

Information paper

Quality and Methodology Information

General details

Title of output:	House Price Statistics for Small Areas
Abbreviated title:	HPSSAs
Designation:	Official Statistics not designated as National Statistics
Geographic coverage:	England and Wales
Date of last QMI:	February 2014
Contact details:	better.info@ons.gov.uk

Executive summary

This document provides information about House Price Statistics for Small Areas (HPSSAs), produced by us (ONS) since February 2015. Elements of these statistics were formerly produced by the Department for Communities and Local Government (DCLG). These statistics report the count, median price, mean price, lower quartile price and tenth percentile price of all residential dwellings sold and registered since 1995. They are calculated using open data from the Land Registry, a source of comprehensive record level administrative data on residential property transactions.

Statistics are reported for a range of subnational geographies in England and Wales. These cover:

1. English regions, England and Wales
2. Local authorities
3. Parliamentary constituencies
4. Middle layer super output areas (MSOAs)
5. Local enterprise partnerships
6. Travel to work areas
7. Rural-urban classification

Statistics are also reported for different types of dwelling, including newly built dwellings and existing dwellings, detached, semi-detached, terraced houses, flats/maisonettes and all house types combined.

The data used to produce HPSSAs is available each month when it is published by the Land Registry. The data are available from 1995 to the most recently completed month, and for the purpose of HPSSAs, the reference periods used are quarterly rolling years beginning with the year ending quarter 4 1995, and quarterly thereafter, covering the most recently completed 4 quarters.

The output contains a statistical bulletin, which supports reference tables containing all data. Interactive maps and charts are also produced where appropriate.

This document contains the following sections:

- Output quality
- About the output
- How the output is created
- Validation and quality assurance
- Concepts and definitions
- Other information, relating to quality trade-offs and user needs
- Sources of further information or advice

Output quality

This section provides a range of information that describes the quality of the data and details any points that should be noted when using the output.

We have developed [Guidelines for Measuring Statistical Quality](#), based upon the 5 European Statistical System (ESS) Quality Dimensions. This document addresses these quality dimensions and other important quality characteristics related to HPSSAs, which are:

- relevance
- timeliness and punctuality
- coherence and comparability
- accuracy
- output quality trade-offs
- assessment of user needs and perceptions
- accessibility and clarity

About the output

Relevance

(The degree to which statistical outputs meet users' needs.)

Following a [review of house price statistics by the National Statistician](#), the HPSSAs were produced specifically for medium to smaller geographic areas, such as local authorities and middle layer super output areas in response to demand from users for an indication of house prices at this small area level. The review also concluded that house price statistics for different geographies should be available on a consistent basis. The HPSSAs provide this consistent geographic structure.

The Land Registry (LR) has provided the publically available, open data used in the production of these statistics. The information covers residential dwelling transactions, together with information on the prices paid and type of dwelling (newly built and existing stock, detached, semi-detached, terraced or flat/maisonette). This is known as Price Paid data. Using these data, we calculate counts of the number of sales, median, mean, lower quartile and tenth percentile house prices and publish them in a format which helps users gain the maximum benefit from a large dataset, helping to increase potential onward use of the statistics.

The LR Price Paid data, used in the production of these statistics, are comprehensive in that they capture changes of ownership for individual residential properties which have sold for full market value and covers both cash sales and those involving a mortgage. LR Price Paid data has been in existence since April 2000, with the LR collecting data back to the year 1995. Every change of ownership that is listed by LR is considered for use in the calculation of HPSSAs; therefore the data are a reliable indicator of actual prices paid.

However, HPSSAs should not be taken as a reflection of housing market values. The definitive set of UK, country and regional house price statistics is available in our [House Price Index](#). The HPSSA and the House Price Index (HPI) follow similar trends, but the HPI is a mix adjusted index, which means it is weighted to reflect the actual mix of dwellings that exists. This HPI methodology is specifically designed to reflect market values, while the HPSSAs report average transactional values and are therefore particularly useful when considering house prices from a social change perspective rather than an economic perspective.

Geographic referencing

Geographic referencing is carried out using the National Statistics Postcode Lookup (NSPL) file, available on the [ONS Open Geography Portal](#). The version used is the most up to date file at the time of production. In accordance with National Statistics Geography Policy each record is allocated to a middle layer super output area (MSOA) and all other geographies on which we report, using its address and postcode.

How the data are collected

The data are provided by the Land Registry (LR), a government department, responsible to the Lord Chancellor for maintaining the Land Register, a record of land ownership in England and Wales. The LR was established in 1862 and now operates under the provisions of the Land

Registration Act of 1925. The purpose is to provide a guarantee of ownership as the basis for free movement of interests in land.

It is a statutory requirement for all the relevant details to be sent to the LR whenever a person transfers ownership of a property or takes a mortgage out against it. This change must be registered with the LR to make it legally effective. Over half of the applications for registration are received electronically and the rest are on paper forms. Data received via paper forms are then keyed into an electronic register. Historic data are retained within the files but are not shown on the Land Register.

[LR Price Paid Data](#) is used to produce our HPSSAs and the data is also released in conjunction with the [LR HPI](#). LR Price Paid data follow the same release schedule as the LR HPI and any changes to the data are reported by the LR.

Exclusions from the data

The LR price paid data only include records for single residential properties sold for full market value (including sales made under the Government's [Help to Buy](#) scheme). In order to avoid systematically biasing house price data, the LR excludes records of sales that were not at full market value. These are not included in the data from the LR and if included, would detract from a meaningful average house price figure. All exclusions cover:

1. All commercial transactions, including operational farms
2. Transfer, conveyances, assignments or leases at a premium with nominal rent which are:
 - “right to buy” sales at a discount
 - subject to a lease
 - subject to an existing mortgage
 - purchased with an identifiable buy to let mortgage
 - to effect the sale of a share in a property
 - by way of a gift
 - by way of exchange
 - under a compulsory purchase order
 - under a court order
 - to trustees
 - vesting deeds
 - transmissions or assents
 - of more than one property
 - leases for 7 years or less

Timeliness and punctuality

(Timeliness refers to the lapse of time between publication and the period to which the data refer. Punctuality refers to the gap between planned and actual publication dates.)

The HPSSAs can be produced and published within 4 months of the latest complete Price Paid data being published by the LR, which itself is made available on a monthly basis. For more details on related releases of LR data, the [GOV.UK release calendar](#) is available online and provides 12 months' advance notice of release dates. In the unlikely event of a change to the pre-announced release schedule, we will notify you and provide the reasons for the change, as set out in the [Code of Practice for Official Statistics](#).

How the output is created

Associated files and preparation

LR Price Paid data from LR are used to produce the HPSSA and the Price Paid data can be accessed for free via the [GOV.UK website](#) as a whole dataset or for individual years or months or by other specific search criteria.

In order to link LR Price Paid data to geographic information, the postcode of each transaction is matched to a postcode lookup file which can in turn be used to determine which MSOA and other geographies each transaction took place in. The postcode lookup used in the production of these statistics is the National Statistics Postcode Lookup (UK) (NSPL). This is a complete list of current and historic postcodes in the UK along with a selection of the statistical geographies they are

situated within. The [NSPL](#) is produced by ONS Geography and is freely available for download. The most recent available version of the NSPL is used when the statistics are produced.

Joining data, geographic referencing and calculation

With datasets prepared, the NSPL dataset is matched to the LR Price Paid data using postcodes to assign the relevant statistical geographies to each house sale. This allows groups of sales with the same geographies to be formed and averages and sale counts to be calculated.

A final list of geographies with sale counts and averages is produced and then matched back to geography lists for completeness. Outputs are made into tables and published as [House Price Statistics for Small Areas](#) on our website.

Disclosure control

The LR data behind HPSSAs are record level data with a number of variables relating to the dwelling type and the sale, but no personal data relating to the seller or buyer. This is explained further on the [LR's guidance page on accessing the Price Paid data](#). There is more detail in the [LR Privacy Impact Assessment Review and the Review report: April 2013](#). The [LR Price Paid data](#) or the [LR Price Paid Report Builder](#) has more information on an individual sale.

Validation and quality assurance

Accuracy

(The degree of closeness between an estimate and the true value.)

Registration of a property transaction with the LR is compulsory for all changes of ownership except leases with less than 7 years to run. Solicitors acting for purchasers invariably register the transaction as quickly as possible after completion. It is the view of the LR that under-recording is negligible. Therefore, such under-recording is unlikely to affect the statistics significantly.

The LR Price Paid data used in the HPSSAs are taken directly from the sale contract and are audited. Deliberate misreporting of price in this documentation would in most cases be fraud and is likely to be insignificant in volume. To minimise errors occurring in transcription, the LR has quality control procedures to check for exceptions in the data capture process such as price band and application type.

The LR Price Paid data contains records of each completed and registered residential dwelling transaction. Recent transactions can be registered with the LR after the publication of the Price Paid data, and so are not included in the HPSSAs. This is known as registration lag and is predominantly the case for transactions which occurred in the most recent quarter. For example, for statistics relating to the year ending quarter 2, 2015, it is the second quarter (April to June) of 2015 which is most likely to contain transactions which were not registered and included in the Price Paid data at the time it was published. Therefore, in this instance the number of house sales for the year ending quarter 2, 2015 in our HPSSAs is lower than the actual number of house sales that took place during this time. However, when the next data are published (year ending quarter 3, 2015) all previous periods (including year ending quarter 2, 2015) will be updated to reflect any additional registrations for transactions which took place during those previous periods. This means the effect of registration lag on the HPSSAs is minimised as much as possible.

Registration lag does not affect a sufficiently large enough number of records to influence the house price statistics in the HPSSAs, nor is it thought to affect one house type more than any other, and so a formal process of revision is not employed.

Information on duplicate records, geographical coverage and unmatched records is also produced alongside the output for quality assurance and analytical purposes. This section contains a brief summary of each of these areas along with tables summarising the number of records affected by these issues. For more detailed statistics, or specific queries regarding unmatched records, duplicate records and geographical coverage, please contact better.info@ons.gov.uk.

Duplicate records

The LR Price Paid data used to produce the HPSSAs contains what may appear to be duplicate records. These can occur when multiple properties (newly built flats for example) are sold as a batch with the same postcode, price and date. These records will still have unique identifiers, but

every other variable may match. A process of identifying and quantifying these apparent duplicates was devised in order to identify spikes in duplicates that may require further investigation or deletion. Checks have been done to ensure that the data are deemed to be free of genuine erroneous duplicates which may distort dwelling prices.

Geographical coverage

It is important to note that the coverage of the published HPSSAs may not be 100% for all reported property types in all geographies. This is because some areas did not have enough house sales for a given property type in a given year to produce a robust median house price. This is most common for smaller geographic areas, such as MSOAs but, for example, affected no more than 25% of MSOAs in 2013. If, for a given year, for house type and area there were fewer than 5 sales records in the LR Price Paid data, the house price statistics are not reported in the HPSSAs. Therefore, coverage can vary throughout the time series. This issue has a negligible impact on coverage at the larger geographies and for the “All Types” category for all geographies.

Unmatched records

Part of the calculation process for HPSSAs is joining LR Price Paid data with the NSPL. This process does not always achieve a 100% match rate as not all house sale records have postcodes attributed to them and some that do, have erroneous or invalid postcodes that do not appear on the NSPL. These unmatched records are infrequent and show no bias to particular areas, price ranges or house types. Such records are not included in the calculation of median house price. Any error in the medians and number of transactions resulting from these unmatched records is negligible and assumed to be random. For example, between 1995 and 2013, 0.16% of records could not be matched to postcodes.

Quality assurance

In addition to assessing the extent of duplicate and unmatched records, and in order to quality assure the figures, a small sample of HPSSAs medians and counts with a broad geographical spread are re-calculated manually using the geographically referenced LR Price Paid raw dataset. These figures are cross checked to ensure the automated process worked correctly and outputs were accurate. Validation checks are also carried out by comparing the data produced with other house price data sources to identify similar trends (Coherence and comparability has further details).

Coherence and comparability

(Coherence is the degree to which data that is derived from different sources or methods, but refer to the same topic, are similar. Comparability is the degree to which data can be compared over time and domain, for example geographic level.)

Our HPSSAs are fully comparable throughout the time series going back to 1995 because a single method was used to produce statistics for all available periods, geographies and property types.

A house price index is a measure of the change in average house prices in an area between 2 points in time. An increase to an index of 1 means that house prices have increased by 1% between the 2 points in time. Currently, there are a number of different house price indices published in addition to HPSSAs. There are differences in the statistics reported by each index because there are differences in both the data and methods used to produce them. Therefore, the HPSSAs are not directly comparable with these other indices.

House Price Index (HPI)

The HPI is a monthly release that contains figures for mix-adjusted average house prices and house price indices for the UK, its constituent countries and the regions of England. The index is calculated using mortgage financed transactions that are collected via the Regulated Mortgage Survey by the Council of Mortgage Lenders. These cover the majority of mortgage lenders in the UK. The HPI complements other measures of inflation we publish, such as the Consumer Price Index, the Producer Price Index and the Services Producer Price Indices. Working with the LR, we are currently in the process of developing a new and improved methodology for the HPI. More information on this can be viewed in the [proposed development of a new, definitive HPI - user consultation](#).

Land Registry House Price Index

All residential property transactions in England and Wales are recorded by the LR and form the LR Price Paid data of single residential properties which have sold for full market value. These transactions are used for calculating the LR index. This index is based on repeat-sales regression, which calculates the change in price of any property transacted twice since 1995. Therefore, new build properties are excluded from the index. LR publishes indices at county, unitary authority metropolitan district and London borough levels and is published on the 20th working day of every month. It refers to all transactions of the preceding calendar month. The [LR HPI](#) can be accessed via the GOV.UK website.

Registers of Scotland Official Quarterly Housing Market Statistics

Registers of Scotland record all the property transactions in Scotland. It produces monthly average house prices based on arithmetic means of these transactions, which is published as the [Quarterly House Price Statistics](#) in the second month after the period to which the figures refer to.

Northern Ireland Residential Property Price Index

The Land and Property Services assisted by the Northern Ireland Statistics and Research Agency publish a quarterly [Residential Property Prices Index \(RPPI\) for Northern Ireland](#). The index measures change in the price of residential property sales recorded by HM Revenue & Customs. This statistic was first published in quarter one of 2012.

Halifax House Price Index and Nationwide House Price Index

Both the [Halifax](#) bank and the [Nationwide](#) building society produces monthly house price indices based on their own mortgage approvals only and therefore, like the HPI, will not include any cash transactions. They both have UK-wide coverage, and because the Halifax and Nationwide use only their own data, they can process the data immediately and do not have to await the receipt of data from other lenders. The statistics can therefore be published sooner after the reference period than other monthly house price statistics.

Concepts and definitions

(Concepts and definitions describe the legislation governing the output and a description of the classifications used in the output.)

Median

The HPSSAs report median figures. The median is the value determined by putting all the house sales for a given year, area and type in order of price and then selecting the price of the house sale which falls in the middle, such that an equal number of transactions lie above and below that value. The median is less susceptible to distortion by the presence of extreme values than is the mean, and it is the most appropriate average given the near 100% sample of house sales.

Mean

The mean is the value determined by adding the values of all transactions for an area together and dividing it by the number of transactions that took place in that area. It is useful for comparing with the median to give an indication of the distribution of house prices in an area and assessing what affect extremely high value transactions have on this distribution. However, the median is the most appropriate measure of average house price.

Lower quartile

The lower quartile is the value determined by putting all the house sales for a given year, area and type in order of price and then selecting the price of the house sale which falls three quarters of the way down the list, such that 75% of transactions lie above and 25% lie below that value. These are particularly useful for assessing housing affordability when viewed alongside average and lower quartile income for given areas.

Tenth Percentile

The tenth percentile is the value determined by putting all the house sales, for a given year, area and type, in order of price and then selecting the price of the house sale which falls nine-tenths of the way down the list, such that 90% of transactions lie above and 10% lie below that value. These give an idea of the prices paid for the very cheapest housing in a given area and helps determine housing affordability for those on particularly low income.

Transaction

A transaction occurs when a change of freeholder or leaseholder takes place regardless of the amount of money involved. Similarly, an exchange of properties would be classed as 2 transactions. The transaction covers both the property and the land on which it stands; this could be extensive, but operational farms and other commercial properties are excluded from the LR Price Paid data. Other exclusions are described on page 3 of this document.

Transaction price and market value

The LR records the transaction price, that is, the actual price for which the property changed hands. This will usually be an accurate reflection of the market value for the individual property, but it is not always the case. For example, if the ownership of a property is transferred from one individual to another within the same family the price may be fixed by agreement at a level quite different from the market value. Similarly, if the transaction was agreed under the legislation which gives council tenants the right to buy their home, the price would have been discounted.

In order to generate statistics that more accurately reflect market values, the LR has excluded records of houses that were not sold at market value from the dataset. The remaining data are considered a good reflection of market values at the time of the transaction. However, the LR Price Paid data are not adjusted to reflect the mix of houses in a given area. Fluctuations in the types of house that are sold in that area can cause differences between the median transactional value of houses and the overall market value of houses. The HPSSAs are an accurate measure of social change with regards to house prices, and not an economic measure of change or inflation.

Other information

Output quality trade-offs

(Trade-offs are the extent to which different dimensions of quality are balanced against each other.)

This output is not subject to scheduled revisions, however, the raw data used to calculate these statistics are occasionally revised to maintain the highest possible quality. When this happens, the latest release of House Price Statistics for Small Areas will be calculated using the revised data and any revised historic data will be included in the latest release.

If a particular area contained fewer than 5 property sales in a given year, the median, mean, lower quartile and tenth percentile are not reported. A minimum of 5 records is deemed necessary to produce a robust median house price, indicative of the prices in a particular area. This was agreed by our methodologists and has been standard practice with DCLG in historic releases. This approach results in some areas lacking 100% coverage, in particular the smaller geographic areas and individual house types.

Summary of the main users of House Price Statistics for Small Areas

In order to identify the users and uses of HPSSA, regular communication with other producers of house price statistics, such as our HPI team, Department for Communities and Local Government (DCLG) and the LR, was carried out to understand the types of request users make for more detailed statistics on house prices. The Neighbourhood Statistics Service team and our customer contact centre were also consulted to help gather more information about the potential users of HPSSA. The following broad groups of users have been identified:

- central government: monitoring housing trends, supply and demand in society, policy making and regulation
- local government: monitoring and developing their housing policies to meet the current and future needs of their areas and to understand how changes and policies at the national level affect housing at the local authority level

- devolved administrations: supporting policy making and monitoring changes at the country level, similar to those requirements of central government; also used for comparisons to wider UK policies
- banks and building societies: house price statistics are used for mortgage lending in order to make decisions on whether to lend, how much to lend, and setting interest rates
- house builders: house builders are interested in whether and where demand for new housing exists, and the returns received on homes built or converted
- housing industry specialists: these include organisations such as larger estate agents seeking information on small area trends
- housing bodies: these include organisations such as the Home Builders Federation
- media: make use of the data to inform and aid journalistic coverage of local housing markets
- social researchers: data may be indicative of social status and may be used in the development of further statistics and social analysis
- academics: include geographers and sociologists who make use of these statistics for study and analysis

Sources of further information or advice

Accessibility and clarity

(Accessibility is the ease with which users are able to access the data, also reflecting the format in which the data are available and the availability of supporting information. Clarity refers to the quality and sufficiency of the release details, illustrations and accompanying advice.)

The recommended format for accessible content is downloadable content and web content for narrative, charts and graphs, with data being provided in usable formats such as CSV and Excel. The data for these statistics are available from our website; users can also download the narrative in PDF format. For further information please refer to the contact better.info@ons.gov.uk.

For information regarding conditions of access to data, please refer to the links below:

- [terms and conditions \(for data on the website\)](#)
- [copyright and reuse of published data](#)
- [pre-release access \(including conditions of access\)](#)
- [accessibility](#)

In addition to this Quality and Methodology Information, [all HPSSA publications](#) are available with analytical content and data.

Useful links

[Land Registry information on LR Price Paid data](#)