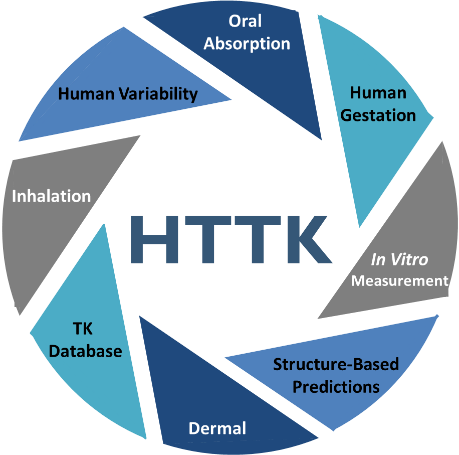
# HTTK Data Curation Standard Operating Procedure



Authors: John Wambaugh, Mark Sfeir, Marci Smeltz, Anna Kreutz, Barbara Wetmore

In order to address greater numbers of chemicals we collect *in vitro*, high throughput toxicokinetic (HTTK) data.[[1]](#footnote-1) HTTK methods have been used by the pharmaceutical industry to determine range of efficacious doses and to prospectively evaluate success of planned clinical trials.[[2]](#footnote-2) The **primary goal** of HTTK is to provide a human dose context for bioactive *in vitro* concentrations from in vitro high throughput screening (that is*, in vitro-in vivo* extrapolation, or **IVIVE**).[[3]](#footnote-3) A **secondary goal** is to provide **open source data and models** for evaluation and use by the broader scientific community.[[4]](#footnote-4)

Statistical analyses are needed to extract the key parameters from multiple in vitro measurements. In order to standardize these analyses, the data from the in vitro experiments must be organized in a way that allows systematic statistical analysis. This standard operating procedure (SOP) describes that process.

## Relevant Definitions

We use the terms “chemical” and “compound” to interchangeably to refer to a substance with a unique DSStox Substance ID (DTXSID)

We use the terms “directory” and “folder” interchangeably to describe a file location on a computer, for example L:\Lab\NCCT\_ExpoCast\ExpoCast20XX\

A “sub-directory” or “sub-folder” is a directory within another directory, for example L:\Lab\NCCT\_ExpoCast\ExpoCast20XX\HTTKNewData

## The HTTK Assays

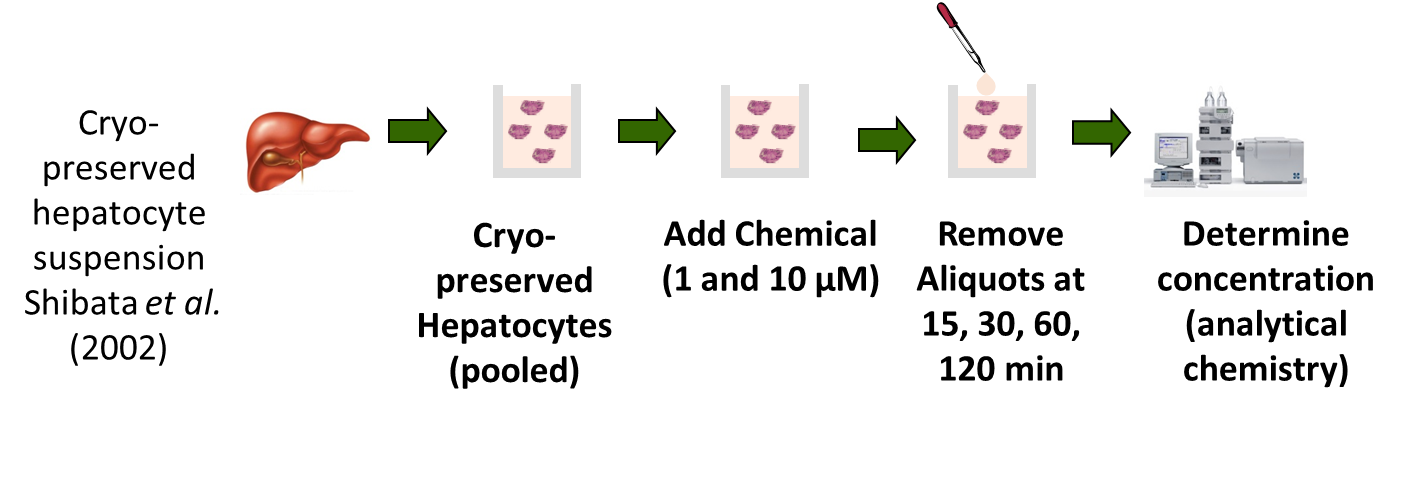
HTTK makes use of multiple in vitro assays that inform toxicokinetics. As more assays become available, new types of measurements may be added. Depending on when a chemical was measured (and what the goals for that set of data were) not all assays may have been conducted.

The data collected to support chemical-specific predictions of toxicokinetics include four types of in vitro TK measurements:

1. Stability of compound when incubated with primary hepatocytes (that is, intrinsic hepatic clearance)
2. Fraction of compound unbound in the presence of plasma protein
3. Caco2 membrane permeability
4. Relative concentration in blood and plasma

All measurements include replicates for statistical analysis. All measurements include control chemicals (for example, propranolol is often used) – these will be repeated in every batch of data to demonstrate that the assay is working. Their identity will be different from the ToxCast chemicals, which are usually only indicated with a DTXSID (<https://www.wikidata.org/wiki/Property:P3117>). We do want to capture this control data because it provides a measure of experimental variability.

### Metabolic Stability[[5]](#footnote-5)



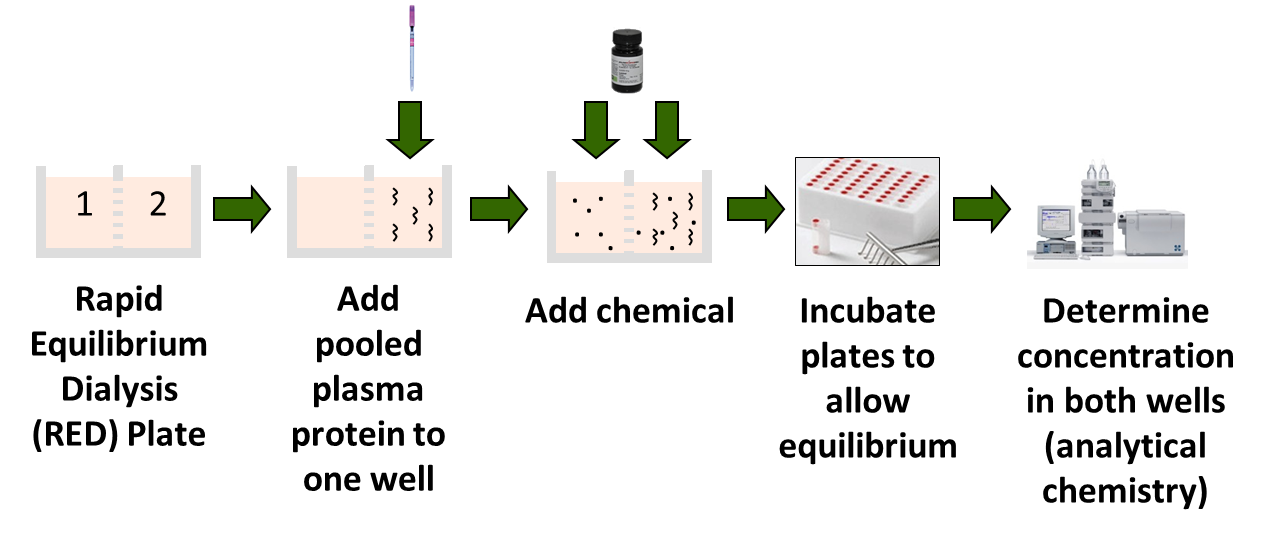
Also know as intrinsic clearance (Clint)

This is a series of peaks areas measured at different times as the chemical incubates with hepatocytes (liver cells that metabolize many chemicals) for up to four hours.

### Fraction Unbound in Plasma

We use two different assays to measure chemical fraction unbound in plasma: 1) rapid equilibrium dialysis and 2) ultracentrifugation.

#### Rapid Equilibrium Dialysis[[6]](#footnote-6)



This is a series of peaks from different fractions of the same experiment to try to assess the ratio of the concentration of chemical without plasma to the concentration in the presence of plasma (Fup).

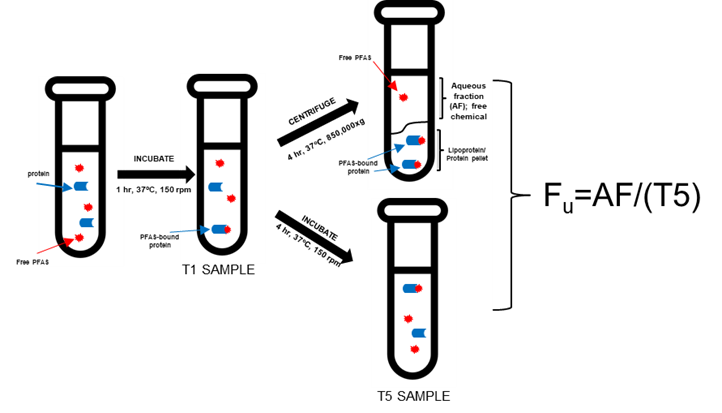
For new measurements HTTK (>200 compounds to data) performed by Cyprotex, we have modified RED protocol to use a titration of plasma protein (10%, 30%, 100%) of physiological concentration

Keeps chemical concentration in the same range

That way highly bound chemicals that are <LOD at 100% are potentially visible at lower concentrations.

Also allows us to estimate binding affinity.

#### Ultracentrifugation



|  |
| --- |
| Additionally, plasma stability time points at 1 hr (initiation of UC) and ~5hr were included |
| Fu = aqueous fraction/T5 hr |
| Dilution corrections: aqueous fraction assay 1:2; stability time points 1:5; for all samples, 1:4 dilution for crash (contains internal stds) and 1:4 dilution for LC-MS analysis in matched mobile phase |
| n-butylparaben was included as referance compound/positive control |
|  |
| **Calibration Curves-** |
| Used mixed matrix (50/50 mixture of human plasma and ultrafiltrate [aqueous fraction of plasma]) as matrix for curve |
| Cal curves range from 0.053 pg/uL (0.00017 μM) - 76.35 pg/uL (0.25 μM) [up to 17 points]; pg/uL notation based off avg MW of 305 g/mol, but subsequently updated into μM for better mass balance |
| Cal curve points were diluted 1:4 for crash and 1:4 for LC-MS analysis |
| Internal standard concentration at instrument is 3 pg/uL (0.010 μM) |
| Curves were prepared either same day or within 2 weeks of assay completion per QAPP |
|  |
| **TAB: Fu Values** |
| This is a summary of PFOA (DTXSID8031865) and K+PFOS (DTXSID8037706) with values that were generated from two different assays (for each analyte) over the last year |
| Based on QAPP, %RSD should be within 25% |
| I have excluded Mix2A from July 2019 for PFOS as it was *identified* to be contaminated, as a more elevated reading was also noted in n-butylparaben |
| The Fu information for n-butylparaben associated with mixes containing these PFAS is also included |
|  |
| **LC-MS Data Interpretation-** |
| Worklists are shown by analyte |
| LC-MS method that was run included examining all analytes across all samples |
| Calibration curves were fit to a quadratic equation with 1/X weighting and an output of the curve for each analyte is shown on side of the data for that particular PFAS |
| Internal standards were assigned based on similarity to PFAS or elution time |
| Sample text provides a description of the sample; CC = calibration curve, QC = quality control, UC\_UF = unbound fraction, UC\_T# = plasma stability time point |
| RT = retention time; response = (peak area analyte/peak area internal standard) \* internal standard concentration |
| Std. conc = theoretical concentration; uM = actual concentration |
| (there's lots of other parameters that can be inputted too if interested to learn what those are) |

### Blood to Plasma Ratio

Ratio of chemical concentration in blood to chemical concentration in plasma

### Caco2 Permeability

Rate of movement of the chemical through a cellular membrane

## Mass Spectrometry Chemical Analysis

The HTTK in vitro assays need to measure differences in chemical concentration. Analytical chemistry via mass spectrometry is a technique used to determine chemical concentration. Mass spectrometers determine a “peak” whose area is proportionate to a chemical of a certain mass in a sample. An

internal standard (ITSD) at a known, fixed concentration is added to the sample –its area fluctuates somewhat with time. The analytical chemist finds the peak that corresponds to the chemical of interest, and then follows the ratio R of the area of the chemical peak to the area of the ITSD peak to determine concentration.

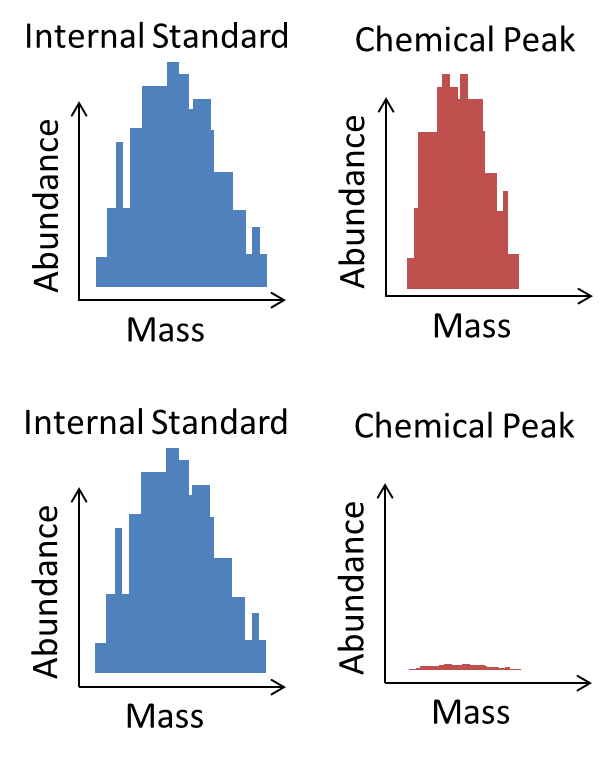


Figure 1: While the internal standard is always selected such that the mass spectrometer has a clear signal, the analyte being studied in that sample may have a less than optimal signal.

There are many types of mass spectrometry, including liquid chromatography (LC/MS) and gas chromatography (GC/MS). Different chemical structures may be detectable by multiple methods, one method, or none. Further, mass spectrometry varies significantly by the specific equipment used and the techniques for amplifying or selecting certain chemical signals.

Samples for mass spectrometry are diluted to keep the analyte signal at a reasonable value and separate out signal from the background. The number of times that a sample is diluted is the “dilution factor”. The dilution factors account for the background “noise” introduced by different matrices in the sample. For instance, in the RED assay we have samples that are just in phosphate-buffered solution vs samples that also have plasma added. We dilute the plasma samples more to try to lower the background introduced by all the different materials in the plasma. For Clint most all the samples have the same background – hepatocytes -- so we dilute them all the same amount.

## Date Generation and Delivery

In vitro data for HTTK are being generated in many places. Dr. Wetmore’s laboratory in the Center for Computational Toxicology and Exposure in the EPA’s Office of Research and Development (ORD) is a primary example. ORD also has many collaborators who either generate the data themselves or employ contractors to do so.

Cyprotex (<https://www.cyprotex.com/admepk>) is currently under contract to ORD to provide these data. EPA issues “task orders” (TOs) specifying a list of chemicals (typically 100-200) and in vitro TK properties to be measured. EPA provides chemical samples to the contractor. The chemical samples are from the ToxCast chemical library. Each sample is assigned a unique identifier by the ToxCast chemical library curators.

Data is typically delivered to us as a Microsoft Excel workbook (.xlsx file). These files often have multiple sheets. The generator may have placed summary data on some sheets and raw data on others.

It can take many months to generate the data on several hundred chemicals, so the data from a single task order may arrive in batches.

## Data Organization and Analysis

Raw data consists of peak areas from mass spec for the chemical under study and the ISTD. We use the “ISTD ratio” ratio of analyte peak area to internal standard peak area.

|  |  |  |
| --- | --- | --- |
| Level 0 | Data As Provided by the Data Generator | Initially in primary data set director (such as “TO1”) then moved to “processed” sub-directory once added to raw file. |
| Level 1 | Data Organized into Standardized “Raw” Files | A single file for each data type with standardized columns. |
| Level 2+ | Results of Statistical Analyses of Data | Often ends up reported as tables in scientific papers. Typically added into R package “httk”. |

Laboratories typically do their own analyses of their data for quality management. They may refer to this as the data “Summary” since it draws from many raw mass spectrometry data measurements. However, once we have the data we do our own analysis of the raw data and typically discard the summary from the data generator. The algorithms used for analysis generate new summary numbers (level 2) using multiple different approaches. In particular, we use a Bayesian framework that includes a model for analytical chemistry[[7]](#footnote-7). The Bayesian approach estimates uncertainty via a credible interval (range of values that would be consistent with the data).

The methods for analyzing hepatic stability and fraction of chemicals were most recently developed and described in Wambaugh et al. (2019)[[8]](#footnote-8). The methods for analyzing Caco2 permeability data are being developed and described by a new Honda et al. manuscript. The methods for analyzing blood:plasma ratio data have not been previously developed for the HTTK project.

## Curation Process

The data files are stored in the directory L:\Lab\NCCT\_ExpoCast\ExpoCast20XX\HTTKNewData

The compiled data files are stored in the main folder, while files received from the contractor are stored in sub-directories according to the source (for example “CyprotexTO1”). When a file from the contractor has been processed (that is, all relevant compiled data files have been updated based on the file from the contractor) then the file from the contractor is moved to a sub-folder named processed (for example “CyprotexTO1/processed”).

We rely on the CompTox Chemicals Dashboard for chemical identities:

<https://comptox-prod.epa.gov/dashboard/dsstoxdb/batch_search>

At this point, we simply cut and paste from the Level 0 files into the Level 1 (“raw”) files

### Compiled Data Files

**all-files-received.xlsx:** Description of all files received by EPA from the contractor

**analytical-methods.xlsx:** Description of the analytical method used for each chemical

**raw-clint-data.xlsx:** Mass spectra data for metabolic stability assays

**raw-ppb-data.xlsx:** Mass spectra data for plasma protein binding assays

**raw-caco2-data.xlsx:** Mass spectra data for the Caco2 permeability assays

**raw-rb2p-data.xlsx:** Mass spectra data for blood 2 plasma assays

**sample-data-summary.xlsx:** Summary of information received on a per-sample basis

The raw files and the analytical methods file will be provided as supplemental material in a peer-reviewed paper associated with the data.

## Example

1. Rotroff, Daniel M., et al. "Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening." Toxicological Sciences 117.2 (2010): 348-358.

   Wetmore, Barbara A., et al. "Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment." Toxicological Sciences 125.1 (2012): 157-174.

   Wetmore, Barbara A., et al. "Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing." Toxicological Sciences 148.1 (2015): 121-136.

   Wambaugh, John F., et al. "Assessing Toxicokinetic Uncertainty and Variability in Risk Prioritization." Toxicological Sciences 172.2 (2019): 235-251. [↑](#footnote-ref-1)
2. Jamei, Masoud, et al. "The Simcyp® population-based ADME simulator." Expert opinion on drug metabolism & toxicology 5.2 (2009): 211-223.

   Wang, Ying-Hong. "Confidence assessment of the Simcyp time-based approach and a static mathematical model in predicting clinical drug-drug interactions for mechanism-based CYP3A inhibitors." Drug Metabolism and Disposition 38.7 (2010): 1094-1104. [↑](#footnote-ref-2)
3. Coecke, Sandra, et al. "Toxicokinetics as a key to the integrated toxicity risk assessment based primarily on non-animal approaches." Toxicology in vitro 27.5 (2013): 1570-1577.

   Wetmore, Barbara A. "Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment." Toxicology 332 (2015): 94-101. [↑](#footnote-ref-3)
4. Pearce, Robert G., et al. "Httk: R package for high-throughput toxicokinetics." Journal of statistical software 79.4 (2017): 1. [↑](#footnote-ref-4)
5. Shibata, Yoshihiro, et al. "Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method." *Drug Metabolism and disposition* 30.8 (2002): 892-896. [↑](#footnote-ref-5)
6. Waters, Nigel J., et al. "Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding." *Journal of pharmaceutical sciences* 97.10 (2008): 4586-4595.

   Barré, Jérrôme, et al. "Equilibrium dialysis, ultrafiltration, and ultracentrifugation compared for determining the plasma-protein-binding characteristics of valproic acid." *Clinical chemistry* 31.1 (1985): 60-64. [↑](#footnote-ref-6)
7. Wambaugh, John F., et al. "Assessing Toxicokinetic Uncertainty and Variability in Risk Prioritization." *Toxicological Sciences* 172.2 (2019): 235-251. [↑](#footnote-ref-7)
8. Wambaugh, John F., et al. "Assessing Toxicokinetic Uncertainty and Variability in Risk Prioritization." Toxicological Sciences 172.2 (2019): 235-251. [↑](#footnote-ref-8)