

---

# Health behaviours in people who respond to a web-based survey advertised on regional news media

Jean Adams, Martin White

**Background:** The internet has become a key tool in health research and is increasingly used for data collection via email and web-surveys. Whilst the demographics of those with and without access to the Internet in the UK are regularly reported, the health behaviours of those who choose to take part in web-based surveys, compared to the wider population, are not known. **Methods:** We compared the health behaviours of those responding to a web-based health and lifestyle survey advertised on regional TV and radio news with regional results from a large national health and lifestyle survey—the Health Survey for England 2003. **Results:** After exclusion of duplicates, 1116 individuals responded to the survey and provided information on age and sex as well as a postcode in the Government Office for the North East region. Those responding to the web-survey were younger and lived in less deprived areas than the regional population. After weighting survey responses for age and deprivation, respondents to the survey reported higher mean BMI, greater fruit and vegetable consumption, as well as differing patterns of physical activity, alcohol consumption and smoking, than regional respondents to the Health Survey for England. **Conclusions:** Conducting a web-based health and lifestyle survey advertised on regional news media was fast, cheap and relatively easy. Given the potential benefits of web-based surveys, further work is justified exploring who responds to web-based health and lifestyle surveys and whether or not more representative samples can be obtained.

**Keywords:** body mass index, diet, Internet, physical activity, smoking

---

## Introduction

The Internet has become a key tool in health research—for searching literature, accessing data, delivering health interventions and, increasingly, collecting data.<sup>1</sup> Web-based quantitative surveys have many attractions, including the speed and low cost of data collection and data entry, increased perceived anonymity and access to large numbers of widely dispersed potential respondents.<sup>1–5</sup> However, web-based surveys also suffer from substantial limitations which have, to date, restricted their use as epidemiological tools.<sup>1–3,5</sup> For example, web-based surveys rely entirely on self reported data, multiple responses from the same individual can be hard to prevent, and, as respondents are often recruited via online advertisements or posts to websites, response rates can be difficult to calculate. Perhaps most importantly, however, web-based surveys require respondents to have both access to the Internet and the necessary computing skills to complete the survey. In line with current variations in Internet use,<sup>6</sup> respondents to web-based surveys are generally younger than the population as a whole.<sup>7,8</sup> Other demographic trends in Internet access—users tend to be more likely to be male and socio-economically affluent<sup>6</sup>—are a potential source of bias but have not been consistently reflected in responses to web-based surveys.<sup>2</sup>

There is evidence that both psychological and behavioural questions asked via web-based surveys are answered similarly to the same questions asked via traditional pen and paper

questionnaires.<sup>9–11</sup> This suggests that any differences seen in web-based surveys compared to traditional questionnaires reflect differences in the people responding to the different data collection methods, rather than differences due to how the questions are asked. Whilst the demographics of those with and without access to the Internet in the UK are regularly reported,<sup>6</sup> the health behaviours of those who choose to take part in web-based surveys, compared to the wider population, are not known.

We compared the health behaviours of those responding to a web-based health and lifestyle survey advertised on regional TV and radio news with regional results from a large national health and lifestyle survey. The results will provide important comparative information for others planning web-based surveys on health topics.

## Methods

A 19 item health and lifestyle questionnaire was developed, which covered basic demographics (age—restricted to those aged 18 or over, gender and postcode) as well as a number of key health characteristics and behaviours [height and weight—used to calculate body mass index (BMI), portions of fruit and vegetables consumed in the last 24 h, 30 min sessions of moderate intensity physical activity in the last week, alcohol consumption in the last week, and cigarette smoking]. This questionnaire was converted to a web-based survey and included within the BBC web pages for the North East and Cumbria. Participants were invited to visit the website and complete the survey following live radio interviews with health experts on BBC regional radio stations in Newcastle, Cleveland and Cumbria and a report on the regional evening news (BBC Look North in the North East and Cumbria—all on the 21 June 2006). Brief mentions of the survey were also made on

**Table 1** Socio-demographic characteristics of those responding to web survey (unweighted) compared to regional population data

	Population of GONE region	Web survey	Test statistic; p-value
n(%) female	1 031 162 (52.2)	574 (51.5)	$\chi^2$ (1)=0.18; 0.67
Mean age (SD)			
Men	46.8 (17.6)	38.7 (14.5)	$t$ =12.98; <0.001
Women	48.8 (19.0)	39.0 (13.7)	$t$ =17.13; <0.001
Mean IMD 2004 <sup>a</sup> (SD)			
Men	11 557 (8452)	14 265 (8674)	$t$ =7.25; <0.001
Women	11 392 (8411)	14 110 (8450)	$t$ =7.70; <0.001

a: Rank of Index of Multiple Deprivation 2004 of super output area of residence, where 1 = most deprived.

the evening edition of BBC Look North on 22 and 27 June 2006. Although local BBC broadcasts were the main method of participant recruitment, there was nothing to stop individuals who had not heard the broadcasts and, instead, found the survey whilst browsing the Internet, from the completing the survey. The survey was 'live' for data collection between 21 June and 6 July 2006 (16 days). Where responses had the same Internet protocol (IP) address, age, sex and postcode as a response previously submitted, they were assumed to be duplicates and excluded from the analysis.

We compared the health behaviours of respondents to the web survey to those of the regional population using results from the Health Survey for England (HSE) 2003. The HSE is a large annual national survey of almost 15 000 adults in England. The majority of data is collected during a face-to-face interview but height and weight are measured directly.<sup>12</sup> The most recent year for which regional results from the HSE have been published—with key results published for participants in each of the nine local government office regions in England—is 2003.<sup>13</sup> As the government office regions do not match exactly with BBC regional news regions, we used postcodes to select for inclusion those respondents to the web-survey living in the Government Office for the North East (GONE) region (see: <http://www.gos.gov.uk/national/>).

We also used postcodes to determine the super output area (lower level) of residence of respondents and, thus, the rank of Index of Multiple Deprivation 2004 score (IMD 2004, the UK government's current preferred method of measuring deprivation)<sup>14</sup> of the super output area of residence (downloaded from: <http://www.communities.gov.uk/index.asp?id=1128440>). Ranks, rather than scores, were used throughout, as suggested in IMD 2004 documentation.<sup>14</sup>

The sex, mean age and mean of IMD 2004 rank of respondents to the web survey were compared to the population of the GONE region as a whole (aged 18 and over) using 2001 census data (downloaded from: <http://www.nomisweb.co.uk/>) and  $t$ -tests and chi-squared tests as appropriate.

Health behaviours are reported here according to the schema of published HSE 2003 regional data.<sup>13</sup> For continuous variables, where means are published for HSE 2003 regional data (BMI and fruit and vegetable consumption),  $t$ -tests were used to compare web-survey and HSE 2003 respondents. Where regional data are only reported in categories in HSE 2003 literature—physical activity, alcohol consumption and cigarette smoking—chi-squared tests were used to compare web-survey and HSE 2003 respondents.

Whilst there was no difference in the proportion of women in the web-survey compared to the regional population as a whole, both men and women responding to the web survey tended to be younger and from less deprived areas than the population of the GONE region as a whole (table 1). In order

**Table 2** Characteristics of those responding to web survey (weighted) compared to regional population data and participants in HSE 2003—continuous variables

	Population/HSE 2003 (GONE region) <sup>a</sup>	Web survey	Test statistic (t); p-value
Mean age (SD)			
Men	46.8 (17.6)	45.0 (25.8)	1.62; 0.11
Women	48.8 (19.0)	46.1 (29.9)	2.16; 0.03
Mean IMD 2004 <sup>b</sup> (SD)			
Men	11 557 (8452)	11 555 (9952)	0.01; 0.99
Women	11 392 (8411)	11 436 (12 072)	0.09; 0.93
Mean BMI <sup>c</sup> (SD)			
Men	27.0 (4.2)	29.8 (10.2)	5.69; <0.001
Women	26.9 (4.8)	29.6 (8.4)	6.34; <0.001
Mean portions of fruit and vegetables per day (SD)			
Men	2.7 (2.3)	5.2 (4.6)	10.75; <0.001
Women	3.1 (2.4)	5.6 (5.0)	10.37; <0.001

a: 2001 census data for age and Index of Multiple Deprivation 2004, Health Survey for England 2003 data for body mass index and fruit and vegetable consumption.

b: Rank of Index of Multiple Deprivation 2004 of super output area of residence, where 1 = most deprived.

c: Body mass index.

to take account of these age and deprivation differences, responses to the web-survey were weighted to reflect the age and IMD 2004 structure of the GONE region—using five age group and five deprivation groups. Sex-specific results are presented throughout.

## Results

A total of 1740 responses to the web-based survey were received. Of these, 1160 (67%) included sex, age and a postcode in the GONE region. After exclusion of duplicates (identified by the same IP address, age, sex and postcode,  $n=46$ ), 1114 responses (64%) were included in the analyses reported here. Completion rates of all information to calculate individual behaviours ranged from 88% ( $n=976$ ) for alcohol consumption (five items) to 99% ( $n=1108$ ) for BMI (two items).

Mean age and rank of IMD 2004 score of super output area of residence of those responding to the web survey, as well as for the population of the GONE region as a whole, are shown in table 2. After weighting, mean IMD 2004 rank of both men and women responding to the survey, as well as mean age of men, was not statistically different from the population of the region as a whole. However, even with weighting, female respondents to the survey were slightly younger than the regional population.

Mean BMI and portions of fruit and vegetables consumed per day for those responding to the web survey and reported for the GONE region in HSE 2003 are also shown in table 2. In both men and women, mean BMI and portions of fruit and vegetable consumed per day were greater in respondents to the web survey, than in HSE 2003 respondents.

The distribution of activity level, alcohol consumption and cigarette smoking status amongst web survey and GONE region HSE 2003 respondents are shown in table 3. There was evidence that the distribution of all of these behaviours differed significantly between the two groups—except for alcohol consumption in the past week in men. Male and female web-survey respondents were more likely to report medium activity levels (1–4 sessions of 30 min or more of moderate intensity activity per week), and less likely to report low activity levels than HSE 2003 participants in the GONE region. Female respondents to the web survey were more likely to have

**Table 3** Characteristics of those responding to web survey (weighted) compared to regional population data and participants in HSE 2003—categorical variables

	Men, n(%)		Women, n(%)	
	HSE 2003 (GONE region) <sup>a</sup>	Web survey	HSE 2003 (GONE region)	Web survey
Physical activity level (30 min sessions of moderate intensity activity per week)				
Low (<1)	132 (36)	82 (16)	166 (40)	101 (17)
Medium (1–4)	96 (26)	304 (54)	146 (35)	333 (56)
High (≥5)	140 (38)	149 (30)	104 (25)	137 (28)
Chi-squared test	$\chi^2$ (2)=92.40; $p$ <0.001		$\chi^2$ (2)=70.72; $p$ <0.001	
Alcohol consumption per week				
Yes	289 (79)	378 (80)	259 (63)	400 (75)
No	77 (21)	89 (20)	152 (37)	109 (25)
Chi-squared test	$\chi^2$ (1)=0.39; $p$ =0.53		$\chi^2$ (1)=26.36; $p$ <0.001	
Smoking status				
Current smoker	95 (26)	40 (7)	120 (29)	51 (9)
Ex-smoker	132 (36)	133 (32)	107 (26)	141 (30)
Never smoker	139 (38)	353 (61)	186 (45)	371 (61)
Chi-squared test	$\chi^2$ (2)=89.68; $p$ <0.001		$\chi^2$ (2)=72.61; $p$ <0.001	

a: Health Survey for England 2003 (Government Office for the North East region).

had an alcoholic drink in the last week compared to HSE 2003 participants from the GONE region. Male and female web-survey respondents reported never smoking more often and current or ex-smoking less often than HSE 2003 participants from the GONE region.

## Discussion

We compared the health behaviours of respondents to a web survey advertised on regional TV and radio news with regional results from a large national face-to-face interview survey (the HSE 2003), weighting web-survey responses for age and deprivation to match the regional population. There was evidence that mean BMI and portions of fruit and vegetables consumed per day by web-survey respondents were greater than estimates for the regional population. In addition, patterns of activity, alcohol consumption (in women only) and cigarette smoking differed between web-survey respondents and the regional population. Web-survey respondents reported medium activity levels more often and low activity levels less often than the regional population. Female web-survey respondents reported not drinking at all less often than the regional population. Web-survey respondents reported being current and ex-smokers less often and never-smokers more often than the regional population. Before population age and deprivation weights were applied, there was also evidence that those responding to the web survey were younger and lived in less deprived areas than the regional population as a whole.

The demographic distribution of respondents to our web-survey echo previous findings that younger individuals are more likely to take part in web surveys.<sup>7,8</sup> Previous data on the socio-economic characteristics of respondents to web surveys is conflicting,<sup>2</sup> but the patterns seen in our work reflects current Internet usage data from the UK with more affluent individuals having greater internet access than more deprived individuals.<sup>6</sup> We did not find any difference in the gender balance of respondents to the web survey compared to the population from which they were drawn and it is possible that the greater tendency for women to complete health-related surveys effectively ‘cancels out’ any preponderance of men amongst Internet users.<sup>2</sup>

Although we applied weights to take account of the differing demographic structure of respondents to the web survey compared to the regional population, these were somewhat crude and did not remove all differences—in particular, female

web-survey respondents were significantly younger than the female regional population, even after weights were applied. It, therefore, remains possible that the differences in health behaviours seen between respondents to the web-survey and participants in the HSE 2003 were due to the socio-demographic differences between the groups. It is also possible that there was response bias with those choosing to respond to the web survey having greater interest in health issues and greater health-related knowledge than the general population. Respondents to the web survey may, thus, have been more likely to know and report the ‘correct’, rather than ‘true’, answer than HSE 2003 participants. This does not explain the BMI differences seen with web-survey respondents having greater mean BMI than HSE 2003 participants. BMI is a calculated variable and the ‘correct’ weight to report for any given height is not generally publicized by health promoters in the same way that the ‘correct’ fruit and vegetable consumption or smoking status is.

The web survey was fast and cheap to produce and conduct. There were no data entry costs. However, the survey was subject to many of the limitations of web surveys that have been extensively described.<sup>1–3,5</sup> We relied on self reported data and could not prevent individuals submitting more than one response, although we did attempt to exclude duplicate responses using by excluding responses that had the same IP address, age, sex and postcode as previously submitted responses. In addition, as we were unable to determine who was exposed to the various TV and radio invitations to take part in the survey, we were unable to calculate a response rate. The characteristics of those who took part are also likely to be influenced by our method of recruitment—invitations on regional TV and radio news. Furthermore, as we did not include a question on how people found out about the survey, we were unable to distinguish between those who were directed to the survey via the TV and radio broadcasts and those who found the survey independently. As such, our results reflect those who choose to take part in web surveys from amongst those who watch and listen to regional TV and radio, or independently found the survey online, and not necessarily those from the whole population who would choose to take part in web surveys.

Although we have assumed throughout that regional HSE 2003 data gives an accurate portrayal of current health behaviours living in the GONE region, this may not be the case. Only around 5% of the population of England, and participants in HSE 2003 (around 750), live in the GONE region.

Although HSE 2003 data are weighted for non-response, more individuals from the region took part in our web survey than in the HSE 2003 and regional behavioural estimates from HSE 2003 data may not be accurate. However, HSE 2003 remains the best available data on health behaviours at a regional level in England.

It is also possible that regional behavioural patterns have changed in the 3 years between the HSE 2003 and the current work. The HSE is the largest regular study that includes comprehensive data on health and lifestyles in England. Although the full survey is annual, regional results are only reported intermittently. Thus, whilst the most recent available HSE data is from 2005, the most recent data available at the regional level are from 2003—that used here. Overall trend data from the HSE shows very small, or no, changes in the behaviours studied here between 2003 and 2005. Where changes have occurred, these are not universally in either the more or less ‘healthy’ direction—for example, whilst mean BMI in women increased from 26.7 to 26.9, mean portions of fruit and vegetables consumed per day in women also increased from 3.5 to 3.8. Such small changes over time are unlikely to change the overall pattern of results seen here.

## Conclusions

We have found that conducting a web-based health and lifestyle survey advertised on regional news media is fast, cheap and relatively easy. We collected usable data and found that respondents were more likely to be younger and live in less deprived areas than the population from which they were drawn. When weights were applied to take account of these socio-demographic differences, web-survey respondents differed from the best available data on the population in terms of BMI, fruit and vegetable consumption, physical activity, alcohol consumption (in women only) and smoking characteristics—although the differences were not universally in the more or less ‘healthy’ direction. Given the potential benefits of web-based surveys, further work is justified exploring who responds to web-based health and lifestyle surveys and whether or not more representative samples can be obtained.

## Acknowledgements

Many thanks to Eileen Kaner, Ashley Adamson, Amelia Lake and Louise Hayes (all of Institute of Health and Society, Newcastle University, UK) who provided useful comments on questionnaire content; and to Phillip McAllister of the BBC for producing the web pages. Anne Kostalis of the BBC initiated the idea for the web-survey and BBC Look North provided some funding. J.A. is supported by a UK Medical Research Council Special Training Fellowship in Health Services and Health of the Public Research. BBC Look North provided some funding towards the cost of this research.

*Conflicts of interest:* None declared.

## Key points

- Conducting a web-based health and lifestyle survey advertised on regional news media was fast, cheap and relatively easy and may be a useful way for public health practitioners to collect data.
- Respondents to the web survey differed from the population in terms of BMI and fruit and vegetable consumption, physical activity, alcohol consumption (in women only) and smoking characteristics—although not universally in the more or less ‘healthy’ direction.
- Further work is justified exploring who responds to web-based health and lifestyle surveys and whether or not more representative samples can be obtained.

## References

- 1 Eysenbach G, Wyatt J. Using the internet for surveys and health research. *J Med Internet Res* 2002;4:e13.
- 2 Rhodes S, Bowie D, Hergenrather K. Collecting behavioural data using the world wide web: considerations for researchers. *J Epidemiol Commun H* 2003;57:68–73.
- 3 Coomber R. Using the internet for survey research. *Sociological Research Online* 1997;2: <http://www.socresonline.org.uk/socresonline/2/2/2.html>.
- 4 Schleyer T, Forrest J. Methods for the design and administration of web-based surveys. *J Am Med Informatics Assoc* 2000;7:416–25.
- 5 Wyatt J. When to use web-based surveys. *J Am Med Informatics Assoc* 2000;7:426–30.
- 6 Pollard M. *First release: internet access, households and individuals*. London: National Statistics, 2006.
- 7 Ross M, Tikkanen R, Mansson S. Differences between internet samples and conventional samples of men who have sex with men: implications for research and HIV interventions. *Soc Sci Med* 2000;51:749–58.
- 8 Soetikno R, Mrad R, Pao V, Lenert L. Quality-of-life research on the internet: feasibility and potential biases in patients with ulcerative colitis. *J Am Med Informatics Assoc* 1997;4:426–35.
- 9 Buchanan T, Smith J. Using the internet for psychological research: personality testing on the world wide web. *Br J Psychol* 1999;90:125–44.
- 10 Jones R, Pitt N. Health surveys in the workplace: comparison of postal, email and world wide web methods. *Occupational Med* 1999;49:556–8.
- 11 Ritter P, Lorig K, Laurent D, Matthews K. Internet versus mailed questionnaires: a randomized comparison. *J Med Internet Res* 2004;6:e29.
- 12 Sproston K, Primatesta P, editors. *Health Survey for England 2003, Volume 3: Methodology and Documentation*. London: The Stationary Office, 2004.
- 13 Sproston K, Primatesta P, editors. *Health Survey for England 2003, Volume 2: Risk Factors for Cardiovascular Disease*. London: The Stationary Office, 2004.
- 14 Noble M, Wright G, Dibben C, et al. *The English indices of deprivation 2004*. London: Neighbourhood Renewal Unit, Office of the Deputy Prime Minister, 2004.

Received 23 May 2007, accepted 26 September 2007