

# **H i C N** Households in Conflict Network

The Institute of Development Studies - at the University of Sussex - Falmer - Brighton - BN1 9RE  
[www.hicn.org](http://www.hicn.org)

## The Reliability of Perception Surveys in Afghanistan

James Khalil<sup>1</sup>

HiCN Research Design Note 15

October 2012

---

<sup>1</sup> James Khalil has designed and analysed numerous perception surveys in locations such as Afghanistan and Iraq for a variety of high profile clients such as the UK MOD, USAID, the US DOD and NATO. He has also recently completed a PhD at the University of Leeds, focussing upon the perception of political actors during the 'People's War' in Nepal. The field research included interviews with politicians and ex-combatants, and a survey with local residents in remote parts of the country.

## Introduction

Surveys are used extensively in Afghanistan to obtain perception information about support for the US-led forces, community safety, levels of corruption, female education, local health provisions, and numerous other themes. As previously observed (Khalil, 2011):

The importance of opinion polling is widely recognized, and prominent funders include security institutions (NATO), government departments (USAID, DfID), and development organizations (numerous NGOs). Due to their varied informational requirements and budgets this research may involve anywhere between a few dozen respondents from a specific community to several thousand individuals from across the country.

The surveys conducted in Afghanistan and other challenging environments tend to polarise opinions, with either too much or too little confidence often being placed in the findings. The arguments used by the sceptics centre around three main assertions that will be considered sequentially in the following sections: (a) that the responses are unreliable, (b) that the samples are unrepresentative of the populace, and (c) that the data is commonly not genuine. Thus, the aim of this article is not to describe how best to undertake perception surveys in Afghanistan *per se*, but to inform about the reliability of the findings, and how the results should be interpreted.

## The Reliability of the Responses

The first commonly-identified concern with perception surveys in Afghanistan relates to the reliability of the responses provided to the researchers.<sup>2</sup> More specifically, it is asserted that:

- Interviewees adapt their responses, either to favour the Taliban and other insurgent groups, or the Afghan Government and the coalition forces, as they fear repercussions in response to the 'wrong answer.'

---

<sup>2</sup> A earlier version of the text on response reliability appeared in Khalil (2011)

- There is a bias towards pro-government responses as individuals ‘tell researchers what they want to hear’, believing that surveys are only undertaken by institutions with loyalties to the state.
- Interviewees manipulate their responses with the objective of obtaining benefits for their community, on the basis that donors are more likely to supply water, electricity, education facilities, etc, where the shortages are deemed the most severe.

The first two of these claims are generally referred to as ‘social desirability bias’, and the most common retort is that the findings frequently place the government and the coalition in a negative light, thus supposedly providing evidence of reliability.<sup>3</sup> However, such assertions are misleading, as even if ninety-nine percent of the respondents claimed to be hostile towards the coalition there would still be a pro-coalition bias if the remaining percent falsely maintained that they were supportive. This line of argument also conveniently overlooks the converse pressure (i.e. the ‘anti-coalition bias’), which is undoubtedly driven in certain locations by the fear of insurgent retaliations.

The most common response to the third claim, that interviewees attempt to manipulate the findings to generate benefits for their community, is equally disingenuous. Specifically, it is observed that many respondents claim that there is adequate access to water, education, etc, and that provisions have improved, supposedly indicating that the responses are reliable.<sup>4</sup> However, again it is not possible to determine the proportion of dishonest replies through focusing solely upon the findings, irrespective of whether these are positive or negative, and if they improve, worsen, or remain the same. Put another way, knowing that the proportion of respondents claiming that the local health facilities were adequate increased from thirty to fifty percent between 2009 and 2012 reveals nothing about how many of the remainder lied.

Rather than denying the impact of social desirability biases and possible attempts to manipulate the research, objective pollsters and policy-makers reliant upon this data must conclude that these phenomena create an unquantifiable distortion, and that it is not possible to have absolute faith in the results. However, greater degrees of confidence can be retained with regard to the comparative conclusions based upon the (admittedly also

---

<sup>3</sup> See, for instance, Langer (2010)

<sup>4</sup> Again, see Langer (2010)

problematic) assumptions that (a) these biases are sufficiently constant between locations and over time, and (b) that the research methods are applied consistently. In other words, findings such as “53 percent of the residents from District X support female education” should be given less credence than comparative conclusions like “residents from District X are more supportive of female education than those from District Y”, or “support for female education has increased in District X over the previous twelve months.”

## The Representativeness of the Sample

A second major concern relates to the perception that the sample from which the data is drawn is unrepresentative of the wider populace, either because certain locations are outside of the reach of the pollsters, or in terms of the demographic characteristics of the respondents more generally.<sup>5</sup> The high concentration of individuals from the Hazara minority in Bamyan province, for instance, is almost entirely inaccessible during winter months. Similarly, a limited number of regions are unreachable due to insecurity, particular in insurgent strongholds such as Helmand. Ironically, it is these locations that most require research as the international community seeks to comprehend the impact of their campaign.

As a result of these limitations the sample is effectively reduced from 'the residents of Region A' to 'the residents of accessible parts of Region A', and this further undermines the absolute findings. However, comparative results over time may again retain value provided that: (a) the research methods have been consistent, and (b) that the data from locations that have on occasions been beyond the reach of the pollsters are excluded during the analysis phase. In other words, comparative conclusion may be drawn about changes in the perceptions of 'residents of the parts of Region A that have consistently been accessible.' However, the extent to which analysts apply this procedure is unclear due to a lack of transparency in the methods.

Whilst the less robust 'convenience sampling' approach remains common amongst organizations operating on a limited budget,<sup>6</sup> the research teams typically seek to gain a representative sample in the selected field locations through more robust randomization processes. Specifically, Kish Grids (or equivalent techniques such as the 'next birthday' approach) are frequently used to identify random individuals from the selected households. In the absence of such procedures the sample attains a bias towards the 'types' of individuals who are more likely to be at home during research hours, often seniors or housewives.

However, whilst many research firms retain loyalty to the Kish Grid approach, the pollsters

---

<sup>5</sup> Whilst this section focuses upon a number of the challenges associated with attaining a representative sample, it is not comprehensive. For instance, another central concern relates to the reliability of the census data used for the sample frame.

<sup>6</sup> Convenience sampling involves drawing information from respondents close to hand, often occurring in bazaars or central locations in selected towns and villages.

themselves commonly report that this often fails to achieve its objective as ‘heads-of-household’ (HOH) insist upon providing responses, either directly or indirectly. One solution is to accept this sample bias on the grounds that HOH have a far greater influence over decision-making than other family members. Indeed, a stronger version of this stance advocates the deliberate selection of HOH, despite the age and gender biases that this entails. For instance, this approach is reportedly applied by the Helmand Monitoring and Evaluation Programme (Ahmar and Kolbe, 2011).

However, unbeknown to many of survey naysayers, it is also possible to correct this and other sample biases through ‘weighting’, a technique that enlarges the unrepresented populations at the expense of those that are overrepresented. For the purposes of demonstration *Figure 1* offers sample data (non-genuine) in which respondents from *Tribe X* are less likely to be satisfied with local education facilities (21 percent), than those from *Tribe Y* (37 percent) and *Tribe Z* (40 percent). The data also indicates that *Tribe X* is underrepresented, as their sample proportion (25 percent) is lower than their population proportion (50 percent). Conversely, *Tribe Y* is overrepresented as their sample proportion (45 percent) is higher than their population proportion (20 percent).

	<b>Population in specified location</b>	<b>Survey sample</b>	<b>‘Very’ or ‘somewhat’ satisfied with local education facilities</b>
<b>Tribe X</b>	2500 (50%)	125 (25%)	21%
<b>Tribe Y</b>	1000 (20%)	225 (45%)	37%
<b>Tribe Z</b>	1500 (30%)	150 (30%)	40%
<b>Total</b>	5000 (100%)	500 (100%)	
<b>Un-weighted</b>			<b>34%</b>
<b>Weighted</b>			<b>30%</b>

Figure 1: Satisfaction with local education facilities

An un-weighted score of 34 percent is calculated through summing the sample sizes of the three tribes multiplied by their respective 'satisfaction' percentages.<sup>7</sup> The weighting process corrects the sample bias through replacing the sample sizes in this calculation with the population sizes, and this reduces the 'satisfaction' percentage to 30%.<sup>8</sup> Whilst adding considerable complexity, the weighting technique can be used simultaneously to remove multiple biases covering gender, age, ethnicity, income, occupation, and so on. However, this is no panacea given the lack of reliable demographic data upon which to base these calculations, and as the technique is inapplicable where the sample size of specific subsets of the populace is 'excessively' small. Whilst beyond the scope of this paper, the technique also complicates the statistical processes upon which the analysis is based.

### **The Reliability of the Pollsters**

The third issue commonly highlighted by individuals seeking to dismiss the findings is that the data may be faked by pollsters or research organizations aiming to avoid the associated security risks or financial costs. In discussing this theme it is helpful to distinguish between *fabricated* and *duplicated* data. The former is artificially created in the absence of interviewees, whereas the latter is copied from previous waves of the same survey, but manipulated to provide 'evidence' of trends in perceptions over time. Whilst fabricated data may be produced by individual pollsters or the local survey teams, data duplication may be performed only by those with access to the dataset.

The research organisations invariably assert that their monitoring systems limit the opportunities for fabrications. Specifically, supervisors are employed to observe a set percentage of the interviews, and these individuals also commonly contact a proportion of households to verify that the interview actually occurred and to check that the respondent provided the recorded answers (a process known as 'back-checking'). Of course, these procedures are unable to prevent collusion amongst the research teams, and neither are they able to guarantee against data duplications. Indeed, a high profile international organisation in Afghanistan relied upon data that had been duplicated between 2007 and 2009.

---

<sup>7</sup> Un-weighted satisfaction rate =  $(25\% * 21\%) + (45\% * 37\%) + (30\% * 40\%) = 34\%$

<sup>8</sup> Weighted satisfaction rate =  $(50\% * 21\%) + (20\% * 37\%) + (30\% * 40\%) = 30\%$

Thus, of perhaps even greater importance are the sanity checks undertaken during the analysis phase. These involve determining whether the expected associations occur, for instance, between (a) local perceptions of security and actual levels of violence, (b) urban residences and the use of specific media forms (specifically television), and (c) the ethnicity or tribe of the respondents and their political loyalties. Whilst it is not possible to detect low-level fraud through this process, wide-scale fabrications are revealed by a consistent failure to identify the expected correlations.

It is also possible to identify duplicated data during the analysis phase through isolating uncommon responses – e.g. claims to be from a rare tribe, or to have an infrequently large number of offspring. For instance, if there are two respondents in the sample claiming to have fourteen children it is necessary to check whether these individuals also ‘coincidentally’ share the same age and occupation, and hold identical perceptions of the coalition forces and female education. Whilst these procedures should be standard practice, the extent to which they are consistently utilized by the research firms again remains unclear due to a general lack of transparency.

## **Conclusion**

The first conclusion to be drawn from the above is that comparative findings such as “the respondents from District X are more supportive of the Karzai regime than those from District Y” are more reliable than absolute results like “34 percent of those from District X support Karzai.” The second is that the concerns frequently associated with survey research can often be overcome during the analysis phase through sanity checks and techniques such as weighting. In a market that will continue to lack adequate regulation over the foreseeable future, improvements in the surveys must revolve around adding rigour to these practices. The numerous funding bodies can assist in this process through becoming suitably informed about the practicalities of surveys, and by ensuring that the research organisations provide sufficient information about their methods. Perhaps even more importantly, surveys remain a blunt instrument through which to develop an understanding of opinions, and it remains necessary for those operating in this environment to place greater value upon the more nuanced insights delivered by qualitative research.

## Bibliography

Ahmar, Samy and Christine Kolbe (2011), 'Innovative Approach to Evaluating Interventions in Fragile and Conflict-Affected States' *Coffey International Development Working Paper*, available at [http://www.coffey.com/Uploads/Documents/HMEPWorkingPaper\\_20110530120625.pdf](http://www.coffey.com/Uploads/Documents/HMEPWorkingPaper_20110530120625.pdf)

Khalil, James (2011), 'Neither Fact nor Fiction', *Small Wars Journal*, <http://smallwarsjournal.com/jrnl/art/neither-fact-nor-fiction-the-validity-of-opinion-polling-in-afghanistan>

Langer, Gary (2010), 'Polling in Afghanistan: An Antidote to Anecdote', *ABC*, <http://blogs.abcnews.com/thenumbers/2010/01/polling-in-afghanistan-an-antidote-to-anecdote.html>