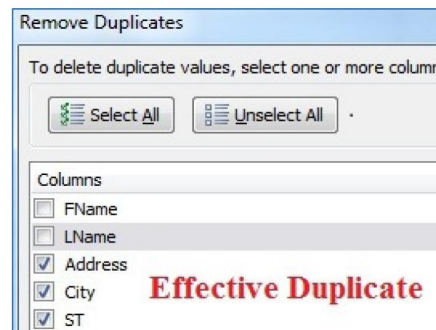


# Remove Duplicates: Payoffs and pitfalls

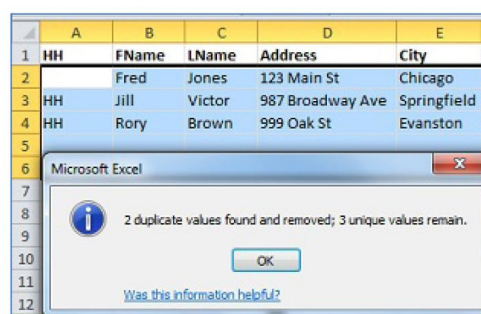
Say you have a list that includes several individuals at one household, but you only want a list of unique households. Excel has a cool tool that can help: a one-button solution to remove duplicates. While it's easier than it has ever been in Excel 2007 and 2010 to remove duplicates from a list based on selected criteria, you'll need to watch out for a few things. Here's how it works and how to avoid potential problems.

## How it works

This tool works best in tables and properly built lists. (One row, one record, no completely blank rows or columns, first row contains column headings.)



1. Select one cell in the middle of your data.
2. On the Data tab, Data Tools group, click the Remove Duplicates button.
3. When the dialog box appears, you may leave all fields selected (true duplicates) or deselect any fields that you do not need (effective duplicates). For example, we mentioned households as opposed to individual names. On this dialog box you would deselect name fields, leaving just the address fields.



## What it does

What records does Excel choose to remove when it finds duplicates? It removes all occurrences of a duplicate record, as you have defined it, after the first one. When you allow all fields to remain selected when removing duplicates, you remove only true duplicates. In other words, it does not matter which ones you keep or delete. If you are trying to remove effective duplicates (some fields not used to determine duplicates), then you'll want to make sure you order your list so the one you want to keep occurs first.

## Strategies to avoid problems

In our example, you could add a column called HH for Head of Household. Next, mark which names represent the head of each household. Now, sort by all the address fields and then by HH. This puts all the HH records first for each address. Finally, when you remove duplicates, the HH records will be kept and the other household members' records will be removed.

To determine which records were deleted, try this strategy.

1. Select both work sheets. Insert a new first column and number sequentially. (Use the AutoFill options box and choose Fill Series.)
2. Create a copy of the work sheet with your data.
3. Tack on the word "source" to the end of the original work sheet tab and "copy" to the other.
4. Apply Remove Duplicates to the Copy work sheet.
5. On the Source work sheet, entering a VLookUp formula in that will show NA for any record deleted on the Copy work sheet.

|   | A   | B | C         | D      | E                |
|---|---|---|-----------|--------|------------------|
| 1 |   |   | FName     | LName  | Address          |
| 2 | =VLOOKUP(B2,'Good List-Copy'!\$A\$2:\$H\$4,1,FALSE) |   |           |        |                  |
| 3 |   | 2 | 2 Miriam  | Brown  | 999 Oak St       |
| 4 | #N/A  |   | 3 Rory    | Brown  | 999 Oak St       |
| 5 |   | 4 | 4 Stephen | Victor | 987 Broadway Ave |
| 6 | #N/A  |   | 5 Jill    | Victor | 987 Broadway Ave |