Are survey measures of medical care utilisation misleading? A comparison of self-reported medical care consumption with actual medical care utilisation

RIC MARSHALL, DAVID GRAYSON, ANTHONY JORM, AND BRIAN O'TOOLE

Ric Marshall is a Visiting Fellow at the Centre for Mental Health Research, Australian National University and Australian Centre for Post-traumatic Mental Health, University of Melbourne.
David Grayson is Associate Professor of Psychology at the University of Sydney. Anthony Jorm is the Director of the Centre for Mental Health Research, Australian National University and Brian O'Toole is a Senior Research Fellow at the School of Community Medicine, University of New South Wales.

Abstract

A substantial over-estimate of medical care consumption cost was found when estimates from self-report data from an epidemiolgical study were compared to actual cost data extracted from administrative records. Even though the few subjects who were actually provided with two or more services in the two-week self-report period substantially under-reported their medical care consumption, a large net over-estimate of medical care consumption was produced by the self-report data. This finding has important implications for use of self-report data from surveys such as the Australian Bureau of Statistics (ABS) National Health Survey for estimating health service consumption.

By combining epidemiological survey data from the Australian Vietnam Veterans Health Study (AVVHS), with data on actual medical care for which the Health Insurance Commission (HIC) or the Department of Veterans' Affairs (DVA) paid benefits, we were able to directly compare self-reported medical care consumption with actual medical care utilisation. The comparison revealed that veterans' self-reports were a valid measure of relative medical care consumption because those who reported care over the past two weeks were much more likely to have been recent consumers than those who did not. This relationship became even stronger if the comparison of self-report was extended to data on benefits paid beyond the two-week self-report period. However, the HIC and DVA data confirmed only 51% of veterans self-reporting medical care consumption during the past two weeks actually received a service.

Introduction

Self-report data are often used to estimate use of health care services (Australian Bureau of Statistics, 1991; Marshall, Jorm, Grayson, & O'Toole, 1998; Stuart, Klimidis, & Minas, 1998; Anie, Jones, Hilton, & Anderson, 1996). However, the precision of self-report surveys is often challenged (Newell, Girgis, Sanson-Fisher, & Ireland, 2000).

In some studies there is the opportunity to obtain both self-report and actual utilisation data (McCallum, Simons, Simons, Sadler, & Wilson, 1995; Shadbolt, McCallum, Bourne, & Singh, 1998). Where these can be obtained for a common period and covering the same set of healthcare services, self-report can be validated against the actual utilisation data as the 'gold standard'.

One area of concern with self-report data arises from the propensity of subjects to over- or under-report as a result of recall errors. Seminal articles on the accuracy of self-report include Eisenhower (1991), Groves (1989), Brown (1985) and Bradburn (1987). All refer to the problem of 'telescoping' when self-report data are used. Telescoping may result in over-reporting, through subjects' recalling events that occurred before the specified period as having occurred within it. It is said to be more common with self-reports relating to the very recent past, and when there are fewer events to report. In such circumstances, the most recent event tends to be reported as having occurred within the reporting period, even if it was some time earlier.

It was such a concern that led us to replicate a self-report survey analysis on health service utilisation costs (Marshall et al., 1998) with an analysis using 'hard' data (Marshall, Jorm, Grayson, & O'Toole, 2000). In the initial study of healthcare consumption, we used a self-reported measure based on a two-week reporting period, which had been developed for the ABS National Health Survey. In the second, we made use of objective administrative data on medical care utilisation to provide a measure that encompassed a substantial time period of twelve months. The reporting period of twelve months was chosen for the second study of factors related to medical care consumption costs because twelve months is the period that is generally used for financial reporting and budgeting. However, the data gathered for this "replication" exercise also gave us the opportunity for a direct comparison between self-report data on medical care utilisation over a two-week period and data on actual services paid for by government healthcare services.

In contrast to the twelve months administrative data used for analysis of medical care consumption, the self-report data used in our earlier analysis (Marshall et al., 1998) covered the two-week period immediately prior to the interview to optimise recall precision. Nevertheless, the self-report data and the objective administrative data both contained subsets of medical care consumption measures covering the two weeks immediately before the interview. This overlap provided an opportunity to also conduct the study reported here on the validity of the self-report data used in earlier analyses. Thus this study reports on the results of the validation of self-report against objective administrative data on medical care costs obtained from the Department of Veterans' Affairs (DVA) and the Health Insurance Commission (HIC).

Self-report data on two-weeks of medical care consumption were obtained from the epidemiological study, where it comprised a subset of all healthcare service utilisation reported. Data on benefits paid for medical care in the same period were obtained from the twelve-month data on each subject provided by DVA and HIC. The two data sources compared were thus overlapping subsets of otherwise non-comparable parent data sets. To obtain these subsets, we selected the two-week data for each subject from each data set that related to the same period of medical care consumption. One dataset represented self-report and the other 'actual' utilisation as indicated by benefits paid by DVA and HIC.

The hypothesis examined in this study is that self-reported medical care consumption costs during two weeks immediately before the interview will be matched by medical care utilisation as measured by benefits paid for medical care actually consumed by the subjects in the same two-week period.

Method

The subjects examined in our epidemiological study were 641 Vietnam veterans interviewed from a random sample of 1000 veterans. Of the 1000 subjects sampled, 213 could not be found, 50 were known to have died, 61 refused interview, and 35 were unable to be interviewed in the time available (O'Toole et al., 1996).

We provided DVA and HIC with a larger sample of 2000 veterans, from which our original sample of 1000 subjects had been drawn, for the data extraction. This provided a level of privacy protection for the identity of the subsample of 641 veterans who had been interviewed and whose data were used for the analyses in this study. It gave the scope to test different approaches to the validation of match extraction and to compare extraction results between subsamples. It also provided a reserve of subjects within which we could, if necessary, investigate

under-matching or bias analysis in the data retrieval without having to approach the subjects already interviewed in our epidemiological study.

DVA and HIC were approached in 1995 and 1997 respectively. This timing difference occurred because it was originally considered possible that the relationships of interest would be detectable within the DVA data alone. However, it soon became apparent that the share of healthcare provided to these veterans through HIC-administered programs was substantial.

Rigorous protocols were developed with the DVA and HIC to ensure the extraction protocols were both thorough and protected the privacy of the individuals involved. To a large extent, both the subject matching and record extraction were iterative processes with several interactions and clarifications between the study investigators and departmental data managers involved.

As interviews had been conducted over a three-year period, four years of HIC and DVA data on benefits paid were extracted covering the period from 1 July, 1989 to 30 June, 1993 - that is, from six months before the first interview to six months after the last interview. This ensured that, for each subject, data on benefits paid were available for the six months periods before and after the subject's interview date.

DVA data were filtered to exclude all but medical consultations. Checking was carried out to exclude the possibility that duplicate records were included on payments once DVA and HIC data sets were combined.

The raw data on each occasion of service were culled so that only those medical services provided within two weeks before each subject's interview date were retained. The HIC and DVA data contained a separate record for each occasion of medical care provided in the two-week reference period. Counts of occasions of medical service and total payments were calculated on each subject's 'actual' medical care utilisation. Cost of medical care actually used in the two weeks was calculated in two ways for comparison with the self-report variable producing three variables for comparison.

Measure 1 was estimated cost derived from the count of occasions of service actually utilised. The number of services was multiplied by $36.33 representing the mean cost of an occasion of medical service at the time the interviews were conducted (Commonwealth Department of Community Services and Health, 1990). This provided a method for estimating cost of medical care consumption that matched an earlier analyses of self-reported healthcare consumption costs of these subjects (Marshall et al., 1998).

Measure 2 was benefits paid for each subject for the two-week period, summed to produce a total cost of medical care utilised. This total would, in fact, represent something more than 85% of the actual cost because it ignores co-payments from those veterans not eligible for DVA care and required to pay about 15% of the fee.

Measure 3 was the self-report medical consumption cost estimated using the same multiplier that was applied to occasions of service in measure 1. This comprised the medical component of the total self-reported healthcare utilisation cost as identified by the ABS Health Survey questionnaire (Australian Bureau of Statistics, 1991).

From this point, the three measures will be referred to as estimated cost, benefit paid and self-report. The total medical care estimated cost and total benefits payments were merged for the following comparisons with self-reported medical care consumption in the two-weeks before the interview. This produced a data set containing one record per subject with the above three measures for the same two-week period.

To examine the telescoping of recalled events into the more recent time period, the self-reported medical care consumption variable was simplified into a dichotomous Y/N variable, indicating whether or not any medical care consumption was self-reported in the past two weeks. The benefits paid data were also examined to produce a data item for each subject identifying the number of days between the date of the last medical service before the interview and the interview date. The percentage of subjects who had actually utilised medical services was then plotted cumulatively on a week by week basis separately for those subjects who self-reported some consumption and those who self-reported no consumption in the past fortnight.

Results

The two-week period of medical care data used in this analysis was taken from a four-year data extraction on each subject covering the full period of the AVVHS study. The four years data served to ensure a high probability of some medical service utilisation by most subjects and hence validate the extraction.

The DVA matching of the sample of 2000 veterans produced 16064 occasions of service from the four years of utilisation data. These comprised 12043 medical consultations, 234 hospital episodes, 184 domicilary nursing occasions of service, 1577 dental services, 2026 allied health consultations. The HIC matching produced 39112 occasions of medical service.

At least some utilisation records were found for 94.9% of the 2000 subjects for whom data were extracted. HIC records were found for 92.0% subjects, DVA records were found for 34.8% of subjects, and only 5.1% of the subjects appeared to be non-users of medical services in the 4 years covered by the initial data extraction. When this envelope was reduced to the six months either side of the interview (twelve months total) 80.0% were represented as users of medical services. When further filtered to the 2 weeks before the interview, 14.4% were represented in the utilisation data.

For the 641 subjects with whom the rest of this study is concerned, 95% were represented in the four-year data extract. As shown in Table 1, in the two-week period before the interview date, which is the period of interest for the self-report comparison, only 91 (14.4%) of the subjects were represented as medical care consumers. This is in contrast to the self-report data where 145 (23%) of the subjects had reported at least one doctor visit in the two weeks before their interview. Nevertheless, with the high percentage of subjects for whom data were retrieved in the four-year window, it was concluded that the subject identification and data retrieval were satisfactory and as accurate as could reasonably be achieved across separately collected data items for both the DVA and HIC extractions.

Table 1. Cross-tabulation: whether self-reported medical care consumption by whether utilisation occurred in two weeks before interview

		Whether medical care utilisation occurred		
		No	Yes	Total
Whether any self-reported medical care consumption	No	479	17	496
	Yes	71	74	145
	Total	550	91	641

Table 1 crosstabulates the self-report medical service consumption for the two-week period before the interview against the medical benefits that were identified as being paid in that period. The largest groups in the table were those where benefit payment data coincide with self-report. A group of 479 subjects self-reported no doctor consultations and, in fact, no benefit payment was found in the two weeks concerned for these subjects. The next largest group (74 subjects) contained those who reported at least one doctor consultation and benefits were identified as having been paid.

Of the other two (discordant) groups, the smaller group (17 subjects) contained subjects who self-reported no doctor consultation but benefits were found to have been paid. This number is plausible as representing a few valid exceptions, as benefits may have been paid for a medical service that did not require a visit from the subject (e.g. a pathology test).

The group of most interest in this analysis contains the 71 subjects who self-reported a doctor consultation but for whom no benefits were paid in the two weeks before the interview. This is a substantial proportion (49%) of those who self-reported a doctor consultation and the group responsible for the over-estimate of medical care consumption. Therefore we decided to compare the actual recent medical care consumption pattern of the group of subjects who self-reported a doctor visit in the past two weeks with that of those who self-reported no doctor visit. The variable used for this measure of the accuracy of self-report was the time period which had

elapsed since the most recent medical consultation for which a benefit actually was paid and the interview.

The measures for self-report medical care consumption cost (Measure 3) and benefits paid over all cases (Measure 2) were correlated at 0.46 (N=641, p< 0.001). For those cases where benefits had been paid , the correlation between benefits (Measure 2) and self-reported consumption (Measure 3) was 0.33 (N=91, p<0.001). Even for the subset of cases where benefits paid for medical care utilisation and self-reported consumption were congruent, the correlation was only a slight improved at 0.42 (N=74, p<0.001) and probably partly attributable to under-reporting by those with multiple visits (See Figure 1).

The correlation between estimated cost of medical care utilisation (Measure 1) and self-reported consumption cost (Measure 3) was 0.50 (N=641, p<0.001). This reflects the indicative nature of the price multiplier as compared with actual benefits paid.

Figure 1 shows the medical care costs, by four user categories, for each of the three measures. The four categories of user were those who used no services, those who attended once during the two weeks, those who attended twice and those who attended more than twice.

Figure 1. Comparison of 3 different medical care cost estimates showing contribution by utilisation category of subjects

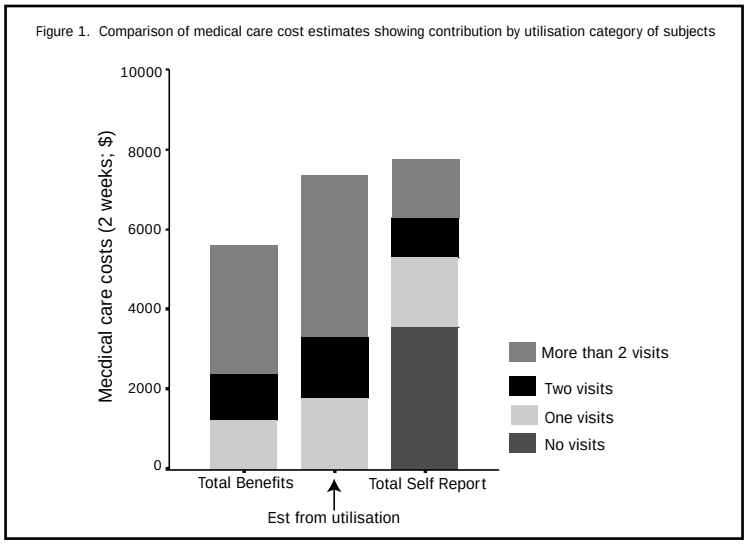


Figure 1 allows a comparison of the contributions of each of the user groups to total medical care costs as measured by the three different methods. Clearly the consumption over-reported by those who, in fact, attended no sessions but reported an attendance (N=71) was the major component of the over-estimate of cost in the self-report measure.

There is a small but important over-report of visits by those subjects (N=550) who actually did not attend in the two-week reporting period. This leads to a mean overestimate of costs of $6.47 per person for this group (95% CI: $4.93, $8.01). However, because of the large number of persons involved, this group creates a very large overestimate that accounted for 46% of the $7,738 estimated as the total cost of self-reported consumption for the sample (55x$6.47/$7,738=46%). In those who attended once or twice, the three measures are similar - indicating either accurate reporting, or compensating over- and under-reporting. The opposite effect occurred in the high utilisation group where the cost estimate comparisons indicated significant net under-reporting in the self-report data with a mean under-estimate of $84.73 per case (N=21, 95% CI: $28.53, $140.93). However, because of the smaller number of cases, this underestimate represented only 23% of the total self-reported cost-estimate. With total benefits paid of $5,619, the net overestimate from self-reported consumption data over all subjects was $2,119 (30%).

Table 2. Comparison of time to most recent consultation of the subjects (N=550) with no consumption in the past two weeks by self-report status.

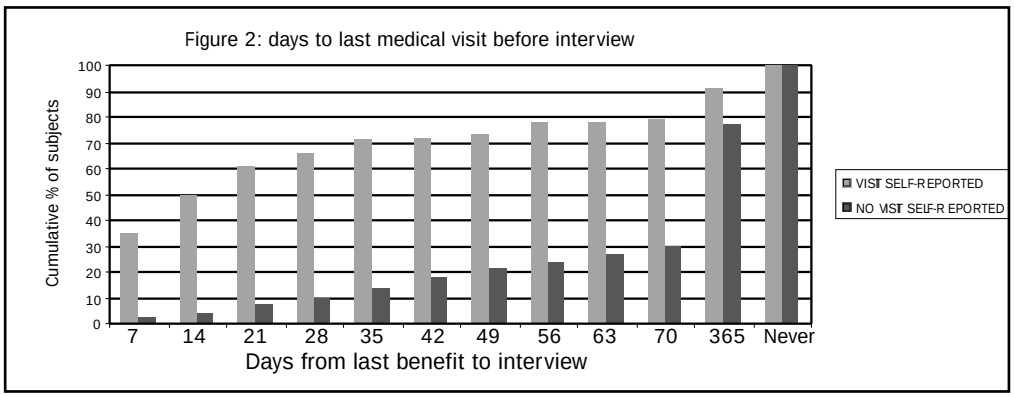
		Time to most recent medical consultation					
	Count (%)	2 to 4 weeks	4 to 8 weeks	8 weeks to 6 months	6 months to 4 years	No consults	Total
Whether any medical care consumption self-reported in past two weeks	NO	31	69	168	93	118	479
		(7%)	(14%)	(35%)	(19%)	(25%)	(100%)
	YES	24	17	13	7	10	71
		(34%)	(24%)	(18%)	(10%)	(14%)	(100%)

Table 2 examines difference in actual recent utilisation between those who self-reported no consultations in the past two weeks and had none (479 subjects) and those who were apparently 'telescoping' recalled consultations from more than two weeks past (71 subjects). The difference in recent utilisation between these groups is significant (Chi squared = 61.13, df = 4, $p<0.0001$).

This confirms that the subjects in the group of 71, while not accurately self-reporting the timing of their most recent consultation in the past two weeks, were nevertheless higher medical care consumers than those who (accurately) self-reported no medical care consumption.

The difference between the groups is also illustrated by Figure 2. A major difference can be seen in the recency of benefits payments for subjects who self-reported a recent visit and those who self-reported no visit in the past two weeks. For those who self-reported a visit in the past two weeks, over 70% had received medical services for which benefits were paid within the past 35 days. This is in strong contrast to those who self-reported no medical service in the last fortnight. Of these, less than 15% had, in fact, received services within the past 35 days. This confirms the view that the apparent over-reporting is largely attributable to 'telescoping' - that is, attributing some more distant albeit still recent services to the nominated reporting period.

Figure 2: days to last medical visit before interview



Discussion

The self-report validity check was achieved using the data collection period common to both the self-report data and the objective administrative data collections. The medical component of the self-reported healthcare consumption was cross-tabulated against the subset of medical services paid for by the HIC and DVA from the two weeks immediately before the interview (Table 1). If self-report of consultations and data retrieval on benefits paid are both precise, they should be very closely related, with some rare logical exceptions.

Legitimate causes for mismatches between a positive answer to "how many times have you seen a doctor in the past two weeks" and benefits actually being paid could include situations where:

- benefits would also have been paid in relation to pathology tests ordered and performed in that period and the subject may not have seen that practitioner
- subject could correctly say that they had seen a doctor for a service such as a follow-up dressing that was part of a service which could be billed to an earlier or later benefit claim
- a subject could have seen a practitioner, paid privately and failed to claim benefits or claimed from workers compensation, accident insurance or other alternative benefit source.

The veteran's recall of the occasion of service appears to have been 'telescoped' so that the veteran reported a service as within the past two weeks that in fact was received up to several weeks previously. This memory distortion is a common phenomenon (Bradburn et al., 1987; Brown et al., 1985; Eisenhower et al., 1991; Groves, 1989). Factors involved in the apparent over-reporting may include:

- wanting to please and so knowingly telescoping (such as to be 'helpful' and provide information that is 'approximately' relevant)
- genuinely thinking the event had happened in last fortnight (e.g. "It seems that recent")
- interpreting 2 weeks liberally as 'recently' (this likelihood would be increased by the absence of an anchor event as a marker like "since some well recognisable event two weeks ago").

It is unlikely that veterans would be more prone to this over-reporting than others. In fact, as demonstrated above, those subjects in the high utilisation group were net under-reporters. This tendency to under-reporting in the high user group creates a dilemma in interpreting the meaning of the large group of 'over-reporters' who actually used no service (N=71). Clearly those who used no services cannot under-report. It is not possible to have less than zero attendances. Therefore any tendency to under-reporting in any members of this group would not be observable. Only over-reporting instances can be manifested in the measure.

Many of the interviewers, when asked for their impressions, considered that the veteran subjects appeared to be answering the health status and service utilisation questions conservatively. Probing usually uncovered a higher level of health problems than volunteered. This apparent stoicism would be more likely to lead to under-reporting or minimisation of health events. It would counteract any tendency to over-reporting in comparison to non-veteran subjects.

While wanting to be 'helpful' and volunteer information would be more likely to lead to telescoping, this project found a predominance of under-reporting in those categories where under-reporting was possible (those who actually had two or more occasions of service). Moreover, those subjects who 'telescoped' self-reported medical care consumption in the past fortnight were more likely to have had a recent consultation than those who reported no medical care. Thus they tended genuinely to be higher users and their recent utilisation history confirmed this.

Conclusions

In summary, this project has provided data on the validity of self-reports for a 2-week health service consumption period, as used in the ABS National Health Survey. It was found that about 49% of subjects self-reporting a consultation in the two-week window appeared to be reporting services that were actually consumed further back in time. However, in the higher consumer groups, this appeared to be counterbalanced by under-

reporting of the number of consultations. In addition to the observed telescoping effects, minor levels of discordance due to billing practices would be expected between accurate self-reported medical care consumption and the data on benefits paid for those consultations by HIC and DVA.

As a scale measure of medical care consumed, self-report appears to be valid in that those who report more services tended to be higher consumers. However, there appears to be up to 30% over-estimating of actual services used in the two-week reference period. Therefore, in estimating the actual rate of medical care consumption from self-report of recent use, a sensitivity discount factor of up to 30% needs to be applied to set a lower limit on the estimate.

Acknowledgements

A Commonwealth Department of Health and Aged Care Casemix Education and Study Grant supported the study. The earlier data collection was supported by grants from the National Health and Medical Research Council, Public Health Research and Development Committee, the Australian Vietnam War Veterans Trust Ltd., the Department of Veterans' Affairs, the Westmead Research Institute and the Australian War Memorial. Assistance from Department of Defence (Army), Canberra with facilities and support for data extraction and analysis is particularly appreciated. The findings and views expressed in this report are those of the authors and are not to be taken as the official opinion of the Department of Defence (Army Office) nor the Department of Veterans' Affairs.

References

Anie K, Jones P, Hilton S & Anderson H 1996, 'A computer-assisted telephone interview technique for assessment of asthma morbidity and drug use in adult asthma'. *Journal of Clinical Epidemiology*, vol 49, pp 653-6.

Australian Bureau of Statistics 1991, '*1989-90 National Health Survey. User's guide*'. *Catalogue No 4363.0*. Canberra: ABS.

Bradburn N, Rips L, & Shevell S 1987, 'Answering autobiographical questions: the impact of memory and inference on surveys'. *Science*, vol 236, pp 157-61.

Brown N, Rips L, & Shevell SK 1985, 'The subjective dates of natural events in very-long-term-memory'. *Cognitive Psychology*, vol 17, pp 139-177.

Commonwealth Department of Community Services and Health 1990, '*Medicare Benefits Schedule Book - Effective 1 May*'. 1990. Canberra: Australian Government Publishing Service.

Eisenhower D, Mathiowetz N & Morganstein D 1991, 'Recall errors: sources and bias reduction techniques'. In P Beimer, R Groves, L Lyberg, et al (Eds), *Measurement Errors in Surveys* (pp 127-144). NewYork: John Wiley.

Groves R 1989, 'The respondent as a source of measurement error', *Survey Errors and Survey Costs* (pp 407-445). NewYork: John Wiley.

Marshall R, Jorm A, Grayson D & O'Toole B 1998, 'Post-traumatic stress disorder and other predictors of healthcare consumption in Vietnam veterans' *Psychiatric Services*, vol 49, pp 1609-11.

Marshall R, Jorm A, Grayson D & O'Toole B 2000, 'Medical-care costs associated with posttraumatic stress disorder in Vietnam veterans'. *Australian and New Zealand Journal of Psychiatry*, vol 34, pp 954-62.

McCallum J, Simons L, Simons J, Sadler P & Wilson J 1995, 'The continuum of care for older people.'. *Australian Health Review*, vol 18, pp 40-55.

Newell S, Girgis A, Sanson-Fisher R & Ireland M 2000, 'Accuracy of patients' recall of Pap and cholesterol screening'. *American Journal of Public Health*, vol 90, pp 1431-5.

O'Toole B, Marshall R, Grayson D, et al 1996, 'The Australian Vietnam veterans health study: I. study design and response bias.'. *International Journal of Epidemiology*, vol 25 no 2, pp 307-18.

Shadbolt B, McCallum J, Bourne M & Singh M 1998, 'Hospital outcomes management: the Care Continuum and Health Outcomes Project'. *Australian Health Review,* vol 21 no 3, pp 150-67.

Stuart G, Klimidis S & Minas I 1998, 'The treated prevalence of mental disorder amongst immigrants and the Australian-born: community and primary-care rates' *International Journal of Social Psychiatry*, vol 44 no 1, pp 22-34.