Trading reliability of recall in the measurement of service utilisation in economic evaluations against research costs

Authors: **Alan Ó Céilleachair**
Takashi Kikuchi
Seow Tien Yeo
Rhiannon Tudor Edwards

Prepared for the Health Economists' Study Group meeting at the University of York - 26-28 July 2006

Trading reliability of recall in the measurement of service utilisation in economic evaluation against research costs

Background:

The measurement of health service utilisation is a routine part of the estimation of costs in economic evaluations. There is a growing literature on the reliability of patient recall as an alternative to accessing GP records, (most recently Mistry et al. 2005)[*]. This is an issue faced by all health economists in the design and conduct of evaluative studies.

Aims:

To discuss to what extent health economists need to consider the characteristics of patient group and any trade-offs between research costs and reliability when designing economic evaluations alongside clinical trials. A comparison of two patients groups will serve as an illustrative example.

Methods:

Data:

Data was drawn from Client Service Receipt Inventory (CSRI) questionnaires and corresponding GP records for three samples: 1) 42 women at intermediate risk of developing familial breast cancer around Wales, 2) 30 parents of children at risk of

[*] Mistry H, Buxton M, Longworth L, Chatwin J & Peveler R on behalf of the AHEAD research team. Comparison of general practitioner records and patient self-report questionnaires for estimation of costs. *Eur J Health Econom*. 2005; **50**: 261-266

developing conduct disorders in North Wales and 3) 30 children whose parents (from sample 2) reported their service use.

Analysis:

We used a weighted κ methodology to measure the level of agreement between GP records and CSRI results for GP, Nurse, outpatient, inpatient and casualty services in all three samples (casualty data was unavailable for sample 1). Paired sample *t*-tests were used to see if the mean cost of services used per patient differed between reporting methods. We also performed an economic costing of the two main methods of data recovery to allow an investigation of research cost/ data reliability trade-offs.

Results:

Agreement between CSRI reported service use and GP records of utilisation ranged from fair to substantial (κ_w =0.27-0.70) for Sample 1 with GP visits displaying the highest level of agreement. There was no significant difference in the mean cost of services used per patient. For Sample 2 agreement was poor to substantial (κ_w =-0.06-0.69) with a significant difference in costs for nurse and casualty contacts at the 5% and 10% levels respectively. In Sample 3 agreement was only poor to moderate (κ_w =0.08-0.55) though no significant difference in mean costs could be found between the two methods.

We calculated the cost of acquiring service utilisation data via postal CSRIs to be between £17.64 and £24.49. The cost of interviewer administered CSRIs was higher

at £45 while we found that accessing GP records was the most expensive method at £70 per record.

Conclusions:

On the whole, we found that for the cohort of women at intermediate risk of familial breast cancer, agreement between GP records and the CSRI was better than for parents in the parenting group whether recalling their own service use or that of their children. This would suggest that choice of data collection method would not bias any ICER calculated for these three samples. This may militate against the use of GP data collection on a cost basis as it was clearly the most expensive of the three methods examined. Finally, there exists a need to perform a broader analysis of this question and many outstanding issues it raises including the appropriate length of recall for CSRIs and the accuracy of GP records.

Introduction

Assessing the cost-effectiveness of new health technologies regularly requires the collection of comprehensive and detailed information on health service utilisation by patients for a range of different services (Drummond and McGuire, 2005). Data describing this service use can be obtained retrospectively using a variety of data collection methods, including consulting service provider records, the administration of patient self-reported service use questionnaires, service utilisation diaries and face-to-face interviews. However, debate continues in the health research community as to the best method for collecting reliable and accurate resource use information. (Mistry et al, 2005; Evans and Crawfords, 2000; Coyle et al, 1999; Drummond et al, 1993).

Two of the most common methods of obtaining data with respect to health service utilisation are the consultation of provider (usually general practitioner) records and the dissemination of client service receipt inventories (CSRIs), that is, questionnaires detailing health service use completed by the patients themselves. To date, a number of studies have compared these two different methods of eliciting patient health service utilisation (Mistry et al., 2005; Richards et al., 2003; Petrou et al., 2002; Evan and Crawford, 2000; Coyle et al., 1999). Based on the findings of these studies it would appear that agreement varies from virtually none to almost perfect dependent on a range of factors including the service being considered, drug usage and internal heterogeneity of patient characteristics.

Coyle et al. (1999) explored the completeness of data collected and the level of agreement between GPs records and patient recall of healthcare resource use in a UK

setting drawing on data from two randomised controlled trials. They found that GP records were less likely to be complete compared to patient-reported service utilisation for a number of community services. This is perhaps unsurprising as GP records do not routinely cover such services. The underlying cost of the services used however did not significantly differ across collection methods. Petrou et al. (2002) found for a sample of first-time mothers at risk of developing postnatal depression that CSRIs tended to under report some community-based services especially when recall was extended over a longer time horizon. However they found that the level of agreement was much closer for hospital services. This is intuitively plausible as hospital visits would appear more likely to leave a lasting impression as an event in patients' minds.

Richards et al (2003) state that while older people are statistically the heaviest users of healthcare resources, the question of the reliability of their recall of service use was poorly understood. They found that chance-corrected agreement was moderate or good, for a sample of older people recently discharged from acute care over periods of less than 12 weeks. Mistry et al's (2005) study of agreement based on a sample of patients presenting to their GP with symptoms of depression found considerable differences in the levels of resource use recorded by GPs compared with CSRIs however agreement based on the weighted κ statistic was still moderate. No significant difference in cost was found, however calculations based on GP records tended to give a higher cost estimate for light service users and a lower cost estimate for heavier service users.

The above studies primarily assess internally homogenous groups with little attempt to directly compare different forms of patient groups. Neither do they explicitly investigate the costs to researchers of the various data collection types. Questions of cost/accuracy trade off have been given little consideration. This study seeks to widen the debate by investigating the implications for health researchers of their chosen project design both in terms of the level of agreement they receive and the costs to them of administering the competing strategies.

We are still left with the following three questions;

1. To what extent does length of recall determine the level of agreement between patients' self-reported service utilisation and provider records?
2. Do different patient groups provide varying levels of agreement between CSRIs and provider records?
3. What are the research cost implications of employing different methods of service utilisation and costs?

This paper addresses questions two and three above based on the analysis of two sets of data from recent studies conducted in Wales.

Methods

Summary of the Samples

Three different cohorts were independently collected alongside evaluations of clinical trials being conducted by the Centre for Economics and Policy in Health, University

of Wales Bangor. The first cohort consists of a geographically stratified sub-sample of 42 women with a mean age of just over 48 years old out of a total sample of 447 women from around Wales at moderate risk of familial breast cancer who were part of the RETRACE programme (2002-2005) funded by the Wales Office for Research and Development service utilisation and resource implications study (RETRACE Team: 2005). These women had high educational attainment. 50% of the sub-group had been educated up to 18 years, with 20% having been involved in education beyond 18.

The second cohort was comprised of a convenience sub-sample of 30 parents from a total sample of 116 participating in a clinical trial of a parenting intervention working with the parents of children at risk of developing conduct disorder (Ó Céilleachair et al, 2006). This intervention was delivered through Sure Start in socio-economically disadvantaged areas across north and mid Wales funded by the Health Foundation. The parents had a mean age of just under 22 at the time of the birth of their first child, well below the national average of 26.9 (Chapell et al, 2005). The median age for exiting education for these women was 16 with only 2 women being in education beyond the age of 18. The average age of parents was just under 30 years of age. The third sample consisted of 30 children whose parents were in the second cohort. In this instance, parents completed a CSRI on behalf of their children.

Data Collection

In both studies full consent was obtained from patients and their GPs to enable access to their GP records. Data on service utilisation were available from two sources for all three samples. The first source was GP records, both electronic and paper records

were consulted. In both studies a research nurse/health economics officer accessed electronic and paper records for patients over the six month period of interest. The second source was CSRI questionnaires. In the RETRACE trial of women at familial risk of breast cancer, each woman completed a CSRI outlining their service use with respect to GPs, Nurses, Casualty attendances, outpatient and inpatient appointments and returned it by post to the research team. In the clinical trial of a parenting programme parents completed a CSRI at home in the presence of a member of the research team detailing both their service use and in a separate CSRI, that of their child over the previous six months for a range of health, social and special educational services (in the case of their children). In both trials CSRIs were issued to all participants (n=116), however only a percentage of these consented to the accessing of their GP records for triangulation purposes.

Statistical Methods

For each study cohort independently we adopted a weighted κ methodology to examine the level of agreement between GP records and CSRIs for service utilisation based on existing literature in this area (Mistry et al, 2005). In this method, the numbers of visits are directly treated as ordinal categories. Kappa (κ) and weighted kappa (κ_w) relate to the interclass correlation and a measure of inter-rater reliability for interval scales, respectively (Fleiss, 1981; Fleiss & Cohen, 1973; Kraemer, 1979). When categories are ordered, some disagreement may be considered more severe than others. The κ_w uses weights bounded from zero to one and describes the closeness of agreement. In common with Mistry et al. (2005) a standard scaling of the level of agreement was employed with results of less than 0.0 being poor agreement, 0.00-0.2 slight, 0.21-0.40 fair, 0.41-0.6 moderate, 0.61-0.8 substantial and 0.81-1.0 almost

perfect. Negative values can occur when agreement is weaker than expected by chance, but this is uncommon (Agresti, 2002).

We used this scale and calculated the values of the weighted kappa and the estimated standard error, following Fleiss & Cohen (1973) and Agresti, (2002). The calculation is carried out by S-Plus and the computing codes for κw and the standard error are kindly provided by Melania Pintilie, Department of Biostatistics, University of Toronto. This was done for the 42 women of the RETRACE trial and for the 30 parents involved in the parenting programme first for their own service use and then for their child's service use over a six-month period.

Costing Methodology

For each of the three sets of data total service use per study participant over the six month period of interest was calculated using national unit and reference costs in 2003-2004 prices (Netten & Curtis, 2004; UK Government, 2004). We were interested in finding out whether there was a statistically significant difference in the mean cost per patient based on the two methods of acquiring service utilisation information: CSRIs administered by an interviewer and the consultation of GP records.

Results

Sample 1: Women at intermediate risk of familial breast cancer (RETRACE) (n=42)

Overall, for this sample of educated and health-aware women at intermediate risk of developing breast cancer there was fair to substantial agreement in service utilisation

between the methods of data collection. Agreement was highest for GP visits and lowest for outpatients. GP visits displayed a substantial level of agreement (κ_w=0.70, SE= 0.15) and the mean number of visits recorded by GP records (1.76) was not statistically different to that of CSRIs (1.83) (*t*=0.3146, *p*=0.7546). Agreement for practice nurse and outpatient contacts was only fair however (κ_w =0.29, SE=0.12 and 0.27, SE=0.14 respectively) though the mean contacts reported again, did not differ significantly between reporting methods (*t*=1.0437, *p*=0.3027 and *t*=0.0779, *p*=0.9383). Finally, inpatient contacts agreed moderately well (κ_w =0.48, SE= 0.13) and once again there was no statistical difference in mean contacts as reported by either method of collection (*t*=-1.1438, *p*=0.1598). These results are summarised in table 1.

Table 1: Sample 1 Women at risk of developing familial breast cancer - Level of GP record/CSRI agreement (n=42)

Service	Mean CSRIS	Mean GP	K_w	SE	*t*	SD	*p*
GP contacts	1.833	1.7619	0.70	0.15	0.3146	2.1059	0.7546
Nurse contacts	0.1190	0.2380	0.29	0.12	-1.0437	0.3952	0.3027
Outpatient appointments	0.881	0.904762	0.27	0.14	-0.0779	1.9025	0.9383
Inpatient appointments	0.0238	0.071429	0.48	0.13	-1.1438	0.1543	0.1598

Sample 2: Parents Own Service use in a parenting programme (n=30)

For these parents living in Sure Start designated socio-economically disadvantaged areas of north and mid Wales, agreement between the two methods of measuring service use ranged from poor to substantial. In contrast to the RETRACE sample, GP visits showed only fair agreement (κ_w =0.24, SE= 0.15). However, there was no

statistical difference in mean reporting (t=-1.6585, p=0.3108). Outpatient and inpatient contacts did display a substantial level of agreement though (κ_w =0.69, SE=0.17 and 0.62, SE=0.17 respectively). Nurse and casualty agreement was slight with especially low weighted κ statistics (κ_w =-0.06, SE=0.15 and 0.072, SE=0.12 respectively). Unsurprisingly mean reported instances of care were significantly different for nurse visits at the 5% level (t=-2.1040, p=0.0442) and for casualty attendances at the 10% level (t=-1.7950, p=0.0831). Table 2 provides a summary of the results for the parenting intervention parent cohort.

Table 2: Sample 2- Parental recall of own service use – Level of GP/CSRI agreement (n=30)

Service	Mean CSRI	Mean GP	K_w	SE	*t*	SD	*p*
GP contacts	1.3333	1.1333	0.24	0.15	-1.6585	2.2338	0.1080
Nurse contacts	0.3	0.6667	-0.06	0.15	-2.1040	1.0372	0.0442
Outpatient appointments	0.1667	0.2	0.69	0.17	-0.2387	0.4611	0.8130
Inpatient appointments	0.1	0.0667	0.62	0.17	0.5708	0.3457	0.5725
Casualty	0.2333	0.1333	0.072	0.12	-1.7950	0.2537	0.0831

Sample 3: Parent's recall of child service use in a parenting programme (n=30)

The parents from sample 2 were next asked to recall their child's service use for the same six-month period. Agreement between the two methods of measuring service use ranges in this instance ranged from poor to moderate. There were only slight levels of agreement between the research methods for GP visits (κ_w =0.08, SE= 0.18) though no statistical difference in mean reporting could be identified (t=-0.4741, p=0.6389). There was moderate agreement between the research methods for inpatient and casualty attendances (κ_w =0.55, SE=0.16 and 0.55, SE=0.16 respectively). Nurse

agreement was fair (κ_w =0.21, SE=0.13) whilst outpatient care was only in slight agreement across methods. For both nurse and outpatient care no statistical difference in the mean was detected (*t*=-1.4588, *p*=0.1554 and *t*=-0.2387, *p*=0.8130). The above results are summarised in table 3.

Table 3: Sample 3 Parental recall of child service use - Level of GP record/CSRI agreement (n=30)

Service	Mean CSRIS	Mean GP	K_w	SE	*t*	SD	*p*
GP contacts	1.90	3.00	0.08	0.18	-0.4741	1.7087	0..6389
Nurse contacts	0.17	0.27	0.21	0.13	-1.4588	0.6513	0.1554
Outpatient appointments	0.40	1.10	0.11	0.18	-0.2387	0.5921	0.8130
Inpatient appointments	0.13	0.20	0.55	0.16	0.5708	0.4026	0.5725
Casualty	0.07	0.27	0.55	0.16	1.3605	0.5040	0.1841

Comparison of mean total service cost per study participant based on CSRI/Access of GP records

For each of the three samples in this paper Table 4 shows there to be no significant difference in mean total cost per participant. For sample 1, women at intermediate risk of developing familial breast cancer for the six month period of interest mean total cost per patient was £176.69 as calculated from GP records while the corresponding figure using the CSRI method was lower at £152.62. However this difference was not statistically significant (*t*=-1.4749, *p*= 0.1479). For sample 2, parents living in Sure Start areas of North and Mid Wales, asked about their own contacts with health services parent sample there was no significant difference in mean cost per patient across data collection methods (*t*=0.1227, *p*= 0.9032) and mean costs per patient for both methods were also quite close (£294.56 and £287.92 respectively). Likewise for

Sample 3, the children of parents in sample 2, the difference in mean total costs per patient was not statistically significant (t=-0.6359, p= 0.5299). The mean cost of services used as reported by the CSRIs was lower than was the case from GP records (£171.06 and £ 232.33 respectively).

Table 4: (National costs 2003/2004)

Sample	Mean Based on CSRI (£)	Mean based on GP records (£)	*t*	SD	*p*
Sample 1: Women at Int. risk of fam. Breast cancer (n=42)	152.62	176.69	-1.4749	237.44	0.1479
Sample 2: Parents of children at risk of developing conduct disorder (n=30)	294.56	287.92	0.1227	507.08	0.9032
Sample 3: Parents of children at risk of developing conduct disorder (n=30). Children's service use	171.06	232.33	-0.6359	446.38	0.5299

Research Costs

(i) Self-reported health service utilisation questionnaires administered by post

The CSRI used for the purposes of this research was based on the Client Service Receipt Inventory designed by Beecham and Knapp (Beecham & Knapp, 1992), however it was amended to account for local conditions. This took approximately 10 hours of research assistant time at an hourly rate of £15 (point 9 on the RA1 scale including national insurance and superannuation but no other overheads, roughly the middle). This cost was then divided over 30 participants at a cost per participant of £5. This figure would obviously reduce accordingly with a greater sample size. In each case postage and materials costs for questionnaires were estimated to be £1 and a further £1 for follow-up phone calls and letters. The research assistant was estimated to have devoted twenty minutes to the preparation and processing of each

questionnaire at a cost of £5. This gives a total cost per questionnaire of £12. As £5 of this figure is an apportioning of more general setup costs, larger samples would reduce this cost significantly.

In the RETRACE study of women at intermediate risk of developing familial breast cancer a response rate of 68% was achieved (RETRACE team, 2005) however, this is somewhat higher than the usual response rates for epidemiological and health service questionnaires with a number of studies positing response rates of between 49-61% (Eaker et al 1998; Cummings, Savitz & Konrad, 2001). Using the response rate from RETRACE as an upper boundary and 49% as a lower boundary the cost of each returned postal CSRI is estimated to be £17.64-£24.49. Table 5 outlines these costs in full.

Table 5: Research costs of CSRI administered by post

Cost Item	Time (hrs)	Unit Cost (£)	Total Cost (£)
Apportionment of Design and Production costs of questionnaire	-	-	5
Postage & materials including SAE	-	1	1
Follow-ups letters including SAE	-	1	1
RA time in preparing and sending forms	0.33	15	5
			12
Response rate adjustment		49%	68%
Cost per returned CSRI		24.49	17.64

(ii) Self-reported health service utilisation questionnaires administered by interviewer

In our study of the cost-effectiveness of a parenting programme for families living in Sure Start designated, socio-economically deprived areas of north & mid Wales, CSRIs were interviewer administered. From a cost perspective this entails an added

cost of the research assistant travelling to administer the questionnaire to the participant. Table 6 presents these costs below. The extra travel and research assistant time brings the total cost to £45. However, several respondents in the one area could bring economies of scale to bear with respect to travel costs.

Table 6: Research costs of CSRI administered by interviewer

Cost Item	*Time (hrs)*	*Unit Cost (£)*	*Total Cost (£)*
Apportionment of costs of design and production of questionnaire	-	-	5
Travel			
Research assistant time	1	15	15
Travel expense	-	10	10
RA time in administering questionnaire	1	15	15
Total cost per CSRI			45

(iii) Accessing of GP records

Table 7 below outlines the costs of accessing GP records. Preparing a protocol and documents for the accessing of GP records including an ethics and data protection statement was estimated to take approximately 22 hours of research assistant time, again at £15 an hour. Letters with an enclosed SAE were sent to participants along with follow-ups where necessary to acquire their written consent.

Contacting practice managers to secure their approval for the accessing of GP records incurred costs of approximately £5. Copies of completed consent forms were forwarded by post to the surgery to complete this process. Many GP surgeries also levy a fee of £10 per record accessed to cover their costs. Letters were sent along with follow-ups where necessary to acquire. A research assistant then had to travel to the

GP surgery to extract the records. This resulted in a total cost per record accessed of £70. Again, we would like to point out that this cost can be lowered by economies of scale, such as several participants having the same GP.

Table 7: Costs of accessing GP records

Cost Item	Time (hrs)	Unit Cost (£)	Total Cost (£)
Apportionment of Administrative setup of record retrieval protocol	-	-	11
Acquisition of patient consent including SAE	-	1	1
Follow-ups letters including SAE	-	1	1
Arrange appointment with practice manager	0.33	15	5
Travel			
Research assistant time	1	15	15
Travel expense	-	10	10
GP charge	-	10	10
Acquisition of GP consent	-	2	2
Extraction of record at GP surgery	1	15	15
			70

Discussion

It appears that levels of agreement from the κ_w analysis we have conducted in this paper are broadly poorer for parents living in Sure Start designated, socio-economically disadvantaged areas in north and mid Wales whether remembering their care or that of their child than for women at intermediate risk of developing familial breast cancer living around Wales. The latter were well educated and generally older. We were surprised that agreement in service utilisation was on the whole poorer for parent's recollection of their child's contacts than for their own. This may be as a result of another family member taking the child to the doctor and that parents may often have more than one child. Children and parents may see other health professionals such as health visitors rather than the GP.

Even with moderate levels of agreement about service utilisation, consistent with Mistry et al (2005) we found that the two data collection methods led to different, but not statistically different mean total costs per patient. This would suggest that ultimately any cost-effectiveness analysis conducted on cost data would be unlikely to be biased as a result of the data collection method chosen in any of the three population subgroups we have presented in this paper. Consequently the question of reliability tradeoffs does not arise for these samples .We estimate that the research costs of accessing GP records directly are 1.5 times greater than interviewer based CSRI and between 3 and 4 times greater than a postal CSRI. Consequently, CSRIs administered by post or administered by interviewers may represent better value for money to researchers.

It would be misleading to view the choice of data collection method in terms of CSRI administration providing a cheaper, less accurate alternative to a "gold standard" of accessing health provider records. Poor or moderate levels of agreement in the weighted κ analysis could result from poor patient recall and/or incomplete GP or other provider records.

The weighted κ coefficient depends greatly on marginal distributions (Argesti, 2002). This means that outliers can have an undue influence and the estimated levels of agreement may not always be accurate. For the weighted κ, a large sample size is preferable to analyse the agreement of the number of visits. Even if the weighted κ value indicates high agreement (and the methodology is assumed to be appropriate), the data of no visits may greatly influence the result because the number of data for no-visit dominates data sets. All three cohorts in this study had relatively small sample sizes and care should be taken in interpreting the results of our study with reference to the above statistical caveats.

While our study had a smaller sample size than much of the literature, we looked at a range of different groups and gained valuable insights into the reporting of service utilisation by parents of their children. Furthermore while some studies have examined the question of agreement of service costs, we have also examined the costs to researchers of various data collection strategies. We hope that this will help to inform future research in estimating its costs more accurately.

To date, all the established literature in this area has come as a result of analysis for samples drawn from up-and-running clinical trials. There is a need for a more systematic investigation of the questions surrounding the issue of agreement/accuracy of the various methods of service-utilisation data collection currently used by researchers. A large sample stratified specifically to address this question would surely go some way to providing evidence that is both internally and externally generalisable. More work is also required on determining the optimum level of the recall period for respondents based on socio-economic and health factors and the issue of patient recall versus the recording of prescribed drugs. Furthermore any such study should look at developing a methodology to comprehensively investigate the issue of accuracy with respect to GP records, while it is widely accepted that GP records are not 100% accurate a more concise picture would appear important in informing researchers.

References:

Agresti A. (2002) *Categorical Data Analysis*. 2nd edition: Wiley

Beecham J & Knapp M. (1992) Costing psychiatric interventions, in G. Thornicroft, C. Brewin & J. Wing (eds) *Measuring Mental Health Needs*. Oxford University Press, Oxford.

Coyle D, Lee K & Drummond M. (1999) Comparison of alternative sources of data on health service encounters. *J Health Serv Res Policy*, **4**(4), 210-214

Cummings SM, Savits, LA & Korad TR. (2001) Reported Response Rates to Mailed Physician Questionnaires. *Health Services Research,* **35,** 1347-1355

Drummond M., Brandt A., Luce B. & Rovira J. (1993) Standardizing methodologies for economic evaluation in health care. *International Journal of Technology Assessment in Health Care*, **9**, 26–36

Drummond, M. and McGuire, A. (2005) *Economic evaluation in health care: merging theory with practice*. Oxford University Press: United Kingdom

Eaker S, Bergström R, Bergström A, Adami, HO and Nyren O (1998) Response Rate to Mailed Epidemiologic Questionnaires: A Population-based Randomized Trial of Variations in Design and Mailing Routines. *Am J Epidemiol,* **147;** 74-82

Evans CJ, and Crawford B. (2000) Data collection methods in prospective economic evaluations: How accurate are the results? *Value in Health,* **3**(4), 277-286

Fleiss JL ., and Cohen J. (1973). The equivalence of weighted kappa and the interclass correlation as measures of reliability. *Educ. Psychol. Meas*. **33,** 613-619.

Fleiss JL. (1981) *Statistical Methods for Rates and Proportions*, 2nd edition: New York: Wiley.

Kraemer HC. (1979) Ramifications of a population model for as kappa as a coefficient of reliability. *Psychometrika* **44**, 461-472.

Mistry, H., Buxton, M., Longworth, L., Chatwin, J. and Peveler, R. on behalf of the Assessing Health Economics of Antidepressants (AHEAD) Team (2005) Comparison of general practitioner records and self-report questionnaires for estimation of costs. *European Journal of Health Economics,* **50**, 261-266

Ó Céilleachair A, Edwards RT, Bywater T & Hutchings J (2006) An Economic Evaluation of the Incredible Years Child-Parenting Programme. Oral presentation at the European Conference on Health economics, July 2006, Budapest.

Petrou, S., Murray, L. & Davidson, L.L. (2002) The accuracy of self-reported healthcare resource utilization in health economic studies. *International Journal of Technology Assessment in Health Care*, **18**(3), 705-740

RETRACE Team (2005) *Women at intermediate risk for familial breast cancer: a study of long-term cancer worries, post-counselling pathways and resource implication for delivery of care*. RETRACE final report for Wales Office for Research and Development in Health and Social Care. Unpublished.

Richards, S.H., Coast, J. and Peters, T.J. (2003) Patient-reported use of health service resources compared with information from health providers. *Health and Social Care in the Community*, **11**(6), 510-518