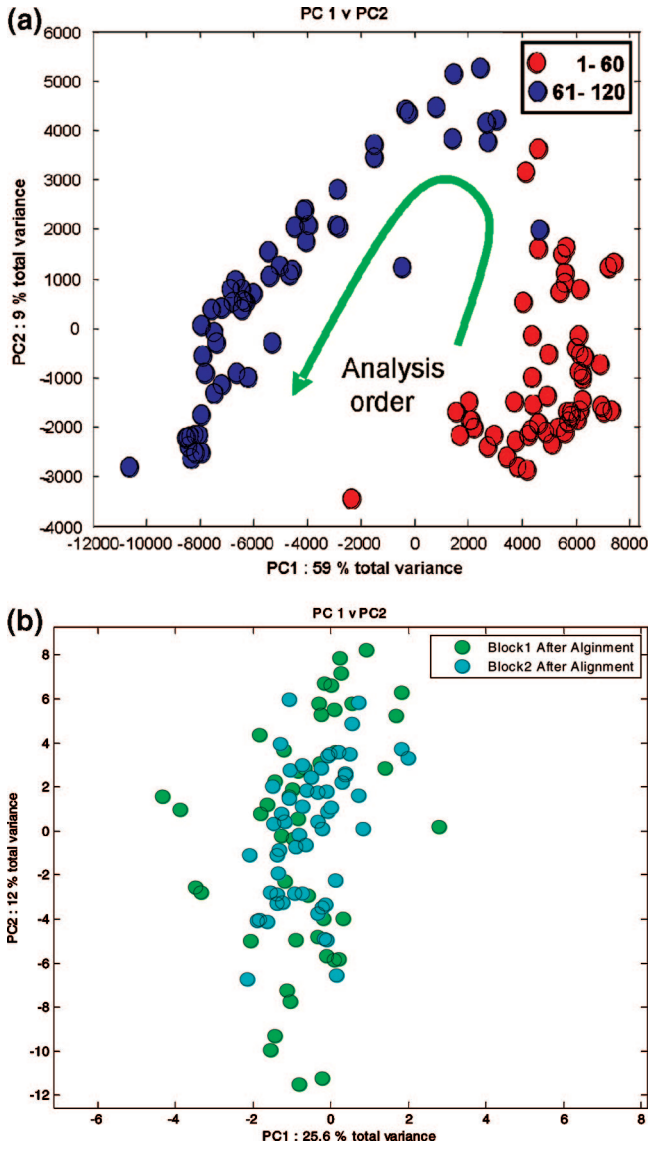

Figure 4. (a) PCA scores plot for a batch of 120 consecutively analyzed Sigma QC samples. The first 60 injections are in red, injections 61−120 are in blue. The green arrow shows the direction of analysis order through the data. There is observable drift connected to analysis order. (b) PCA scores plot for 2 blocks of 60 Sigma QC samples with a cleaning step performed between each batch. Each of the batches is clustered together even though there is an observable difference between them. Injections 1−10 were removed before PCA as these injections are used for conditioning of the analytical platform.

In the first experiment (Expt-A), 120 QC serum samples were injected consecutively onto a previously unused UPLC column. In the second experiment (Expt-B), two blocks of 60 samples were injected, also on to a previously unused UPLC column, with a preventive maintenance step undertaken between these two blocks (the mass spectrometry source and the analytical column were cleaned).

The results from Expt-A demonstrated that the batch consisting of 120 samples was too long and negatively affected the repeatability of the obtained data. This is illustrated in Figure 4a, which shows a PCA plot for data from the experiment consisting of 120 consecutive injections (color coded for clarity, injections 1−60 in red and injections 61−120 in blue). There is an observable nonlinear drift across PC1 vs PC2 (69% of variance). Examination of these data showed that the first 60 injections were randomly clustered, and therefore the data are repeatable when compared to injections 90−120 where the majority of the drift is observed. This inconsistency, which was due to changes in peak area over time between different peaks, is demonstrated in Supplementary Figure 2 in the Supporting Information which shows Shewhart's charts (peak area vs order of instrumental analysis) for two representative peaks. In contrast, Figure 4b shows the PCA plot (PC1 vs PC2) for Expt-B. Two tightly clustered groups, which are observed in the same PCA space for batches 1 and 2, are observed showing that 60 injections appear to be suitable for our application, whereas 120 injections were too high. This highlights that multiple blocks can be compared in long-term experiments and is possible by the use of the QC samples to correct for drift in instrument performance. A secondary use of the QC samples, other than for assessing analytical reproducibility, is to correct for any peak-wise signal attenuation due to order of injection. Here a low-order nonlinear locally weighted spline (LOESS) is fitted to the QC data with respect to the order of injection. A correction curve for the whole block of injections is then interpolated, to which the total data set for that peak is normalized. A similar methodology has recently been presented.[53] As the data has now been normalized to the QC samples, multiple blocks can be readily combined into one data set for statistical analysis.

The conclusions from the PCA results, suggesting that splitting a long experimental run into two short blocks leads to acquisition of more reproducible data, are further reinforced by Figure 5 which compares the RSDs of peak areas for all the peaks detected in an experiment with 120 consecutive injections (Expt-A) and an experiment with two blocks of 60 injections (Expt-B). The reproducibility of injections 1−60 for both experiments is similar. However, the distribution for injections 61−120 in Expt-A is significantly shifted to a higher mean RSD, whereas in Expt-B injections 61−120 (second block, after preventive maintenance) has a comparatively similar distribution to data for injections 1−60 (first block). This highlights that cleaning of the instrument after 60 injections significantly increases the longer-term high-quality of data. Supplementary Figure 3 in the Supporting Information shows the Shewhart's charts for two randomly selected peaks. It is obvious that the repeatability significantly improved in this setup as all the measured areas of these peaks fall within the 20% tolerance limits.

Testing the Optimal Analytical Batch Length for a Set of 30 Serum Samples from a Human Population. To ensure that we can detect differences in the metabolomes of healthy individuals and those metabolomes containing atypically high concentrations of metabolites, a further experiment was performed. Samples from 30 healthy individuals were analyzed and compared to a pooled sample of these individuals spiked with a range of metabolites. Two types of QC samples were included in this study to assess within-batch repeatability of (a) commercially available serum obtained from Sigma-Aldrich and (b) pooled serum from the 30 individuals employed in this study. In short-term studies, a pooled serum QC of the individuals involved in the study is

(53) van der Greef, J.; Martin, S.; Juhasz, P.; Adourian, A.; Plasterer, T.; Verheij, E. R.; McBurney, R. N. *J. Proteome Res.* **2007**, *6*, 1540–1559.

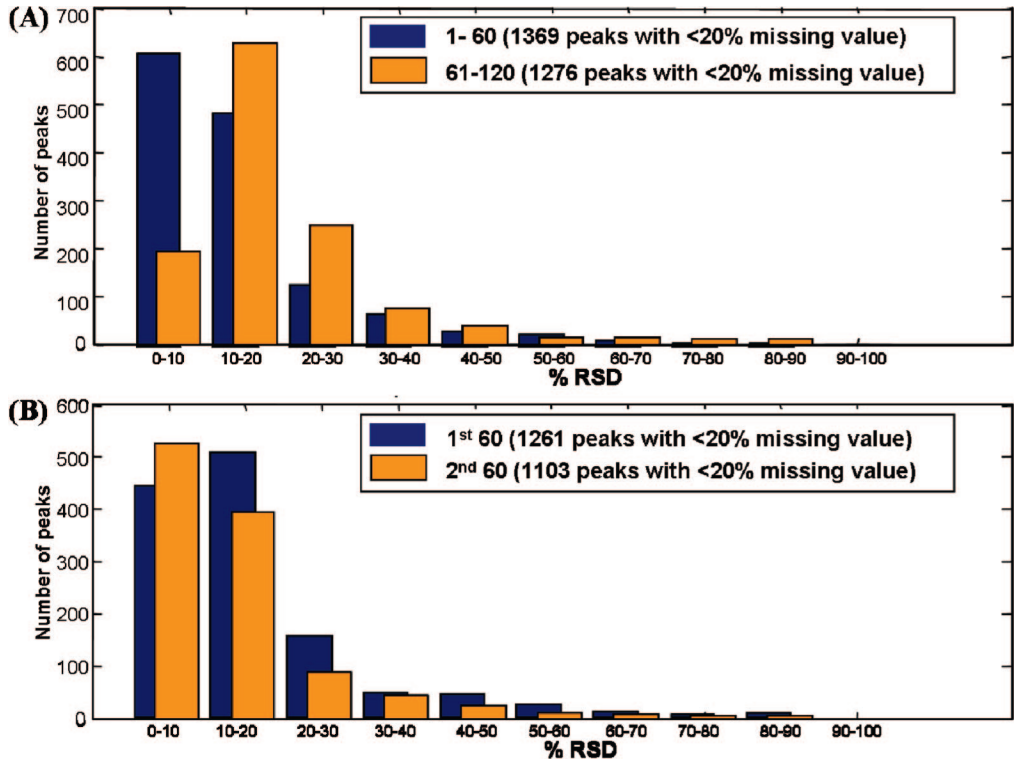

Figure 5. Comparison of the distributions of relative standard deviations of peak areas for (A) injection of 120 samples vs (B) injection of two blocks of 60 samples with instrument cleaning between blocks. The distribution of RSDs for the first 60 samples (in blue) from both experiments are comparable. RSDs of the peak areas from the second half of each sample set (in yellow) differ in distribution; in part A the RSD distribution is shifted to higher RSDs whereas for part B the RSD distribution is similar for both blocks of 60 samples.

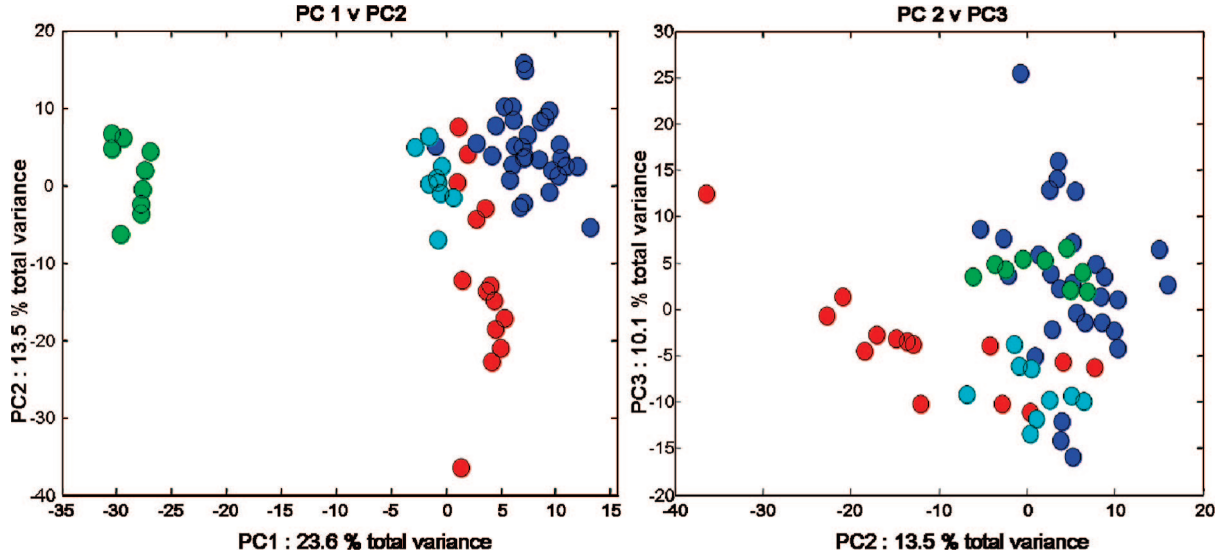

Figure 6. PCA scores plot for the comparison of serum from a healthy population ($n = 30$) (in dark blue), QC serum (in red), and spiked QC serum (in light blue). In PC1 vs PC2, the Sigma serum QC (in green) separates from other samples. The spiked pooled QC samples are separated from healthy subject samples (this shows the potential for using this approach in biomarker discovery). There is strong drift in the first 10 injections on a new UPLC column. The first 10 injections on the column are QC samples only. The only observable feature in the PC2 vs PC3 plot is the drift in the first 10 injections on the analytical columns (this emphasizes the importance of conditioning the column before the analysis is commenced).

applicable. However, for long-term studies, a pooled serum distinct from the individuals being studied is required.

Figure 6 shows the PCA plot for all samples analyzed in this experiment. There is a clear separation of the samples from healthy individuals compared to a spiked pooled sample (spike solution contained metabolites present in increased concentrations in inborn errors of metabolism (alanine, glycine, leucine, phenylalanine, citric acid, and glutamic acid) and also metabolites commonly present in samples due to lifestyle (caffeine and nicotine). The differences in the metabolomes of these two sample classes were observed in this study; therefore, the analytical platform applied here is appropriate (with satisfactory discriminatory power) for detecting metabolic differences between the serum metabolomes of healthy individuals and

those metabolomes from individuals in a different metabolic state caused by inborn errors of metabolism or dietary intake.

The pooled QC samples showed a time-related drift for injections 1−10. Previous research, particularly on urine analysis, e.g.,[39−41] has shown that "conditioning" of the analytical system is required by the injection of a QC sample multiple times, a process that equilibrates the system so that further analyses provide minimal further disturbance to the analysis. In a previous study looking at the HPLC−MS analysis of urine, it was reported that injection of five QC samples was necessary for the stabilization of the system before the analysis is commenced.[39] Our data suggest that the stabilization requirements prior to analysis are technology as well as sample-dependent. Our recommendation is that for serum analysis at least 10 conditioning injections of QC (or similar) samples are required before the commencement of the analytical run.

The Sigma QC sample clusters away from all other samples in PC1 vs PC2, but this separation is not observed in PC2 vs PC3. Of interest is that of 1214 peaks detected in ESI+ for the Sigma QC samples, just 687 are repeatibly detected in both the Sigma QC and healthy subject samples, showing a distinct compositional difference between these samples.

The RSD of peak area for all peaks detected in different sample types shows that a lower median RSD is observed for multiple injections of the same sample (QC samples) when compared to single injections of samples taken from a population of healthy individuals. This highlights that the technical variation is lower than the biological variation, which increases the chance of true discovery as well as decreasing the chance of false discovery.[45] In order to make the subsequent data analysis viable, in total 90 injections were performed in this experiment. The stability of the QC response was monitored and found to be acceptable.

The Sigma QC samples dispersed throughout the analytical run (of length 40 h) all clustered together and therefore it can be assumed that samples do not change metabolically during storage at 4 °C in the UPLC autosampler for at least 40 h (at least at a level detectable by this type of statistical analysis). In addition there was no observable drift related to analysis order which may indicate sample (or instrument) instability. It is recommended to analyze all samples within 40 h of reconstitution as a standard, though samples may be stable for longer periods. In our protocols we prepare the first 10 QC samples and start an analytical batch to ensure both analytical platforms are operating adequately before preparation of all other samples, so as not to waste important samples if an instrument problem is observed.

CONCLUSIONS

The data presented here have demonstrated the complexity of analytical operations performed in metabolomic investigations, specifically sample analysis of serum by UPLC−MS. A range of sources of variation were observed and it is imperative to minimize these variations consistently. We have shown here that, with appropriate control methods in place, UPLC−MS has the potential to serve the required role of a hypothesis-generating holistic acquisition of data related to human serum for thousands of metabolite-related ions. These control measures employ biologically identical quality control (QC) samples which are analyzed periodically throughout the analytical block to assess repeatability. We recommend the injection of 10 QC samples at the start of each block to condition the UPLC−MS system followed by the analysis of a QC sample every fourth injection. Data from the analysis of QC samples describe repeatability of the response and/or retention time across an analytical block for each metabolite and allows the removal of data not deemed "fit for purpose". Here we apply a ruling of removing data with a RSD greater than 20% for all QC samples. A UPLC column test mixture should also be analyzed after the initial 10 QC samples, in the middle, and at the end of the analytical block. This allows chromatographic performance to be assessed and allows poor quality data sets to be detected in a quick and efficient manner before further time-consuming data analysis.

The block size is hugely important to the repeatability of the data. This will always be a compromise between technical reproducibility and statistical validity in that statistical power is obviously dependent on sample number. For serum studies described here, fewer than 90 injections in the testing of the analytical block size experiment (60 samples and 30 QC samples) has been shown to provide results of adequate repeatability. Substantial and unacceptable drift was observed with an analytical block of greater size (greater than 100 injections for the analytical platform discussed). Studies employing urine as the biofluid have used 77 injections per block[48] or 114 per block.[40] In this study where 120 subject samples will be analyzed per week, two blocks of 90 injections (60 samples and 30 QC samples) with preventative maintenance in-between blocks was chosen and showed higher repeatability than a single block of 120 injections. This is our standard operating procedure employed in all UPLC−MS metabolomic studies. The results have allowed the authors to prepare SOPs for "fit for purpose" long-term metabolomic studies, such as are being employed in the HUSERMET project which allow the acquisition of data that will allow comparison of data acquired across many months or years.

ACKNOWLEDGMENT

The authors E.Z., W.D., D.B., S.F.-M., and K.C. contributed equally to this work. W.D., K.M.C., I.W., and D.B.K. wish to thank BBSRC and EPSRC for financial assistance. S.F.-M., P.B., and A.H. wish to thank BBSRC, EPSRC, AstraZeneca and GlaxoSmithKline for financial assistance of the HUSERMET project.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review September 12, 2008. Accepted December 11, 2008.

AC8019366