Understanding Society

Working Paper Series

No. 2018 – 03

April 2018

The Longitudinal Item Count Technique: a new technique for asking sensitive questions in surveys

Alessandra Gaia

City, University of London

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Non technical summary

Social surveys often collect data on intimate and personal topics, such as, for example, sexuality, personal finances, and mental health. Collecting data on these topics is needed to answer important research questions and, ultimately, guide public policies.

For example, data on sexual orientation are used to shed light on discrimination and inequalities suffered by the gay, lesbian and bisexual population; data on sexual behaviours are used to study sexually transmissible diseases.

Collecting data on intimate and personal topics is important, but it does entail difficult challenges. When faced with a survey question that queries intimate topics, belonging to the respondents' private sphere, respondents may feel embarrassed and may be not willing to answer, or not willing to answer truthfully. This may lead to poor data quality on these important topics.

Furthermore, sensitive questions may be a threat to the success of an entire scientific study, if respondents decide to stop participating in the study after the unpleasant experience arising from being asked to disclose information that they regard as too personal.

Over the last decades, several researchers have developed techniques to collect data on intimate topics, in an attempt to lower the respondents' embarrassment and to obtain truthful answers. These techniques have several advantages, but also some limitations.

In this research I propose a new technique to ask intimate and personal questions in surveys; I apply this technique to ask questions on sexuality on a sample of respondents in Great Britain, and I evaluate the effectiveness of this technique comparing it with other methods of data collection. In order to do this test, I use data from a large scale study, *Understanding Society*: The UK Household Longitudinal Study.

The Longitudinal Item Count Technique: a new technique for asking sensitive questions in surveys

Alessandra Gaia

Abstract

The Item Count Technique is used to improve measurement of sensitive topics, reducing social desirability bias. I propose a variation of the Item Count Technique: the Longitudinal Item Count Technique (LICT). I describe the features of this innovative technique and the underlying assumptions, I provide guidance on its implementation, discuss its limitations, and the ethical implications associated with it. Furthermore, I present an empirical application of the method estimating the prevalence of the gay, lesbian, and bisexual population in the United Kingdom.

Keywords: Item Count Technique; sensitive questions; social desirability; hard-to-reach population; longitudinal data; LGBT research.

JEL classifications: C80, C81, C90.

Author contact details: alessandra.gaia@city.ac.uk

Acknowledgements: I am grateful to Noah Uhrig for contributing to the study design and to the development of the Longitudinal Item Count Technique. I am also thankful for comments from Peter Lynn and Annette Jäckle on the preliminary draft of this paper.

.

The Longitudinal Item Count Technique: a new technique for asking sensitive questions in surveys

1. Introduction

The Item Count Technique (also called "Unmatched Count Technique" or "List Experiments") is used to improve the measurement of sensitive topics, reducing social desirability bias. This is a promising technique, as it allows complete protection of respondents' privacy. However, as other indirect questioning methods, this technique does not allow to derive individual level estimates.

To overcome this limitation, I propose a variation of the Item Count Technique: the Longitudinal Item Count Technique (LICT). In this paper, I describe the features of this innovative technique and the underlying assumptions, I provide guidance on its implementation, discuss its limitations, and the ethical implications associated with it. Furthermore, I present an empirical application of the method estimating the prevalence of the gay, lesbian, and bisexual population (sexual minorities) in the United Kingdom.

This substantive topic of analysis (gay, lesbian, and bisexual populations) is chosen, for both the importance and the complexity of obtaining reliable estimates in this area. Indeed, providing sound statistical information on the lesbian gay or bisexual population (also called "sexual minorities) is needed to inform policy makers on disadvantage and discrimination. However, obtaining good quality data is methodologically challenging, as sexuality is one of the most sensitive topics in surveys; also, classification is complex as "sexual orientation" is a multidimensional construct involving three different dimensions: romantic attraction, sexual behaviour, and self-identification (Laumann *et al.*, 1994).

"Heterosexual/homosexual/bisexual attraction" measures whether a person is sexually attracted by someone of the same sex, of the opposite sex, or of both sexes. "Heterosexual/homosexual/bisexual behaviour" measures whether someone has had sexual experiences with someone of the same sex, opposite sex, or of both sexes. And sexual identity measures self-identification into "heterosexual", "homosexual", "bisexual", or "other" sexual identities.

Classification of the population could occur along any of these three dimensions (sexual attraction, behaviour, and identity) or amongst any combination of them, and it is not clear which are most relevant for population estimation much less equalities monitoring (Aspinal 2009). To the best of my knowledge, up to now, large scale multi-purpose UK studies have measured sexual identity, as self-identification into "heterosexual", "homosexual", "bisexual", or "other" sexual identities, for equality monitoring purposes.

Given the sensitive nature of the topic and the complexity of measuring sexual orientation, I consider the estimation of the all three dimensions of sexual orientation (attraction, behaviour, and identity), as an interesting case study for the first implementation of the Longitudinal Item Count Technique.

The implementation of the method is conducted using first hand experimental data from a large scale nationally representative survey of the UK population, the Innovation Panel of *Understanding Society*: the UK Household Longitudinal Study. I compare experimentally the estimates derived from the Longitudinal Item Count Technique with estimates derived from the "traditional" Item Count Technique; also, I compare these results with the estimates obtained from direct questions, administered in different modes of data collection: face-to-face Computer Assisted Personal Interview (CAPI) with showcards, and Computer Assisted Self-Interview (CASI).

This paper provides a novel contribution to the literature by describing an innovative research method, providing guidance on its implantation, and testing it empirically against other data collection techniques, i.e. direct questions implemented in different modes.

2. Research Questions and Hypothesis

This paper has two aims. First, it describes the Longitudinal Item Count Technique. Second, it tests an application of the technique to the estimation of the prevalence of the lesbian, gay, and bisexual population in the UK.

In this latter empirical part, I address the following research questions:

1. What is the estimated prevalence of: homosexual and bisexual attraction, homosexual and bisexual behaviour, and of self-identification into lesbian, gay, bisexual and "other" sexual identity obtained with an indirect questioning method, such as the Item Count questioning Technique (ICT)? And what is the estimate obtained using the Longitudinal Item Count Technique (LICT)?

2. Does face-to-face interviewing with a showcard lead to a significant different estimate of lesbian, gay, and bisexual sexual identity compared to the estimate produced with a computer administered self-interview (CASI)? And how does these two estimates compare with the estimate produced with the Item Count Technique (ICT) and the Longitudinal Item Count Technique (LICT)?

In relation to research question 2, I hypothesize that the Item Count Technique and the Longitudinal Item Count Technique lead to more accurate reporting of sexual identity than direct questioning technique. Data quality will be assessed in this case using the "more is better" approach, *i.e.* given that sexual identity is considered to be (at least in some contexts) a socially undesirable behaviour to report, it is expected that a higher estimate of the size of the gay, lesbian and bisexual population is an indicator of higher data quality.

3. Methodology – a longitudinal application of the Item Count Technique

The Item Count Technique, introduced by Droitcourt *et al.* (1991), is an indirect questioning technique to ask sensitive questions in surveys. Instead of inferring the population prevalence of a sensitive behaviour by asking respondents whether they engaged in that behaviour (as in direct questioning), through the Item Count Technique the researcher can extrapolate this information experimentally.

Specifically, the Item Count Technique works as following: sample members are randomly divided into two groups; respondents in each group are presented with a list of items and asked to count how many items apply to them. Each group's list is identical but for the sensitive item appearing only in one of them. The difference in the mean of the items counted in the two lists gives an estimate of the prevalence of the sensitive behaviour in the population.

Formally, the estimated prevalence of the sensitive item is calculated as following:

$$\hat{p}_{ICT} = \bar{x}_{\mathrm{a+s}} - \bar{x}_{\mathrm{a}} \quad (1)$$

where:

$\bar{x}_{\mathrm{a+s}}$ is the average number of items counted in list *a* plus the sensitive item;

$\bar{x}_{\mathrm{a}}$ is the average number of items counted in list *a*.

In this paper I propose a variation to the Item Count Technique: the Longitudinal Item Count Technique (LICT). Instead of splitting the sample in two groups, all respondents are presented with the list which includes the sensitive item in one survey wave and the list that does not include the sensitive item in another survey wave. The estimated prevalence of the sensitive item is calculated as the difference between the mean of the items counted in one survey wave and the mean of items counted in another survey wave. Formally, the estimated prevalence of the sensitive item is calculated as following:

$$\hat{p}_{LICT} = \bar{x}_{\mathrm{a+s,\ t}} - \bar{x}_{\mathrm{a,\ t+1}} \quad (2)$$

where:

$\bar{x}_{\mathrm{a+s,\ t}}$ is the average number of items counted in list *a* plus the sensitive item (s), at wave *t*;

$\bar{x}_{\mathrm{a,\ t+1}}$ is the average number of items counted in list *a*, at wave *t*+1.

The sensitive item can either be placed in the first of the two survey waves, or in the second. If the sensitive item is placed in the second wave, then formula (2) is adapted accordingly:

$$\hat{p}_{LICT} = \bar{x}_{\mathrm{a+s,\ t+1}} - \bar{x}_{\mathrm{a,\ t}} \quad (3)$$

Finally, and most importantly, the researcher can derive an individual level indicator of the sensitive item, by subtracting the number of items reported by the respondent in the list including the sensitive item at wave t+1, from the number of items reported by the respondent in the list that does not include the sensitive item at wave t. The formula for deriving the individual level indicator of whether the sensitive item applies to the respondent is the following:

$$\hat{p}_{LICT\,i} = x_{\mathrm{i},\ \mathrm{a}+\mathrm{s},\ \mathrm{t}+1} - x_{i,\ \mathrm{a},\ \mathrm{t}} \qquad \forall\ \mathrm{i} \in \mathrm{I}$$

where:

if $\hat{p}_{LICT\,i} = 0$ the sensitive item does not apply to the respondent,

if $\hat{p}_{LICT\,i} = 1$ the sensitive item applies to the respondent,

I is the set of respondents.

In comparison to the "traditional" Item Count Technique, the Longitudinal Item Count Technique has the advantage of producing individual level estimates of the prevalence of the sensitive item; such micro-data can be used in statistical analysis beyond the reporting of aggregate figures from the Item Count Technique.

Also, it has the advantage of not requiring to randomly assign respondents to two different groups. Randomly assigning respondents to different groups has the disadvantage that theoretically random allocation may not lead to allocation that are empirically random, especially for rare characteristics, such as belonging to a sexual minority[1].

In terms of implementation guidelines, it is advised to choose items that are relatively time invariant – e.g. items that refer to past events, like where the respondents grew up ("I have grown-up in the country-side"), dates in the past which are significant to the respondents, like birthdays of significant others ("My father's birthday is in October"), *etc*.

If the selected items are not time invariant (e.g. "I have travelled to Spain"), the event may occur between the two data collection waves. If that is the case, respondents answering the survey question accurately would report a higher number of "non-sensitive" items in the second wave compared to the first survey wave.

Variations over time in the occurrence of the non-sensitive items result in an higher counting of the "non-sensitive" items in the second wave, compared to the first wave;

[1] According to data from *Understanding Society*: the UK Household Longitudinal Study (UKHLS), wave 3, the 3% of the general population in the United Kingdom in 2011-2012 self-identify in gay, lesbian, bisexual.

this would result in an overestimate of the prevalence of the sensitive item (if the sensitive item is placed in the second survey wave) and an underestimate of the sensitive item (if the sensitive item is placed in the first survey wave). I, thus, recommend the use, as much as possible, of time invariant "non-sensitive" items.

Other guidelines for the design of the traditional Item Count Technique are relevant also for the Longitudinal Item Count Technique. A recent summary of these guidelines can be found in Glynn (2013).

In the application of the Longitudinal Item Count Technique, researchers need to consider whether an ethical approval is needed for data collection. Indeed, the Longitudinal Item Count Technique poses more challenges than the Item Count Technique from an ethical point of view. In fact, in the Item Count Technique the privacy of the respondent is fully protected: the respondents does not reveal whether the sensitive behaviour applies to him/her, and only aggregated figures are available for the researcher; nor the researchers, nor the interviewers, nor anyone else accessing the data, can derive the individual level prevalence of the sensitive item. Conversely, in the Longitudinal Item Count Technique, respondents are revealing their sensitive behaviours, and it is very likely that they are not aware of revealing them. Different ethical boards may consider differently this circumstance.

The Longitudinal Item Count Technique (LICT) can be applied also to a variation of the Item Count Technique: the Two Lists Longitudinal Item Count Technique – for a detailed description, see Droitcourt *et al.* (1991) and Tourangeau and Yan (2007), and for an application of the method, see Biemer and Brown (2005).

As in the Item Count Technique, also in the Two Lists Item Count Technique, respondents are randomly assigned to one of the two groups, but every individual receives two lists. For one group the sensitive item is included in the first list, but not the second, for the other group the sensitive item is included in the second list, but not the first (Tourangeau and Yan, 2007).

The estimated prevalence of the sensitive item in the Two List Item Count Technique can be formalised as:

$$\hat{p}_{2ICT} = (\hat{p}_{s1} + \hat{p}_{s2})/2$$

Where:

$$\hat{p}_{s1} = \bar{x}_{\mathrm{a+s}} - \bar{x}_{\mathrm{a}}$$

$$\hat{p}_{s2} = \bar{x}_{\mathrm{b+s}} - \bar{x}_{\mathrm{b}}$$

$\bar{x}_{\mathrm{a+s}}$ is the average number of items counted in list *a* plus the sensitive item;

$\bar{x}_{\mathrm{a}}$ is the average number of items counted in list *a*;

$\bar{x}_{\mathrm{b+s}}$ is the average number of items counted in list *b* plus the sensitive item;

$\bar{x}_{\mathrm{b}}$ is the average number of items counted in list *b*.

In comparison with the original Item Count Technique, the Two Lists Item Count Technique utilises the full-sample to estimate population proportions; it provides lower variance than the one list Item Count Technique, but higher respondents' burden, since respondents are asked to answer two survey questions instead of one.

The Longitudinal Item Count Technique (LICT) can be applied in conjunction with the Two Lists Item Count Technique. At the first survey wave the design of the Two Lists Longitudinal Item Count Technique is the same as the design of the Two Lists Item Count Technique. Then, in the two lists Longitudinal Item Count Technique, the design is again repeated at the subsequent survey wave. Respondents allocated to Group 1 at the first wave are again allocated to Group 1 and respondents allocated to Group 2 are again allocated to Group 2. However, the treatment that respondents receive varies across waves. Respondents allocated to Group 1 receive at the first survey wave two lists: list *a* with the addition of the sensitive item, and list *b* (Treatment 1). At the subsequent wave, respondents allocated to Group 1 receive list *a* (without the sensitive item), and list *b* with the addition of the sensitive item (Treatment 2).

Similarly, Group 2 receives at the first survey wave two lists: list *a* (without the sensitive item), and list *b* with the addition of the sensitive item (Treatment 2). At the subsequent wave, respondents allocated to Group 1 receive list *a* with the addition of the sensitive item, and list *b* without the sensitive item.

Four estimates of the prevalence of the sensitive item can be produced: one from Group 1 and list *a*, one from Group 1 and list *b*, one from Group 2 and list *a*, and one from Group 2 and list *b*. The average of these estimates provides the Two Lists Longitudinal Item Count Technique estimated prevalence of the sensitive item.

Formally, the estimated prevalence of the sensitive item in the Two Lists Longitudinal Item Count Technique can be formalised as:

$$\hat{p}_{2LICT} = (\hat{\mathrm{p}}_{\mathrm{a},\ \mathrm{Group\ 1}} + \hat{\mathrm{p}}_{\mathrm{a},\ \mathrm{Group\ 2}} + \hat{\mathrm{p}}_{\mathrm{b},\ \mathrm{Group\ 1}} + \hat{\mathrm{p}}_{\mathrm{b},\ \mathrm{Group\ 2}})/4$$

Where:

$$\hat{p}_{\mathrm{a},\ group\ 1} = \bar{x}_{\mathrm{a+s},\ \mathrm{Group\ 1},\ \mathrm{t}} - \bar{x}_{\mathrm{a},\ \mathrm{Group\ 1},\ \mathrm{t+1}}$$

$$\hat{p}_{\mathrm{a},\ group\ 2} = \bar{x}_{\mathrm{a+s},\ \mathrm{Group\ 2},\ \mathrm{t}} - \bar{x}_{\mathrm{a},\ \mathrm{Group\ 2},\ \mathrm{t+1}}$$

$$\hat{p}_{\mathrm{b},\ group\ 1} = \bar{x}_{\mathrm{b+s},\ \mathrm{Group\ 1},\ \mathrm{t+1}} - \bar{x}_{\mathrm{b},\ \mathrm{Group\ 1},\ \mathrm{t}}$$

$$\hat{p}_{\mathrm{b},\ group\ 2} = \bar{x}_{\mathrm{b+s},\ \mathrm{Group\ 2},\ \mathrm{t+1}} - \bar{x}_{\mathrm{b},\ \mathrm{Group\ 2},\ \mathrm{t}}$$

and,

$\bar{x}_{\mathrm{a+s},\ \mathrm{t+1}}$ is the average number of items counted in list *a* plus the sensitive item;

$\bar{x}_{\mathrm{a},\ \mathrm{t}}$ is the average number of items counted in list *a*;

$\bar{x}_{\mathrm{b+s}}$ is the average number of items counted in list *b* plus the sensitive item;

$\bar{x}_{\mathrm{b}}$ is the average number of items counted in list *b*.

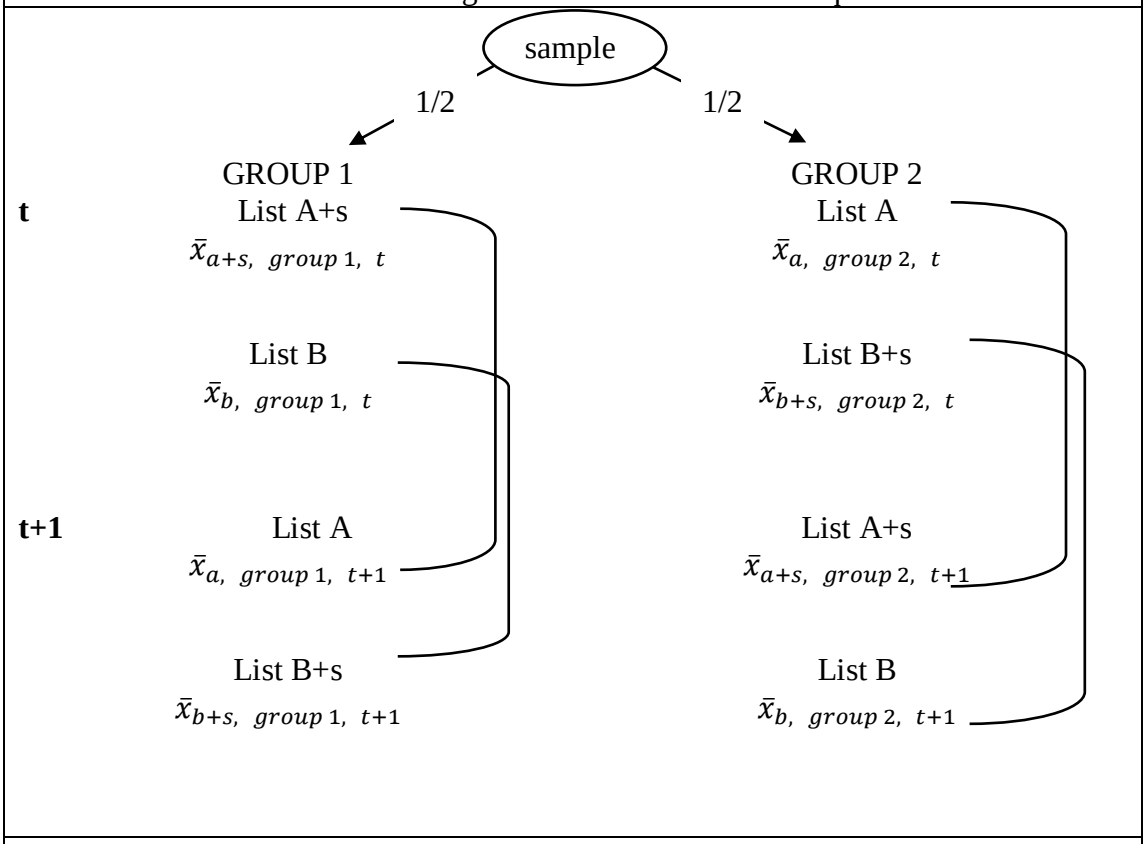
Table 1 summarises the Item Count Technique and its variations with formulas.

Table 1: The Item Count Technique and its variations

<table>
<tr><td>Item Count Technique</td></tr>
<tr><td>sample
1/2 1/2
GROUP 1
List A+s
$\bar{x}_{a+s,\ group\ 1}$
GROUP 2
List A
$\bar{x}_{a,\ group\ 2}$
$\hat{p}_{ICT} = \bar{x}_{\mathrm{a+s,\ Group\ 1}} - \bar{x}_{\mathrm{a,\ Group\ 2}}$</td></tr>
<tr><td>Longitudinal Item Count Technique</td></tr>
<tr><td>sample
t
List A+s
$\bar{x}_{a+s,\ t}$
t+1
List A
$\bar{x}_{a,\ t+1}$
Aggregate level estimate
$\hat{p}_{LICT} = \bar{x}_{\mathrm{a+s,\ t}} - \bar{x}_{\mathrm{a,\ t+1}}$
Individual level estimate
$\hat{p}_{LICT\ i} = x_{\mathrm{i,\ a+s,\ t+1}} - x_{i,\ \mathrm{a,\ t}} \quad \in \mathrm{I}$
where:
if $\hat{p}_{LICT\ i} = 0$ the sensitive item does not apply to the respondent
if $\hat{p}_{LICT\ i} = 1$ the sensitive item applies to the respondent</td></tr>
<tr><td>Tow Lists Item Count Technique</td></tr>
<tr><td>sample
1/2 1/2
GROUP 1
List A+s
$\bar{x}_{a+s,\ group\ 1}$
List B
$\bar{x}_{b,\ group\ 1}$
GROUP 2
List A
$\bar{x}_{a,\ group\ 2}$
List B+s
$\bar{x}_{b+s,\ group\ 2}$</td></tr>
<tr><td>$\hat{p}_{2ICT} = (\hat{p}_{\mathrm{a}} + \hat{\mathrm{p}}_{\mathrm{b}})/2$
Where:
$\hat{p}_a = \bar{x}_{\mathrm{a+s,\ Group\ 1}} - \bar{x}_{\mathrm{a,\ Group\ 2}}$
$\hat{p}_b = \bar{x}_{\mathrm{b+s,\ Group\ 1}} - \bar{x}_{\mathrm{b,\ Group\ 2}}$</td></tr>
</table>

Table 1 (continued)

Two Lists Longitudinal Item Count Technique



$$\hat{p}_{2LICT} = (\hat{p}_{a,\ group\,1} + \hat{p}_{a,\ group\,2} + \hat{p}_{b,\ group\,1} + \hat{p}_{b,\ group\,2})/4$$

Where:

$$\hat{p}_{a,\ group\,1} = \bar{x}_{\text{a+s, Group 1, t}} - \bar{x}_{\text{a, Group 1, t+1}}$$

$$\hat{p}_{a,\ group\,2} = \bar{x}_{\text{a+s, Group 2, t+1}} - \bar{x}_{\text{a, Group 2, t}}$$

$$\hat{p}_{b,\ group\,1} = \bar{x}_{\text{b+s, Group 1, t+1}} - \bar{x}_{\text{b, Group 1, t}}$$

$$\hat{p}_{b,\ group\,2} = \bar{x}_{\text{b+s, Group 2, t}} - \bar{x}_{\text{b, Group 2, t+1}}$$

The implementation of the Two Lists Longitudinal Item Count Technique allows the derivation of multiple estimates of the sensitive item. As shown in Table 2, by analysing the Two Lists Longitudinal Item Count Technique cross-sectionally, researchers derive two different estimates of the sensitive item (one from wave 1 and another one from wave 2). By analysing the Two Lists Longitudinal Item Count Technique, researchers obtain the Longitudinal Item Count Technique twice, once analysing the two lists between waves 1 and 2, within one group and once analysing the two lists between waves 1 and 2 in another random group.

Table 2 summarises the relationship between the Item Count Technique, the Longitudinal Item Count Technique, and the two list version Item Count Technique.

Table 2: Item Count Technique and Longitudinal Item Count Technique

	Wave 1	Wave2	
Group 1	List	List+Sensitive Item	Longitudinal Item Count Technique
Group 2	List+Sensitive Item	List	Longitudinal Item Count Technique
	Item Count Technique	Item Count Technique	

4. Data

I use data from an experiment implemented in the *Understanding Society* Innovation Panel waves 8 and 9 (IP8 and IP9)[2]. *Understanding Society: the UK Household Longitudinal Study* (UKHLS) is a multidisciplinary study that focuses on a wide range of topics such as living arrangements, fertility, housing, economic activity, income, health, and political attitudes. *Understanding Society* includes an Innovation Panel (IP). This is a separate sample used to test methodological innovations in longitudinal surveys, in general, and *Understanding Society*, in particular.

The Innovation Panel target population are adults (aged 16+) living in Great Britain. The study aim is to interview each adult member of the household and individuals are followed when they move to other parts of Great Britain. Sample members are interviewed every 12 months. The Innovation Panel mirrors *Understanding Society* in its design and it is a stratified, clustered, probability sample.

In the Innovation Panel a direct question on sexual identity was asked to all adult (aged 16 years old or older) at waves 3 in the self-completion section (computer assisted personal interviewing, CASI) of a face-to-face interview. At wave 5 the question was again asked but only to young adults (aged 16-21); in this wave the question was asked experimentally either by CASI or by web.

[2] Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service.

In IP8 and IP9 sexual orientation was measured experimentally using direct questions and using the Two Lists Longitudinal Item Count Technique. Details of the experimental design are presented in the section below.

5. Experimental design

In this paper I implement the Two Lists Longitudinal Item Count Technique to measure all three dimensions of sexual orientation (attraction, behaviour, and identity). The Two Lists Longitudinal Item Count Technique thus implies administering two lists for every elicited behaviour (sexual attraction, behaviour, and identity), thus, 6 lists in total. The lists are then repeated at the subsequent survey wave to derive the longitudinal element of the Item Count Technique. Table 3 describes the design of the Item Count Technique.

Table 3: the Two Lists Longitudinal Item Count Technique implemented at IP8 and IP9 to measure homosexual attraction, homoerotic behaviour, and sexual identity

	IP8	IP9
Group 1	List A List B + S1 List C List D + S2 List E List F + S3	List A + S1 List B List C + S2 List D List E + S3 List F
Group 2	List A + S1 List B List C + S2 List D List E + S3 List F	List A List B + S1 List C List D + S2 List E List F + S3

Note: S1 refers to being sexual attracted from someone of the same sex, S2 refers to having had homoerotic sexual experiences (sexual experiences with someone of the same sex), and S3 refers to self-identifying as gay, lesbian, or bisexual.

The question wording for the Item Count Technique is presented in Appendix 1. In the Item Count Technique, the ordering of item counts (i.e. the different lists) was randomised across respondents. The statements which respondents counted also had a randomised order across respondents.

The Innovation Panel IP8 and 9 follow an experimental sequential mixed-mode design, including web, face-to-face, and (few) telephone interviews (details of this design are presented in Jäckle *et al.*, 2017); given the complexity of the Item Count Technique question, this was not administered in the telephone interviews.

The wording of the Item Count Technique questions was designed with the aim of mixing non-sensitive items that were expected to be high prevalence with non-sensitive items that were expected to be low prevalence; this is consistent with the indication of the literature (see Glynn, 2013).

Indeed, if all items in the list are of a high prevalence, respondents belonging to the gay, lesbian, and bisexual population may count all items in the list, and thus self-identify themselves as gay, lesbians, and bisexuals; this phenomenon is called "ceiling effect"; conversely, if all "non-sensitive" items are very rare (and perceived by respondents as being more rare than belonging to the gay, lesbian and bisexual

population), respondents may fear that by counting one item, they would similarly self-identify themselves; this phenomenon is called “floor effect”.

Thus, I combined items that I expected to be low prevalence (e.g. “I would describe myself as being disabled”), with items that I expected to be high prevalence (e.g. “I would describe myself as being British”).

When items were designed, in early 2014, items: “I consider myself as being British” (list E) and “I consider myself as being European” (list F) were considered non-sensitive high prevalence items. However, the debate on the United Kingdom European Union membership (which developed in conjunction with the referendum, held on 26th June 2016) pervaded public opinion during the fieldwork for IP9 (summer 2016). It, thus, may have increased the sensitive nature of these two items, and altered the estimating prevalence of the two items at IP9, undermining the longitudinal comparison with IP8.

In addition to the Item Count Technique, also direct questions were tested. Specifically sample members were randomly allocated to two different Protocols for asking a question on sexual identity. These two Protocol are currently adopted in two large scale studies in the United Kingdom, i.e. *Understanding Society*: the UK Household Longitudinal Study (UKHLS) and the Integrated Household Survey (IHS). I will refer to these two Protocols as “UKHLS” and “IHS”.

Specifically, the two Protocol entails:

- Protocol 1 – UKHLS:

  The question is asked in self-completion either by Computer Assisted Self-Interview (CASI) or by Web.

- Protocol 2 – IHS:

  The question is asked Face-to-Face (in Computer Assisted Personal Interview, CAPI) with the aid of a showcard, or by telephone.

The question wording for the two Protocols, the showcard, and the interviewer instructions are presented in Appendix 1.

Sample members were randomly allocated to receive either Protocol 1 or Protocol 2. The experimental allocation was fully crossed with the allocation to the two lists Item Count Technique groups. Table 4 below summarises the 2x2 design.

Table 4: The 2x2 experimental design

	Group 1		Group 2	
"UKHLS" Protocol	Lists: A, B+S1, C, D+S2, E, F+S3	UKHLS direct question	Lists: A+S1, B, C+S2, D, E+S3, F	UKHLS direct question
"IHS" Protocol		IHS direct question		IHS direct question

Deviations to the experimental allocations presented in Table 4 were implemented to accommodate the mixed-mode nature of the survey design. Specifically, respondents completing the survey by Web answered the question according to the self-completion UKHLS Protocol, regardless of their original allocation. Telephone respondents (which are a small fraction of the entire sample) answered the question on sexual orientation following the IHS telephone protocol, regardless of their original allocation.

In the analysis section, I will show for descriptive purposes the prevalence of sexual orientation measured by Web interviewing; however, given that respondents self-selected into different modes, it is not possible to disentangle whether any difference in the estimates of sexual identity is due to different modes or to self-selection into modes.

If respondents answered the survey face-to-face the original randomised allocation was kept. Thus, differences in the estimated prevalence of sexual identity between respondents answering the survey question by face-to-face (CAPI) with showcards, or by self-completion (CASI) are to be attributed to mode effects. In the comparison with the Two Lists Item Count Technique, to avoid the confounding effect of selection into mode, CAPI with showcards and CASI estimates will be compared with estimates obtained from the Two Lists Longitudinal Item Count Technique, among face-to-face respondents only (excluding web and telephone respondents).

The Two Lists Item Count Technique and the direct questions random groups were allocated independently from the IP8 and IP9 mixed-mode experiments (which are described in detail in Jäckle *et al.*, 2017). The Item Count Technique questions were separated from the direct sexual identity question in the questionnaire in order to avoid carry-over effects between these survey tasks.

6. Results

Overall, sample members reacted well to the Two List Item Count Technique questions, both in IP8 and in IP9. On all questions, refusal was low, ranging from 3.4% (n=37) of respondents in IP8 on a question on sexual behaviour to <0.1% (n=6) of respondents in IP9 on a question on sexual identity. Also "don't know" answers were rare, to levels lower than 1% (n=8) in all items and waves.

One method to evaluate the quality of the Item Count Technique is to analyse the "ceiling" and "floor" effects. In the fields of attraction (lists A and B) and behaviour (lists C and D), the relative majority (over 32%) of respondents, in both waves, reported that none of the items presented applied to them; thus, I have evidence of a "floor effect"; conversely, in the identity questions (lists E and F) the "floor effect" was not problematic, as "none of the statements are true" was selected by only a tiny percentage of respondents (<5%).

The evidence on the "ceiling effect" is mixed; while lists A (attraction), list C and D (behaviour) and E (identity) resulted well designed, with only a tiny proportion of respondents selecting that all "four statements are true"; conversely, in lists B (attraction) and F (identity), the prevalence of respondents reporting that all four behaviours range between 16% and 20%, indicate that a non-ignorable fraction of respondents may have not revealed the sensitive item in the full list to avoid disclosing the sensitive attribute.

Another possible source of measurement error is that respondents may have made mistakes in counting the items in IP8 or IP9 or both, resulting in a higher number of items counted in the short list compared with the long list. While the items were

carefully designed and pre-tested[3], ex post rationalisation led to the conclusion that items might have been not sufficiently salient for the respondent to minimise measurement error.

Measurement error – either due to "ceiling" and "floor" effects or to suboptimal questionnaire design – may have influenced the estimates of the "attraction" and "identity" items, where, unexpectedly, in both IP8 and IP9 the average number of items counted is lower in the list with the sensitive item (list B+S, list E+S, list F+S), then in the list without the sensitive items – i.e. list B, list E, and list F (see Table 5 and 6).

Vice versa, and consistently with expectations, in both IP8 and IP9, in the "behavioural" questions the average number of items counted was higher in the lists which include the sensitive item (list C+S and list D+S), compared with the list that excludes the sensitive item (list C and list D). The resulting estimated prevalence of the population having had a homoerotic experience (i.e. experience of a sexual kind – for example kissing, cuddling or sexual intercourse – with a person of the same sex) is 9.9% at IP8 and 9.1% at IP9.

[3] Items were pre-tested among staff members of the Institute for Social and Economic Research of the University of Essex.

Table 5: The estimates from the Item Count Technique IP8

Average List A	average List A+S	average List A+S – average List A
1.15	1.26	0.10
average List B	average List B+S	average List B+S – average List B
1.39	1.36	N.A.
Estimated prevalence of homosexual/bisexual attraction: N.A.		
average list C	average list C+S	average list C – average list C+S
0.93	1.04	0.11
average list D	average list D+S	average list D – average list D+S
1.03	1.12	0.08
Estimated prevalence of homosexual/bisexual experience: 9.9%		
average list E	average list E+S	average list E+S – average list E
1.95	1.90	N.A.
average list F	average list F+S	average list F+S – average list F
2.56	2.43	N.A.
Estimated prevalence of homosexual/bisexual identity: N.A.		

Table 6: The estimates from the Item Count Technique IP9

Average List A	average List A+S	average List A+S – average List A
1.40	1.46	N.A.
average List B	average List B+S	average List B+S – average List B
1.77	1.54	0.23
Estimated prevalence of homosexual/bisexual attraction: N.A.		
average list C	average list C+S	average list C – average list C+S
1.17	1.11	0.06
average list D	average list D+S	average list D – average list D+S
1.35	1.23	0.12
Estimated prevalence of homosexual/bisexual experience: 9.1%		
average list E	average list E+S	average list E+S – average list E
1.90	1.35	0.56
average list F	average list F+S	average list F+S – average list F
2.43	2.56	N.A.
Estimated prevalence of homosexual/bisexual identity: N.A.		

In order to derive estimates for the Longitudinal Item Count Technique, respondents must have answered to the Item Count Technique at both survey waves; thus, I excluded from this analysis sample members that did not participated in one survey wave; also, in the analysis of each longitudinal list, I excluded those sample members that in at least one of the waves provided a "don't know" or "refusal" answer.

In the next paragraphs, I analyse the Two Lists Longitudinal Item Count Technique, starting with the experiment in which the list with the sensitive item was implemented at the first survey wave (IP8) and the list without the sensitive item was implemented at the second survey wave (IP9).

As in the Item Count Technique, also in the Longitudinal Item Count Technique some sample members reported a higher number of items in the lists not containing the sensitive items compared to the lists containing the sensitive item. This phenomenon may be due to the already recalled sources of measurement error, or to another source of error, typical of the Longitudinal Item Count Technique: the "non-time invariance of non-sensitive items". While care was taken to present time invariant items, some items were not time invariant, thus they may have occurred between IP8 and IP9, and respondents may have reported more items in IP9 (in the short list) compared to IP8 (in the long list). As a result of this error, when the list with the sensitive item (long list) was presented at IP8 and the list without the sensitive item (short list) was presented at IP9, the average number of items counted in the short list was higher than the average number of items counted in the long list, leading to unreliable estimates, in all concepts measured.

When the list including the sensitive item (long list) was presented at IP9 (rather than IP8) I don't record any measure of attraction or behaviour where the average number of items counted in the list including the sensitive item is smaller than the average number of items counted in the list that does not include the sensitive item. However, it is not possible to conclude whether this evidence is a sign of good data quality or whether it is driven from the non-time invariance of the non-sensitive items, which in this case, may have led to higher averages in the longer list compared to the short list.

The average number of items counted in the list including the sensitive item in IP9 is however smaller than the average number of items counted in the list excluding the sensitive item at IP8, in the estimates of sexual identity (List E and List F); these two latter lists include two items "I would describe myself as British" (List E) and "I would describe myself as European", which might have been of problematic implementation. Given that the survey at IP9 was fielded during the United Kingdom European Union Membership Referendum, respondents may have changed their answers from the previous wave.

Tables 7 show the Two list Item Count Technique estimated prevalence of homosexual/bisexual attraction, homosexual/bisexual behaviour, and self identification into gay, lesbian and bisexual sexual identities, when the list without the sensitive item is presented at IP8 and the list with the sensitive item is presented at IP9. The Longitudinal Item Count Technique leads to an estimate of the prevalence of sexual same sex sexual attraction in the study population of 31.9% and the prevalence of homosexual/bisexual sexual experience of 28.4%. This latter estimate is higher than the estimate obtained with the Item Count Technique, which ranged between 9.9% and 9.1%.

Table 7: The two list ICT derived estimated prevalence of estimated prevalence of homosexual/bisexual attraction, homosexual/bisexual behaviour, and self identification into gay, lesbian and bisexual sexual identities (IP8 short list and IP9 long list)

	Group 1	Group 2	average
Estimated prevalence of homosexual/bisexual attraction	0.28	0.36	0.32
Estimated prevalence of homosexual/bisexual sexual experience	0.28	0.29	0.28
Estimated prevalence of homosexual/bisexual sexual identity	N.A.	N.A.	N.A.

In the next paragraphs I analyse how the results from the Item Count Technique compare with results from direct questioning. Table 8 display the estimated prevalence of same sex sexual identity by mode of data collection in IP8 and IP9. The estimated prevalence of the population that self-identify as gay, lesbian bisexual or "other" is: 3.7% in IP8 and 4.3% in IP9[4]. At IP8, this figure does not vary by mode of data collection, when I compare web respondents with respondents participating in the survey face-to-face (both protocols, UKHLS and IHS); conversely, at IP9, a higher fraction of web respondents (4.7%) self-identify as gay, lesbian bisexual or "other" than the face-to-face respondents (4.0%) it should be noticed, however, that respondents self-select into mode, and thus it is not possible to disentangle the mode effect from the selection effect.

Given that respondents in the Face-to-Face group are randomly allocated to two different data collection protocols (i.e. the IHS Face-to-Face with showcards protocol and the UKHLS Computer Assisted Self-Interview protocol) it is possible to analyse mode effect for this subgroup of respondents. While in both waves a higher fraction of respondents self-identity themselves as gay, bisexual or "other" under the UKHLS protocol compared to the IHS protocol (5.8 *vs*. 3.7 at IP8 and 3.9% *vs*. 3.5% at IP9), differences are not statistically significant (as shown in Table 8).

[4] A small fraction of respondents were interviewed by telephone (N=29 at IP8 and N=31 at IP9); as detailed below, due to the small sample size, figures for telephone respondents are not shown separately (but the cases are included in the general population estimate).

Table 8: Estimated prevalence of self-reported sexual identity using direct questioning, all modes and protocols IP8 and IP9

	All sample			Web (UKHLS Protocol)			Face-to-Face (IHS and UKHLS Protocols)			UKHLS (CASI)			IHS (Face-to-Face with showcards)		
IP8															
	%	95% C.I.		%	95% C.I.		%	95% C.I.		%	95% C.I.		%	95% C.I.	
Heterosexual	92.8	91.8	93.9	93.3	91.6	95.1	92.5	91.1	92.5	90.8	88.6	93.0	94.0	92.3	95.7
Gay or Lesbian	1.6	1.1	2.1	1.9	0.9	2.8	1.4	0.8	1.4	1.5	0.6	2.4	1.2	0.4	2.0
Bisexual	1.4	0.9	1.8	1.6	0.8	2.5	1.2	0.6	1.2	0.9	0.2	1.6	1.5	0.6	2.4
Other	0.8	0.4	1.1	0.1	-0.1	0.4	1.1	0.6	1.1	1.5	0.6	2.4	0.8	0.2	1.5
Prefer Not to Say	2.3	1.7	2.9	2.1	1.1	3.2	2.4	1.6	2.4	5.1	3.4	6.8	N. A.	N. A.	N. A.
Refused	0.7	0.3	1.0	0.9	0.2	1.5	0.6	0.2	0.6	0.2	-0.1	0.4	0.9	0.2	1.7
Don't Know	0.5	0.2	0.8	0.0	0.0	0.0	0.8	0.3	0.8	0	0	0	1.5	0.6	2.4
Total	100			100			100			100			100		
N	2,222			793			1400			663			737		
IP9															
Heterosexual	92.8	91.8	93.9	93.3	91.6	95.1	92.5	91.1	92.5	90.0	87.3	92.7	92.0	89.7	94.4
Gay or Lesbian	1.6	1.1	2.1	1.9	0.9	2.8	1.4	0.8	1.4	1.7	0.5	2.8	1.8	0.6	2.9
Bisexual	1.4	0.9	1.8	1.6	0.8	2.5	1.2	0.6	1.2	2.7	1.3	4.2	1.6	0.5	2.6
Other	0.8	0.4	1.1	0.1	-0.1	0.4	1.1	0.6	1.1	1.5	0.4	2.5	0.4	-0.2	0.9
Prefer Not to Say	2.3	1.7	2.9	2.1	1.1	3.2	2.4	1.6	2.4	4.0	2.2	5.7	N.A.	N.A.	N.A.
Refused	0.7	0.3	1.0	0.9	0.2	1.5	0.6	0.2	0.6	0.2	-0.2	0.6	2.1	0.9	3.4
Don't Know	0.5	0.2	0.8	0.0	0.0	0.0	0.8	0.3	0.8	0	0	0	2.1	0.9	3.4
Total	100			100			100			100			100		
N	2,222			793			1400			480			513		

Note: Telephone respondents (N=29 at IP8 and N=31 at IP9) are included in columns “all sample” but not shown in subsequent columns due to small sample size; the category “prefer not to say” is not displayed in the IHS version a this was not one of the response option.

7. Discussion and Conclusions

This paper describes a new technique for collecting data on sensitive topics in surveys: the Longitudinal Item Count Technique. The method has the promising advantage of deriving an individual level indicator on whether the socially undesirable and stigmatised characteristics apply to respondents.

The paper presents, what is, to the best of my knowledge, the first application of the Longitudinal Item Count Technique. The method has been used to derive estimates of same-sex sexual attraction, behaviour and self-identification as homosexual or bisexual.

In this particular circumstance, the application of the method lead to estimates that seemed not fully reliable and not comparable with estimates from direct questioning methods. This is possibly due to measurement error in the Item Count Technique, as a result of "ceiling" and "floor" effects, suboptimal questionnaire design and non-time invariance of the items presented.

Survey practitioners are advised to consider both the standard questionnaire design recommendations as well as the advices on the design of the Item Count Technique (e.g. see Glynn, 2013) in the design of the Longitudinal Item Count Technique. In addition, the successful design of the Longitudinal Item Count Technique implies the design of non-sensitive items which are as much as possible time invariant, to avoid overestimation or underestimation of the socially undesirable item.

This paper also compares direct questions on sexual identity administered in different modes. Specifically, a question administered with Face-to-Face with showcard is compared with a question administered with Computer Assisted Self-Interviewing. The two methods do not lead to statistically significant differences in the estimate of homosexual, bisexual and "other" sexual identities.

Future research may provide further applications of the Longitudinal Item Count Technique in different context and compare the results derived with other data collection methods.

References

Aspinal, P. J. (2009). Estimating the size and composition of the lesbian, gay, and bisexual population in Britain. in *Equalities and human rights commission research report 37*. Manchester: Equalities and Human Rights Commission.

Biemer, P., & Brown, G. (2005). Model-based estimation of drug use prevalence using item count data. *Journal of Official Statistics*, *21*(2), 287.

Dalton, D. R., Wimbush, J. C., & Daily, C. M. (1994). Using the unmatched count technique (UCT) to estimate base-rates for sensitive behavior. *Personnel psychology*, 47, 817-828.

Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsley, T. L., Visscher, W., & Ezzati, T. M. (1991). The Item Count Technique as a method of indirect questioning: A review of its development and a case study application. *Measurement errors in surveys*, 185-210.

Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly* 77, 159-172.

Laumann, E. O., Gagnon, J. H., Michael, R. T. & Michaels, S. (1994). *The social organization of sexuality: sexual practices in the United States*. Chicago: University of Chicago Press.

Tourangeau, R. & Yan, T. (2007) Sensitive questions in surveys. *Psychological bulletin* 133(5), 859-883.

Jäckle, A., Gaia, A., Al Baghal, T., Burton, J. and Lynn, P. (eds) (2017). *Understanding Society The UK Household Longitudinal Study Innovation Panel, Waves 1-9, User Manual*. Colchester: University of Essex.

University of Essex. Institute for Social and Economic Research. (2017). Understanding Society: Innovation Panel, Waves 1-9, 2008-2016. [data collection]. 7th Edition. UK Data Service. SN: 6849

Appendix 1: Question wording

Item Count Technique (CASI & WEB)

Group 1

Item count list A

I have at least once been sexually **attracted** to someone who …
• has a disability
• is fit and muscular
• grew up with me in my local area
• is ten or more years older than me

How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements

Item count list B + sensitive item

I have at least once been sexually **attracted** to someone who …
• is the same sex as me
• wears the latest trends and fashions
• has a tattoo or body piercing
• is of a different ethnicity to me
• is from a different class background to me

How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements
Five statements

Sexuality item count list C

I have at least once had an **experience** of a sexual kind – for example kissing, cuddling or sexual intercourse – with a person who …
• has a disability
• is fit and muscular
• grew up with me in my local area
• is ten or more years older than me
How many statements are true for you?
None are true

One statement
Two statements
Three statements
Four statements

Item count list D + sensitive item
I have at least once had an **experience** of a sexual kind – for example kissing, cuddling or sexual intercourse – with a person who …
• is the same sex as me
• wears the latest trends and fashions
• has a tattoo or body piercing
• is of a different ethnicity to me
• is from a different class background to me
How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements
Five statements

Sexuality item count list E
I would describe myself as **being** …
• stylish and fashionable
• disabled
• patient
• British
How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements

Sexuality item count list F + sensitive item
I would describe myself as **being** …
• gay, lesbian or bisexual
• healthy
• tolerant
• European
• working class

How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements
Five statements

Group 2

Sexuality item count list A + sensitive item

I have at least once been sexually **attracted** to someone who …

• is the same sex as me
• has a disability
• is fit and muscular
• grew up with me in my local area
• is ten or more years older than me

How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements
Five statements

Sexuality item count list B

I have at least once been sexually **attracted** to someone who …

• wears the latest trends and fashions
• has a tattoo or body piercing
• is of a different ethnicity to me
• is from a different class background to me

How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements

Sexuality item count list C + sensitive item

I have at least once had an **experience** of a sexual kind – for example kissing, cuddling or sexual intercourse – with a person who …

• is the same sex as me
• has a disability
• is fit and muscular
• grew up with me in my local area
• is ten or more years older than me

How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements

Five statements

Sexuality item count list D
I have at least once had an **experience** of a sexual kind – for example kissing, cuddling or sexual intercourse – with a person who …
• wears the latest trends and fashions
• has a tattoo or body piercing
• is of a different ethnicity to me
• is from a different class background to me
How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements

Sexuality item count list E + sensitive item
I would describe myself as **being** …
• gay, lesbian or bisexual
• stylish and fashionable
• disabled
• patient
• British

How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements
Five statements

Sexuality item count list F
I would describe myself as **being** …
• healthy
• tolerant
• European
• working class

How many statements are true for you?
None are true
One statement
Two statements
Three statements
Four statements

Direct questions:

Protocol 1 – IHS

Mode: Face-to-Face with showcard

Question wording: "Which of the options on this card best describes how you think of yourself? Please just read out the number next to the description."

SHOWCARD
27. Heterosexual / Straight
21. Gay / Lesbian
24. Bisexual
29. Other

Note: "Don't Know" and "Refuse" were not displayed in the showcard. Interviewers recorded "Don't Know" and "Refuse" if those where spontaneous answers of the respondent.

Mode: Telephone
Question wording: "I will now read out a list of terms people sometimes use to describe how they think of themselves: "Heterosexual or Straight", "Gay or Lesbian", "Bisexual", or "Other". As I read the List Again please say 'yes' when you hear the option that best describes how you think of yourself.
Heterosexual or Straight
Gay or lesbian
Bisexual
Other"

Interviewer Instruction: on first reading, read list to end without pausing. Note that "heterosexual or straight" is one option "gay or lesbian" is one option. On second reading, please pause briefly after each option.

Protocol 2 – UKHLS

Mode: WEB or CASI

"Which of the following options best describes how you think of yourself?
Heterosexual or Straight
Gay or Lesbian
Bisexual
Other
Prefer not to say"