Methodology
All 'don't knows' are not created equal

Chuck Chakrapani

All 'don't knows' are not created equal
Most marketing research tables contain a line at the bottom-'Don't know/No answer'. The percentage of people who fall into this category can range from 0% to well over 50% in some cases. While a small proportion of 'don't knows' may not affect the results, a large proportion of 'don't knows' in response to a question can pose problems: Money can be wasted, or even worse, data can be misinterpreted.

Yet the issue of 'don't knows' is one of the least discussed (perhaps one of the least understood) aspect of survey responses. Although the author usually points out a large proportion of 'don't knows' are usually pointed out in the report that accompanies such tables, the implications are seldom explored. In fact, in most marketing research tables 'don't know', 'not stated' and 'refused' are generally combined, as though these categories have similar implications in all cases. The fact is that 'don't know', 'not stated' and 'refused' may or may not have similar implications, depending on the context. "Don't knows', in particular, can have widely divergent meanings depending on the type of question asked. In some contexts don't know is tantamount to refusal while in other contexts it might be exactly the answer the researcher is looking for. In yet other contexts don't know might indicate confusion in the minds of respondents.

Three different meanings
There are at least three different meanings attached to 'don't know' responses produced by a respondent, these are: *1) valid 'don't knows' 2) refusal 'don't knows' and 3) special purpose 'don't knows'*.

'Don't know' responses				
valid 'don't knows'	*deliberate 'don't knows '*	*special case 'don't knows'*		
		differences too small to identify	no real preference	undifferentiated behaviour

1. Valid 'don't knows'
'Don't know' response to a question like

Who makes Brand A beer?	
Labatts	5%
Molson	33
Other	7
Don't know	55

would indicate that the respondent genuinely does not know who manufactures Brand A. In this case, a large proportion of 'don't knows' indicates that consumers are unaware of the identity of the manufacturer.

For instance, if Brand A is manufactured by Molson, the data shows that only a third of the respondents can identify the manufacturer correctly and that more than one-half of all respondents cannot guess-even incorrectly-who the manufacturer is. This is a valid response to the question and a large proportion of 'don't knows' does not warrant changing the question or the method of obtaining the data.

2. Deliberate 'don't knows'
Consider question like this:

What is your total household income?	
Under $ 25K	20%
$25K-49.9K	16
$50K-74.9K	13
$75K-99.9K	8
$100K+	3
Don't Know/NS	40

In a case like this, it is more likely that 'don't know' answers are deliberate and that many people may not be willing to divulge their income to a stranger. In theory, then, the percentage of people earning over $100,000 cannot be assumed to be 3% but a percentage ranging anywhere between 3% and 43% (3%+40%). A similar statement can be made for any category-for instance the percentage of people earning under $25,000 cannot be assumed to be 20% but a percentage ranging anywhere between 20% and 60% (20%+40%). While valid 'don't knows' noted in the previous section gave us some valuable information, 'don't knows' in this context pose great difficulty in interpreting the results.

One common way of handling such 'don't knows' is to exclude them and re-percentage those who responded with a figure. In this case we will have

Under $ 25K	33%
$25K-49.9K	27
$50K-74.9K	22
$75K-99.9K	13
$100K+	5

(Excluding Don't Know/NS)

This procedure has, however, an implied assumption. It assumes that the proportion of 'don't knows' is the same in all income groups. This is unlikely to be a valid assumption in this case. For instance, it is quite likely that people from the upper income groups are more reluctant to reveal their income than people from the middle income groups. By excluding 'don't knows' we lose this distinction.

Here is the problem. When we include 'don't knows', we understate the percentage of people who fall into each category; we don't know by how much.

When we exclude 'don't knows', we assume a uniform non-response level in each subgroup, thereby hiding the underlying problem of interpretation. Neither is a satisfactory solution.

How do we handle this problem? There are a few alternatives some of which-although not totally satisfactory-may provide a better solution than the commonly used procedures described above.

Ascription. This procedure consists of identifying each respondent who gave a 'don't know' response to a question (income in the example above) and matching the respondent with another with similar predetermined characteristics (such as education, area of residence, gender, age etc.). The latter's income is then 'ascribed' or attributed to the former.

Behavioural matching. This procedure is somewhat less 'scientific' but when it works can give us insight into the non-responders.

Consider the data presented in Table 1. There seems to be a clear relationship between income and the use of the premium brand. The higher the income, the greater the use of the premium brand.

Table 1

	Household Income					
Brand Last Bought	*>25K*	*25-49.9K*	*50-74.9K*	*75-99.9K*	*100K+*	*DK*
Premium	19	28	32	47	68	65
Regular	81	72	68	53	32	35

The use of the premium brand among 'don't knows' is very similar to its use in the higher income groups. We can hypothesize therefore (and may have to confirm by a non-responder survey) that most non-responders to the income question come from the upper income groups. In this context, 'don't know' and 'refused' may have a similar meaning.

3. Special case 'don't knows'
Consider the following question:

Did you drink more coffee last week than the week before?	
Yes	20%
No	18
DK	62

Here the problem may be that most people may drink approximately the same quantity each week. Although they may have drunk more in one week than in another, the difference may be too small for them to note and remember Here, the researcher has to rethink the purpose of this question: Is s/he trying to capture any minor fluctuations in drinking from week to week or is s/he trying to identify a trend or the effect or an ad campaign? If the former, for example, a diary method would probably provide a better indication. If the latter, then a pre- vs. post-measure could be considered.

A variation of the theme is the following:

Which brand of beer do you prefer?		
	Regular drinkers	Casual drinkers
Brand A	20%	61
Brand B	18	17
DK	62	20

The clue to interpreting 'don't knows' in this case comes from the fact that the proportion of 'don't knows' is high both among regular and casual beer drinkers. It could mean that the two brands are not well differentiated and consumers equally like (or dislike) both brands.

Another special case of 'don't know' deals with a valid don't know that may not be of interest to the researcher. For instance, commodity type of products such as salt and sugar may have low brand awareness and loyalty.

So if asked 'Which brand of salt are you using now?' the answers may look like:

Brand A	10%
Brand B	18
DK	72

Here the 'don't knows' may or may not be valid responses depending on the objective of the question.

If the objective of the question is to find out the extent to which consumers know about brands, this is a valid way of collecting the data.

On the other hand, if the objective is to find out the extent to which consumers use each brand, it is an ineffective procedure. If the researcher wants to know which brand is used, the respondent could be asked to actually check the pantry to identify the brand.

Different ways of handling don't knows
'Don't know' responses mean different things and have different implications for data analysis depending where they occur.

All 'don't know' responses, except valid don't knows', require special treatment, unless they form a small proportion of the total responses.

Such special treatments can include data modification through ascription and behavioural matching (if 'don't knows' mean withholding information) or different ways of collecting information such as diary method (if the differences are too small to be perceived through recall techniques) or physical verification (if the product is bought as a commodity rather than as a branded product). It is quite misleading to treat all 'don't knows' the same way, without regard to the context.

It is also important to remember that 'don't know', 'refused' and 'not stated' mean the same thing in some contexts but not in other contexts.

Whenever we encounter a large number of 'don't knows', it is important to identify the nature of 'don't knows' so the results can be interpreted properly.

Table 2
How to handle 'don't knows'

Type of 'don't know'	*Example*	*Possible remedy*
Valid 'don't knows'	Who makes product x?	None needed
Deliberate 'don't knows'/ not stated	What is your household income?	Ascription Behavioural matching
Special case 'don't knows'		
1) Small differences	Did you drink more coffeethis week compared to last week?	Other methods of data collection
2) No real preference	Which brand do you prefer?	Careful interpretation/re-checking
3) Undifferentiated behaviour	Which brand of salt do you use?	If brand awareness, none needed If brand usage, use other method of data collection.

Dr Chuck Chakrapani of Standard Research Systems is a Toronto-based international consultant, author, and lecturer.

Acknowledgement: *This discussion is heavily influenced by the ideas of Professor Hans Zeisel, a man whom*

I consider to be an extremely insightful data analysts.