Representation of Unsold Properties – The Biggest Problem You Don't Know You Have!

By Kevin Keene, Keene Mass Appraisal Consulting kevin@keenemac.com and
Alex Raju, Modeling Supervisor, Office of Property Assessment, City of Philadelphia alex.raju@phila.gov

As mass appraisers, we attempt to estimate something unknown (the values of a universe of properties) based on the prices and attributes of properties that have sold. We use different methods, such as cost, comparable sales, sales and income in regression models and even AI models to estimate values. The basis of all of our techniques is the assumption that properties that have sold represent the properties that have not sold. Representation of unsold properties is one of the most common, yet least understood, problems that all assessors face.

A well-represented sales file is the foundation for reliable and effective mass appraisal. It enhances data quality, feature engineering and validation, ultimately leading to more accurate property value estimates.

Whenever we use ratio studies to analyze assessment performance, or use sales-based methods to estimate property value, we are assuming:
That properties that sell are similar to properties that do not sell
- AND -
That representation of unsold properties is proportionate to sales activity.

Is it safe or reasonable to make these assumptions?

How can we test the validity of these assumptions?

There are no standards, and very little in the body of knowledge that does more than hint at the magnitude of the issue. There are statistical tests to determine the degree of variation in a sales file, but no widely accepted methods for *directly* comparing the observations in the sales file to the wider universe of unsold properties.

In an efficient market, most or all types of properties in the inventory will be represented by sales, but markets are not equally efficient. In my experience, there is a great deal of variation in the degree of representation of unsold properties from one market to the next. Many properties in disadvantaged communities or properties that are not "typical" are not represented by sales, which can easily lead to errors in valuation and/or analysis. When properties are not directly represented by sales, our valuation processes have to generalize to estimate values, and the likelihood of error increases as the degree of generalization increases. In addition, our sales ratio studies tell us little or nothing about those properties that are not represented by sales. Can we truly draw valid conclusions about assessment performance and quality across the entire inventory if some or many properties are not represented?

It is important for all assessors to understand how well sales represent their respective inventories. Better yet, techniques that can *precisely* identify properties that are not represented by sales can be very helpful in improving assessment performance and equity. Properties that are not represented by sales present higher risks for overvaluation or undervaluation.

Some of the dimensions in which properties might not be represented by sales include neighborhoods, property types, condition of improvements, building or lot sizes, construction quality, age and even price class or value class.

Group Summary Method

The Group Summary Method assigns properties to groups using a common schema called a Group Identifier (Group ID). Every transaction in the sales file and every property in the inventory file will be assigned a Group ID.

Using these groups, we can summarize and present data about each group, and directly compare sales to the wider inventory. This is not a new technique. I built the first Group IDs in Philadelphia in the late 1990s. They have been used to great effect ever since.

Heuristic Measure of Representation (HMR)
The Heuristic Measure of Representation is similar to the Group Summary Method in that it assigns properties to groups. The primary difference is that the groups are defined by regression models, with a different schema for each model that is used to estimate values. The method derives a score for each property that reflects the degree of generalization from the model used to estimate value. This method is relatively new, and has yet to be comprehensively applied.

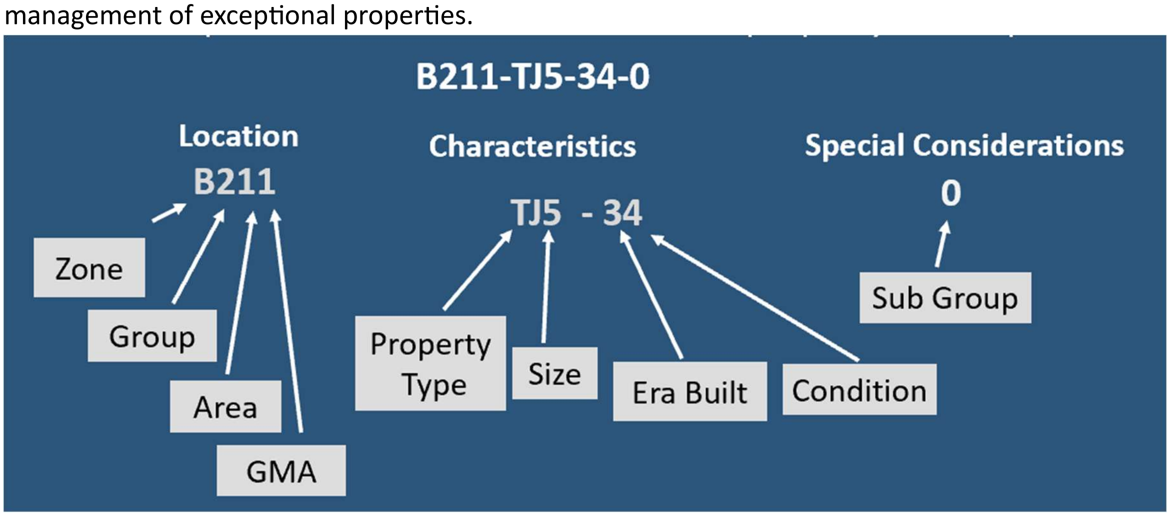
Both methods allow us to gain more precise insight into model and assessment performance, facilitate review of valuation projections and can identify submarkets or even specific properties that are not well served by the valuation process.

First, let's look at the Group Summary Method.

Why Use Grouping IDs? It is not particularly difficult to make good decisions in appraising properties. The hard part of mass appraisal is making sure that, when decisions are applied, they affect
Every property that *should* be affected - and
No properties that should *not* be affected.

Building Group IDs
Group IDs are built by identifying the five or six most important contributors to value. These are usually the attributes that determine comparability. These attributes are transformed into codes that are concatenated into a text string. Most categorical attributes are already in the CAMA system as codes, so transformation might not be necessary. Numeric attributes, such as building square footage or lot size, need to be transformed into categorical bins to which codes can be assigned. The important attributes may vary from market to market, and certainly from one property type to the next. Group IDs provide a 'snapshot' of a property, bringing together the most important elements that describe a property in one place.

As an example, in Philadelphia, we built Group IDs from Location (neighborhood); Building Design; The Relative Building Size code (building SF transformed into 5 categorical bins, ranging from smallest to largest); the year the property was built (transformed into seven categorical time periods); and the condition of the improvements. A sixth element allowed for recognition of any special circumstance that would make the property different in some way, which simplifies the management of exceptional properties.



Properties in the group B211-TJ5-34-0 would be similar, but not necessarily identical. Some might have garages; others not. Some might have central air conditioning. They could be on different size lots. But they would all be from the same neighborhood, have the same design, fall within a range of square footage that would allow them to be considered similar, have been built at around the same time period and would be in the same condition.

One of the advantages of the method is its great flexibility. I have built Group IDs for jurisdictions that used other attributes, such as Quality of Construction, the Number of Stories, or the class of the building or complex. Each market can define Group IDs in its own way. Group Ids can be built for all types of properties – from vacant land to Office Buildings to Condos.

There are many advantages to using Group Ids and Group summaries.

Group Ids allow us to designate properties as members of groups and *make decisions* at the group level. This ensures that all properties in the group are affected equally. It also allows us to have different methods, adjustment coefficients and techniques for different groups of properties. We can also keep aggregate or summarize data for all groups and easily publish that data to our constituents.

- Group Ids make databases much more efficient, avoiding multi-key joins between tables and simplifying retrieval of data through queries.
- Group IDs support direct comparison between sold and unsold properties.
- Group summaries can greatly improve our understanding of our markets and the performance of our assessments.
- Group IDs are very useful for reviewing market value estimates both within and between groups.
- Group IDs can greatly simplify identification and selection of comparable sales.
- Group IDs can be built for any sales and inventory file regardless of the valuation method used. Anyone can use this technique!

The Power of Persistent and Consistent Groups

Groups created by Group IDs are both consistent and persistent.

Consistent means that the group is always the same, no matter who accesses the data.
Persistent means that the group, and data about the group, exists at all times.

Every account has a Group Id. There are around 59,000 distinct Group Ids in the Philadelphia data.

Every transaction has a SGroup Id (Group ID at time of sale). There are around 23,000 groups with 1 or more validated sales in the Philadelphia data.

Because the groups are both Consistent and Persistent, we can maintain data about Group IDs and SGroup IDs.
Every account can be linked to the data about its Group ID and SGroup ID groups, including the number of accounts, number of valid sales, average size, market value or sale price per square foot, average sale price, median ratio of assessment, and many others.

We can compare any given account to what is typical for the group, allowing us to find those that are at significant variance. How well does the sale price of a new transaction match what we know about what is typical for the group? How well does a specific value align with other properties in the group? These questions become easy to answer.

When we run a query, the result is a set of records that meet the parameters that were input. We can examine or analyze the records that were pulled, but we can't analyze the records that were not pulled. With Group ID summaries, we can also compare attributes of a set of records to attributes of records that are NOT in the dataset. This is called 'what is' to 'what is not' comparison.

You can't do this if you don't group, or if you group 'on the fly'!

Summary Tables

These tables store information about persistent groups. Every account can be linked to these tables by Group ID, so that any account can be compared to the summary data and all accounts in a group can be identified

GroupId	Accounts	TotalProjMV	AvgProjMV	MinProjMV	MaxProjMV	ProjMVRange	AvgProjMVSF	WtdAvgProjN	MinProjMVSF	MaxProjMVS	ProjMVSFRange	TotalBldSF	AvgBldSF	MinBldSF	MaxBldSF	BldSFRange	AvgLotSF
A111LRSA540	2	$94,000	$47,000	$41,500	$52,500	$11,000	$23.73		$23.73	$27.93	$4.20						2002
A111M2B0514	2	$411,100	$205,550	$200,400	$210,700	$10,300	$169.97	$169.81	$169.97	$173.51	$3.54	2421	1210	1155	1266	111	2382
A111M2B0523	2	$707,100	$353,550	$338,800	$368,300	$29,500	$220.97	$220.97	$220.97	$230.19	$9.22	3200	1600	1600	1600	0	4026
A111M2B0524	27	$6,691,700	$247,841	$231,700	$273,700	$42,000	$156.63	$156.57	$156.63	$179.44	$22.82	42738	1583	1440	1600	160	2581
A111M2B0533	1	$348,500	$348,500	$348,500	$348,500	$0	$176.19	$176.19	$176.19	$176.19	$0.00	1978	1978	1978	1978	0	2597
A111M2B0534	125	$31,755,300	$254,042	$230,000	$280,100	$50,100	$142.49	$142.21	$142.49	$161.07	$18.58	223296	1786	1680	2026	346	2210
A111M2B0535	1	$149,300	$149,300	$149,300	$149,300	$0	$88.87	$88.87	$88.87	$88.87	$0.00	1680	1680	1680	1680	0	1800
A111M2C0514	2	$340,800	$170,400	$166,400	$174,400	$8,000	$152.14	$152.14	$152.14	$155.71	$3.57	2240	1120	1120	1120	0	2278
A111M2C0515	1	$111,000	$111,000	$111,000	$111,000	$0	$85.65	$85.65	$85.65	$85.65	$0.00	1296	1296	1296	1296	0	1318
A111M2C0534	3	$759,300	$253,100	$226,300	$273,000	$46,700	$143.08	$143.10	$143.08	$151.67	$8.59	5306	1769	1618	1888	270	3859
A111M2C0547	1	$104,800	$104,800	$104,800	$104,800	$0	$44.11	$44.11	$44.11	$44.11	$0.00	2376	2376	2376	2376	0	2950
A111M3B0000	1						$0.00		$0.00	$0.00	$0.00						
A111M3B0533	1	$393,700	$393,700	$393,700	$393,700	$0	$234.35	$234.35	$234.35	$234.35	$0.00	1680	1680	1680	1680	0	3050
A111M3B0534	2	$636,000	$318,000	$317,300	$318,700	$1,400	$173.96	$173.96	$173.96	$174.34	$0.38	3656	1828	1828	1828	0	2855
A111M3C0524	4	$1,012,500	$253,125	$248,600	$259,100	$10,500	$159.40	$159.40	$159.40	$163.57	$4.17	6352	1588	1584	1600	16	2398
A111M3C0534	12	$3,384,700	$282,058	$260,100	$297,500	$37,400	$156.40	$156.14	$156.40	$168.71	$12.30	21678	1806	1732	2000	268	2561
A111M4B0554	7	$3,361,600	$480,229	$464,100	$525,300	$61,200	$123.86	$123.30	$123.86	$127.43	$3.57	27264	3895	3642	4572	930	6022
A111RPF2532	2	$535,600	$267,800	$266,800	$268,800	$2,000	$239.11	$239.11	$239.11	$240.00	$0.89	2240	1120	1120	1120	0	1512
A111RPF2533	4	$887,100	$221,775	$218,100	$226,300	$8,200	$198.01	$198.01	$198.01	$202.05	$4.04	4480	1120	1120	1120	0	1475
A111RPF2534	44	$6,835,000	$155,341	$151,900	$158,000	$6,100	$138.70	$138.70	$138.70	$141.07	$2.37	49280	1120	1120	1120	0	1480
A111RPW2434	1	$152,300	$152,300	$152,300	$152,300	$0	$125.25	$125.25	$125.25	$125.25	$0.00	1216	1216	1216	1216	0	1474
A111RPW2531	1	$225,400	$225,400	$225,400	$225,400	$0	$201.25	$201.25	$201.25	$201.25	$0.00	1120	1120	1120	1120	0	1755
A111RPW2532	60	$18,097,500	$301,625	$275,500	$353,700	$78,200	$261.07	$260.61	$261.07	$315.80	$54.73	69444	1157	1120	1296	176	1791
A111RPW2533	[illegible]	$41,441,400	$243,773	$212,800	$306,500	$93,700	$213.50	$213.23	$213.50	$252.59	$39.09	[illegible]	[illegible]	1120	1296	176	1776
A111RPW2534	2029	$344,459,200	$169,768	$124,700	$295,400	$170,700	$148.50	$148.30	$148.50	$263.75	$115.25	2322678	1145	1120	1296	176	1765
A111RPW2535	[illegible]	[illegible]	$99,363	$95,700	$105,500	$9,800	$86.80	$86.74	$86.80	$91.25	$4.45	[illegible]	[illegible]	1120	1260	140	1517
A111RPW2536	2	$147,200	$73,600	$72,200	$75,000	$2,800	$64.82	$64.79	$64.82	$66.96	$2.15	2272	1136	1120	1152	32	1584

This table summarizes property data by Group ID. Group A111RPW2534 contains 2029 properties.

SGroupId	ValidSales	TotalTASP	AvgTASP	MinTASP	MaxTASP	TASPRange	AvgTASPSF	WtdAvgTASP	MinTASPSF	MaxTASPSF	TASPSFRange	TotalProjMV	AvgProjMV	MinProjMV	MaxProjMV	ProjMVRange	AvgProjMVSF	WtdAvgProjN
A111M2B0514	1	$213,172	$213,172	$213,172	$213,172	$0	$190.33	$0.00	$190.33	$190.33	$0.00	$180,989	$180,989	$180,989	$180,989	$0	$190.33	$0.00
A111M2B0524	4	$1,116,508	$279,127	$202,206	$310,237	$0	$179.84	$0.00	$126.38	$215.44	$0.00	$1,057,391	$264,348	$258,165	$273,393	$0	$179.84	$0.00
A111M2B0534	10	$2,790,917	$279,092	$219,571	$344,030	$0	$159.86	$0.00	$126.77	$199.79	$0.00	$2,568,317	$256,832	$248,002	$269,052	$0	$159.86	$0.00
A111M2B0535	1	$102,232	$102,232	$102,232	$102,232	$0	$60.85	$0.00	$60.85	$60.85	$0.00	$149,151	$149,151	$149,151	$149,151	$0	$60.85	$0.00
A111M2C0534	1	$133,698	$133,698	$133,698	$133,698	$0	$82.63	$0.00	$82.63	$82.63	$0.00	$226,235	$226,235	$226,235	$226,235	$0	$82.63	$0.00
A111M3B0533	1	$500,458	$500,458	$500,458	$500,458	$0	$297.89	$0.00	$297.89	$297.89	$0.00	$393,368	$393,368	$393,368	$393,368	$0	$297.89	$0.00
A111M3B0534	1	$278,912	$278,912	$278,912	$278,912	$0	$152.58	$0.00	$152.58	$152.58	$0.00	$317,042	$317,042	$317,042	$317,042	$0	$152.58	$0.00
A111M3C0534	2	$613,452	$306,726	$248,251	$365,202	$0	$169.86	$0.00	$139.94	$199.78	$0.00	$557,068	$278,534	$262,715	$294,353	$0	$169.86	$0.00
A111RPF2532	2	$479,327	$239,663	$239,155	$240,172	$1,017	$213.99	$213.99	$213.53	$214.44	$0.91	$535,646	$267,823	$266,819	$268,828	$2,009	$239.13	$239.13
A111RPF2533	3	$642,845	$214,282	$198,520	$229,679	$31,158	$191.32	$191.32	$177.25	$205.07	$27.82	$670,908	$223,636	$218,108	$226,464	$8,356	$199.67	$199.67
A111RPF2534	5	$924,394	$184,879	$153,157	$208,312	$55,155	$165.07	$165.07	$136.75	$185.99	$49.25	$772,205	$154,441	$152,130	$157,823	$5,693	$137.89	$137.89
A111RPF2535	2	$275,392	$137,696	$125,894	$149,499	$23,605	$122.94	$122.94	$112.40	$133.48	$21.08	$182,724	$91,362	$91,349	$91,375	$26	$81.57	$81.57
A111RPW2532	58	$14,863,195	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$301,978	$276,855	$353,722	$76,867	$260.17	$259.58
[illegible]	75	$16,992,123	$226,562	$182,838	$261,961	$79,123	$198.38	$198.11	$163.25	$233.89	$70.65	$18,340,705	$244,543	$212,880	[illegible]	$65,061	$214.17	$213.83
A111RPW2534	118	$21,920,940	$185,771	$108,899	$229,297	$120,398	$163.70	$163.62	$97.23	$204.73	$107.50	$20,108,748	$170,413	$151,358	$201,852	[illegible]	$150.21	$150.10
A111RPW2535	[illegible]	[illegible]	[illegible]	$99,899	$154,268	$54,369	$107.62	$107.58	$89.20	$137.74	$48.54	[illegible]	[illegible]	[illegible]	$114,659	$22,610	$87.96	$87.81
A111RPW2536	1	$92,361	$92,361	$92,361	$92,361	$0	$82.47	$82.47	$82.47	$82.47	$0.00	$75,480	$75,480	$75,480	$75,480	$0	$67.39	$67.39

This table summarizes sales data by Group ID. Group A111RPW2534 is represented by 118 sales.

GroupId	Accounts	ValidSales	TotalProjMV	AvgProjMV	MinProjMV	MaxProjMV	ProjMVRange	AvgProjMVSF	WtdAvgProjN	MinProjMVSF	MaxProjMVS	ProjMVSFRar	TotalBldSF	AvgBldSF	MinBldSF	MaxBldSF	BldSFRange
A111RPF2533	4	3	$887,100.00	$221,775.00	$218,100.00	$226,300.00	$8,200.00	$198.01	$198.01	$198.01	$202.05	$4.04	4480	1120	1120	1120	0
A111RPF2534	44	5	$6,835,000.00	$155,340.91	$151,900.00	$158,000.00	[illegible]	$138.70	$138.70	$138.70	$141.07	$2.37	49280	1120	1120	1120	0
A111RPW2434	1		$152,300.00	$152,300.00	$152,300.00	$152,300.00	$0.00	$125.25	$125.25	$125.25	$125.25	$0.00	1216	1216	1216	1216	0
A111RPW2531	1		$225,400.00	$225,400.00	$225,400.00	$225,400.00	$0.00	$201.25	$201.25	$201.25	$201.25	$0.00	1120	1120	1120	1120	0
A111RPW2532	60	58	$18,097,500.00	$301,625.00	$275,500.00	$353,700.00	$78,200.00	$261.07	$260.61	$261.07	$315.80	$54.73	69444	1157	1120	1296	176
A111RPW2533	170	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	194350	[illegible]	1120	1296	176
A111RPW2534	2029	118	$344,459,200.00	$169,767.96	$124,700.00	$295,400.00	$170,700.00	$148.50	$148.30	$148.50	$263.75	$115.25	2322678	[illegible]	1120	1296	176
A111RPW2535	[illegible]	[illegible]	[illegible]	$99,362.50	$95,700.00	$105,500.00	$9,800.00	$86.80	$86.74	$86.80	[illegible]	[illegible]	9164	[illegible]	1120	1260	140
A111RPW2536	2	1	$147,200.00	$73,600.00	$72,200.00	$75,000.00	$2,800.00	$64.82	$64.79	$64.82	$66.96	$2.15	2272	1136	1120	1152	32
A111RPW2542	1	1	$318,600.00	$318,600.00	$318,600.00	$318,600.00	$0.00	$220.33	$220.33	$220.33	$220.33	$0.00	1446	1446	1446	1446	0
A111RPW2543	16	7	$4,202,300.00	$262,643.75	$239,600.00	$328,500.00	$88,900.00	$187.65	$187.70	$187.65	$218.42	$30.76	22388	1399	1332	1504	172
A111RPW2544	177	11	$32,449,700.00	$183,331.64	$166,900.00	$255,800.00	$88,900.00	$133.38	$133.22	$133.38	$164.18	$30.80	243575	1376	1332	1596	264
A111RPW2552	1	1	$357,700.00	$357,700.00	$357,700.00	$357,700.00	$0.00	$223.56	$223.56	$223.56	$223.56	$0.00	1600	1600	1600	1600	0
A111RPW2553	1		$293,500.00	$293,500.00	$293,500.00	$293,500.00	$0.00	$175.54	$175.54	$175.54	$175.54	$0.00	1672	1672	1672	1672	0
A111RPW2554	3		$611,300.00	$203,766.67	$186,600.00	$216,700.00	$30,100.00	$114.90	$113.62	$114.90	$133.93	$19.03	5380	1793	1618	2070	452
A111RPW3533	1	1	$252,800.00	$252,800.00	$252,800.00	$252,800.00	$0.00	$199.68	$199.68	$199.68	$199.68	$0.00	1266	1266	1266	1266	0
A111RPW3534	4		$695,100.00	$173,775.00	$168,500.00	$176,400.00	$7,900.00	$137.26	$137.26	$137.26	$139.34	$2.07	5064	1266	1266	1266	0
A123M2C0315	1		$105,100.00	$105,100.00	$105,100.00	$105,100.00	$0.00	$91.23	$91.23	$91.23	$91.23	$0.00	1152	1152	1152	1152	0

This table combines sales and property data summaries into one table.

GroupId	Parcel_ID	ParldNum	OPA_ACCOUI	PROPID	SEC_FLD	BLOCK_ID	ADDRESS	CENSUS_TRA	CENSUS_BLO	Zone	NBHD
A111RPW2534	1001483344	1001483344	344166100	7262007312	3257	7262007300	7312 SHERWOOD RD	098	104	A	A111
A111RPW2534	1001675471	1001675471	343278700	8947001334	3257	8947001300	1334 N 75TH ST	098	207	A	A111
A111RPW2534	1001675468	1001675468	343278400	8947001328	3257	8947001300	1328 N 75TH ST	098	207	A	A111
A111RPW2534	1001675472	1001675472	343278800	8947001336	3257	8947001300	1336 N 75TH ST	098	207	A	A111
A111RPW2534	1001675473	1001675473	343278900	8947001338	3257	8947001300	1338 N 75TH ST	098	207	A	A111
A111RPW2534	1001675474	1001675474	343279000	8947001340	3257	8947001300	1340 N 75TH ST	098	207	A	A111
A111RPW2534	1001675475	1001675475	343279100	8947001342	3257	8947001300	1342 N 75TH ST	098	207	A	A111
A111RPW2534	1001675476	1001675476	343279200	8947001344	3257	8947001300	1344 N 75TH ST	098	207	A	A111
A111RPW2534	1001675477	1001675477	343279300	8947001346	3257	8947001300	1346 N 75TH ST	098	207	A	A111
A111RPW2534	1001675478	1001675478	343279400	8947001348	3257	8947001300	1348 N 75TH ST	098	207	A	A111
A111RPW2534	1001106925	1001106925	343213100	1926007422	3257	1926007400	7422 BROOKHAVEN RD	098	604	A	A111
A111RPW2534	1001106921	1001106921	343212800	1926007416	3257	1926007400	7416 BROOKHAVEN RD	098	604	A	A111
A111RPW2534	1001099866	1001099866	343221700	1874007530	3257	1874007500	7530 BRENTWOOD RD	098	606	A	A111
A111RPW2534	1001099868	1001099868	343221800	1874007532	3257	1874007500	7532 BRENTWOOD RD	098	606	A	A111
A111RPW2534	1001675811	1001675811	343304500	8950001429	3257	8950001400	1429 N 76TH ST	098	305	A	A111
A111RPW2534	1001675679	1001675679	343304800	8950001300	3257	8950001300	1300 N 76TH ST	098	301	A	A111
A111RPW2534	1001675681	1001675681	343304900	8950001302	3257	8950001300	1302 N 76TH ST	098	301	A	A111
A111RPW2534	1001675683	1001675683	343305000	8950001304	3257	8950001300	1304 N 76TH ST	098	301	A	A111
A111RPW2534	1001675685	1001675685	343305100	8950001306	3257	8950001300	1306 N 76TH ST	098	301	A	A111

These are just some of the over 2,000 properties in Group A111RPW2534. They can be directly linked and compared to the summary tables.

Once unrepresented properties are identified, we can use a variety of cluster or tabular analytics to better understand the relationships between properties that sell and those that do not.

Using Summaries to Understand Representation

Properties will not be proportionately represented by sales, as sales will represent different numbers of accounts. Some groups of properties may be represented by few or no sales, as illustrated in this sample from a Group summary report. You can directly compare time adjusted prices to values, examine variance in both price and value within the group and even see how well the properties that sold compare to the unsold properties in terms of building and lot size.

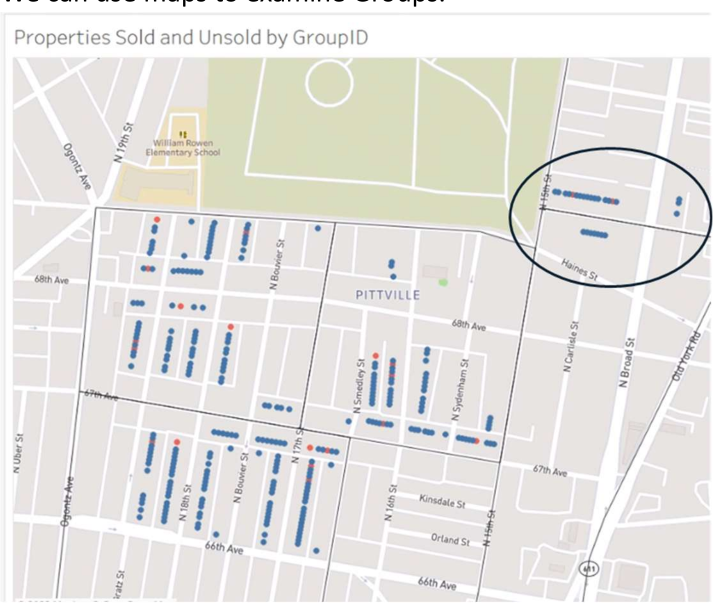
This report supports direct comparison of sold and unsold properties both *within* and *between* groups. There are no sales for the 30 properties in group M111RPW2454, but they are larger than the properties in group M111RPW2444. We would expect slightly higher values, but possibly lower MV per square foot rates in M111RPW2454 when compared to M111RPW244 – which is exactly the case. Comparing group M111RPW2445 (fair condition) to group M111RPW2444 (average condition), we can see if the adjustment coefficient from the model is producing the correct effect.

GroupId	Accounts Valid Sales	Agv MV Agv TASP	Min MV MinTASP	Max MV MaxTASP	MVRange TASPRange	WtdAvgMVSF WtdAvgTASPSF	Min / Max MVSF Min / Max TASPSF		*Range*	AvgBldSF AvgSBldSF	AvgLotSF AvgSLotSF
M111RPF2344	301	$165,418	$155,200	$185,900	$30,700	$117.22	$117.41	$130.34	$12.93	1,411	1,522
	17	$170,057	$139,064	$191,300	$52,236	$116.82	$96.36	$131.19	$34.83	1,456	1,529
M612TOS3353	43	$531,742	$440,300	$816,700	$376,400	$192.35	$196.39	$242.49	$46.10	2,764	4,072
	22	$500,055	$366,316	$658,954	$292,638	$192.46	$123.13	$272.45	$149.33	2,598	4,169
M621TOS3354	49	$473,682	$432,500	$603,400	$170,900	$200.34	$200.74	$216.00	$15.26	2,364	3,303
	2	$463,211	$453,694	$472,728	$19,034	$204.42	$202.18	$206.61	$4.43	2,266	3,225
M111RPW2444	16	$179,244	$178,400	$179,300	$900	$119.18	$119.18	$119.22	$0.04	1,504	2,400
	2	$176,882	$174,114	$179,650	$5,536	$117.61	$115.77	$119.45	$3.68	1,504	2,400
M111RPW2445	1	$111,100	$111,100	$111,100	$0	$73.87	$73.87	$73.87	$0.00	1,504	2,400
	1	$109,275	$109,275	$109,275	$0	$72.66	$72.66	$72.66	$0.00	1,504	2,400
M111RPW2454	30	$187,770	$181,800	$204,700	$22,900	$112.57	$112.68	$120.30	$7.61	1,668	2,707

This report presents inventory data on the top row and sales data for the same group on the bottom row.

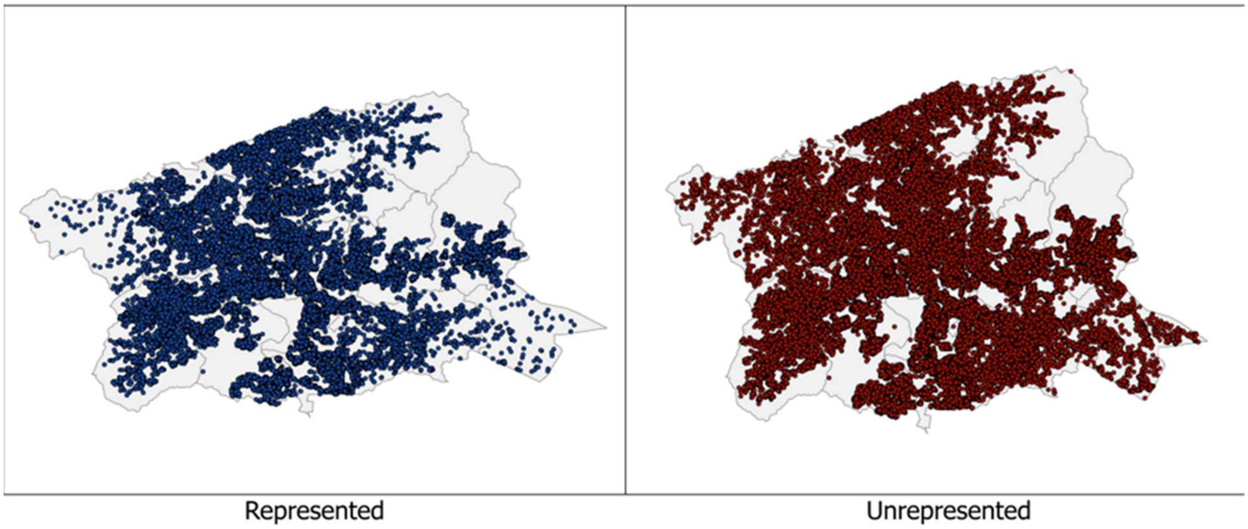
Mapping Representation

We can use maps to examine Groups.



All of these properties are in the same Group. Blue dots are unsold properties, while sales are represented by red dots. The circled cluster has no sales, but all of these properties are comparable.

Buncombe County Represented vs Unrepresented Properties



This map shows the locations of all properties that are in groups that are represented by at least one sale juxtaposed with properties that are in group that are not represented by any sales. In many cases, they are interspersed with each other, but we can see that there are some neighborhoods in the county where there are no sales.

Representation Summary

Here's a sample breakdown of representation by groups of similar properties. There is a lot of information in this table, but perhaps the most telling is that there are 21,000 groups representing almost 60,000 properties that are not represented by any sales – 14% of the inventory - in this dataset. I have worked with datasets where the percentage of unrepresented accounts is closer to 50%.

We can identify every property that is in any of these groups.

	Groups	Accts	Sales	Pct of Groups	Pct of Accts	Pct of Sales	Representation Pct
SF Total	35,233	422,996	54,966				13%
1 account	15,708	15,708	4,094	45%	4%	7%	26%
Lt 5 Accounts	25,245	40,836	10,905	72%	10%	20%	27%
100 or more Accts	845	201,453	16,030	2%	48%	29%	8%
500 or more Accts	65	47,401	3,530	0.2%	11%	6%	7%
No Sales	20,951	59,763	-	59%	14%	0%	0%
At least 1 sale	14,282	363,233	54,966	41%	86%	100%	15%
At least 3 sales	5,002	291,239	43,126	14%	69%	78%	15%
Less than 3 sales	30,231	131,757	11,840	86%	31%	22%	9%
10 or more sales	1,141	172,065	24,895	3%	41%	45%	14%
Condition 7	1,337	3,715	745	4%	1%	1%	20%
Condition 6	1,088	2,094	954	3%	0%	2%	46%
Condition 5	3,234	11,889	4,044	9%	3%	7%	34%
Condition 4	18,109	352,270	26,092	51%	83%	47%	7%
Condition 3	6,902	31,406	10,861	20%	7%	20%	35%
Condition 2	3,625	14,007	8,552	10%	3%	16%	61%
Condition 1	898	7,572	3,718	3%	2%	7%	49%

This table examines representation by value class. We can easily see how relatively underrepresented are the lower value classes.

ValueClass * Represented2 Crosstabulation

			Represented2 No	Represented2 Yes	Total
ValueClass	Below 120k	Count	7288	2777	10065
		% within ValueClass	72.4%	27.6%	100.0%
	120k to 175k	Count	6123	3648	9771
		% within ValueClass	62.7%	37.3%	100.0%
	175k to 215k	Count	4963	5088	10051
		% within ValueClass	49.4%	50.6%	100.0%
	215k to 250k	Count	4651	5866	10517
		% within ValueClass	44.2%	55.8%	100.0%
	250k to 282k	Count	3883	5585	9468
		% within ValueClass	41.0%	59.0%	100.0%
	282k to 322k	Count	4363	5833	10196
		% within ValueClass	42.8%	57.2%	100.0%
	322k to 374k	Count	4643	5329	9972
		% within ValueClass	46.6%	53.4%	100.0%
	374k to 460k	Count	4791	5289	10080
		% within ValueClass	47.5%	52.5%	100.0%
	460k to 640k	Count	4820	5063	9883
		% within ValueClass	48.8%	51.2%	100.0%
	640k+	Count	4504	3670	8174
		% within ValueClass	55.1%	44.9%	100.0%
Total		Count	50029	48148	98177
		% within ValueClass	51.0%	49.0%	100.0%

This table examines representation by condition of improvements. Notice the low percentages of representation in less than normal condition properties.

Condition * Represented2 Crosstabulation

		Represented2 No	Represented2 Yes	Total
Condition	Fair	2775	201	2976
		93.2%	6.8%	100.0%
	Good	4923	5871	10794
		45.6%	54.4%	100.0%
	Normal	36784	39281	76065
		48.4%	51.6%	100.0%
	Poor	1102	52	1154
		95.5%	4.5%	100.0%
	Renovated	3020	2737	5757
		52.5%	47.5%	100.0%
	Unsound	260	6	266
		97.7%	2.3%	100.0%
Total		50029	48148	98177
		51.0%	49.0%	100.0%

Group Ids are a good way to understand representation. They are easy to implement and can help you better understand your market.

Now let’s look at another way to measure and understand representation.

Heuristic Measure of Representation (HMR)

Multiple Regression Analysis (MRA) is used in many jurisdictions to develop Models for Mass Appraisal. Ratio Study metrics are used to measure performance such as Uniformity and Equity. Seldom does one check if the data used to build models are representative of the population that we are trying to model. In the world of (re)assessment, for any given year the Population (Master Roll) for any jurisdiction is fixed. The following is an effort to measure / quantify representation of your data used in your model to the population that you are trying to model - **A Heuristic Measure of Representation (HMR).**

In a multiple regression model, continuous variables always develop adjustment coefficients. What happens to binary or categorical variables that do not develop coefficients? They get treated like the base. The contributory value is generalized for that attribute. Why does this happen? The variable is under-represented or unrepresented in the data whereby it does not develop a coefficient with significance. If we can identify and quantify the degrees of generalization, we can formulate a metric to measure / quantify representation.

In the simplest form, the scoring algorithm can be:

- Identify the base binaries for each categorical variable.
- Identify the binaries that developed coefficients.
- Assign a score for each binary variable in the Population that is
    - NOT used as the base

      AND
    - DID NOT generate a coefficient.
- Tally scores for each case

The Final Score = Degree of Generalization

Each parcel to which the model is applied will get a score. Scores will range from "0" to "*n*", where n is the number of unrepresented binary or categorical attributes. A group of properties with the same characteristics will have the same score. Higher the score, higher the degree of generalization which indicates a **lower degree of representation** in the model and possibly lower degrees of accuracy in the estimates.

Scoring Example

ID	ADDR	Locale	Condition	Style	PredMV	L_Score	C_Score	S_Score	TOT_SCORE
8386000300	322 WINONA ST	M111	3	R	$ 280,000.00	0	0	0	0
5136000600	607 LOCUST AVE	M112	5	T	$ 110,000.00	1	0	1	2
2390000100	126 E CLIVEDEN ST	M113	3	S	$ 229,900.00	1	0	0	1
1746007400	7426 BEVERLY RD	M131	1	S	$ 245,000.00	0	1	0	1
3108001800	1825 W ELEANOR ST	M132	3	R	$ 140,000.00	1	0	0	1
8196001400	1412 E WEAVER ST	M133	4	R	$ 115,000.00	0	0	0	0
2296005500	5518 CHEW AVE	M141	4	T	$ 131,500.00	0	0	1	1
3666006300	6375 GERMANTOWN AVE	M143	1	R	$ 242,500.00	1	1	0	2
2330000000	65 E CLAPIER ST	M143	4	R	$ 275,000.00	1	0	0	1
1650006400	6415 N BEECHWOOD ST	M132	4	T	$ 149,900.00	1	0	1	2
2366006200	6215 CLEARVIEW ST	M133	3	R	$ 70,000.00	0	0	0	0
5136000600	617 LOCUST AVE	M131	5	T	$ 104,000.00	0	0	1	0
2532000100	151 E COULTER ST	M132	3	S	$ 272,000.00	1	0	0	1
5180001500	1518 W LOUDON ST	M133	5	S	$ 50,000.00	0	0	0	0
8817005900	5986 N 20TH ST	M141	1	R	$ 235,000.00	0	1	0	1
8817007300	7347 N 20TH ST	M141	4	R	$ 179,900.00	0	0	0	0
7308004500	4514 N SMEDLEY ST	M123	4	T	$ 129,900.00	0	0	1	1

There are three unrepresented attributes – Condition, Location and Style – in this model. Scores will range from "0" to "3". M1125T generates a score of "2". All properties in that group will have the same score, and the same degree of generalization in the model.

Additionally, scoring can be scaled or weighted based on variable importance, adapted for different machine learning model types. Furthermore, continuous variables can be binned and scored as an additional facet. Scoring can be used in both feedback & feedforward pipeline to improve models. Scores and related groups may be used to transplant baked in intelligence into (e.g.: Comparable Engine; Neighborhood Definitions etc.). Ratio studies and other performance metrics can be run on binned groups of HMR scores to get additional insights.

HMR scoring allows us to:

- Recognize parcel groups in the population that are under-represented or unrepresented in the data used to create the model.
- Respecify / recalibrate your variables to improve your models.
- Identify parcel groups that may need additional review before finalization.
- Gain more precise insight into model and assessment performance
- Identify submarkets that are not well served by your valuation process
- Better understand representation through maps and visualizations

Conclusions

The founding fathers of mass appraisal devised some great tools and methods for modern practitioners. IAAO standards and education have promoted use of these tools, established best practices and helped improve assessment performance in many jurisdictions. Understanding the degree to which sales represent unsold properties is critical in improving and validating assessment performance, and represents a significant blind spot in our body of knowledge. Any study of assessment equity should include analysis of representation. In some markets, it may be appropriate to use alternative methods of estimating values for sub-markets, groups or clusters of properties that are not represented by sales and about which sales tell us little or nothing.

Beyond the methods presented in this article, other methods and techniques should be discussed, developed and used to advance the understanding of representation in the industry. The authors encourage further engagement and discussion.