cellphone use, including the number of SMS texts they sent daily, or weekly or monthly. They could estimate a range, if preferred. Many students chose to estimate the extent of their texting for all three periods. This led, serendipitously, to the analysis presented in this paper.

During data entry we noticed a common tendency for individuals' weekly and monthly estimates to be very low in comparison with their daily texting estimates. For instance, one participant estimated 10 daily, 35 weekly and 150 monthly, and another recalled 20 daily, 50 weekly and 150 monthly. At first, we thought this might reflect poor arithmetic skills, but one teacher informed us that the class selected for participation was a top-stream one: students' science grades all exceeded 85%. Despite this, they showed the same estimation tendency. Consequently, we explored the literature on magnitude estimation.

Magnitude estimation is a basic area of research, currently considered in the field of neuroscience. In 1834, Weber observed what change in weight was needed for the person lifting it to notice. He realised that 'the extent to which two stimuli can be discriminated is determined by their ratio'.[6] Fechner developed this theory, 'postulating that the external stimulus is scaled into a logarithmic internal representation of sensation'.[7] These concepts came to be called the Weber-Fechner law whereby linear change in sensation (S) is proportional to the logarithm of the stimulus magnitude (m): $S=k \times \log(m)$, where k is a constant. It has been shown to apply generally to the way our senses perceive environmental stimuli (eg, light intensity, volume and length). Over the last few decades, research has suggested that a logarithmic mental number line is also consistent with the estimation of observed numerical quantity (referred to as the numerosity).[8] [9]

Here, we explore our data for indicators of the mental process behind estimating a number of past events—specifically, the extent of cellphone texting and cordless calls made weekly. We checked whether the consistent, but unexpected, tendencies we had observed in participants' texting estimates were explained by the Weber-Fechner law. We sought to find explanations for commonly observed features of recall and use these to improve analytical procedure in epidemiological risk analyses which use numerical recall data. Results based on such data provide indications of public health risk from environmental exposures or medical interventions, therefore it is important to minimise bias in the analytical methods and resulting inferences.

## METHODS

The methodology evolved during examination of the data. The analysis was undertaken using data from our cross-sectional survey of New Zealand adolescents' wireless phone habits. The study population has been described previously.[10] Briefly, it was representative of the region for school type and decile (socioeconomic ranking of schools by their area), and included the capital city through to rural areas. Years 7 and 8 students (N=373; 207 male, 165 female and 1 transgender) from around the region participated. The median age was 12.3 years. There was an 85% participation rate. Ethical approval was given by the Victoria University of Wellington Human Ethics Committee. Informed consent was obtained from principals of participating schools and parents of participating students. Students could choose to opt out.

## Primary-independent variables

We examined the following variables: recalled and billed weekly texts sent, pairs of recalled and billed weekly texts sent from those on 500 and 2000/month plans, and the estimated number of cordless phone calls made weekly.

Participants retrieved their remaining text balance on their prepaid monthly plan from their provider. This allowed us to calculate their daily actual use ('billed') pro rata by dividing the used portion by the number of days since billing, and multiplying this by 7 for the weekly rate.

## Statistical analyses
### Distribution of the estimation data

We considered two aspects of the distribution of the estimation data. First, that of the estimates themselves, overall and within each order of magnitude, which could reasonably be expected to reflect the distribution of actual use. Second, that of the leading digits, which we would expect to be randomly and uniformly distributed if the mental processes involved in recollection were linear. Analyses were undertaken using the statistical programmes SPSS V.17.0.1, Chicago, Illinois, 2008, and Microsoft Excel, 2010.

The distribution of estimated and billed weekly texts sent was examined with cumulative distribution plots using raw and log-transformed data. These raw text data were plotted on three-dimensional column graphs (for the orders 1–9, 10–99 and 100–999; second order of each at figure 1). This was to enable us to examine the nature and extent of rounding within each order of magnitude, and the distribution of data across each order. We calculated the extent of rounding to fives/tens and fifties/hundreds in the second and third orders of magnitude, respectively. The percentage of datapoints in the lower 31.6% of each order of magnitude was calculated, 31.6% being the half-way point on a logarithmic scale for 10 (base 10; geometric mean $(1,10)=\sqrt{10}=3.16$). Regression plots were used to assess 'daily' versus 'weekly' and 'billed' versus 'estimated' texts sent. We checked whether there was a tendency towards overestimation or underestimation with increasing numerosity (in the texting data) by regressing the difference of the log-recalled and log-billed against the log-billed. The explanation for this variation to the Bland and Altman approach[11] is given elsewhere.[12]

## a Distribution of # recalled texts sent 10-99
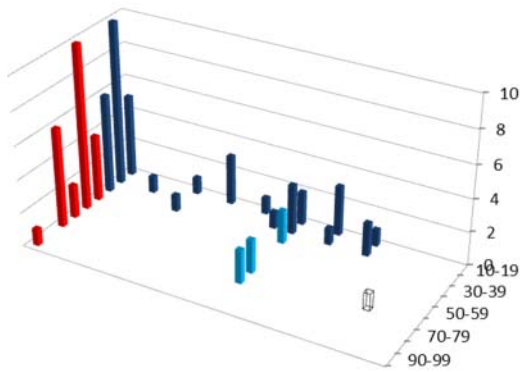


NAVY: estimates in lower 31.6% of order
RED: rounded estimates in top 68.4% of order
Blue: rounded estimates in top 68.4% of order
Black outline: unrounded estimates in top 68.4% of order

## b Distribution of # of billed texts sent 10-99
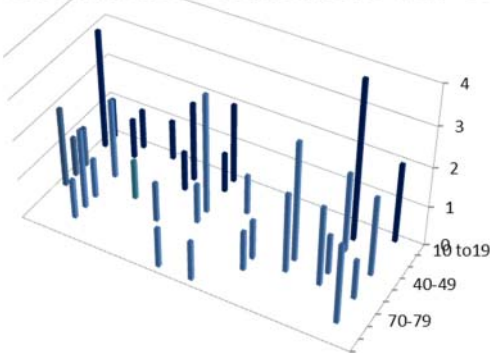


**Figure 1** (A) Distribution of weekly texting estimation data (second order): 61% of estimates fell in the lower 35% of the order, and there was a strong rounding effect. There were only three unrounded estimates in the upper 65% of all orders (1 in the second order). (B) Distribution of weekly billed texts (second order) shows a homogeneous distribution despite the overall data being log normally distributed; 36% of estimates fell in the lower 35% of the order. All specific (ie, non-range) estimates are shown, with columns representing the number of participants who gave each estimate. Read from the back-left across each row, working forward in rows.

We assigned the geometric mean to responses given as a range (explanation below); these were included in overall distribution reporting but excluded from digit analysis in this paper as we focused on specific estimates when exploring the mental process of estimation.

Valid zeros were included in reporting the overall distributions, but not in the calculations of mean and SD of the log-transformed data.

### Distribution of the first digits in estimation data

We assessed first-digit distribution in estimates of weekly texts sent and cordless phone calls made. For comparison, we did the same for a set of random numbers drawn from the same distribution. We began by removing all data given as a range, and all estimates of zero. We sorted those remaining into nine groups, one for each digit from 1 to 9. Each was then allocated into two groups: those with only a single non-zero leading digit (eg, 2, 20 and 200) and the remaining estimates starting with that digit (in the example case 2). These were displayed as stacked columns, with each two-part column representing the percentage of estimates starting with that digit (figure 2).

## RESULTS
### Descriptive statistics

At least one cellphone was owned by 285 (76%) of participants, while 331 (89%) currently used one. Most
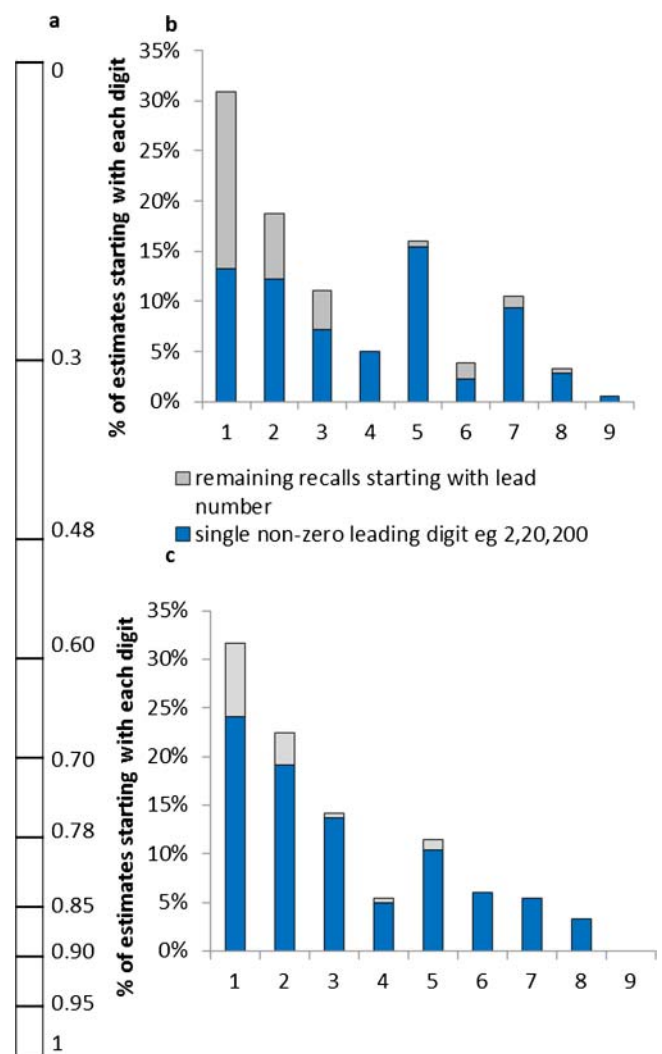


**Figure 2** (A) Marks the distribution of 'tenths' from 0 to 1 on a log scale (equivalent of 1–10 on a linear scale). (B) Distribution of leading digits of participants' estimated number of texts sent weekly, n = 181, range 1–1800, and (C) cordless calls made weekly, n = 183, range 1–150. The columns add up to 100% of specific estimates made. All columns are split into participants' estimates with single non-zero digits (eg, 2, 20 and 200) and the remaining ones for each leading digit (eg, 23, 25 and 270).

**Table 1** Texting rates and percentage of texting estimations in the lower 31.6%* of each order

| | Daily | Weekly | Monthly |
|---|---|---|---|
| **Estimated number of texts sent over different periods†** | | | |
| Total N | 248 | 240 | 247 |
| n (%) who provided a range | 66 (27) | 51 (21) | 55 (22) |
| Mean of estimated texts sent | 37.04 | 146.90 | 643.44‡ |
| **Percentage of estimated and billed weekly texts in lower 31.6%* of each order of magnitude** | | | |
| | n | Estimated (%) | n | Billed (%) |
| First order (0–9) | 40 | 50 | 18 | 72 |
| Second order (10–99) | 71 | 48 | 55 | 33 |
| Third order (100–999) | 74 | 58 | 75 | 64 |

The second order of magnitude is most relevant as there are no outside influences on the distribution. It is not clear whether first-order values are estimated on a mental linear or logarithmic scale, and the third order is influenced by the group who had only 500 texts available monthly: their weekly estimates will fall in the lower half of the order, and are more likely to be less than about 150.
*This represents the half-way point on a logarithmic scale.
†Includes data given as a range with the geometric mean applied.
‡617.60 with top outlier excluded.

participants had a cordless phone at home which they used (341, 91%). We retrieved paired estimated and billed texting data from 108 participants (38% of cell-phone owners). Other relevant descriptive statistics can be seen in table 1.

### Overall distribution of estimation data
Recalled estimates of recent texts sent were right skewed. The variance of estimates increased by a fixed ratio with increasing estimated numerosity. Once the data were log transformed, the regression of estimated daily-to-weekly texts became linear (Pearson's r 0.91 p<0.01; figure 3), showing a systematic tendency to underestimate use over a week compared to that estimated for a day. The log estimated to billed texts (Pearson's r 0.78 p<0.01) revealed a large, but homogeneous, variance of the residuals of log-to-log regression (random error).

Log-transformed data from all those who sent texts followed a normal distribution (not shown), while the influence of a plan with a known prepaid quantity of texts monthly (500 or 2000) appeared to have a calibrating effect on daily and weekly estimates. This was evident in each plan's data, which had a distribution closer to exponential, that is, $f(b) \cong (1/\mu) \exp(-b/\mu)$ where $\mu$ was the population mean use, estimated by the sample mean. The mean of estimated texts sent weekly for the 2000/month plan fell within the 95% CI of four times that of those with the 500/month plan.

Two types of systematic error existed in recall. The first resulted in a trend significantly different from zero, moving from overestimation by those who sent few texts towards underestimation by those who sent many (figure 4). The second systematic error was apparent when comparing recalled texts sent over different periods (figure 3). The ratios of individual recall (daily: weekly and daily:monthly) were both only a little over half that expected (0.58 and 0.54), while that of weekly: monthly was 0.90. This applied, both between and within participants, in data which ranged from 0 to >1000.

### Distribution of estimated and billed texts within each order
About half or more of participants' estimates fell in the lower 31.6% of each order of magnitude (table 1). This represents the half-way point on a logarithmic scale. The billed data was homogeneously spread (illustrated for

**Figure 3** Regression of participants' weekly-to-daily text estimates (log-transformed data). The best-fit line indicates that on average weekly use is underestimated compared to estimates of daily use. For instance, on average, estimates of 10 or 100 daily were allotted 40 or 340 weekly, respectively (blue gridlines).
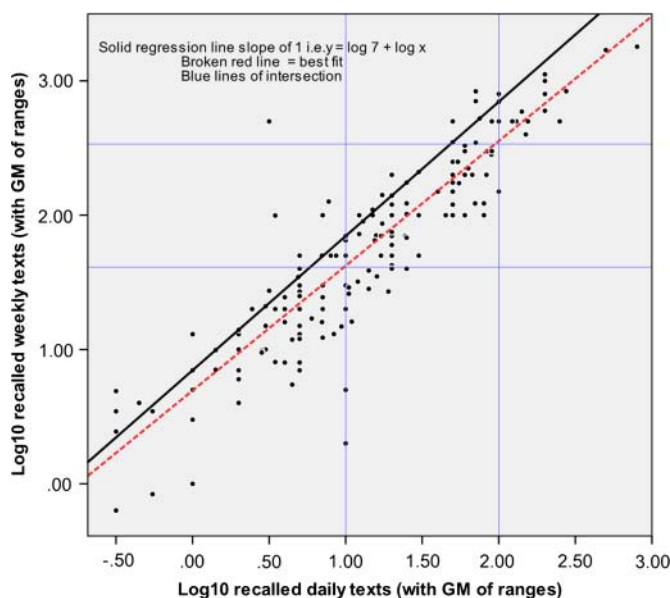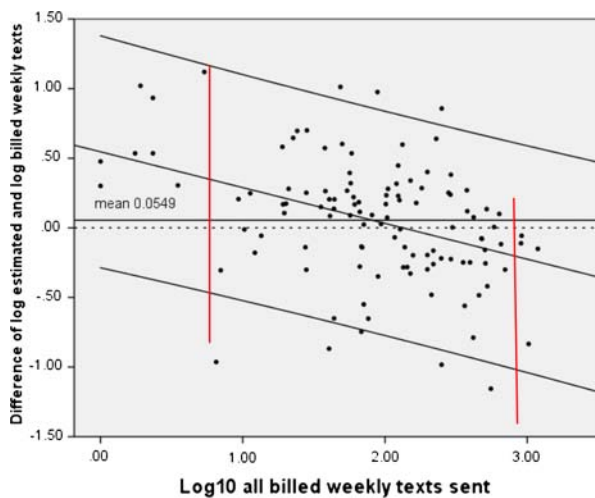
**Figure 4** Bland and Altman plot displaying the difference of the logged estimation and billed weekly texting data against the log-billed. Accurate estimates would all fall on the dotted line. There is a clear and significant trend from overestimation of little use to underestimation of extensive use. All lowest and highest estimates to the left and right of the red lines were too high/too low, respectively.

the second order of each at figure 1). The second order of magnitude (10–99) is most relevant for comparison as there were no outside influences on its distribution.

### Distribution of digits and rounding effect
The leading digits of texting and cordless phone-call estimation data were distributed very unevenly, with proportions of each digit from 1 to 9 resembling those of the intervals on a log scale (figure 2A).

There were several rounding effects. Responses became less specific as numerosity increased within *each order* of magnitude, and as the leading digit increased from 1 to 9 (figure 2B,C). Distribution of leading digits showed that 123 (68%) of weekly texting responses were rounded by participants to a single non-zero digit, as were 158 (86%) of cordless call responses.

There was an additional rounding effect to final digits of 5 and 50 in the upper 68.5% of the second and third orders, respectively; these are visible as bright blue columns in figure 1. Only three texting responses (5%) greater than 35 or 350 in the second and third orders, respectively, were not rounded thus, being 68, 525 and 839; for cordless phone calls, only 2 (15%) were not rounded (being 53 and 59).

We can only hypothesise about the spike in estimates starting with five and seven (figure 2B). An excess of leading fives is probably related to the rounding effect and is shown in fewer leading fours. (The same applies to almost no nines in preference for rounding to a final zero.) However, the excess of leading sevens may reflect a more linear approach from a quarter of those estimating 1, 10 or 100 texts daily, whose weekly estimates were seven

times greater. This explanation is supported by there being no excess of leading sevens for the estimated number of cordless phone calls weekly (figure 2C).

Although 50% of first-order texting estimates fell in the lower 31.6% (table 1), we could not resolve whether the mental process for estimating very low numerosity is better described as linear or logarithmic, but speculate that the first order of magnitude is transitional towards the latter.

### DISCUSSION
We report for the first time that the way numerosity of recent events (specifically cellphone use) is recalled conforms to the Weber-Fechner law. In other words, there appears to be a mental logarithmic scale consistent with that found in the estimation of observed numerosity. This provides a new direction for understanding human magnitude estimation, as, rather than a mental representation of an environmental stimulus, it is the outcome of an internally generated (ie, recalled) stimulus.

Let us examine the evidence for this. Texting estimation data were very unevenly distributed, but with strong similarities in each order. First, the majority of estimates fell in the lower 31.6% of each order, possibly related to a mental logarithmic scale, as well as consistent with the estimations accurately representing the log normal or exponential distribution of the billed data. Second, there was a strong rounding effect; data were almost exclusively rounded in the upper 68.5% of each order, reflecting a logarithmic mental estimation scale. This is clearly visible in figures 1 and 2. Further, the pattern of leading digits in the estimation data did not match that of the leading digits of random numbers drawn from this distribution. This only occurred (in the first digit after the decimal) after log transformation.

If estimation were carried out linearly for data which, overall, formed a log normal distribution, then we might expect more than half of all estimates to be *evenly distributed* through the lower 31.6% of the full range 1–1000, with the balance being *evenly distributed* through the remainder. This is what we saw in the billed data (figure 1).

The neuroscience literature describes a numerical magnitude effect: 'discrimination of two numerosities of a given numerical distance becomes more difficult as the absolute values of the two sets get higher' (ref.[13], p.4). Our data show that this applies *within each order*. There needed to be an appreciable-imagined difference (stimulus) in numerosity on a log scale for it to be acknowledged in the resulting estimate. This is evidenced in the rounding effect within orders. Testing of visual estimation of numerosity has generally been limited to the first two orders of magnitude (1–9 and 10–99) so rounding appears only to have been commented upon by Krueger[14] who reported 89% of estimates being rounded to a last digit of 5 or 0 when participants were shown arrays of Xs numbering 25–300.

If the mental estimation process were linear we would expect all leading digits to be equally represented, but their distribution closely resembled the intervals of a logarithmic scale. This also applied to the distribution of leading digits in recalled cordless call data. Integers with single non-zero digits were vastly over-represented (figure 2). Looking to the remaining digits in estimations, these were also far from being evenly distributed. The rounding effect was so strong that estimates in the top 65% of the second order were almost exclusively rounded to tens or fives (figure 1), and in the third order to hundreds or fifties. These effects are all consistent with estimation on a logarithmic mental scale.

Most of the phenomena we have reported are consistent with the estimation of *observed* numerosity, but estimation of *recalled* numbers of events over recent months has not previously been reported in the cognitive science literature. Estimation of observed numerosity is one of the several foci of magnitude estimation. When these ratio-based estimations are log-transformed they become linear.[14] This mental process reflects logarithmically compressed number-neurons operating like a slide-rule by ensuring accuracy proportional to the size of the numbers being processed,[7] thus maximising neuronal efficiency. In humans, this neuronal activity has been traced to the horizontal segment of the intraparietal sulcus.[15] It has been suggested that this logarithmic method of weighing the comparative value to ascribe to a large numerosity may be 'deeply embedded' as the default method in humans,[8] a prelinguistic in-born approach to number.[16]

The logarithmic mental process has been shown to result in increasing numerosity progressively being assigned proportionally lower comparative values, with high numbers commonly underestimated.[6] This applied to our data, that of Inyang *et al*[4] and to the CEFALO study.[17]

Hollingsworth *et al*[9] reported the same tendency in a psychological test resulting in mean overestimation of an array of <130 dots and underestimation of large arrays up to 650 dots. Several cellphone studies have found the opposite tendency, with high values overestimated. Since much of the literature on magnitude estimation has adult participants, we doubt that this 'reverse' trend is a feature of age, but suggest that it may result from the elapsed period since that being recalled, as cellphone studies often ask participants to recall their phone use over periods up to 10 years. The Interphone study reported greater over-reporting in this situation.[2]

Psychological studies of observed numerosity-estimation have resulted in the hypothesis of a consistent variance of the residuals once the data are log-transformed[6] thus providing a common probabilistic range at any given point on the line. We found this applied to recalled numerosity, as has been reported in other cellphone studies.[1 18] However, recalled estimation has an important difference from the visual estimation process as the variance of the residuals in recall estimation reported in this and other cellphone studies is routinely much wider than when numerosity is observed. It appears that this is a function of recall, introducing greater random error.

## Implications for epidemiology

Our findings have implications for other cellphone studies and other epidemiological studies involving recalled numbers of events. A high proportion of rounded estimates could affect categorisation. Specifically, if quantile-cuts occur at round numbers (particularly those starting with 1 and 5), there may be many same-value digits. Forming cut-points before or after these would form irregularly sized quantiles. Arbitrarily allotting same values to different quantiles is not viable as it would return different results when analysed against other variables depending on how the dataset was ordered prior to categorisation. This would be true independent of sample size.

The mental process of estimation affects how given ranges of data should be averaged. The geometric rather than arithmetic mean is likely to align better with single value estimates as this is equivalent to averaging the logarithms of the values and back transforming. It would thus avoid introducing bias which would occur by mixing specific estimates made on a logarithmic scale with the arithmetic mean of a range, which is appropriate for a linear process. The geometric mean would also be better when imputing missing central data between two provided estimates. Typically in cellphone research, these situations have been allocated the arithmetic mean or median.[19–21] An example from our study of the possible outcome being strongly affected is when the range is wide and starts at a low number, for instance, 1–70. Here the arithmetic mean is 35.5, while the geometric mean is 8.37. A quarter of all weekly text estimates were provided as a range. Recording their geometric means instead of arithmetic means resulted in the mean of all the data being 10% lower.

There is some evidence from the cognitive neuroscience literature that it may be possible to reduce recall inaccuracy by providing a calibration point.[6] Variability in our study was smaller where participants knew the monthly maximum available on their account compared to those with no account. This is also applied to two Interphone studies[1 22] where location questions may have acted as contextual prompts.[18] Variability was considerably broader in the MoRPhEUS study[4] where no prompts were given, and in the UK Interphone validation study[3] that was conducted by postal questionnaire. The possible beneficial influence of a calibration point suggests that supplying participants in case–control studies with an accurate record of their recent cellphone use may allow them to better judge their earlier levels of use. This could be tested in further research.

In summary, *recalled* numerosity of recent events appears to be processed in the brain in a very similar way as is

*observed* numerosity. This finding extends the cognitive science literature on estimation of numerical quantity, and lends some predictability to epidemiological studies involving recalled numerosity: Numerical recall estimated on a logarithmic mental scale means that as numerosity increases, estimations reduce comparatively. This trend from overestimation to better estimation or underestimation in recall of the extent of recent events is of great importance for epidemiology, as is the large variance in the residuals of recalled data. If these aspects are not allowed for during analysis, it may introduce error or bias, leading to overestimation or underestimation of relative risk for those with extremes of cellphone use. Bias or error may also be introduced as the high incidence of rounding could affect categorisation.

We offer some solutions. First, the rounding effect and a logarithmic mental process imply that recalled numbers should be log-transformed prior to analysis. This is usual, but our study provides empirical justification. Second, recalled number ranges and imputed missed data between given estimates are better represented by the geometric rather than arithmetic mean. Third, informing study participants of their correct current level of use over a short period may improve estimation of use over somewhat longer periods. These steps should help reduce random and systematic bias in cellphone studies, but we anticipate that they will also be applicable to other research which relies on recalled estimations of recent numbers of events.

## REFERENCES

1. Vrijheid M, Cardis E, Armstrong BK, *et al*. Validation of short term recall of mobile phone use for the interphone study. *Occup Environ Med* 2006;63:237–43.
2. Vrijheid M, Armstrong B, Bedard D, *et al*. Recall bias in the assessment of exposure to mobile phones. *J Expo Sci Environ Epidemiol.* 2009;19:369–81. Epub 21 May 2008.
3. Parslow RC, Hepworth SJ, McKinney PA. Recall of past use of mobile phone handsets. *Radiat Prot Dosimetry* 2003;106:233–40. Epub 24 December 2003.
4. Inyang I, Benke G, Morrissey JJ, *et al*. How well do adolescents recall use of mobile telephones? Results of a validation study. *BMC Med ResMethodol* 2009;9:36–45.
5. Aydin D, Feychting M, Schuz J, *et al*. Mobile phone use and brain tumors in children and adolescents: a multicenter case-control study (CEFALO). *J Natl Cancer Inst* 2011;103:1264–76.
6. Izard V, Dehaene S. Calibrating the mental number line. *Cognition* 2008;106:1221–47. Epub 1 June 2007.
7. Dehaene S. The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends Cogn Sci* 2003;7:145–7.
8. Dehaene S, Izard V, Spelke E, *et al*. Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science* 2008;320:1217–20.
9. Hollingsworth WH, Simmons JP, Coates T, *et al*. Perceived numerosity as a function of array number, speed of array development, and density of array items. *Bull Psychon Soc* 1991;29:448–50.
10. Redmayne M, Smith A, Abramson M. Adolescent in-school cellphone habits: a census of rules, survey of their effectiveness, and fertility implications. *Reprod Toxicol* 2011;32:354–9. Epub 6 September 2011.
11. Bland M, Altman D. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–60.
12. Redmayne M, Smith E, Abramson M. A forecasting method to reduce estimation bias in self-reported cellphone data. *J Expo Sci Environ Epidemiol*The online publication date is 18 July 2012 doi:10.1038/jes.2012.70.
13. Nieder A. Counting on neurons: the neurobiology of numerical competence. *Nat Rev Neurosci* 2005;6:177–190.
14. Krueger LE. Single judgments of numerosity. *Percep Psychophy* 1982;31:175–82.
15. Piazza M, Izard V, Pinel P, *et al*. Tuning curves for the approximate numerosity in the human intraparietal sulcus. *Neuron* 2004;44:547–55.
16. Nieder A, Miller EK. Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 2003;37:149–57.
17. Aydin D, Feychting M, Schuz J, *et al*. Predictors and overestimation of recalled mobile phone use among children and adolescents. *Prog Biophys Mol Biol* 2011;107:356–61.
18. Tokola K, Kurttio P, Salminen T, *et al*. Reducing overestimation in reported mobile phone use associated with epidemiological studies. *Bioelectromagnetics* 2008;29:559–63. Epub 2 June 2008.
19. Cardis E, Richardson L, Deltour I, *et al*. The INTERPHONE study: design, epidemiological methods, and description of the study population. *Eur J Epidemiol* 2007;22:647–64.
20. Abramson MJ, Benke GP, Dimitriadis C, *et al*. Mobile telephone use is associated with changes in cognitive function in young adolescents. *Bioelectromagnetics* 2009;30:678–86. Epub 30 July 2009.
21. Aydin D, Feychting M, Schüz J, *et al*. Impact of random and systematic recall errors and of selection bias in case-control studies on mobile phone use and brain tumours in adolescents (CEFALO study). *Bioelectromagnetics.* 2011;32:396–407.
22. Vrijheid M, Deltour I, Krewski D, *et al*. The effects of recall errors and of selection bias in epidemiologic studies of mobile phone use and cancer risk. *J Expo Sci Environ Epidemiol* 2006;16:371–84.