Data: All three valuation methods used the entire universe of home sales in Philadelphia from 1980 through 2011Q1. The dataset was created by combining deeded title transfers from the Department of Records with tax roll data from the Office of Property Assessment by merging on the common variable, the unique parcel ID. Data were cleaned to remove sales between family members, transactions where the buyer or seller was not a private individual (e.g. bank sales, sheriff sales), and any sales where the price did not reflect a true "market" value.

Value1: Value1 was derived via a regression model that decomposed the total sales price of a house into the prices of its individual components. The model was computed by regressing the sales price of each dwelling on the characteristics of the structure (e.g. square footage, number of stories, physical condition), the characteristics of the location (e.g. distance from Center City, proximity to parks or transit), the tax treatment of the dwelling (e.g. if abated or not), and what year the dwelling transacted in. The regression contains over 50 different variables in an attempt to capture as much variation in housing characteristics as possible. Although this is the most sophisticated of the three different valuation approaches, it is also the most resource-intensive, and hence prone to error if the resources are insufficient. Unfortunately, this is a real issue in Philadelphia, where the problems associated with the recorded characteristics of individual properties by the former BRT are both well-known and well-documented. If the regression model is given an erroneous value, such as an incorrect lot size or number of stories, then it will return an erroneous valuation. Since it is not known to the user which particular properties may have incorrectly recorded characteristics, we derived two other methods of valuation that are less sophisticated, but are also less susceptible to erroneous data.

Value2: Value2 was computed by indexing each subject property to the median price in its Census Tract. Using only home sales in the 2010-2011 period, the price per square foot of each home was computed by dividing the square footage of the dwelling into the recorded sales price. The median price/sqft was then computed for each Census Tract. Value2 was then computed by multiplying this median price/sqft times the square footage of each subject property. For example, if the median price/sqft in a particular tract is $85/ft, and the size of a subject property is 1,200 square feet, then its predicted value is computed as $85x1,200=$102,000. This approach tends to produce a smaller range of valuations than Value1 because the variation of house prices/sqft within an area as small as a Census Tract tends to also be quite small. While the benefit of this approach is that it is not as susceptible to data-driven errors as the approach in Value1, its flaw is that it implicitly assumes that the physical styles and conditions of homes in each neighborhood (Tract) are basically the same, and that homes within a given neighborhood differ only in size.

Value3: Value3 was estimated by marking each home's value to movements in the overall housing market since its last transaction. Using the regression in Value1, a house price index was derived that tracked the overall changes in Philadelphia house values, by neighborhood, from 1980 through 2011Q1. The percent change in the index over time represents the average percent change in house values in that neighborhood during that same period. This percent change was then applied to the last recorded transaction price of the home to update its historic value to a current value. For example, suppose a home in South Philadelphia last sold for $82,000 in 1990, and that the house price index for South Philadelphia had grown by a 120% since 1990. Then the current value of the house would be computed

as $82,000x(1+1.2)=$180,400. The benefit of this approach is that it makes use of the last, actual known market value of the dwelling, while the flaw is that it assumes that the physical condition of the property has not changed since that last transaction. For example, if a homeowner has made significant improvements in a dwelling since its last transaction, but house prices in that neighborhood have generally fallen during that same period, then this approach will generate a predicted value that is lower than the previous transaction price, when in fact the value of the dwelling may have gone up, due to the improvements.

In the event that a property has a missing value for one of the three valuation approaches, it is for the following reasons:

- **Value1:** Since this approach used a regression model that takes into account all of the dwelling's characteristics, then it requires that all of the dwelling's characteristics in the data to be populated with correct values. If even one field is not populated or has a non-credible value, then the model cannot produce a valuation.
- **Value2:** Since this approach computes each dwelling's value by looking at the prices of recent sales in the subject dwelling's Census Tract, then it requires for there to be a reasonable number of recent, arms-length sales in the tract. If there were no recent sales, or the sales were not arms-length (e.g. blanket, nominal or sheriff) then it is not possible to produce a valuation.
- **Value3:** Since this approach generates a value by updating a dwelling's past sales price to today, based upon average house price movements since the last sale, then a valid past sales price that occurs within the span of our sales data (1980-2011) is required. If the property last transacted prior to 1980, and/or its last sales price was not an arms-length market value (e.g. it was an inter-family transfer or bank sale), then this approach cannot produce a valuations.

Arriving at a single value: It is necessary for the purpose of deriving the fiscal implications of unpaid taxes to arrive at a single number for each property's value, in order to apply the city's taxation formula to this single value. While we computed all three values (where possible) for each tax-delinquent property, we generally found that two of these three values would be quite close to each other, with one of the values being an outlier. In these cases, we averaged the two values that were close and ignored the third. For example, if Value1 and Value2 were similar, with Value3 being an outlier, this would typically represent a property that had accurate characteristic data, but whose change in value over time was different than house price movements in its neighborhood. Such an example would be a property that had been renovated substantially since its purchase (and hence had an abatement), and so had risen in value by more than average house values in its neighborhood. As another example, if Values2 and Values3 were close to each other, with Value1 being an outlier, this would typically represent a property with some erroneous characteristics data, but one whose characteristics were still similar to those of other homes in its neighborhood. Such an example would be a dwelling that had an erroneous number of stories recorded for it, but whose condition had not changed much since its last transaction, so its change in value was very similar to other house price changes in its neighborhood due to its similarity to other houses in its neighborhood.