June Survey of Agriculture and Horticulture

What are these statistics?

This is an annual sample survey carried out between June and October each year which ask farmers about land usage, livestock numbers and people working on the farm. The first survey was carried out in 1867 and remained a census of all farms until the mid 1990s. Indeed the survey is still (mistakenly) referred to as the June or Agricultural Census in some quarters today.

The June Survey is the primary source for information about agricultural land, livestock and farm labour covering all known farms. This is in contrast to administrative systems, such as the agricultural subsidy payments, which give definitive information about a restricted range of farms.

Typically the survey is sent to about 11,000 farms each year though in some years a larger sample is required. This is when more detailed information is required on particular subjects (eg animal housing) and specified level of coverage needs to be achieved.

The results of the survey are usually published in November following the survey. They also feed into the results for the UK as a whole and these are published in December.

This is followed by an imputation exercise which produces imputed results for each individual farm who were either not included in the sample or who were sent a form but did not respond.

The imputation process then allows for the building of a data set at individual farm level. This offers the maximum flexibility for analysing the data. This also underpins our other statistical publications which are based on the survey:

- Agricultural Small Area Statistics
- Farming Facts and Figures
- StatsWales

The key trade off is thus between the level of detail being sought and the robustness of the resulting estimate. This is discussed further in Missing Data later in this report. The imputed data does not purport to be accurate at the individual farm level. Each farm will have its own characteristics which will not be represented in the imputation process. These differences become less pronounced as data from more farms becomes aggregated together.

Users of the survey results

The survey results are used by a variety of users. The main users known to us can be classified into the following groups.

Other areas of Welsh Government

Survey estimates are used widely within Welsh Government both from the main regular publications and from ad-hoc analyses to address specific policy questions. The survey contributes in many ways including:

- As baseline information showing what the sector is currently like and how it is changing;
- To help develop new policy by showing the likely scale of issues and their impact;
- To help monitor existing policy by showing how things change over time;
- To help draw representative samples of farms in Wales to support evaluation of existing policies, particularly when trying to reach farms not in particular grant schemes;
- To link with other data sources to give as full a picture of Welsh agriculture as possible. For example, linking with payments information from Rural Payments Wales.

Data is also used in the area of animal health in assessing the potential effects of various disease outbreaks or in limiting the spread of diseases.

The data is also used extensively by Welsh Government economists in their work in assessing and analysing the agricultural economy in Wales (eg through the Farm Business Survey).

Other Government bodies in Wales

- Local government. The number of farms is a factor used in the Local Government revenue distribution formula. Helping authorities understand the nature of farming in their area.
- Health and Safety Executive. Helping to plan events promoting safe working on farms.
- Natural Resources Wales. Helping to understand the distribution of diffuse pollution sources from agriculture and how this may affect the health of various river catchments.
- National Parks. General background on activity within the National Parks.

Government outside Wales

- Survey results are provided in the compilation of UK results by the Department for the Environment, Food and Rural Affairs (DEFRA). Survey data is used periodically in assorted projects and studies into a range of aspects of agriculture. These are commissioned by DEFRA at a UK, Great Britain or England and Wales level;
- The data are required by the international organisations (such as the United Nations and World Trade Organisation) in order to assess the scope of global food production. Historically detailed statistics have also been provided to Eurostat for analysis of agricultural statistics across the European Union.

Agricultural sector bodies

The key sector bodies who make regular use of the survey results are:

Farming unions – primarily the National Farmers Union (Wales) and the Farmers Union of Wales. Generally the statistics enable them to maintain an overall knowledge of the current state of the agricultural sector as well as being aware of the latest trends.

Meat Promotion Wales / Hybu Cig Cymru – their role is to monitor the state of the red meat industry in Wales. Survey livestock numbers provide one of a number of inputs that they require in order to carry this out.

The Agricultural Development and Advisory Service (ADAS) undertake consultancy, research and policy advice on the areas of the environment and rural development. They are an independent company but are frequently commissioned by the Welsh Government and other government departments. Agricultural survey data is often an input required to allow them to undertake this.

Other, more specialist, sector bodies also require data but this is on a more infrequent basis.

Media

First releases of survey results are provided to contacts in the farming press. The interest from the general media is more infrequent, but in the past have come from television, radio and the written media.

Researchers

Often academics and/or specialists who have a focus on a specific topic. In many cases the agricultural survey data is combined with data from other sources for modelling purposes. Recent examples include statistics on numbers of cattle for studying greenhouse gasses, and details on the growing of arable crops in assessing the effect of pesticide usage.

Animal and Plant Health Agency (APHA)

Formerly known as the Animal Health and Veterinary Laboratories Agency (AHVLA). This body are provided with detailed survey data in order to carry out modelling work in the event of a serious disease outbreak or other emergency.

General enquiries

A whole range of enquiries have been received in the past. These tend to be a combination of phone calls and e-mails to our general inbox (stats.agric@gov.wales). This category includes individual farmers, schoolchildren, postgraduate students and other members of the public with an interest in agriculture or in Wales as a whole.

Data Processing Cycle

Sampling

The sampling for the survey is based on first stratifying the population (all active farms) based on the size of the farm. In this context size means economic size rather than geographic area.

Farm size is measured by European Size Unit (ESU). This is a measure of the economic turnover of the holding. All active holdings will have data associated with it for land usage and livestock numbers and thus an ESU. The use of ESU as a measure of farm size is a historic one and the current measure of farm size is Standard Output (a measure of profit). We are currently reviewing survey methodology with a view to basing it on this measure instead. However, the principles outlined below are still applicable.

The ESU for each holding is calculated by applying a coefficient to each variable in that holding's data, the outcome being the coefficient multiplied by the quantity of that variable present. In other words, a weighted sum of the number of livestock and areas of crops.

To help interpret this measure, the following table shows how many animals or how much land of a specific use is required to give a value of one European Size Unit. This is a brief selection of some of the main variables, intended to give an illustration.

Item	Required for 1 ESU	Measure
Sheep - breeding ewes	29.3	headcount
Cows - dairy breeding females	1.0	headcount
Cows - beef breeding females	3.1	headcount
Pigs – breeding sows	2.8	headcount
Pigs - others	63.2	headcount
Laying hens	271.3	headcount
Table chicken (broilers)	563.4	headcount
Wheat	1.4	hectares
Barley	1.6	hectares
Oats	1.2	hectares
Potatoes	0.3	hectares

Holdings are then grouped into 6 size groups (defined below) and sampled at the following rates:

Size group	Criteria (in ESU)	Sampling rate
New holdings	n/a	100%
Zero ESU	0	0%
Very small	>0 and <8	33%
Small	=>8 and <40	40%
Medium	=>40 and <100	60%
Large	=>100 and <200	100%
Very large	=>200	100%

New holdings are mainly those who have been registered in the previous 12 months. Since nothing is known about these holdings, they are all included as part of the sample.

This framework excludes holdings with a zero ESU from the sample as, by definition, zero ESU means there is no agricultural activity on the holding (and thus little point in sending a form).

Holdings that specialise in pigs and/or poultry are relatively small in number (about 250 larger units) and so can be targeted with a separate form which focusses on these activities. These holdings are excluded from the main sampling frame detailed aboveData is primarily collected via a paper-based survey. The numbers of cattle have been sourced from the Cattle Tracing System since 2007. We are continuing to explore other areas where administrative data can be used in order to reduce the burden on farmers.

Online returns were introduced for the 2020 survey and showed and encouraging start with 1 in 5 returns being submitted online. Further information on this introduction can be found in 2020 survey results release and in a Digital and Data blog on the subject.

Validation and verification

All survey forms are returned to the Welsh Government where they are checked for basic errors (eg areas being given in acres and not hectares) and corrected where necessary. They are then sent for data capture and this data is then loaded for further validation checks.

Any data failing any of the validation rules is checked. After checking, the data may be corrected or it may found to be correct (and just outside the parameters of what would usually be expected). In the first case, it is hoped the correction will mean the amended data now passes validation. In the second case, the data value(s) are noted as accurate and allowed to stand.

At the close of the survey there will be some data which still have outstanding validation errors. Usually these will be where it has proved to be not possible to contact the farmer to resolve the matter. This data is excluded from the estimation process.

Estimation

The data taken from the survey forms can be compared with previously known data for the same farm. Data from similar farms can be aggregated to calculate trends and these trends can be extended to produce estimates for the population – a process known as raising.

However, this method does not prove suitable for some questions on the form. These are in areas where there is a limited amount of sample data and/or base data. In these cases, a very small number of "outliers" (observations that are markedly different from their base data values) can have a profound effect on the estimate. (These would tend to be cancelled out or at least have a much smaller influence where larger amounts of data were available).

Therefore the variables are split into two groups. Those where there is sufficient data to be able to use the original raising method and those which are not. These groups are shown below – in most cases listed by section than individual question. There is no numerical threshold in terms of the number of observations required to fall into one group or the other. The data itself neatly splits into areas where there are several hundred or thousand of observations and areas where there are relatively few (in most cases less than 100 and often less than 50).

Sections/questions that use raising method (with many observations)	**Sections/questions that use most recent observations(with few observations)**
Arable crops (13 questions)	Horticulture (6 questions)
Grass and other non-arable land (5 questions)	Pigs (12 questions)
Sheep (6 questions)	Poultry (8 questions)
Horses (2 questions)	Goats (3 questions)

The simpler process is that using most recent observations so this method is described first. For each individual question, two lists of holdings are calculated. All holdings in the current year who have any value for that variable and an equivalent list from the previous year. These lists are then amalgamated. Any holdings that are no longer active are removed. Each holding is then assigned a value for that variable according to the following process:

The data provided in the current year is taken where it is available

If there is no data for the current year then the value from the previous year is taken

Note that holdings with a (non-zero) value in the previous year who return a zero in the current year will now be assigned that zero value. The final estimate is simply derived by totalling across all holdings. Note that no trend is applied (unlike in the raising method) because the paucity of data makes it impossible to calculate a sensible one.

The estimates produced by the raising method are calculated as follows. First all holdings that have outstanding validation errors are excluded from the raising process. Raising is carried out for individual questions at a time. It should be noted that only holdings with validation errors in that

particular area being raised are excluded. That is to say, for example, if it was an estimate for sheep being raised, holdings with validation errors for crops would be included in the process.

In a similar manner to the validation errors, holdings that are thought to be outliers are excluded from the raising process. As mentioned above, outliers are holdings on which the change between the base value and the observed value may have a disproportionate effect on the raised estimate. Again only data relevant to the question being estimated are excluded. As this process is based on a relatively large number of observations, the effect of outliers is much diminished. Those outliers that are identified fall into two categories:

i) those holdings where the base data is one of the largest values for that variable and the observed value is zero or a tiny proportion of the base value (or the other way round)

ii) holdings in the larger size groups which exhibit very significant differences between the base and observed values. The larger size groups are focussed on in more detail as they contain far fewer holdings and an element of the raising concerns raising by individual strata (see below).

As mentioned previously, the raising is carried out for each individual item on the survey. After validation errors and outliers have been removed, the raising process involves producing two estimates by alternative methods.

The first estimate is derived by splitting holdings into their size group or strata and raising an individual estimate for each stratum. These estimates are then aggregated to produce the overall estimate for the item. The second estimate raises a single estimate for the item (irrespective of strata).

The two estimates, along with their associated standard errors, are then compared and the better (in most cases that with the lower error) is chosen. The data not included in the raising (validation errors and outliers) are then added to the selected raised estimate to produce the final estimate.

Imputation

The estimation process produces results for Wales as a whole and these are published as soon as possible – usually in mid-November. However there is a wide demand for assorted, more detailed analysis of the survey (listed in the Users section of this report). In order to provide these, we need to produce imputed values at individual holding level for those farms who are still active but who either weren't sampled or who did not respond in the most recent survey. (NB this doesn't apply for cattle as the Cattle Tracing System data covers the entire population). For these holdings, the trend observed on those holdings with returns in the current year is applied to the base value for each question in turn. Thus in aggregate, when combined with the actual observed values, the sum across all holdings will be the total for Wales.

This then provides a data set which offers maximum flexibility for any cross-cutting analyses. However, care must be taken as the process is not intended to produce an accurate imputed value at individual holding level. Changes on individual farms will not be uniform. Thus any analysis undertaken should be at a suitable level of aggregation so that any result derived will be sufficiently robust.

Quality information

The Code of Practice for Statistics requires the release of data at a level of aggregation that is sufficient to ensure that the results are adequately robust and so "fit for purpose". Since the results are derived from a sample survey they are subject to various sources of uncertainty. The key sources of uncertainty are:

- Missing data
- Register of farms
- Geographical location of farms

Missing data

The June Agricultural Survey is a sample survey of farms. Not all farms are selected for the sample. Not all farms that are selected complete the questionnaire. The overall Wales totals are estimated from the survey results by assuming that the increase from last year for the population as a whole is the same as that we see in the sample. Given that we have taken a random sample of farms this assumption is supported by statistical theory. Analysis of actual results also shows that *at the aggregate level* these estimates are adequately robust.

These estimates will be at their best when there are a lot of similarly sized farms that are, broadly speaking, following the same patterns. This is usually the case for cattle and sheep farms in Wales. The estimates are more problematic when an activity is dominated by a few large farms or when there is great volatility. Specialist chicken producers would be an example of the former while very small or non-commercial farms would be an example of the latter.

For holdings who were not sampled or who did not respond we attempt to impute for these missing values. A limitation of this methodology is the assumption that an imputed farm will have the same set of agricultural activities as it did in the base year. For the individual holding there can be no changes to the types of crops grown or livestock kept – although the areas and numbers will change.

Clearly these imputed farm level results are not particularly accurate *for individual farms*. Their usefulness comes in being a flexible building block to construct a very wide range of aggregate estimates. The accuracy of these aggregate estimates comes from combining a large number of individual farms. Statistical theory shows how this process of aggregations leads to more accurate estimates.

Register of farms

There is no compulsory register of farms in the UK. The voluntary nature of the register means that it will always have issues about missing farms (false exclusions) and farms that have ceased to operate but that are still on the register (false inclusions). If nothing else there will be issues of how quickly we are informed of changes.

There are two main routine sources of information for the register.

- Links with administrative systems. The register is updated with information from various administrative systems. The main sources are the subsidy payment system and the

compulsory register of cattle keepers (BCMS). Data are received from time to time from other administrative systems.

- Correspondence through running the various agricultural surveys. This provides updated name and address details plus a number of forms that are returned undelivered.

There are a number of reasons why a farm would need to be registered. The registration code (CPH number) is used for tracking animal movements and when animals are sold or slaughtered. This means that farms wanting to sell their animals have a strong motivation to be registered. In Wales farms without livestock are rare, although they do exist. Farms where the animals are kept for non-commercial reasons have a weaker motivation to register.

There is little motivation for a farmer to de-register when a farm ceases to operate. These closures are picked up in a number of ways though the Survey and various administrative systems. However, we expect the coverage is less complete than for new farms. The implication of this is that we will continue to estimate land for the closed farms while actually the land has been sold or rented out, leading to double counting.

From these sources it is clear that register information will be at its best when we have regular contact with the farms. The coverage will be at its best for farms that are within the agricultural subsidy system because they have the strongest reason to be registered. In Wales the majority of farms are claimants.

For holdings outside the subsidy system there are four key types of holding that are known to have coverage problems.

- Specialist poultry producers.
- Specialist pig producers.
- Specialist horticulturists.
- The very smallest holdings including non-commercial concerns.

The specialist producers are a particular concern because there tend to be few of these in Wales and these sectors are dominated by a small number of farms. The most extreme is for chicken producers where more than 90% of the birds are on less than 1% of the farms with chickens.

The smallest farms do not make a significant contribution to most agricultural variables. Typically these holdings will have well below 1% of the Wales total. Note that the very smallest farms contribute a significant proportion of the total number of horses in Wales.

There are, however, large numbers of these very small farms. This may be important if what we want to know is all holdings with a particular activity – for example livestock keepers at the time of the Foot and Mouth outbreak.

Geographical location of farms

The Welsh Agricultural Survey collects information at the level of the individual farm. These farms are collections of land farmed together as a unit. *They are not necessarily contiguous sets of fields.* The fields on a farm may be scattered over a wide area. Therefore, while we will know the number

of animals on a farm we do not know where the animals graze. Similarly while we know the total area of crops we do not know where the crops are grown.

It is possible to make an estimate of the general location of a farm. These estimates can use the digital field boundaries where these are available through the subsidy system and address information otherwise. We can then build up geographical estimates by assuming that all the activity on a farm takes place at this estimated location. *Clearly this an estimate and a great simplification of reality.*

The accuracy of the geographical estimates depends mainly on using areas that contain a sufficiently large number of farms. With a large number of farms the overall net error will decrease as a false inclusion on one farm is balanced by a false exclusion on another. However, for very specific location, with small numbers of farms involved, the level of error because of the location issues must be more significant.

Strengths and Weaknesses

Strengths

The survey is the only source of information that covers all farming in Wales. The data gathered by Rural Payments Wales (RPW) in order to administer their payment schemes is more extensive in some areas (eg land use) but is subject to a number of limitations:

- RPW only hold information for claimants. This would cover around 16,000 farms in Wales. The farming register that underpins the survey comprises around 24,000 farms.
- Not all types of farms will claim payments. Specialist horticulture, pig and poultry holdings are all examples where this is the case.
- The livestock figures gathered by RPW can sometimes be limited. For example, a farmer would only indicate if he kept over 500 ewes on RPW. The survey collects actual numbers which allow more accurate trends to be calculated.
- The survey is the only source of information on people working on the farm.

The survey has been carried out annually since 1867. This provides a time series for some of the key variables going back over 150 years. This historical context can often be important when viewing numbers today. For example the number of pigs in Wales today is only one-tenth the figure it was 50 years ago. This historical time series is included in the data that accompanies the survey results.

An equivalent survey is carried out in England, Scotland and Northern Ireland. The geographical differences (eg land quality, climate) between different parts of the United Kingdom mean that farming practices vary in different parts of the country. So whilst the survey does have some variations in the detail of what is collected in each country, it allows comparisons in farming practices (current and historical) between different parts of the UK.

Weaknesses

As covered in the missing data section, the results of the survey are estimates based on the responses received. Statistical theory tells us that the more responses that are received, the more robust the estimate will be (ie the smaller will be the margin of error).

The primary sectors of farming in Wales are the rearing of cattle and sheep. This is primarily dictated by the geography of Wales being very restrictive for arable farming. Thus estimates for these livestock sectors will have much lower associated margins of error than other farming sectors that are less common in Wales.

As has been observed with other government surveys, this survey has suffered from a declining response rate over a number of years. In addition, farmers are already required to complete detailed paperwork for payment/subsidy and animal health purposes, which may adversely affect the response rate for a statistical survey.

Various methods have been attempted to address this problem. These include:

- Reviewing the form content and layout to make it easier to understand.
- Providing an online option for completing the form.
- Publicising the importance of the survey through the farming press.
- Posting reminder cards to non-respondents.
- Contacting farmers directly via RPW's Broadcast Message facility.

Survey Changes

From time to time the detail of how the survey is conducted is changed, The reason for these changes are two-fold:

- Planned changes. These are almost always changes to the question set that is used for the survey. Examples where this has been in attempting to simplify the questions relating to arable crops and for questions on people working on the farm. Such changes are kept to a minimum as it avoids any disruption in time series data comparing results from different years. Making these changes can also lead to uncertainty in how data is reported. If farmers are asked the same questions each year then we can be confident that the numbers the report are on the same basis.
- Unplanned changes. Occasionally events will arise which force us to consider the fundamentals (eg sample size, timing) about carrying out the survey in a particular year. Indeed whether the survey is even carried out. Such events are relatively rare. Prior to 2020, the last occurrence was the Foot and Mouth Outbreak in 2001. The COVID-19 pandemic resulted to a delay in carrying out the 2020 survey for 10 weeks. More detail on the effects of COVID-19 on the survey can be found in the 2020 survey results release.

Consistency of UK statistics

Each of the four counties in the UK run a similar agricultural survey in June each year. Publication of UK data is co-ordinated by the Department for Food, Environment and Rural Affairs (Defra).

The definitions of livestock types agree well between the four surveys. There are some issues with crop types. In the main this is due to an east/west split in Great Britain. Land in the western side (including Wales) is mainly upland and does not lend itself to arable production. In contrast, land in the east is predominately lowland. Thus there is a big contrast between the limited areas suitable for growing crops in Wales and the huge areas of crops found on the eastern side of England.

There are some differences in the target population of farms. England does not sample the very smallest farms. Describing the population as of “significant farms”. This will only really impact on the farm numbers since the thresholds are set at a low level so that even in aggregate the farms contribute few animals and little land. Scotland include all farms claiming payments and then sample a selection of non-claimants.

Evaluation

All official statistics should comply with all aspects of the Code of Practice for Statistics. This is assessed by the Office for Statistics Regulation (the UK Statistics Authority’s regulatory arm). These survey results last underwent a compliance check against the Code of Practice in 2018.

Compliance Check for Agricultural Statistics

Following this review, we have continued to comply with the Code of Practice for Statistics. This includes:

- Introducing the facility to complete the survey online.
- Continuing to use clear, non-technical language to present the statistics and their context to as wide an audience as possible.
- Headline “latest data” is presented at the top of each section as these will be of most interest to our key users.
- Historical data is provided to accompany this release along with commentary to explain the underlying factors behind historic trends.
- Expanding the detail on how the data is used in other areas of Welsh Government.
- Using other Welsh Government and Farming Union magazines which are circulated to farmers in order to encourage them to return their survey forms.
- In early 2020 to reduce the length of this Release in future by having the Methodology Annex accessible as a stand-alone document on our website which can be referenced via a hyperlink from this document.

We welcome any feedback on any of our statistics. stats.agric@gov.wales

Produced by Knowledge and Analytical Services, Welsh Government

Last updated: January 2021