

## ORIGINAL ARTICLE

## Validation of short term recall of mobile phone use for the Interphone study

M Vrijheid, E Cardis, B K Armstrong, A Auvinen, G Berg, K G Blaasaas, J Brown, M Carroll, A Chetrit, H C Christensen, I Deltour, M Feychting, G G Giles, S J Hepworth, M Hours, I Iavarone, C Johansen, L Klæboe, P Kurtio, S Lagorio, S Lönn, P A McKinney, L Montestrucq, R C Parslow, L Richardson, S Sadetzki, T Salminen, J Schüz, T Tynes, A Woodward, for the Interphone Study Group

*Occup Environ Med* 2006;**63**:237–243. doi: 10.1136/oem.2004.019281

See end of article for authors' affiliations

Correspondence to:  
Dr M Vrijheid,  
International Agency for  
Research on Cancer  
(IARC), 150, Cours Albert  
Thomas, 69372 Lyon  
Cedex 08, France;  
vrijheid@iarc.fr

Accepted 27 October 2005

**Aim:** To validate short term recall of mobile phone use within Interphone, an international collaborative case control study of tumours of the brain, acoustic nerve, and salivary glands related to mobile telephone use.

**Methods:** Mobile phone use of 672 volunteers in 11 countries was recorded by operators or through the use of software modified phones, and compared to use recalled six months later using the Interphone study questionnaire. Agreement between recalled and actual phone use was analysed using both categorical and continuous measures of number and duration of phone calls.

**Results:** Correlations between recalled and actual phone use were moderate to high (ranging from 0.5 to 0.8 across countries) and of the same order for number and duration of calls. The kappa statistic demonstrated fair to moderate agreement for both number and duration of calls (weighted kappa ranging from 0.20 to 0.60 across countries). On average, subjects underestimated the number of calls per month (geometric mean ratio of recalled to actual = 0.92, 95% CI 0.85 to 0.99), whereas duration of calls was overestimated (geometric mean ratio = 1.42, 95% CI 1.29 to 1.56). The ratio of recalled to actual use increased with level of use, showing underestimation in light users and overestimation in heavy users. There was substantial heterogeneity in this ratio between countries. Inter-individual variation was also large, and increased with level of use.

**Conclusions:** Volunteer subjects recalled their recent phone use with moderate systematic error and substantial random error. This large random error can be expected to reduce the power of the Interphone study to detect an increase in risk of brain, acoustic nerve, and parotid gland tumours with increasing mobile phone use, if one exists.

Widespread and increasing use of mobile phones over the past decade has raised concerns about their possible health effects. This has prompted a series of epidemiological studies, particularly focusing on the risk of brain tumours related to mobile phone use. Studies of brain cancer risk related to mobile phone use have used mainly the case control approach, as summarised in several recent reviews.<sup>1–2</sup> Exposure assessments in most case control studies have relied on participants' self reports of phone use, as it is usually impossible to obtain long term independent records of phone use. Very few studies have, however, attempted to validate recalled phone use.<sup>3</sup>

Interphone is an international collaborative case control study investigating whether mobile telephone use (and particularly radio frequency exposure from this use) is related to risk of tumours of the brain, acoustic nerve, and parotid gland.<sup>4–5</sup> Thirteen countries worldwide are participating following a common protocol. Exposure assessment in Interphone is primarily based on a standardised personal interview and includes a full history of mobile phone use. The validity of this method of assessing exposure is important to the interpretation of Interphone results, which are now starting to be published.<sup>6–10</sup> Validation studies, comparing recalled phone use with independently recorded data, have been carried out in most countries participating in Interphone.

Reports of two national components of the Interphone validation studies in the Northern UK and Germany, have

recently been published and show only moderate correlation between recorded and self reported phone use (correlation coefficients between 0.5 and 0.6).<sup>11–13</sup> The UK-North study, moreover, reported a high level of overreporting of both number and duration of calls (by a factor of 1.7 and 2.8 on average respectively) and a high level of variability in recall.<sup>11</sup>

Here, we report the combined findings of the Interphone validation studies in the 11 countries that have conducted the studies. The studies compare recalled phone use against phone use recorded by operators and/or software modified mobile phones.

## METHODS

Data were collected in 11 countries (table 1) following a common core protocol. In the UK-North, two separate studies were carried out, one using only software modified phones and one using only mobile phone operators' records to collect data on individual phone use. The latter followed a somewhat different protocol.<sup>11</sup>

## Selection of volunteers

In each country, between 40 and 100 volunteers were recruited to the study. In most countries (all but Australia and the UK-North operator study), volunteers were selected from colleagues and acquaintances of the investigators. It

**Abbreviation:** CAPI, computer assisted personal interview; SMP, software modified phone

**Table 1** Recalled phone use compared to actual phone use: number of calls

Country	Subjects (n)	Mean calls/month (n)	Correlation coefficient	Weighted kappa (95% CI)*	Ratio of recalled to actual number of calls		
					Geometric mean ratio (95% CI)	95% limits of agreement†	% subjects with ratio 0.5–2
Australia	44	154	0.78	0.59 (0.38–0.80)	1.06 (0.83–1.35)	0.21–5.42	61%
Denmark	44	75	0.50	0.35 (0.16–0.54)	0.80 (0.67–0.96)	0.23–2.75	73%
Finland	77	168	0.62	0.42 (0.27–0.57)	1.07 (0.90–1.27)	0.23–5.06	65%
France	65	135	0.63	0.41 (0.25–0.57)	0.90 (0.72–1.13)	0.14–5.97	59%
Germany	75	54	0.53	0.34 (0.17–0.50)	0.78 (0.60–1.02)	0.08–7.89	45%
Israel	39	351	0.59	0.34 (0.13–0.56)	0.91 (0.66–1.24)	0.12–6.83	54%
Italy	89	156	0.55	0.38 (0.24–0.52)	1.10 (0.89–1.36)	0.15–8.38	58%
New Zealand	19	68	0.79	0.29 (0.02–0.57)	0.42 (0.28–0.64)	0.06–2.77	42%
Norway	45	84	0.51	0.21 (0.06–0.36)	0.45 (0.35–0.59)	0.08–2.62	44%
Sweden	44	139	0.76	0.50 (0.30–0.70)	0.85 (0.64–1.12)	0.13–5.63	64%
UK-North SMP study	40	78	0.71	0.39 (0.18–0.60)	0.62 (0.47–0.80)	0.11–3.30	53%
UK-North operator study	91	33	0.44	0.34 (0.20–0.48)	1.61 (1.23–2.10)	0.12–21.82	54%
Total all countries	672		0.69	0.50 (0.44–0.55)	0.92 (0.85–0.99)	0.12–7.85	57%
Excluding countries not using CAPI‡	497		0.70	0.49 (0.43–0.55)	0.86 (0.79–0.94)	0.12–5.98	57%
Excluding UK-North operator study	581		0.70	0.48 (0.43–0.54)	0.84 (0.78–0.91)	0.13–6.01	57%

\*Weighted kappa statistic for quintiles of the distribution of actual number of calls.

†95% limits of agreement: mean plus/minus two standard deviations.<sup>15,16</sup>

‡Sweden and the UK are the excluded countries.

CAPI, computer assisted personal interview.

was not feasible to recruit random samples of mobile phone users in most countries because of the valuable equipment (that is, the software modified mobile phones) used in these studies. The UK-North operator study recruited volunteers through advertising in local newspapers, advertising to local council and university staff, and distribution of letters to areas profiled for their socioeconomic status.<sup>11</sup> In Australia, subjects were chosen from controls who had already taken part in the main Interphone case control study and who had expressed enthusiasm to participate in further research. In each of the countries an effort was made to select volunteers to correspond broadly to the Interphone study population with respect to age and sex.

The studies were approved by the IARC Ethical Review Committee and by the relevant ethical committees of the participating countries.

### Actual phone use

Data on actual phone use were collected from network operators in countries where this was possible (all except Sweden, Denmark, and New Zealand) and through the use of software modified phones (SMPs). Operators recorded date, time, and duration of each call made or received for periods of 3–8 months. The UK-North operator study included outgoing calls only.

SMPs are normal mobile phones whose software has been modified to record the date, starting time, and duration of each call, as well as information on power levels used for radio frequency exposure assessment purposes in Interphone. Volunteers used the SMPs with their own phone SIM card, thus maintaining their own phone numbers and subscriptions. The SMPs were used for a period of about one month by about 40 volunteers in each country. In New Zealand, 15 out of 35 volunteers who originally used the phones had to be excluded because of errors that prevented the matching of subjects to their SMP data ( $n = 7$ ) and because a number of subjects had their phones stolen ( $n = 2$ ) or moved overseas and could not be interviewed ( $n = 6$ ).

When both operator and SMP data were available for a subject, network operator data were used as the preferred measure of actual phone use as they were collected for longer periods of time; SMP data were used otherwise. Detailed comparisons of the two data sources in countries and time

periods where both were available have shown near perfect agreement.

Actual phone use was calculated as the average number of calls per month (made and received), and the average total (per person) duration of calls per month (in minutes), from operator or SMP data.

### Recalled phone use

Data on recalled phone use was collected by use of the Interphone computer assisted personal interview (CAPI),<sup>5</sup> which was administered at least six months and up to 12 months after the end of the monitoring period. The mobile phone section of the CAPI includes an “event history calendar”, structured according to factors that may change patterns of phone use. For each period of mobile phone use, more specific questions are asked concerning average duration and number of calls made or received. A shorter postal validation questionnaire was developed for use in countries that did not have the resources to interview participants in person (UK-North and Sweden). In the UK-North operator study, questionnaires were sent out directly after the end of the monitoring period.

Recalled phone use was calculated from questionnaire responses for the period overlapping the operators’ recordings or SMP use.

### Analysis

Analyses compared phone use reported by the volunteers in the questionnaires (recalled use) and phone use recorded by the operators or SMPs (actual use) for two measures of phone use: the number of calls made and received per month, and the total duration of call time per month (in minutes). Agreement between recalled and actual phone use according to these two measures was analysed on both a categorical and continuous scale.

Agreement between categories of number and duration of phone use was tested using the kappa statistic, which measures the amount of agreement between two measures beyond that expected by chance.<sup>14</sup> Kappa statistics range from  $-1$  (complete disagreement) to  $1$  (perfect agreement). A value of zero indicates the level of agreement expected by chance. Quintiles of the actual phone use distribution were used to categorise the data. Each country had data in at least

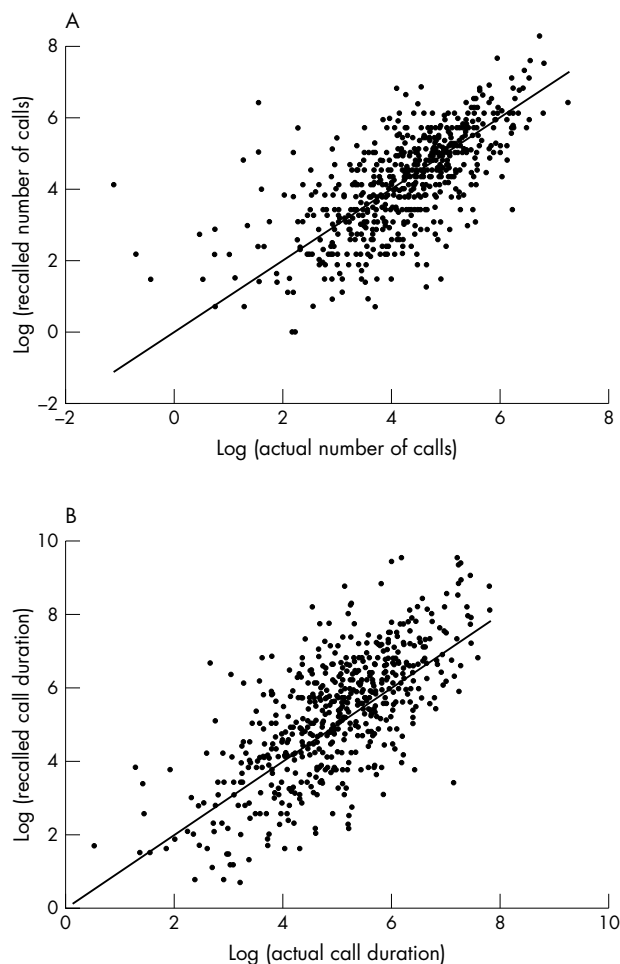
four of the five quintiles, most in all five. Weighted kappa values and their 95% confidence intervals were calculated.

On the continuous scale, all analyses used log-transformed data as both number of calls and duration of calls have very skewed distributions. Pearson correlation coefficients were calculated between recalled and actual phone use. The level of agreement between the continuous measures of phone use was measured by calculating the ratio of recalled phone use to actual phone use. The mean of this ratio represents the average level of over- or underestimation, and its spread provides a measure of the variation between individuals and therefore of the random error in recall. A ratio was of one indicates absence of bias in recall, or complete agreement. For presentation of results, the mean and confidence limits of the log-transformed data were exponentiated to the arithmetic scale and the geometric mean of the ratio was presented. Differences in the ratio of recalled to actual phone use between sexes, age groups, and other variables of interest were tested using analysis of variance and always adjusted for country. The graphical method of Bland and Altman was used to illustrate the mean ratio and its limits of agreement, and to assess the relation between the extent of over or underestimation and the level of use.<sup>15 16</sup> Following the Bland-Altman method, the ratio of the two measures (recalled to actual phone use) was plotted against their average on a log scale and the 95% limit of agreement calculated as the mean ratio plus or minus two standard deviations. Linear regression was used to examine the relation between the extent of over- or underestimation (ratio of the two measures) and level of use (average of the two measures), adjusting for country. The residuals of this regression were regressed against the level of use to examine whether variance in the error increased with increasing use.

Analyses for all countries combined were carried out both with and without the UK-North operator study as this study followed a somewhat different protocol. Additionally, analyses were carried out excluding the countries that did not use the CAPI questionnaire (UK and Sweden).

## RESULTS

A total of 672 subjects with available operator or SMP data completed questionnaires (table 1). Of these, 663 subjects completed information on call duration (table 2). Figures 1A and B show, respectively for the number and the duration of calls, phone use reported in the questionnaire against actual



**Figure 1** Scatter plot of (A) number of calls and (B) duration of calls (in minutes) reported in the questionnaire against the actual use recorded by operator or SMP (including line of equality).

phone use recorded by operator or SMP on the log scale. Correlation coefficients for number of calls (table 1) ranged from 0.44 (in UK-North operator study) to 0.79 in New Zealand, with a correlation of 0.69 for all countries combined.

**Table 2** Recalled phone use compared to actual phone use: duration of calls

Country	Subjects (n)	Mean duration (min)/month	Correlation coefficient	Weighted kappa (95% CI)*	Ratio of recalled to actual call duration		
					Geometric mean ratio (95% CI)	95% limits of agreement†	% subjects with ratio 0.5–2
Australia	44	305	0.81	0.54 (0.34–0.75)	1.39 (1.04–1.85)	0.20–9.65	45%
Denmark	44	255	0.55	0.40 (0.21–0.59)	1.34 (1.01–1.78)	0.19–9.25	48%
Finland	76	910	0.56	0.35 (0.20–0.51)	1.53 (1.19–1.79)	0.16–14.30	43%
France	64	656	0.50	0.40 (0.24–0.56)	1.45 (1.09–1.92)	0.14–14.45	45%
Germany	75	187	0.52	0.27 (0.12–0.42)	1.03 (0.74–1.44)	0.05–19.89	33%
Israel	38	2015	0.57	0.28 (0.04–0.52)	1.43 (0.93–2.19)	0.10–21.24	39%
Italy	89	1071	0.65	0.39 (0.26–0.51)	2.18 (1.71–2.77)	0.21–22.47	40%
New Zealand	20	104	0.77	0.56 (0.26–0.86)	0.70 (0.48–1.04)	0.12–4.10	40%
Norway	45	193	0.54	0.33 (0.15–0.50)	0.56 (0.40–0.78)	0.06–5.67	40%
Sweden	43	514	0.70	0.37 (0.17–0.56)	1.46 (1.03–2.08)	0.14–15.35	35%
UK-North SMP study	39	285	0.80	0.63 (0.41–0.85)	1.06 (0.78–1.43)	0.16–7.12	69%
UK-North operator study	86	206	0.58	0.35 (0.22–0.48)	2.47 (1.84–3.31)	0.15–39.45	36%
Total all countries	663		0.69	0.49 (0.44–0.55)	1.42 (1.29–1.56)	0.12–17.37	42%
Excluding countries not using CAPI‡	495		0.70	0.48 (0.41–0.54)	1.31 (1.18–1.46)	0.11–15.47	41%
Excluding UK-North operator study	557		0.70	0.48 (0.42–0.54)	1.31 (1.18–1.44)	0.12–14.73	43%

\*Weighted kappa statistic for quintiles of the distribution of actual number of calls.

†95% limits of agreement: mean plus/minus two standard deviations.<sup>15 16</sup>

‡Sweden and the UK are the excluded countries.

**Table 3** Recalled versus actual phone use by categories of phone use for number of calls

Actual number of calls/month	Recalled number of calls/month					Total
	<30	30-60	60-100	100-175	>175	
<30	101	12	13	6	4	136
30-60	64	28	26	11	6	135
60-100	33	19	38	26	19	135
100-175	12	11	33	34	42	132
>175	6	4	14	25	85	134
Total	216	74	124	102	156	672

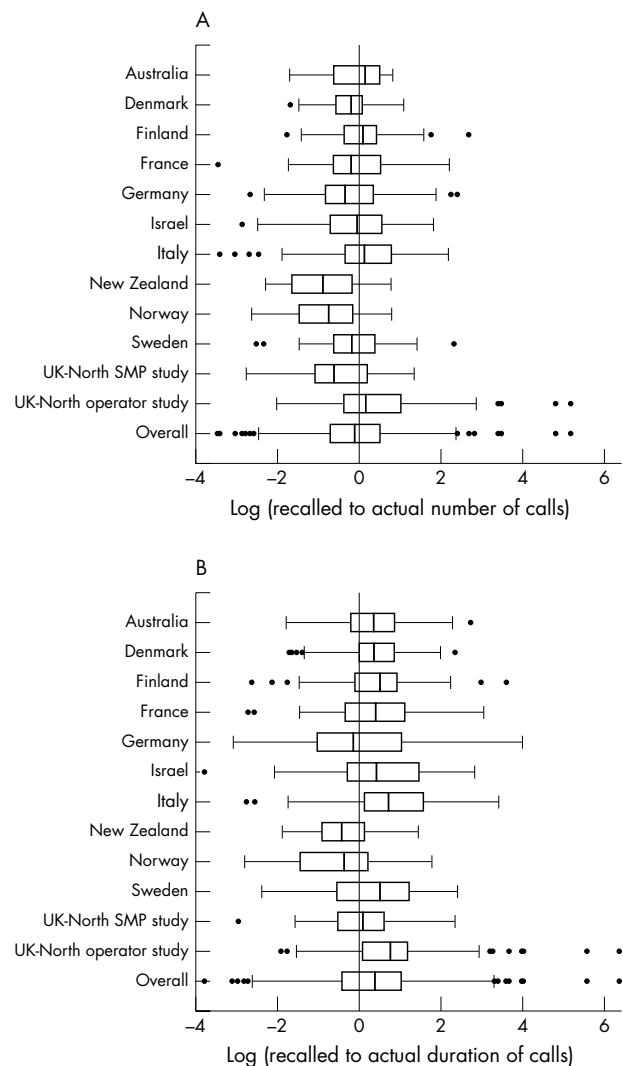
For duration of calls (table 2) correlations were similar and ranged from 0.50 (in France) to 0.81 (in Australia) with an overall correlation of 0.69.

There was agreement between quintiles of recalled and actual phone use in 43% of subjects for the number of calls and 41% for duration (tables 3 and 4). Disagreement of two quintiles or more was found in 20% of subjects for number of calls and 21% for duration of calls. The weighted kappa statistics for quintiles of number of calls (table 1) ranged from 0.21 in New Zealand to 0.59 in Australia; for all countries combined the kappa was 0.50. For duration of calls (table 2) kappa values ranged from 0.27 in Germany to 0.63 in the UK operator study, with a value of 0.49 for all countries combined.

For number of calls, the average level of over- or underestimation, as measured by the ratio of recalled to actual number of calls, ranged across countries from underestimation by a factor of 2.4 (geometric mean ratio = 0.42) in New Zealand, to overestimation by a factor of 1.6 (ratio = 1.61) in the UK-North operator study (table 1). Box plots (fig 2A) show the distribution of ratios in each country and demonstrate very wide variation between individuals. Differences between countries in the mean ratio were statistically significant ( $p < 0.0001$ ). For all countries combined, the ratio for number of calls was 0.92 (95% CI 0.85 to 0.99), indicating 9% underestimation on average. The mean ratio decreased slightly when restricting to countries that used CAPI (ratio = 0.86), or when excluding the UK-North operator study (ratio = 0.84). The 95% limits of agreement for the mean ratio of 0.92 ranged from 0.12 to 7.85 (table 1, fig 3A), indicating very wide variation between volunteers from around eight times under to around eight times overestimation. Table 1 includes a column with the 95% limits of agreement in each country; the widest interval is seen in the UK-North operator study (0.12 to 21.82). Also shown in Table 1 is the percentage of subjects who had a ratio of recalled to actual phone use between 0.5 and 2—that is, subjects who under- or overestimated their phone use by less than a factor of 2. This percentage ranges from 42% in New Zealand to 73% in Denmark and is 57% overall.

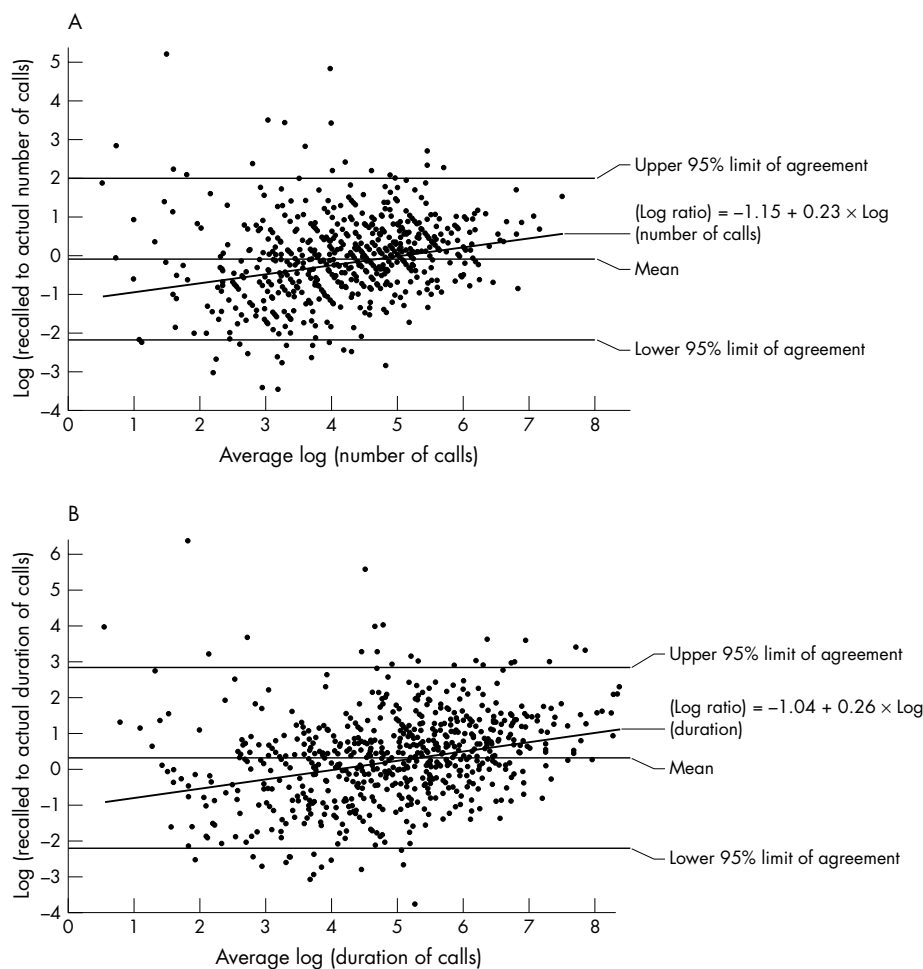
**Table 4** Recalled versus actual phone use by categories of phone use for duration of calls

Actual duration/month (minutes)	Recalled duration/month (minutes)					Total
	<43	43-102	102-186	186-377	>377	
<43	78	37	6	4	9	134
43-102	35	33	28	22	16	134
102-186	13	25	26	37	32	133
186-377	5	12	18	32	64	131
>377	1	5	6	15	104	131
Total	132	112	84	110	225	663

**Figure 2** Box plots of (log) ratio of recalled to actual phone use in each country for (A) number of calls and (B) duration of calls; with box limits indicating the 25th and 75th percentiles, line indicating the median, adjacent lines and whiskers indicating the adjacent values, and dots indicating outlying values.

Linear regression showed an increase in the ratio of recalled to actual numbers of calls as the average of the two increased (fig 3A, regression coefficient 0.23,  $p < 0.0001$ ). This regression equation predicts a ratio of 0.59 at the 10th percentile of use (15 calls/month) and a ratio of 1.13 at the 90th percentile (257 calls/month). Variance did not increase with level of use on the log scale ( $p = 0.999$ ), implying that variance increased on the arithmetic scale. Regression analyses in each country separately demonstrated positive regression slopes, indicating increasing ratios with increasing numbers of calls, in all but two countries (UK-North operator study and Australia).

The ratio of recalled to actual duration of calls was larger on average than that for number of calls (table 2). It ranged from 0.56 in Norway (that is, underestimation by a factor 1.8) to 2.47 in the UK-North operator study (that is, overestimation by a factor 2.5). In most countries (all but Norway and New Zealand), subjects tended to overestimate their call duration. Box plots (fig 2B) demonstrate the distribution of ratios in each country. Differences between countries were highly significant ( $p < 0.0001$ ). For all countries combined, duration of calls was overestimated by 42% (ratio = 1.42, 95% CI 1.29 to 1.56). A ratio of 1.31 was found



**Figure 3** Bland-Altman plot: ratio of recalled to actual phone use against mean phone use (log-transformed data) with lines indicating the mean ratio, the 95% limits of agreement, and the regression line of ratio against the mean; for (A) number of calls and (B) duration of calls (minutes).

when restricting to CAPI using countries, or when excluding the UK-North operator study. Individual variation was larger for duration of calls than for number of calls; the limits of agreement around the ratio ranged from 0.12 to 17.37 for all countries combined (table 2, fig 3B) and this interval narrowed only slightly when the UK-North operator operators' study was excluded. Only 42% of subjects under or overestimated their total duration of calls by less than a factor 2 (table 2).

The ratio of recalled to actual duration of use increased with increasing duration of phone use (fig 3B, regression coefficient 0.26,  $p < 0.0001$ ). This regression predicts a ratio of 0.78 at the 10th percentile of duration (21 min/month) and a ratio of 2.00 at the 90th percentile (788 min/month). Again, variance in the (log) ratio did not increase with level of use ( $p = 0.999$ ), meaning that variance on the arithmetic scale did increase. Similar to the analyses of number of calls, analyses in each country separately indicated increasing ratios with increasing duration of calls in all countries except the UK-North operator study and Australia.

In CAPI, duration of calls is reported either by giving the average duration of a call, or by estimating the total duration of calls per day, week, or month. Subjects who reported their phone use per call showed greater levels of overestimation of call duration (ratio = 1.47; 95% CI 1.31 to 1.64) than those who reported per day/week/month (ratio = 0.88; 95% CI 0.73 to 1.07). No important differences in under- or overestimation of number or duration of calls were found between different sex and age groups, or between short term ( $\leq 1$  year) and long term ( $> 5$  years) mobile phone users.

**DISCUSSION**

This study compares recalled mobile telephone use with assumed accurate data recorded by network operators or SMPs, thereby providing information essential to the interpretation of studies of mobile phone use in which exposure assessments are based on recall. The findings give important indications of the level of systematic and random error associated with recalling of phone use.

Correlations between recalled and actual phone use were moderate to high, ranging from 0.5 to 0.8 in the different countries. They are generally somewhat lower than those reported in a US study of recall compared to billing records (0.74).<sup>3</sup> These correlations do not give information on the level of numerical agreement between the two measures.

Kappa statistics give a measure of agreement between categories of a variable. There are no absolute standards for interpreting kappa values.<sup>17</sup> Values below 0.2 have been suggested to represent slight agreement, values of 0.2–0.4 fair agreement, values of 0.4–0.6 moderate agreement, and values of over 0.6 good agreement.<sup>18</sup> The kappa values reported in this study (between 0.2 and 0.6 across countries) can thus be described as fair to moderate. It is common practice in epidemiological studies of mobile phone use to categorise continuous measures of phone use, such as cumulative number of calls and cumulative duration of calls, using quintiles (or other percentiles) of their distribution.<sup>6–10</sup> This study shows that, using quintile categories, the exposure of up to 60% of subjects may be misclassified, albeit that most (40%) are misclassified to the adjacent quintile.

On a continuous scale, levels of under- and overreporting in recalled phone use found in this study show that, even though the systematic error in recalled phone use was relatively small on average, variation between subjects was very large and a substantial proportion of subjects considerably over- or underestimated their phone use. This indicates a large random error in recalled phone use. Random errors in recall of phone use can have an important impact on risk estimates based on them. Non-differential random errors usually bias risk estimates for dichotomous and continuous exposures and trend effects for ordered polytomous exposures towards the null (no effect), and increase their uncertainty, making it more likely that real associations are not detected.<sup>19</sup> Such errors do not normally induce spurious associations if in fact no association exists between exposure and outcome. However, for comparisons of exposure categories within polytomous exposures, bias can be in either direction.<sup>19–22</sup> Systematic under- or overestimation of phone use leads to a bias of risk estimates upwards or downwards: if all subjects overestimate, measured relative risk estimates will be lower than the true relative risk, and vice versa. Differential recall errors in cases and controls may also lead to a bias, the direction of which depends on the direction of the differences between cases and controls.

The errors estimated in the current validation studies have been used in simulations to assess their impact on tumour risk estimates. These simulations show that when random errors are large (of the level found in these validation studies) they have a large impact, biasing risk estimates for continuous exposure towards a null effect (Vrijheid *et al*, unpublished data). The simulations also show that random errors have a larger impact on the risk estimates than do systematic errors, even when relatively extreme systematic errors are modelled and when the systematic errors simulated differ between cases and controls.

That error in recall of phone use varied between countries requires some consideration. We find a wide range of effects: both for number of calls and duration of calls the average error ranges from underestimation by a factor of about two in some countries to overestimation by a similar factor in others. The high levels of underestimation found in New Zealand and Norway are notable; there are no clear differences in the protocol that may explain them, although differences between convenience samples may affect comparability between countries. New Zealand and Norway have some of the lowest average levels of use, possibly explaining the higher levels of underestimation. The UK-North operator study shows a higher level of overestimation than the other studies. It used a somewhat different protocol: in particular, the study recruited subjects through advertisement, was limited to outgoing phone calls, and used a short postal questionnaire administered directly after the monitoring period.<sup>11</sup> Other countries that used a postal questionnaire (Sweden and the UK-North SMP study) did not show similar levels of overestimation however. It is particularly notable that the two validation studies carried out in the UK at different time periods and with different protocols, but in the same region, showed different levels of error. The main difference between the two studies is the method used to record actual phone use (operators *v* SMPs). Although these methods record the same data (date, time, and duration of each call), the use of the SMP may have influenced a subject's use during the monitoring period. However, when subjects who reported a change of use in the monitoring period were excluded, the results did not change. The UK-North operator study was also based on relatively small numbers of calls as the study was carried out in 2000, when phone use was lower than in more recent years.<sup>11</sup> In all other countries, tendencies for a small underestimation of number of calls and for a larger overestimation of duration of calls are found nevertheless.

We find that the level of systematic error is related to the level of phone use. For both number and duration of calls, it appears that light users underestimated and heavy users overestimated their mobile phone use. This tendency was found in almost all countries. The random error in recall also increased with increasing use. These effects may have important implications for the exposure-response analysis in Interphone. In particular, it is expected that larger overestimation and random error in heavy users will lead to an underestimation of tumour risk in this group of users if these errors are non-differential (that is, not related to case or control status).

The results also suggest that the level of overestimation is related to the option chosen to report call duration (per call or per time period). This is not surprising for two reasons. Firstly, for those who reported duration per call, the total duration of calls was calculated from two values (number of calls and duration per call), both of which are subject to error. For those who reported duration per day, week, or month, only one value was estimated. Secondly, the questionnaire gave only two options for duration of calls: minutes or hours. Subjects who made many short calls of less than a minute might have overestimated their duration by having to answer in minutes per call.

These validation studies were generally carried out with healthy volunteer subjects who may not have the same age, sex, and socioeconomic status as cases and controls in Interphone, and who may have been more motivated to give accurate accounts of their phone use. We also only tested recall over a six month period, whereas historical phone use extending years or decades in the past is an important component of exposure assessment in the Interphone study. These differences might suggest there will be less accurate recall in Interphone subjects *per se*, although longer term recall was found to be more accurate than short term recall in one study in France (Hours, personal communication). Volunteers in the validation studies were generally younger than Interphone cases and controls. Analyses by age, however, do not indicate appreciable differences in the agreement by levels of this variable. The validation studies in Australia used volunteers from the whole Interphone control group. They had already completed CAPI once, but before the period of SMP use or operator recording. Their recall error was similar to that of all countries combined, and their correlation coefficient and kappa measure of agreement were among the highest in the different countries. In order to address questions regarding the long term recall of Interphone cases and controls, billing records of Interphone subjects are also being collected from phone service providers. This is only possible in a few countries, but will provide important insights into the possibility of differential recall between brain tumour cases and healthy controls.

In summary, validation study subjects recalled relatively recent phone use with moderate systematic error and substantial random error. These large random errors are likely to lead to important reductions in the power of the Interphone study to detect an increase in risk if one exists. Results of this study, together with those of the simulation studies and of validation studies using billing records of Interphone subjects, will play an important role in interpretation of the results of the Interphone study and of similar studies in which exposure assessments are based on recall of mobile phone use.

## ACKNOWLEDGEMENTS

The authors thank the volunteers and the network operators in all countries, as well as the SMP manufacturers, for their participation. Emilie Combalot and Monika Moissonnier assisted in the preparation of the data at IARC. Angus Cook (University of Western Australia) contributed to the design of the validation study protocol and questionnaire, and Cara Marshall coordinated the distribution of the SMPs in New Zealand.

## Main messages

- On average, volunteers tend to slightly underestimate the number of mobile phone calls they make, but overestimate total duration of calls by about 40%.
- There is large variation between individuals in the level of under- or overestimation of mobile phone use and variation increases with level of use.
- The random errors quantified in this study are likely to reduce the power of the Interphone study to detect an increase in risk if one exists.

## Policy implications

- These findings will play an important role in the interpretation of results of studies of mobile phone use in which exposure assessments are based on recall.

## Authors' affiliations

**M Vrijheid, E Cardis, I Deltour, L Richardson**, International Agency for Research on Cancer, Lyon, France

**B K Armstrong, J Brown, M Carroll**, School of Public Health, The University of Sydney, Sydney, Australia

**A Auvinen, T Salminen**, School of Public Health, University of Tampere, Tampere, Finland

**A Auvinen, P Kurtio, T Salminen**, STUK – Radiation and Nuclear Safety Authority, Helsinki, Finland

**G Berg**, School of Public Health, University of Bielefeld, Bielefeld, Germany

**K G Blaasaas**, Institute of Epidemiology, Norwegian Armed Forces, Oslo, Norway

**A Chetrit, S Sadetzki**, The Cancer and Radiation Epidemiology Unit, The Gertner Institute for Epidemiology and Health Policy Research, Chaim Sheba Medical Centre, Tel-Hashomer, Israel

**H C Christensen, C Johansen**, Institute of Cancer Epidemiology, The Danish Cancer Society, Copenhagen, Denmark

**M Feychting, S Lönn**, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

**G G Giles**, Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Australia

**S J Hepworth, P A McKinney, R C Parslow**, Centre for Epidemiology and Biostatistics, University of Leeds, Leeds, UK

**M Hours, L Montestrucq**, Unité Mixte de Recherche Épidémiologique et de surveillance Transport Travail Environnement (INRETS), Université Lyon 1-InVS, Lyon, France

**I Iavarone**, Department of Environment and Primary Prevention, Istituto Superiore di Sanità, Rome, Italy

**L Klæboe**, The Cancer Registry of Norway, Oslo, Norway

**S Lagorio**, National Centre for Epidemiology, Surveillance and Health Promotion, Istituto Superiore di Sanità, Rome, Italy

**S Sadetzki**, The Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

**J Schüz**, Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), University of Mainz, Mainz, Germany

**T Tynes**, Norwegian Radiation Protection Authorities, Østerås, Norway

**A Woodward**, School of Population Health, University of Auckland, New Zealand

Funding sources: we acknowledge funding from the European Union Fifth Framework Program, "Quality of Life and Management of living Resources" (contract QLK4-CT-1999-01563), the International Union against Cancer (UICC), and national funding sources. The UICC received funds for this purpose from the Mobile Manufacturers' Forum and GSM Association. Provision of funds to the Interphone study investigators via the UICC was governed by agreements that guaranteed Interphone's complete scientific independence. These agreements are publicly available at <http://www.iarc.fr/pageroot/UNITS/RCA4.html>. Funding sources for the national validation studies included: Australia: National

Health and Medical Research Council (EME Grant 219129), Bruce Armstrong is supported by a programme grant from the University of Sydney Medical Foundation; Finland: Emil Aaltonen Foundation, TEKES National Technology Agency, and Academy of Finland; France: French Association for Research on Cancer (ARC); Germany: Ministry for Environment, Agriculture and Consumer protection of North Rhine Westphalia; New Zealand: Health Research Council of New Zealand, New Zealand Cancer Society, Waikato Medical Research Foundation. UK: Department of Health, contract ref RRX51.

Competing interests: as stated above, the UICC received funding for Interphone from the Mobile Manufacturers' Forum and the GSM Association, and Interphone's complete scientific independence from these funders was guaranteed. The University of Leeds (authors: SJH, RCP, PAMcK) and the University of Lyon (author MH) received some financial support for the main Interphone study (not the validation study), from, respectively, the UK Network Operators (O2, Orange, T-Mobile, Vodafone, '3') and French Network Operators (Orange, SFR, Bouygues) under legal signed contractual agreements which ensure complete independence for the scientific investigators. These funders had no involvement in the study design, data collection, statistical analysis or interpretation of the data, report writing, or decision to submit this paper. Individual authors have no competing interests to declare.

The other members of the Interphone Study Group include: Canada: Professor Daniel Krewski, Dr Mary McBride, Dr Marie-Elise Parent, and Professor Jack Siemiatycki; Japan: Dr Naohito Yamaguchi; UK: Professor Anthony Swerdlow.

## REFERENCES

- 1 **Habash RWY**, Brodsky LM, Leiss W, *et al*. Health risks of electromagnetic fields. Part II: Evaluation and assessment of radio frequency radiation. *Crit Rev Biomed Eng* 2003;**31**:197–254.
- 2 **Boice JD**, MacLaughlin JK. Epidemiologic studies of cellular telephones and cancer risk. Stockholm Sweden: Swedish Radiation Protection Authority, 2002 (Publication no 2002, 16).
- 3 **Funch DP**, Rothman KJ, Loughlin JE, *et al*. Utility of telephone company records for epidemiologic studies of cellular telephones. *Epidemiol* 1996;**7**:299–302.
- 4 **Cardis E**, Kilkenny M. International case-control study of adult brain, head and neck tumours: results of the feasibility study. *Radiat Prot Dosimetry* 1999;**83**:179–83.
- 5 **Cardis E**, Kilkenny M. INTERPHONE International Case Control Study of Tumours of the Brain and Salivary Glands, Protocol rev 1, IARC Internal Report 01/002. Lyon: International Agency for Research on Cancer, 2001.
- 6 **Christensen HC**, Schüz J, Kosteljanetz M, *et al*. Cellular telephone use and risk of acoustic neuroma. *Am J Epidemiol* 2004;**159**:277–83.
- 7 **Lönn S**, Ahlbom A, Hall P, *et al*. Mobile phone use and risk of acoustic neuroma. *Epidemiology* 2004;**15**:653–9.
- 8 **Christensen HC**, Schüz J, Kosteljanetz M, *et al*. Cellular telephone use and risk for brain tumors. *Neurology* 2005;**64**:1189–95.
- 9 **Lönn S**, Ahlbom A, Hall P, *et al*. Long-term mobile phone use and brain tumor risk. *Am J Epidemiol* 2005;**161**:526–35.
- 10 **Schoemaker MJ**, Swerdlow AJ, Ahlbom A, *et al*. Mobile phone use and risk of acoustic neuroma: results of the Interphone case-control study in five North European countries. *Br J Cancer*, 2005; advance online publication; [www.bjccancer.com](http://www.bjccancer.com); doi:10.1038/sj.bjc.6602764.
- 11 **Parslow RC**, Hepworth SJ, McKinney PA. Recall of past use of mobile phone handsets. *Radiat Prot Dosimetry* 2003;**106**:233–40.
- 12 **Samkange-Zeeb F**, Berg G, Blettner M. Validation of self-reported cellular phone use. *J Expo Anal Environ Epidemiol* 2004;**14**:245–8.
- 13 **Berg G**, Schüz J, Samkange-Zeeb F, *et al*. Assessment of radiofrequency exposure from cellular telephone daily use in an epidemiological study: German Validation study of the international case-control study of cancers of the brain—INTERPHONE-Study. *J Expo Anal Environ Epidemiol* 2005;**15**:217–24.
- 14 **Cohen J**. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;**20**:37–46.
- 15 **Bland JM**, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *The Lancet* 1995;**346**:1085–7.
- 16 **Bland JM**, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;**8**:135–60.
- 17 **Altman DG**. Practical statistics for medical research. London: Chapman and Hall, 1995:396–439.
- 18 **Landis JR**, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
- 19 **Armstrong BG**. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med* 1998;**55**:651–6.
- 20 **Dosemeci M**, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 1990;**132**:746–8.
- 21 **Birkett NJ**. Effect of nondifferential misclassification on estimates of odds ratios with multiple levels of exposure. *Am J Epidemiol* 1992;**136**:356–62.
- 22 **Dosemeci M**, Stewart PA. Recommendations for reducing the effects of exposure misclassification on relative risk estimates. *Occup Hyg* 1996;**3**:169–76.