

Briefing Note: confidence intervals and statistical significance within the Active People Survey

The Active People Survey is a **sample-based** survey. It is a telephone¹ market research survey interviewing a random sample of the English adult population (16 plus). The results are then used to draw conclusions about the English population as a whole.

How does interviewing c363,000 people give an accurate picture for a total population of over 40 million?

The Active People Survey uses a genuine random sample of the England population so the findings of the survey are therefore representative of the England population. This is the basic principle of sample-based surveys and is founded on mathematical probability theory.

So, interviewing 500 people out of a 40 million population is as statistically valid as interviewing 500 out of 100,000 people?

Yes, if the sample drawn is genuinely random then statistically, the underlying population total is irrelevant if it is a large population.

How therefore does a sample-based survey differ from interviewing the whole population?

A census interviews everybody in a given population, for example the UK Census 2001, so there is no statistical chance the findings do not represent the whole population.

All sample-based surveys contain a level of uncertainty to the findings because they are based on a proportion (a sample) of the population, not the whole population. However, we are able to measure how likely it is the survey would show the same results if the whole population was interviewed. This 'likelihood' (also known as 'uncertainty', 'sampling error' or 'confidence') in results from a sample survey can be measured through using the standard statistical tests summarised below:

- **Confidence level** – the Active People Survey results are presented using the 95% confidence level. This means there is only a one in 20 chance the findings of the survey are not a true representation of the population; in other words, there is a 95% chance that the confidence interval covers the true, underlying population mean. The 95% confidence level is the most commonly used confidence level in social research.
- **Confidence interval** - this calculates the range within which the 'true' figure will sit if the whole population was interviewed. For example, if Active People Survey 1 found that 22.0% of respondents in a local authority with 1000 sample size undertook 3x30², the associated confidence interval with this

¹ See appendix A for information on the inclusion of 'mobile phone only' respondents within Active People.

² Active People Survey KPI 1 – 3x30 – 3 days a week, 30 minutes moderate intensity sport and active recreation

result is $\pm 2.6\%$. This means we are 95% sure (confidence level), if the whole local authority population was interviewed, the 3x30 result would be between 19.6% and 24.6% ($22.0\% \pm 2.6\%$).

- Another way of describing the confidence interval is to refer to the level of precision. The smaller the confidence interval the greater the level of precision.

What does the confidence interval depend on?

Whilst confidence intervals are most strongly affected by the sample size, the population size can also have an effect on it. If the population is finite, then the actual confidence interval would be somewhat narrower than it would otherwise be if the population was infinite. This effect only normally occurs where the population size is 10 times (or less) as large as the sample; otherwise it can be ignored.

This effect is not critical with respect to the Active People Survey as it based on the entire adult population in England. Therefore a 1000 sample for a national population of 40 million will still have, to all intents and purposes, the same confidence interval as a 1000 sample for a local authority with a population of 800,000.

The confidence intervals for results from the Active People Survey depend on two main factors: a) the sample size used within the survey and b) the result (or prevalence) obtained from the survey:

- **The sample size**

For any given random sample as the sample size of the survey increases (i.e. the number of respondents interviewed), the confidence interval decreases. This is the same as saying that if you interview more people you get a more accurate result. For example Table 1 shows that 95 times out of 100, the *true* value for the whole population would lie between 22.4% and 22.6% (a confidence interval of plus or minus 0.1%). Likewise, at local authority level, with a smaller sample size, 95 times out of 100, the *true* value for the whole population would be between 19.9% and 25.1% (a confidence interval of plus or minus 2.6%). The larger confidence interval for the local authority in this example is solely due to the much smaller sample size of 1,000.

Table 1

	Active People Survey 1 (95% confidence level)		
	%	Sample size	Confidence interval
National	22.5%	363,000	$\pm 0.1\%$
Local Authority 1	22.5%	1,000	$\pm 2.6\%$

- **The prevalence (percentage obtained) (%)**

Confidence intervals decrease as the survey result becomes polarised, away from 50%. For example a 99% prevalence has a smaller confidence interval than a 50% prevalence. In Table 2, the sample size for each area is the same (1,000)

so the difference in the associated confidence intervals is due to the different prevalence obtained for each area (92.5% compared with 22.5%).

Table 2

	Active People Survey 1 (95% confidence level)		
	%	Sample size	Confidence interval
Local Authority 1	22.5%	1,000	±2.6%
Local Authority 2	50%	1,000	± 3.1%
Local Authority 3	92.5%	1,000	±1.6%

What are the advantages of a larger sample size?

A larger sample size provides a greater level of precision within any population and in turn allows analysis of more sub-groups within the survey. When comparing cross sectional results (two independent surveys carried out at different points in time) (discussed later), the scale of the Active People Survey enables detailed analysis of the results by demographic subgroups and at all geographies, down to a Local Authority level. The overall sample of Active People Survey 1 of c.363,000 is exceptional compared to other national population surveys, such as the Health Survey for England, which has an annual sample size of 15,000 and the Taking Part Survey (27,000 per year). By way of comparison, National Opinion Polls conducted during election periods are usually based on a national sample of approximately 1,500/2,000 people. The larger sample size of Active People enables an increased level of precision when analysing results.

How do confidence intervals apply when interpreting results within a single survey, for example Active People Survey 1 or Active People Survey 2?

When comparing results within a single survey, for example Active People Survey 1 or Active People Survey 2, it is necessary to consider the confidence intervals. The confidence intervals may vary for different geographical areas, for example a local authority and a region, mainly due to sample size differences, as explained above. What the confidence interval tells you is a measure of how close your sample-based result is likely to be to that of the population. From a purist point of view, if 100 experimenters independently generated samples of the same size from the same population and calculated the results with an associated confidence interval, then 95 of the confidence intervals would cover the population result. In this case 2 or 3 of the confidence intervals would have their upper limits falling below the population result and 2 or 3 would have their lower limits falling above the population result. From a pragmatic point of view, one can consider a 95% confidence interval as having a 95% chance of covering the population result.

Figure 1 shows a scenario where there is no statistically significant difference between the two results from a single survey. This is illustrated by the confidence interval for Result 1 overlapping with the confidence interval for Result 2, therefore the overall difference between the two results is not statistically significant. Conversely, Figure 2 shows the confidence interval for Result 3 does not overlap with the confidence interval for Result 4, therefore there is a statistically significant difference between the results.

Figure 1

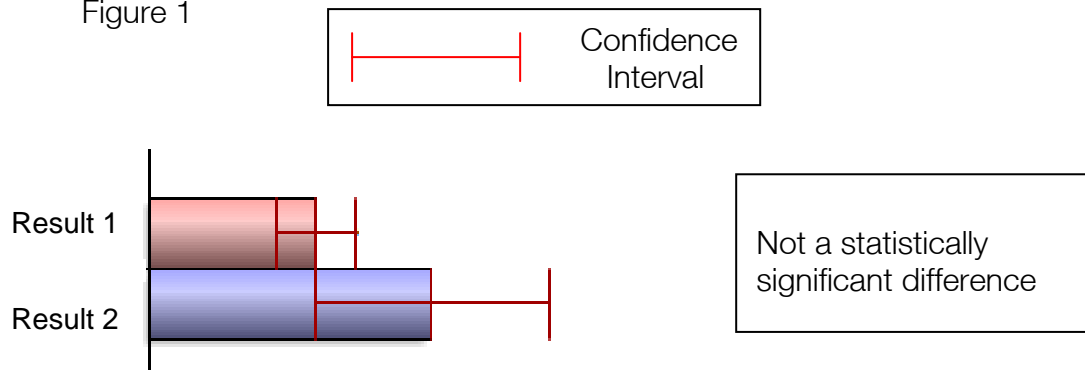


Figure 2



Confidence intervals are only normally used to describe results of single surveys, or to compare a single survey against a target / baseline. When comparing two independently sampled surveys, one should not use / compare confidence intervals of each of the two surveys, but should instead perform tests of statistical significance, which is described in the following paragraph. Confidence intervals to compare two survey waves may be used, but this is restricted to the context of calculating a confidence interval around the **difference** in the results between the two surveys.

How do we measure statistical significance when measuring change between two surveys, e.g. changes from Active People 1 and Active People 2?

Just as results from a single sample survey are subject to a confidence level and confidence intervals, the difference between results in different surveys (cross sectional) is subject to the same principle of statistical testing.

When measuring a change between two independent survey samples (Active People 1, 2005/06 and Active People 2, 2007/08) it is necessary to statistically account for the confidence interval for each estimated result. The measured change between the results therefore has a standard statistical 'error of the difference' as well as an associated confidence interval. This is termed a **joint confidence interval or comparative error** and determines whether cross sectional results are statistically significant. The comparative error is again dependent on

the sample size and prevalence (percentage result) obtained from each survey result.

Measuring whether a difference between Active People 1 and Active People 2 is **statistically significant** can therefore be calculated by whether the difference (Active People 2 result minus Active People 1) exceeds the associated **joint confidence interval or comparative error**. For example, in Table 3, from the three local authorities presented, only local authority 2 has changed sufficiently to ensure the difference between their Active People 1 and Active People 2 results is a *real* or **statistically significant** increase. In this example, the sample sizes are the same for each local authority, so the difference in the prevalence for the Active People Survey 2 result (percentage result) explains the variation in the joint confidence intervals.

Local Authority 1 has achieved a 2% change between its Active People 1 and Active People 2 survey result. The joint confidence interval is 4.6% which means the change Local Authority must achieve in order for it to be statistically significant is plus or minus 4.6%. Therefore the change it has achieved is not statistically significant.

Local Authority 2 has achieved a change of 5% between its Active People 1 and Active People 2 survey result. The joint confidence interval is 4.7% which means the change Local Authority 2 must achieve in order for it to be statistically significant is plus or minus 4.7%. Therefore the change it has achieved is a statistically significant change.

Table 3

	Active People 1 (sample size 1,000)	Active People 2 (sample size 500)	Difference	Comparative error/ joint confidence interval	Statistically significant difference
Local authority 1 3x30	22.5%	24.5%	+2.0%	±4.6%	No
Local authority 2 3x30	22.5%	27.5%	+5.0%	±4.7%	Yes
Local authority 3 3x30	22.5%	19.0%	- 3.5%	±4.3%	No

Reporting accuracy of Active People Survey results

When comparing results from Active People 1 and Active People 2, Sport England will indicate whether the difference is statistically significant at the 95% confidence interval.

When partners are quoting and reporting results from the survey it is important that only results which are **statistically significant** are reported and highlighted as

'real' differences. For example from table 3 above, local authority 2 can report as follows:

"Local Authority 2 has seen a statistically significant increase in participation (3x30) from Active People Survey 1 (2005/6) to Active People Survey 2 (2007/8).

However, local authority 1 should report their result using different language and with no conclusions drawn about the apparent change. For example, local authority 1 could report as:

" In Active People 1 22.5 % of the population undertook 3x30 while in Active People 2 the equivalent result was 24.5%. This does not show a statistically significant change however it may still indicate participation is moving in the right direction."

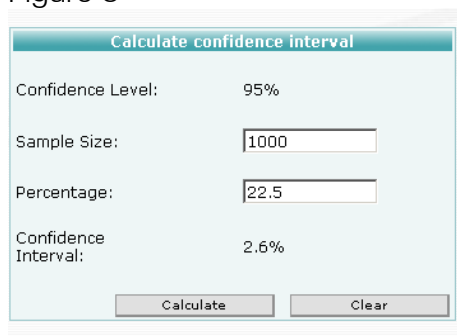
When using results for measuring performance and for strategic planning it is essential to consider whether a change is statistically significant.

How do I work out the confidence interval?

It is possible to calculate whether survey results are significantly different by using the calculator on Sport England's Active People Diagnostic (within the Information section of the site). The online calculator is available showing confidence intervals for results within a single survey (figure 3).

AA calculator is also provided for calculating whether comparative results from Active People 1 and Active People 2 are statistically significant (figure 4).

Figure 3



The image shows a web-based calculator titled "Calculate confidence interval". It has a light blue header and a white body. The interface includes the following elements:

- Confidence Level:** 95%
- Sample Size:** A text input field containing the value "1000".
- Percentage:** A text input field containing the value "22.5".
- Confidence Interval:** 2.6%
- At the bottom, there are two buttons: "Calculate" and "Clear".

Figure 4

The figure displays two identical-looking calculator interfaces side-by-side, both titled "Calculate significant difference".

Left Calculator:

- Confidence Level: 95%
- First Sample Size: 1000
- First Percentage: 22.5
- Second Sample Size: 500
- Second Percentage: 27.5
- Result: Difference is significant at the 95% level

Right Calculator:

- Confidence Level: 95%
- First Sample Size: 1000
- First Percentage: 22.5
- Second Sample Size: 500
- Second Percentage: 24.5
- Result: Difference is **not** significant at the 95% level

Both calculators have "Calculate" and "Clear" buttons at the bottom.

Below the calculators is a red banner with logos for "Powered by Archway", "Ipsos MORI Public Affairs", and "SPORT ENGLAND".

Research Team, November 2008

Useful Links:

Sport England's Active People Diagnostic:
<http://www.webreport.se/apd/login.aspx>

Research Glossaries:
<http://www.mrs.org.uk/mrindustry/glossary.htm>

http://www.marketresearchworld.net/index.php?option=com_glossary&Itemid=26

Appendix A

Are mobile phone numbers included within the Active People Survey?

No, the Active People Survey interviews a random sample of the adult population in England from a sample generated by random digit dialling of landline telephone numbers.

Large scale telephone survey samples do not include mobile phone numbers as standard. This applies to all market research surveys, not just the Active People Survey. Issues generated by excluding mobile phone numbers from a sample are therefore an issue related to the whole market research industry and not unique to Active People Survey. The main reasons for excluding mobile phone numbers are:

- Using mobile phones as an interviewing channel would mean the cost of the Active People survey is prohibitive (the cost could be up to 6 times the current value)
- Landline numbers are prefixed by a geographical code e.g. 0207, 0121 which provides a first basis for generating a random sample within different local authority areas, through random digit dialling (although in the Active People Survey the respondent location is still verified by the respondent). Generating a random sample of the population through mobile phone numbers would require a much more complex methodology and again significantly increase the cost of the survey.
- Respondents could be answering survey questions outside of their home in an environment which does not allow for the best quality of responses (e.g. shopping, social situation or at their work place).
- There is no evidence that including mobile phone numbers would consequently generate a “different” sample of the population. This incorporates the question whether people within the England population who are only contactable by mobile are fundamentally different from those who can be contacted by landlines. For people within a peer group (e.g. same gender, age, socio-economic status, and ethnicity group) does having or not having a landline impact on their sports participation levels? For example, higher education students who only have a mobile phone telephone number (e.g. halls of residence) are not necessarily *different* in sports participation terms from students contactable by landline (e.g. living in private accommodation). Halls of residence are not included within the Active People Survey sample due to the inability to randomly select a respondent from the halls of resident. Other ‘institutions’, for example care homes, are excluded from the sample for the same reason. This is a standard market research technique.
- Although a large percentage of the overall population can be contactable by mobile phones this is still a much smaller percentage than people who can be contactable by landlines. It is more likely people within a younger age group (including students) only have a mobile phone number; however there is also the likely population change that people are living in their parental home for a longer period of time. The Active People Survey sample will capture full time students located with their parents all year round,

undertaking visits at weekend, or located there during the 22 weeks of the year outside of a university term.

- The Active People Survey data is 'weighted' to ensure the people interviewed are representative of the English population. Data 'weighting' is based on many categories (see briefing note for further information www.sportengland.org/research) including a specific weighting for the 16-24 age group who are less likely to be accessible by a telephone landline.

Does excluding mobile phone numbers from the Active People Survey affect the robustness of the results?

The main barrier for not including a sample of respondents by mobile phone within the Active People Survey sample is that of cost (it is not standard for national market research survey interviews to include a sample of mobile phones). The scale of the Active People Survey ensures the results are completely robust and credible for market research results.