1999

The accuracy of self-reported health behaviours and risk factors relating to cancer and cardiovascular disease in the general population: a critical review

Sallie Newell
Southern Cross University

Afaf Girgis
University of Newcastle

Rob William Sanson-Fisher
University of Newcastle

Nina Johanna Savolainen
Commonwealth Grants Commission

Publication details

Post-print of: Newell, S, Girgis, A, Sanson-Fisher, RW & Savolained, NJ 1999, 'The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population1: a critical review', *American Journal of Preventive Medicine*, vol. 17, no. 3, pp. 211-229.
Published version available from:
http://dx.doi.org/10.1016/S0749-3797(99)00069-0

The accuracy of self-reported health behaviours and risk factors relating to cancer and cardiovascular disease in the general population: A critical review.

Sallie A. Newell* (PhD),

Afaf Girgis* (PhD),

Rob W. Sanson-Fisher# (PhD),

Nina J. Savolainen (BA Hons).

NSW Cancer Council Cancer Education Research Program,

Locked Bag 10, Wallsend, NSW 2287, Australia

* Sallie Newell and Afaf Girgis are also conjoint lecturers in the Faculty of Medicine & Health Sciences at the University of Newcastle, Australia.

\# Rob Sanson-Fisher is currently the Dean of the Faculty of Medicine & Health Sciences at the University of Newcastle, Australia.

Address for Correspondence

The Secretary, NSW Cancer Council Cancer Education Research Program, Locked Mail Bag 10, WALLSEND NSW 2287, Australia. Tel: +61 249 246 368, Fax: +61 249 246 209.

Email: sallien@mail.newcastle.edu.au

Word count: 4,900 (introduction - conclusions)

Abstract

Objective: To critically review the literature concerning the accuracy of self-reported health behaviours and risk factors relating to cancer and cardiovascular disease among the general population.

Method: A literature search was conducted on three major health research databases: Medline, Healthplan and Psychlit. The bibliographies of located articles were also checked for additional relevant references. Studies meeting the following five inclusion criteria were included in the review:

- They were investigating the accuracy of self-report among the general population, as opposed to among clinical populations.
- They employed an adequate and appropriate gold standard.
- At least 70% of respondents consented to validation, where validation imposed minimal demands on the respondent; 60% consent to validation was considered acceptable where validation imposed a greater burden.
- They had a sample size capable of estimating sensitivity and specificity rates with 95% confidence intervals of width □ 10%.
- The time lag between collection of the self-report and validation data for physical measures did not exceed one month.

Results: 24 of 66 identified studies met all the inclusion criteria described above. In the vast majority, self-report data consistently underestimated the proportion of individuals considered "at-risk". Similarly, community prevalences of risk factors were considerably higher according to gold standard data sources than they were according to self-report data.

Conclusions: This review casts serious doubts on the wisdom of relying exclusively on self-reported health information. It suggests that caution should be exercised both when trying to identify "at-risk" individuals and when estimating the prevalence of risk factors among the general population. The review also suggests a number of ways in which the accuracy of individuals' self-reported health information can be maximised.

Keywords: self-report, accuracy, validation, cancer, cardiovascular disease, risk assessment.

Introduction

In the field of health promotion, we need to assess both individuals' and populations' health risk status for a variety of risk factors and behaviours. Reliable, valid and appropriate measures are prerequisites to any descriptive research or intervention assessment[1]. Without appropriate measures, it is impossible to accurately assess the prevalence of a target behaviour, to identify and describe the characteristics of those individuals at greater risk or to assess the efficacy of interventions aimed at reducing risk factors.

Self-report is one of the easiest, cheapest and most widely used methods of collecting data about individuals' health and risk factor status[2-6]. Many health studies use self-reported data to assess the prevalence of given risk factors (eg: smoking, lack of exercise, etc) or health behaviours (eg: having Pap tests, wearing sunscreen, etc) in the community or to evaluate the success or failure of health promotion interventions. For example, a MEDLINE[7] search of 1997 alone, located 31 studies which reported rates of Pap testing: 22 (71%) described the prevalence or predictors of screening[8-29]; 10 (32%) assessed the effectiveness of interventions to increase screening rates[26;29-37]; and one (3%) assessed the accuracy of self-reported screening rates[38]. Excluding the one validation study, 19 studies (63%) used self-report as the screening outcome measure, highlighting that almost two thirds of current research into the prevalence and predictors of pap testing and the effectiveness of interventions to increase Pap testing rates rely on self-report data [8-22;25;28-30].

As Pap testing is a health behaviour with clear recommendations, where alternative, objective measures exist, in the form of health insurance records, and where questions about the accuracy of self-report have been raised for a number of years, the authors consider it likely that at least a

similar degree of reliance on self-report data is likely to be found in the literature investigating many other health risk factors and behaviours. With such a reliance on self-reported health data, it is imperative that these data are assessed to be valid and reliable measures.

In order for such prevalence estimates and outcome measures to be useful, the self-report items must provide an accurate measurement of that which they are supposed to be measuring. Inaccurate self-report could lead to underestimation or overestimation of the prevalence of risk factors or health behaviours in the community or to the misclassification of risk status at the individual level, which could obscure causal relationships between risk factors and subsequent disease.

Of the pap testing studies mentioned above, the validation study found that self-reported screening prevalence did not accurately reflect the prevalence obtained from pathology records[38]. These findings are consistent with those of earlier studies investigating the accuracy of self-reported pap testing history[39;40]. Serious questions were raised about the accuracy of self-reported health data nearly 30 years ago[41-43]. Despite such doubts about self-report accuracy, of the 19 prevalence and intervention studies published in 1997, that employed self-reported pap testing as an outcome measure: nine (47%) did not raise the issue of self-report accuracy at all[10;18-22;25;29;30]; one (5%) raised the issue, attempted no validation but assumed it was sufficiently accurate[16]; eight (42%) raised the issue and accepted it as a limitation but did not discuss to what extent it may have affected the results[8;9;11;13-15;17;28]; and one (5%) validated the self-report of respondents against their health billing records, finding that many women had overestimated their pap screening frequency[12].

The potential consequences of this avoidance of the issue of self-report accuracy are significant. Previous validation studies have indicated that approximately 45%-65% of women inadequately screened for cervical cancer are not identified by self-report[3;39;40]. These studies have also found that only 70%-78% of self-report claims of being adequately screened could be verified. Assuming similar inaccuracy rates in the pap testing studies described above, it would seem likely that any conclusions drawn about the prevalence of pap testing, the predictors of being adequately screened or the effectiveness of interventions to increase screening rates should be treated with a high degree of scepticism.

Why Conduct a Review of this Literature?

The failure of these studies to adequately address the issue of self-report accuracy could be partly due to the fact that many studies have investigated the accuracy of self-reported health information and drawn varying conclusions, depending on the behaviour being investigated, the quality of the methodology employed to investigate them and the "gold standard" against which the self-report was compared. Whatever the explanation, it is of concern that an extensive body of literature regarding the accuracy of self-reported health data is seldom addressed in the subsequent descriptive and intervention-based research. Therefore, the authors considered that conducting a critical review which attempted to establish some consensus from the literature in this area would serve to emphasise the importance of considering accuracy issues when conducting health research using self-report data. As secondary aims, this review also explores the characteristics found to be predictive of increased accuracy of self-report and of consenting to have one's self-report validated.

Review Methodology

Search Strategy

This paper represents an extensive review of data-based studies, published between 1983 and 1997 which investigated the accuracy of self-report of health risk, preventive, screening or early detection behaviours for cancer and cardiovascular disease in the general population. These diseases were selected as they are responsible for approximately three quarters of all deaths[44,45]; they represent major target areas for current health promotion initiatives[46,45]; and because individuals' health risk and screening status for these diseases are frequently measured in self-report surveys.

Only studies assessing the accuracy of self-report among general, as opposed to clinical, populations were included in this review. It was considered likely that clinical populations, who are usually receiving some form of intervention aimed at changing the behaviour being studied, may feel more compelled to respond in socially desirable ways[47-49] and we wished to estimate the maximum accuracy of self-reported health behaviours in community surveys.

The literature search, which was limited to data-based studies published in English, was conducted on three major health research databases: MEDLINE, HealthPLAN and PsychLit. Additional relevant studies were identified from the bibliographies of studies located during the literature search, in an iterative fashion, until no new references were being identified.

This process located 66 studies which had investigated the accuracy of self-report of a variety of cancer and cardiovascular risk factors and screening behaviours, namely: height, weight and body mass index[4;50-66]; alcohol intake[67-71]; smoking status[4;60;72-81]; physical activity level[2;82-85];

blood pressure and hypertension[4;60;86-91]; diabetes[4;60;88]; cholesterol level and hypercholesterolemia[4;60;88]; nevi[92;93]; diet[94]; Pap testing[3;39;40;95-101]; mammograms[3;86;98;100-105]; clinical breast examinations[3;98]; sigmoidoscopies[3;98]; faecal occult blood tests[3;98]; digital rectal examinations[3;86;98]; gastric photofluorographs for stomach cancer[106]; and family history of cancer and coronary heart disease[107-110].

Critical Review Process

To ensure confidence in the conclusions of any systematic literature review, it is necessary to set a priori criteria for assessing the internal validity of the individual studies potentially available for inclusion in the review[111;112]. Threats to internal validity in the studies contributing to this review could have lead to the formulation of erroneous conclusions about the accuracy of self-report. Therefore, to maximise the validity of our conclusions only studies meeting the following four minimal inclusion criteria were included in this review.

First, the study should have employed an adequate gold standard. For the purpose of this review, an adequate gold standard was one which: validly and reliably assessed the behaviour of interest; allowed categorisation of individuals into "at risk" and "not at risk" groups; and was appropriate to the self-report question being validated. For example, as cotinine levels can be used to detect smoking in only the last four days, it would not have been acceptable to say that a low cotinine level confirmed that a person had abstained from smoking for six months. Table 1 shows the gold standards employed in the studies located, indicating which were considered to be acceptable and unacceptable. Thiocyanate was considered to be an unacceptable gold standard for assessing smoking status as levels can be contaminated by a number of dietary and other sources[113-115] and the differences in levels between smokers and non-smokers is much

smaller than with other biochemical markers, making misclassification more likely[116;117]. Physicians' records were considered an unacceptable gold standard for assessing screening status and family histories as the accuracy of such records is doubtful[118]. Structured interview was considered unacceptable for assessing alcohol intake as it is still a subjective measure and vulnerable to reporting error. Similarly, the Harvard Alumni Activity Scale, another subjective measure, was considered unacceptable for assessing physical activity. In addition, prospective diaries were considered unacceptable gold standards for assessing physical activity or diet as they are likely to lead to reactivity, whereby, the subject feels obliged to behave in more "compliant" ways than they would otherwise. Pool attendance records were also considered unacceptable due to insufficient information about their accuracy. It should be pointed out that those gold standards considered acceptable are not necessarily without limitations: they were merely considered adequate according to the above criteria.

Insert Table 1 about here

Second, the proportion of respondents consenting to validation should have been at least 70% where minimal effort was required from the respondent or at least 60% where validation placed greater demands on the respondent, such as where it formed part of a detailed and time-consuming physical examination which required the respondent to travel to a testing centre. High consent to validation is important since a potential reason for non-consent may be the individual's knowledge that the self-reported response was inaccurate. Indeed, some smoking cessation validation studies assume that self-reported quitters not consenting to validation are actually still smoking and include them in their analyses as such[119;120]. However, given the comparatively low demand characteristics inherent in the community survey, such an assumption was not made in this review. The lower consent to validation rates were considered acceptable where validation

imposed a greater burden on the respondent as it was considered likely that a larger proportion of the non-consenters would have refused due to the additional burden imposed rather than due to the knowledge that their self-report was inaccurate.

Third, the sample size employed in the study should have allowed the estimation of sensitivity and specificity at a 95% confidence interval of width ± 10%. The sample size necessary for this degree of precision is dependent on the sensitivity and specificity estimates. The required sample size can be calculated using the following formula where N = the required sample size, SE = the sensitivity estimate, SP = the specificity estimate and ↙ = the width of the confidence interval[121].

$$N = \frac{1.96^2 * SE * (1 - SE)}{↙^2} + \frac{1.96^2 * SP * (1 - SP)}{↙^2}$$

Given the worst case scenario, where both sensitivity and specificity are 50%, a sample size of 192 would be required to estimate sensitivity and specificity at a 95% confidence interval of width ± 10%. Therefore, only studies with at least 192 subjects were included in this review. Studies with larger sample sizes or with higher or lower sensitivity and specificity estimates would permit more accurate estimation of sensitivity and specificity.

Fourth, the time between self-report and validation of physical measures, such as smoking status and weight, should not exceed an average of one month as the likelihood of a respondent's status having changed increases as the time between self-report and validation increases, leading to inaccurate conclusions about the accuracy of self-report.

Table 2 shows how the 66 studies located performed on each of the inclusion criteria described above and which studies were ultimately included in the next stage of this review.

Insert Table 2 about here

Of the 66 studies reviewed, 25 (38%) used inadequate gold standards[2;62;65;68-72;74;82-85;90;94-97;100;101;105;107-110]; 18 (27%) had inadequate consent rates to validation[61;83;87] or did not report the consent rate to validation[52;53;66;70-72;79-81;85;91;96;107;108;110]; 19 (29%) employed sample sizes incapable of estimating sensitivity and specificity at a 95% confidence interval of width ± 10%[39;57;62;64;66;67;71;78;80;83;90;93;95;96;105;107-110]; and 12 (29%) of the 42 studies involving physical measurements allowed more than one month between self-report and validation[56;57;61;66;83;92;94] or did not report the time lag from self-report to validation[52;62;68;72;91].

Only 24 (36%) studies[3;4;40;50;51;54;55;58-60;63;73;75-77;86;88;89;98;99;102-104;106] met all of the relevant inclusion criteria and were included in the next stage of the review. The failure of more than half the studies to meet or even address these minimal criteria suggests a need for a more standard way of investigating and reporting this type of data. Even with the studies which reported the information required for the inclusion criteria, it was often necessary to refer back to one or more previous publications in order to gather the required information.

Of particular concern was the application of inappropriate gold standards in over a third of the studies located. This is a cause for concern as the use of inappropriate gold standards can lead only to erroneous conclusions about the accuracy of self-report and, therefore, provides no beneficial knowledge to researchers. For example, a number of studies investigating the accuracy of self-reported cancer screening employed physicians' records as the gold

standard[90;100;101]. As the accuracy of physicians' records is questionable, especially in relation to normative, minor events[118], it would be impossible to know what proportion of any discrepancy between self-report and physicians' records to attribute to inaccurate self-report, as opposed to inaccurate recording in the physicians' records.

The lack of adequate gold standards in many areas, especially in the assessment of dietary intake and physical activity level, makes it difficult for researchers wishing to explore the accuracy of self-reported information. A number of gold standards have been suggested in these areas, such as the doubly-labelled water method for measuring total energy expenditure, which has been developed and extensively tested on small animals by Lifson et al[122;123]. However, the testing of this and other energy expenditure assessment methods on humans has largely been restricted to studies involving only a handful of self-selected subjects who are often studied under controlled laboratory conditions and further work is required if such methods are to become acceptable gold standards when assessing energy intake and expenditure among the general population.

Review Results

The Accuracy of Individuals' Risk Factor Status Derived from Self-Report

The next stage was to review the accuracy of individuals' self-report found in these studies. Self-report data can be used in two ways: to establish an individual's risk factor status or to establish the prevalence of a given risk factor in the community. One of the problems with the literature in this area is that a number of different statistical methods have been employed when reporting on the accuracy of self-report. Some of these methods can lead to erroneous conclusions about the accuracy of self-report. For example, some studies reported a kappa statistic which indicates the

degree of agreement between two sources of data[124]. However, a "good" degree of agreement in a kappa statistic does not necessarily translate to high sensitivity and specificity and it gives little information about where the disagreement lies. Therefore, in order to allow for easier comparison between the studies, in this review, individual accuracy was assessed in a number of ways. Each of these methods is described below, accompanied by an example analysis and interpretation based on the hypothetical data included in Figure 1.

Insert Figure 1 about here

1) Sensitivity [a□(a+c)] represents the proportion of true positives[125] - in the example, sensitivity = 98%, meaning that 98% of the women who had truly had a pap test correctly self-reported their screening status.
2) Specificity [d□(b+d)] represents the proportion of true negatives[125] - in the example, specificity = 45%, meaning that 45% of the women who had truly not had a pap test correctly self-reported their screening status.
3) Positive predictive value (PPV) [a□(a+b)] represents the proportion of false positives[125] - in the example, PPV = 72%, meaning that 72% of the women who self-reported having had a pap test correctly reported their screening status.
4) Negative predictive value (NPV) [d□(c+d)] represents the proportion of false negatives[125] - in the example, NPV = 93%, meaning that 93% of the women who self-reported not having had a pap test correctly reported their screening status.
5) Percentage misclassified [(b+c)□(a+b+c+d)] represents the proportion of people classified differently on the self-report and the gold standard - in the example, percentage misclassified = 24%.

6) Percentage of those at risk missed represents the proportion of people truly "at risk", according to the gold standard, who were not identified by self-report. This was calculated as [c□(a+c)] for risk behaviours such as smoking and as [b□(b+d)] for preventive behaviours such as pap testing - in the example, the percentage of those at-risk missed = 55%, meaning that 55% of the women truly at-risk were not identified by self-report.

Table 3 summarises the accuracy data for each of the studies considered eligible for inclusion in this stage of the review.

Insert Table 3 about here

The accuracy of self-reported health risk status varied from study to study and from behaviour to behaviour. Overall, when investigating risk behaviours, such as smoking, between 7% and 22% of "at-risk" individuals were not identified by self-report. When investigating screening histories and risk factor status, the proportion of "at-risk" individuals not identified by self-report increased to between 4% and 76%.

The Accuracy of Risk Factor Prevalences Derived from Self-Report

While acknowledging the existence of problems with assessing individuals' risk factor status, some studies have claimed that self-report data can still provide accurate prevalence estimates of the investigated risk factors at the community level[53;57;95]. Table 4 indicates the estimated prevalence of risk factors based on the self-reported and the gold standard information.

Insert Table 4 about here

As there were sufficient grounds to predict that self-reported risk factor prevalences would be less than gold standard prevalences, one-tailed z tests were conducted to assess whether significant differences existed between the prevalences estimated by self-report and by the corresponding gold standards. As shown in Table 4, the prevalence of risk factors, according to the gold standards, was significantly underestimated by self-report in the majority of cases. In only five cases did the self-report prevalence estimate exceed the gold standard prevalence estimate: the prevalence of smoking was overestimated by 2%[75]; need for a digital rectal exam by 3-13%[98]; hypertension by 3-5%[88;89]; and elevated cholesterol by 8%[88]. Therefore, on the whole, considerable underestimations of risk factor prevalence would have occurred had estimates been based solely on self-reported information. Of all the risk factors investigated, the self-reported prevalence estimates for smoking were the most accurate. This is most likely due to the fact that an individual's current smoking status is a relatively stable and concrete risk factor which requires no recall of distant events and to the low "demand" setting in the studies reviewed.

Predictors of Inaccurate Self-Report

Of the 24 studies included in the latter stages of this review, 15 (63%) discussed predictors of inaccurate self-report[3;4;50;51;54;55;59;63;73;89;98;99;102;103;106]. These findings are summarised in Table 5.

Insert Table 5 about here

A variety of variables were found to correlate, in a variety of directions, with self-report accuracy, making it difficult to draw any solid conclusions. However, some trends did emerge. Gender and education level were found to have little effect on accuracy[3;86;89;102;103;106], while for studies investigating weight, women tended to more consistently underestimate[4;51;54;55;59;63] whereas men

tended to be more inaccurate overall, but in both directions[51;54;59]. Being male and older were characteristics frequently associated with increased inaccuracy in self-reported height[4;50;51;54;55]. Increases in elapsed time from the date of screening also contributed to increased inaccuracy in self-report[89;99;103].

Predictors of Inaccurate Self-Report

Of the 24 studies included in the latter stages of this review, only 6 (25%) discussed predictors of non-consent to validation[4;60;75;99;104;106] finding quite different results. These results are summarised in Table 6.

Insert Table 6 about here

Discussion

With the exception of current smoking and diabetic status, this review found significant, and often substantial, differences between community prevalences of health behaviours and risk factors estimated from self-report data and from corresponding gold standard data. When assessing individuals' risk factor status, the proportion of individuals at-risk who would not have been identified by their self-report data was of concern for all the behaviours explored and quite substantial for many behaviours.

These findings have implications for public health research. At the individual level, for example, only 40% – 79% of women's claims to be adequately screened for cervical cancer could be verified by the gold standard data. In other words, 21% - 60% of women claiming to be adequately screened apparently were not. Therefore, where a population-based study, using self-report data alone, reported that, say, 70% of women were adequately screened for cervical cancer, using the above results to adjust this figure to compensate for the inaccuracy of self-report would result in the prevalence of truly adequately-screened women dropping to only 28% - 55%. This is quite a sizeable difference and would have quite different implications for public health researchers than the 70% figure.

Why is Self-Report Inaccurate?

The large variation in the accuracy of self-report in these studies could be due to a number of factors. First, respondents may have trouble remembering exactly when they last engaged in a specific behaviour. For example, questions about screening history, which often require respondents to recall events which may have occurred a number of years earlier, may be prone to "telescoping error", whereby an event is recalled to have happened more recently than it actually did[126;127]. This explanation is supported by the studies located which assessed the accuracy of self-reported screening history[3;40;99;102].

Second, respondents may lack the knowledge to accurately answer the questions posed. For example, questions about screening history require the respondent to have knowledge of both the screening test at issue and of whether or not they have undergone that test. For example, Pap tests may be conducted, without the patient realising, as part of other gynaecological investigations. Similarly, when asked their weight or height, someone who has not measured or weighed themself for some time is forced to estimate. This has been suggested to explain the preference shown for rounding-off self-reported weights and heights to the nearest five or ten units[55].

Third, poorly designed survey instruments could result in respondents not fully comprehending the questions posed. For example, inadequate or confusing response options may leave respondents unsure of how to answer questions.

Fourth, the demand characteristics inherent in the survey situation may result in respondents being untruthful in an attempt to respond in a socially desirable manner. Such socially desirable

responses have been widely demonstrated in many survey situations[5;47-49;116]. In an attempt to minimise the impact of this factor, only studies involving relatively low demand situations were included in this review. However, the current trend of public education programs about ways in which people can reduce their risk of developing cancer and cardiovascular disease may be a sufficient demand characteristic to reduce the accuracy of self-reported health status in these areas. This has been described as the "tyranny of health promotion"[128], whereby individuals who engage in health risk behaviours or who do not engage in preventive health behaviours are made to feel responsible for any subsequent disease, making them less likely to admit to being "at-risk".

Fifth, very few, if any, gold standards have 100% sensitivity and specificity, meaning that measurement errors in the gold standards considered acceptable for this review are likely to account for some of the discrepancies between the self-reported and the gold standard data. However, given the restrictions placed on gold standards in this review, it is unlikely that such measurement errors in the gold standards could account for the magnitude of the discrepancies found.

Another potential source of error in the estimation of risk factor status and prevalence lies in the potential differences between responders and non-responders and, in validation studies, between respondents consenting to and not consenting to validation. Only a quarter of the reviewed studies explored the predictors of non-consent to validation, finding rather contradictory results. Therefore, as with the predictors of inaccurate self-report, it is difficult to draw any firm conclusions. However, it is an area that warrants further investigation.

Recommendations for the Future

This review has identified a number of deficits in self-reported health data which should be addressed in future research. For studies estimating the prevalence of health risk factors in the general population, it is recommended that self-report data should not be used as the only data source. Researchers should investigate alternative methods of estimating prevalences and, where possible, utilise existing objective data sources. In Australia, for example, all medical services provided by family physicians are uniquely coded and recorded on computer by a government body called the Health Insurance Commission, as part of our Medicare billing process. This database, which is quite separate from physicians' own, usually paper-based, patients' records, can be accessed by researchers to provide aggregate data on specific health services, such as pap tests conducted, for specified regions. Of course, safeguards are employed to ensure that both individual patients' and individual physicians' confidentiality is preserved. Many government bodies have similar types of databases accessible to researchers in similar ways.

Relying on existing government databases becomes more difficult when researchers are interested in exploring predictors of increased risk factor status or trying to identify individuals considered "at-risk". This is difficult as few of the databases contain much information on individuals' demographic or other characteristics. However, this problem could possibly be overcome by conducting surveys to obtain the necessary additional information and asking respondents for written, informed consent to access their existing medical or pathology records. Although this would undoubtedly extend the time taken to collect data, the improvement in data quality may be worth it.

Where no alternatives to self-report data exist, it is strongly recommended that researchers explore the existing literature for previous studies which have investigated the accuracy of self-report for the risk factors of interest, consider factors which may affect the accuracy of respondents' self-report and employ strategies which have been shown to maximise the accuracy of self-reported information. These strategies include ensuring that respondents fully understand the questions posed[126;127]; phrasing questions in such a way as to minimise socially desirable responses[5;47-49;116]; using bounded recall to improve respondents' recall of more temporally-distant events[126;127]; encouraging respondents to provide exact rather than rounded-off answers to questions regarding continuous variables, such as weight and height[129]; ensuring the questions have clear, exhaustive, mutually exclusive response options which assess the behaviour of interest[126;127]; and employing "bogus pipeline" techniques whereby respondents are deceived, prior to answering, into believing that their self-reported response to some behavioural question will be objectively verified[130-134]. The latter technique may prove one of the most effective as it has been found to produce higher, and presumably more accurate, self-reported estimates of smoking status[131-134] and alcohol intake[130;132;133]. Although limited to risk factors accessible to some form of objective assessment, one could certainly envisage this technique being applied in the field of dietary intake and energy expenditure.

In addition, when employing self-reported health information, it is strongly recommended that researchers routinely report on the data collection methods employed, highlighting why it was necessary to use self-report, the steps taken to maximise self-report accuracy and the adjustments made to compensate for the inaccuracies or an indication of how inaccuracy is likely to have affected the data collected.

Wherever possible, researchers using self-reported health information should include validation sub-studies. Not only would this allow the researcher to adjust their overall findings to account for the degree of inaccuracy found in the sub-study, it would also provide invaluable reference data for future researchers whose budget or target group prohibited validation studies or the use of objective data sources. Therefore, researchers conducting any type of self-report validation study should be encouraged to routinely report on the adequacy of the gold standard employed, the rate of consent to validation, predictors of consent to validation and, where appropriate, the time lag from self-report to validation.

Following on from this point, a valuable direction for future research into the accuracy of self-reported health information would be to conduct large-scale health risk factors validation studies among the general population. Such studies could provide data for use in the development of age- and gender-specific correction factors to be applied to self-reported continuous variables such as height, weight and recency of screening tests. Although adequate gold standards do not exist for all risk factors, such data available for pap tests could be used to give a general guide to "telescoping" errors when reporting the recency of screening tests.

Finally, other areas where further research could improve the quality of research in the field of health promotion include more research into ways to maximise the accuracy of self-report, more research into alternative, creative methods of assessing risk factor status and more research into additional valid and reliable gold standards against which to evaluate the accuracy of self-report.

Conclusions

This review casts serious doubts on the wisdom of relying exclusively on self-reported health information and concludes that caution should be exercised both when trying to identify "at-risk" individuals and when estimating the prevalence of risk factors in the general population. In summary, this review has suggested a number of ways in which the accuracy of individuals' self-reported health information can be maximised but strongly recommends the use of validation techniques wherever possible, as well as the use of objective data wherever possible for determining community prevalence estimates. Furthermore, it stresses the need for further research to allow the calculation of correction factors for community-based prevalence estimates, to work towards the development of more valid and reliable gold standards and to develop more standard ways of investigating and reporting studies assessing the accuracy of self-reported health information.

Acknowledgements

The preparation of this paper was supported jointly by the National Health and Medical Research Council and the New South Wales Cancer Council Cancer Education Research Program, directed by R. Sanson-Fisher until November 1997. The views expressed are not necessarily those of the Cancer Council.

References

1. Hawe P, Degeling D, Hall J. Evaluating Health Promotion: A Health Worker's Guide, Sydney, MacLennan & Petty; 1990.

2. Washburn RA, Goldfield SRW, Smith KW, McKinlay JB. The validity of self-reported exercise-induced sweating as a measure of physical activity. Am J Epidemiol 1990;132(1):107-113.

3. Gordon NP, Hiatt RA, Lampert DI. Concordance of self-reported data and medical record audit for six cancer screening procedures. J Natl Cancer Inst 1993;85(7):566-570.

4. Bowlin SJ, Morrill BD, Nafziger AN, Jenkins PL, Lewis C, Pearson TA. Validity of cardiovascular disease risk factors assessed by telephone survey: the behavioural risk factor study. J Clin Epidemiol 1993;46(6):561-571.

5. Strecher VJ, Becker MH, Clark NM, Prasada-Rao P. Using patients' descriptions of alcohol consumption, diet, medication compliance and cigarette smoking: the validity of self-reports in research and practice. J Gen Intern Med 1989;4:160-166.

6. Baranowski T. Methodologic issues in self-report of health behaviour. J Sch Health 1985;55(5):179-182.

7. United States' National Library of Medicine. MEDLINE on Silver Platter database. 1994. Bethseda, Maryland, United States' National Library of Medicine.

8. Reid SE, Simpson JM, Britt HC. Pap smears in general; a secondary analysis of the Australian morbidity and treatmet survey 1990 to 1991. Aust J Public Health 1997;21:257-264.

9. Peragallo NP, Alba ML, Tow B. Cervical cancer screening practices among Latino women in Chicago. Public Health Nursing 1997;14:251-255.

10. Strauss KF, Mokdad A, Ballew C, et al. The health of Navajo women: Findings from the Navajo health and nutrition survey. J Nutr 1997;suppl.:2128s-2133s.

11. Solberg LI, Brekke ML, Kottke TE. Are physicians less likely to recommend preventative services to low-SES patients. Prev Med 1997;26:350-357.

12. Stuart GCE, McGregor E, Duggan MA, Nation JG. Review of the screening history of Alberta women with invasive cervical cancer. Can Med Assoc J 1997;157:513-519.

13. Cassard SD, Weisman CS, Plichta SB, Johnson TL. Physician gender and womens preventive services. Journal of Womens Health 1997;6:199-207.

14. Rakowski W, Clark MA, Pearlman DN, et al. Integrating pros and cons for mammography and pap testing; Extending the construct of decisional balance to two behaviors. Prev Med 1997;26:664-673.

15. McPhee SJ, Bird JA, Davis T. Barriers to breast and cevical cance screening among Vietnamese-American women. Am J Prev Med 1997;13:205-213.

16. Ruchlin HS. Prevalence and correlates of breast and cervical cancer among older women. Obstet Gynecol 1997;90:16-21.

17. O'Malley AS, Mandelblatt J, Gold K, Cagney KA, Kerner J. Continuity of care and the use of breast and cervical cancer screening services in a multiethnic community. Arch Intern Med 1997;157:1462-1470.

18. Lantz PM, Weigers ME, House JS. Education and income differentials in breast and cervical cancer screening. Med Care 1997;35:219-236.

19. Suarez L, Roche RA, Nichols D, Simpson DM. Knowledge, behaviour and fears concerning breast and cervical cancer among older low income Mexican-American women. Am J Prev Med 1997;13:137-142.

20. Branoff.R, Santi K, Campbell JK, Roetzheim R, Oler M. A family practice residency cervical screening project: Perceived screening barriers. Fam Med 1997;29:119-123.

21. Sung JF, Blumenthal DS, Coates RJ, Alema-Mensah E. Knowledge, beliefs,attitudes, and cancer screening among inner-city African-American women. Journal of the National Medical Association 1997;89:405-411.

22. Lazcano-Ponce EC, Najera-Aguilar P, Buiatti E, et al. The cervical cancer screening program in mexico: Problems with access and coverage. Cancer Causes and Control 1997;8:698-704.

23. Lavin C, Goodman E, Perlman S, Kelly lS, Emans SJ. Follow-up of abnormal papanicolou smears in a hospital based adolescent clinic. J of Ped Adolesc Gynecol 1997;10:141-145.

24. Fox P, Arnsberger P, Zhang X. An examination of differential follow-up rates in cervical cancer screening. J Community Health 1997;22:199-209.

25. Burak LJ, Meyer M. Using the health belief model to examine and predict college womens cervical cancer screening beliefs and behavior. Health Care Women Int 1997;18:251-262.

26. Miller MM, Siejak KK, Schroeder CM, Lerman C, Hernandez E, Helm W. Enhancing adherence following abnormal pap smears among low income minority women: a preventive telephone counselling strategy. J Natl Cancer Inst 1997;89:703-708.

27. Lurie N, Margolis K, McGovern PG, Mink PJ, Slater JS. Why do patients of female physicians have higher rates of breast and cervical cancer screening. J Gen Intern Med 1997;12:43.

28. Sanchez V, Rohlfs I, Borrell C. Migration within Spain, level of education, and cervical cancer screening. European Journal of Cancer Prevention. 1997;6:31-37.

29. Reding DJ, Lappe K, Krueger M, Kolehouse BL, Steneil DLRA. Cancer Screening and prevention in rural Wisconsin:The greater Marshfield experience. Wisconsin Medical Journal 1997;August:32-37.

30. Sung JF, Blumenthal DS, Coates RJ, Williams JE, Alema-Mensah E, Liff JM. Effect of a cancer screening intervention conducted by lay health workers among inner city women. Am J Prev Med 1997;13:51-57.

31. Pengsaa P, Vatanasapt V, Scriamporn S, et al. A self-administered device for cervical cancer screening in northeast Thailand. Acta Cytologica 1997;41:749-754.

32. Kant AC, Palm BTHM, Wentink E, van Weel C. General practitioner based screening for cervical cancer: Higher participation of women with a higher risk? J Med Screen 1997;4:35-39.

33. Palm BTHM, Kant AC, Visser EA, Vooijs GP, Van Den Bosch WJHM, van Weel C. The effect of the family physician on improving follow-up after an abnormal PAP smear. Int J for Quality in Health Care 1997;9:277-282.

34. Somkin CP, Hiatt RA, Hurley LB, Gruskin E, Ackerson L, Larson P. The effect of patient and provider reminders on mammography and papanicolou smear screening in a large health maintenance organisation. Arch Intern Med 1997;157:1658-1664.

35. Buehler SK, Parsons WL. Effectiveness of a call/recall system in improving compliance with cervical cance screening: a randomised controlled trial. Can Med Assoc J 1997;157:521-526.

36. Mak D, Straton JAY. Effects and sustaiability of a cervical screening program in remote Aboriginal Australia. Aust J Public Health 1997;21:67-70.

37. Mitchell H, Hirst S, Mitchell A, Staples M, Torcello N. Effect of ethnic media on cervical cancer screening rates. Aust J Public Health 1997;21:265-267.

38. Bowman JA, Sanson-Fisher R, Redman S. The accuracy of self-reported pap smear utilisation. Soc Sci Med 1997;44:969-976.

39. Bowman JA, Redman S, Dickinson JA, Gibberd R, Sanson-Fisher RW. The accuracy of pap smear utilisation self-report: a methodological consideration in cervical screening research. Health Serv Res 1991;26(1):97-107.

40. Fruchter RG, Rones K, Roth T, Webber CA, Camilien L, Boyce JG. Pap smear histories in a medical clinic: accuracy of patients' self-report. NY State J Med 1992;92:421-424.

41. Young DW. Comparison of information collected by a questionary with that in the patient hospital record. Methods Inf Med 1972;11:20-22.

42. Glass R, Johnson B, Vessey M. Accuracy of recall of histories of oral contraceptive use. Br J Prev Soc Med 1974;28:273-275.

43. Pernanen K. Validity of survey data on alcohol use, in Gibbens RJ (ed): Research Advances in Alcohol and Drug Problems, Volume 1. New York, Wiley; 1974.

44. Australian Bureau of Statistics: Causes of Death, Australia 1992. (ABS Catalogue number: 3303.0), Canberra, Australian Government Publishing Service; 1992.

45. United States' Preventive Services Task Force. Guide to Clinical Preventive Services, Baltimore, Williams & Wilkins; 1996.

46. Nutbeam D, Wise M, Bauman A, Harris E, Leeder S. Goals and Targets for Australia's Health in the Year 2000 and Beyond (Report prepared for the Commonwealth Department of Health, Housing and Community Services), Canberra, Australian Government Publishing Services; 1993.

47. Beach DL, Mayer JA. The effects of social demand on breast self-examination self- report. J Behav Med 1990;13(2):195-205.

48. Delamater AM, Kurtz SM, White NH, Santiago JV. Effects of social demand on reports of self-monitored blood glucose in adolescents with Type 1 Diabetes Mellitus. J Appl Psychol 1988;18(6):491-502.

49. Furnham A. Response bias, social desirability and dissimulation. Person Indiv Diff 1986;7(3):385-400.

50. Stewart AW, Jackson RT, Ford MA, Beaglehole R. Underestimation of relative weight by use of self-reported height and weight. Am J Epidemiol 1987;125(1):122-126.

51. Waters A-M. Assessment of self-reported height and weight and their use in the determination of body mass index: analysis of data from the 1989 Risk Factor Prevalence Survey, Canberra, Australian Institute of Health & Welfare; 1993.

52. Lackland DT, Wheeler FC. The need for accurate hutrition survey methodology: the South Carolina experience. J Nutr 1990;120:1433-1436.

53. Tienboon P, Wahlqvist ML, Rutishauser IHE. Self-reported weight and height in adolescents and their parents. J Adolesc Health 1992;13:528-532.

54. Jalkanen L, Tuomilheto J, Tanskanen A, Puska P. Accuracy of self-reported body weight compared to measured body weight: a population survey. Scand J Soc Med 1987;15:191-198.

55. Rowland ML. Reporting bias in height and weight data. Stat Bull Metrop Insur Co 1989;70(2):2-11.

56. Kuskowska-Wolk A, Bergstrom R, Bostrom G. Relationship between questionnaire data and medical records of height, weight and body mass index. Int J Obes 1991;16:1-9.

57. Le Marchand L, Yoshizawa CN, Nomura AMY. Validation of body size information on drivers' licenses. Am J Epidemiol 1988;128(4):874-877.

58. Kuskowska-Wolk A, Karlsson P, Stolt M, Rossner S. The predictive validity of body mass index based on self- reported weight and height. Int J Obes 1989;13:441-453.

59. Schmidt MI, Duncan BD, Tavares M, Polanczyk CA, Pellanda L, Zimmer PM. Validity of self-reported weight: a study of urban Brazilian adults. Rev Saude Publica 1993;27(4):271-276.

60. Bowlin SJ, Morrill BD, Nafziger AN, Lewis C, Pearson TA. Reliability and changes in validity of self-reported cardiovascular disease risk factors using dual response: The behavioural risk factor survey. J Clin Epidemiol 1996;49:511-517.

61. Roberts JG. Can self-report accuracy describe the prevalence of overweight. Public Health 1995;109:275-284.

62. Weaver TW, Kushi LH, McGovern PG, et al. Validation study of self-reported measures of fat distribution. Int J Obes 1996;20:644-650.

63. Hall TR, Young TB. A validation study of body fat distribution as determined by self-measurement of waist and hip circumference. Int J Obes 1989;13:801-807.

64. Freudenheim JL, Darrow SL. Accuracy of self-measurement of body fat distribution by waist, hip and thigh circumferences. Nutr Cancer 1991;15:179-186.

65. Stevens J, Keil JE, Waid R, Gazes PC. Accuracy of current, 4-year, and 28-year self reported body weight in an elderly population. Am J Epidemiol 1990;132:1156-1163.

66. Kushi LH, Kaye SA, Folsom AR, Soler JT, Prineas RJ. Accuracy and reliability of self-measurement of body girths. Am J Epidemiol 1988;128:740-748.

67. Meier SE, Brigham TA, Handel G. Accuracy of drinkers' recall of alcohol consumption in a field setting. J Stud Alcohol 1987;48(4):325-328.

68. Cutler SF, Wallace PG, Haines AP. Assessing alcohol consumption in general practice patients - a comparison between questionnaire and interview (findings of the Medical Research Council's general practice research framework study on lifestyle and health). Alcohol Alcohol 1988;23(6):441-450.

69. Hoyer G, Nilssen O, Brenn T, Schirmer H. The Svalbard study 1988-89: a unique setting for validation of self-reported alcohol consumption. Addiction 1995;90:539-544.

70. Gronbaek M, Heitmann BL. Validity of self-reported intakes of wine, beer and spirits in population studies. European Journal of Clinical Nutrition 1996;50:487-490.

71. Duckert F, Johnsen J, Amundsen A, Stromme J, Morlands J. Co-variation between biological markers and self reported alcohol consumption. Alcohol Alcohol 1992;27:545-555.

72. Kornitzer M, Vanhemeldonck A, Bourdoux P, De Backer G. Belgian heart disease prevention project: comparison of self- reported smoking behaviour with serum thiocyanate concentrations. J Epidemiol Community Health 1983;37:132-136.

73. Wagenknecht LE, Burke GL, Perkins LL, Haley NJ, Friedman GD. Misclassification of smoking status in the CARDIA study: a comparison of self-report with serum cotinine levels. Am J Public Health 1992;82(1):33-36.

74. Coultas DB, Howard CA, Peake GT, Skipper BJ, Samet JM. Discrepancies between self-reported and validated cigarette smoking in a community survey of New Mexico Hispanics. Am Rev Respir Dis 1988;137:810-814.

75. Pierce JP, Dwyer T, DiGiusto E, et al. Cotinine validation of self-reported smoking in commercially run community surveys. J Chron Dis 1987;40(7):689-695.

76. Fortmann SP, Rogers T, Vranizan K, Haskell WL, Solomon DS, Farquhar JW. Indirect measures of cigarette use: expired air carbon monoxide versus plasma thiocyanate. Prev Med 1984;13:127-135.

77. Suadicani P, Hein HO, Gyntelberg F. Serum validated tobacco use and social inequalities in risk of ischaemic heart disease. Int J Epidemiol 1994;23:293-300.

78. Wills TA, Cleary SD. The validity of self-reports of smoking:Analyses by race/ethnicity in a school sample of urban adolescents. Am J Public Health 1997;87:56-61.

79. van Vanakis H, Tashkin DP, Simmons M, Hilda BG, Clark VA. Relative sensitivity and specificity of salivary and serum cotinine in identifying tobacco-smoking status of self-reported nonsmokers and smokers of tobacco and/or marijuana. Archives of Environmental Health 1989;44:53-58.

80. Slattery ML, Hunt SC, French TK, Ford MH, Williams RR. Validity of cigarette smoking habits in three epidemiologic studies in Utah. Prev Med 1989;18:11-19.

81. Haddow JE, Palomaki GE, Knight GJ. Use of serum cotinine to assess the accuracy of self-reported non-smoking. BMJ 1986;293:1306-1306.

82. Smith KW, McKinlay SM, McKinlay JB. The validity of heart risk appraisals for coronary heart disease: results from a randomised field trial. Am J Public Health 1991;81(4):466-470.

83. Arroll B, Jackson R, Beaglehole R. Validation of a three-month physical activity recall questionnaire with a seven-day food intake and physical activity diary. Epidemiol 1991;2(4):296-299.

84. Chase DR, Harada M. Response error in self-reported recreation participation. J Leisure Res 1984;16(4):322-329.

85. Cousins SO. Validity and reliability of self-reported health of persons aged 70 and over. Health Care Women Int 1997;18:165-174.

86. Brown JB, Adams ME. Patients as reliable reporters of medical care process: recall of ambulatory encounter events. Med Care 1992;30(5):400-411.

87. Giles WH, Croft JB, Kreenan NL, Lane MJ, Wheeler FC. The validity of self-reported hypertension and correlates of hypertension awarness among blacks and whites within the stroke belt. Am J Prev Med 1995;11:163-169.

88. Robinson JR, Young TK, Roos LL, Gelskt DE. Estimating the burden of disease: Comparing administrative data and self-reports. Med Care 1997;35:932-944.

89. Vargas CM, Burt VL, Gillum RF, Pamuk ER. Validity of self-reported hypertension in the national health and nutrition examination survey III, 1988-1991. Prev Med 1997;26:678-685.

90. Horwitz SM, Prados-Torres A, Singer B, Bruce ML. The influence of psychological and social factors on accuracy of self-reported blood pressure. J Clin Epidemiol 1997;50:411-418.

91. Ford ES, Harel Y, Heath G, Cooper RS, Caspersen CJ. Test characteristics of self-reported hypertension among the hispanic population:Findings from the hispanic health and nutrition survey. J Clin Epidemiol 1990;43:159-165.

92. Ernstoff LT, Thorn M, Tosteson TD, et al. The accuracy of skin self-examination for atypical nevi. Epidemiol 1996;7:619-623.

93. Little P, Keefe M, White J. Self screening for risk of melanoma: validity of self mole counting by patients in a single general practice. BMJ 1995;310:912-916.

94. De Vries MJ, Zock PL, Mensink RP, Katan MB. Underestimation of energy intake by 3-d records compared with energy intake to maintain body weight in 269 nonobese adults. Am J Clin Nutr 1994;60:855-860.

95. Coulter A, Baldwin A. Survey of population coverage in cervical cancer screening in the Oxford region. J R Coll Gen Prac 1987;37:441-443.

96. Sawyer JA, Earp JA, Fletcher RH, Daye FF, Wynn TM. Accuracy of women's self-report of their last pap smear. Am J Public Health 1989;79(8):1036-1037.

97. Walter SD, Clarke EA, Hatcher J, Stitt LW. A comparison of physician and patient reports of pap smear histories. J Clin Epidemiol 1988;41(4):401-410.

98. Hiatt RA, Perez-Stable EJ, Quesesnberry C, Sabogal F, Otero-Sabogal R, McPhee SJ. Agreement between self-reported early cancer detection practices and medical audits among hispanic and non-hispanic white health plan members in northern california. Prev Med 1995;24:278-285.

99. Bowman JA, Sanson-Fisher R, Redman S. The accuracy of self-reported pap smear utilisation. Soc Sci Med 1997;44:969-976.

100. Suarez L, Goldman DA, Weiss NS. Validity of pap smear and mammogram self-reports in a low-income Hispanic population. Am J Prev Med 1995;11:94-98.

101. Paskett ED, Tatum CM, Mackay J, Hoen H, Case L, Velez R. Validation of self-reported breast and cervical cancer screening tests among low income minority women. Cancer Epidemiology, Biomarkers and Prevention 1996;5:721-726.

102. Degnan D, Harris R, Ranney J, Quade D, Earp JA, Gonzalez J. Measuring the use of mammography: two methods compared. Am J Public Health 1992;82(10):1386-1388.

103. Zapka JG, Bigelow C, Hurley T, et al. Mammography use among sociodemographically diverse women:The accuracy of self report. Am J Public Health 1996;86:1016-1021.

104. Crane LA, Kaplan CP, Bastani R, Scrimshaw SCM. Determinants of adherence among health department patients reffered for a mammogram. Women & Health 1996;24:43-64.

105. Fulton-Kehoe D, Burg MA, Lane DS. Are self-reported dates of mammograms accurate? Public Health Rev 1992;93:233-240.

106. Tsubono Y, Fukao A, Hisamichi S, Hosokawa T, Sugawara N. Accuracy of self-report for stomach cancer screening. J Clin Epidemiol 1994;47:977-981.

107. Parent M-E, Ghadirian P, Lacroix A, Perret C. Accuracy of reports of familial breast cancer in a case-control series. Epidemiol 1995;6:184-186.

108. Silberberg J, Wlodarczyk J, Hensley WJ, et al. The accuracy of reported family history of heart disease: The impact of "don't know" responses. Aust NZ J Med 1994;24:386-389.

109. Aitken J, Bain C, Ward PH, Siskind V, MacLennan R. How accurate is self-reported family history of colorectal cancer? Am J Epidemiol 1995;141:863-871.

110. Kerber RA, Slattery ML. Comparison of self-reported and database-linked family history of cancer data in a case-control study. Am J Epidemiol 1997;146:244-248.

111. Mulrow C, Oxman A (Eds) Cochrane Collaboration Handbook [updated September 1997], Oxford, The Cochrane Collaboration; 1997.

112. Liddle J, Williamson M, Irwig L: Method for Evaluating Research and Guideline Evidence, Sydney, NSW Health Department; 1996.

113. Velicer WF, Prochaska JO, Rossi JS, Snow MG. Assessing outcome in smoking cessation studies. Psych Bull 1992;111(1):23-41.

114. Pojer R, Whitfield JB, Poulos V, Eckhard IF, Richmond R, Hensley WJ. Carboxyhaemoglobin, cotinine and thiocyanate assay compared for distinguishing smokers from non-smokers. Clin Chem 1984;30(8):1377-1380.

115. Prignot J. Quantification and chemical markers of tobacco-exposure. Eur J Respir Dis 1987;70:1-7.

116. Lee PN. Misclassification of Smoking Habits and Passive Smoking: A Review of the Evidence, Berlin, Springer-Verlag; 1988.

117. Stookey GK, Katz BP, Olson BL, Drook CA, Cohen SJ. Evaluation of biochemical validation measures in determination of smoking status. J Dent Res 1987;66(10):1597-1601.

118. Aaronson LS, Burman ME. Use of health records in research: reliability and validity issues. Res Nurs Health 1994;17:67-73.

119. Jamrozik K, Vessey M, Fowler G, Wald N, Parker G, Van Vunakis H. Controlled trial of three different antismoking interventions in general practice. BMJ 1984;288:1499-1502.

120. Windsor RA, Orleans CT. Guidelines and methodological standards for smoking cessation intervention research among pregnant women: improving the science and art. Health Educ Q 1986;13:131-161.

121. Dobson AJ. Calculating sample size. Trans Menzies Found 1984;7:75-79.

122. Lifson N, Gordon GB, McClintock R. Measurement of total carbon dioxide production of means of De2O18. J Appl Physiol 1955;7:704-710.

123. Lifson N, McClintock R. Theory of use of the turnover rates of body water for measuring energy and material balance. J Theor Biol 1966;12:46-74.

124. Fleiss JL. Statistical Methods for Rates and Proportions, New York, Wiley; 1981.

125. Lemeshow S, Hosmer DWJ, Klar J, Lwanga SK. Adequacy of Sample Size in Health Studies, Chichester, John Wiley & Sons; 1990.

126. Sudman S, Bradburn NM. Asking Questions: A Practical Guide to Questionnaire Design, San Francisco, Jossey-Bass Publishers; 1983.

127. Sudman S, Bradburn NM. Response Effects in Surveys: A Review and Synthesis, Chicago, Adline Publishing Company; 1974.

128. Becker MH. The tyranny of health promotion. Public Health Rev 1987;14:15-25.

129. Rowland ML. Self-reported weight and height. Am J Clin Nutr 1990;52:1125-1133.

130. Lowe JB, Windsor RA, Adams B, Morris J, Reese Y. Use of a bogus pipeline method to increase accuracy of self- reported alcohol consumption among pregnant women. J Stud Alcohol 1986;47(2):173-175.

131. Hansen WB, Malotte CK, Fielding JE. The bogus pipeline revisited: the use of the threat of detection as a means of increasing self-reports of tobacco use. J Stud Alcohol 1985;70(4):789-792.

132. Botvin EM, Botvin GJ, Renick NL, Filazzola AD, Allegrante JP. Adolescents' self-reports of tobacco, alcohol and marijuana use: examining the comparability of video tape, cartoon and verbal bogus pipeline procedures. Psychol Rep 1984;55(2):379-386.

133. Werch CE, Gorman DR, Marty PJ, Forbess J, Brown B. Effects of the bogus pipeline on enhancing validity of self- reported adolescent drug use measures. J Sch Health 1987;57(6):232-236.

134. Aguinis H, Pierce CA, Quigley BM. Conditions under which a bogus pipeline procedure enhances the validity of self-reported cigarette smoking: a meta-analytic review. J Appl Soc Psychol 1993;23(5):353-373.

Table 1: The acceptability, by risk factor, of the gold standards used in the studies located.

Risk factor	Acceptable gold standards	Unacceptable gold standards
Smoking status	* Cotinine * Carbon Monoxide	* Thiocyanate
Cancer screens & family history	* Pathology laboratory records * Radiology centre records * Health insurance records	* Physicians' records
Height, weight & BMI	* Measurements	___
Alcohol intake	* Breathalyser	* Interview
Physical activity	___	* Harvard Alumni Activity Scale * Prospective diary * Pool attendance records
Blood pressure & hypertension	* Measurements	* Physicians' records
Nevi	* Physician examination	___
Diet	___	* Diary
Diabetes	* Measurements	___
Cholesterol level & hypercholesterolemia	* Measurements	___

Table 2: Performance on the inclusion criteria of studies of self-report accuracy, published 1983 - 1997.

Study	Risk factor(s) investigated	Adequate gold standard?	Adequate consent to validation?	Adequate sample size?	≥1 month self-report to validation?	Study included?
Kornitzer et al. (1983)[72]	Smoking	X	N/R	✓	N/R	No
Fortmann et al. (1984)[76]	Smoking	✓	✓	✓	✓	Yes
Haddow et al.(1986)[81]	Smoking	✓	N/R	✓	✓	No
Pierce et al. (1987)[75]	Smoking	✓	✓	✓	✓	Yes
Coultas et al. (1988)[74]	Smoking	X	✓	✓	✓	No
Slattery et al. (1989)[80]	Smoking	✓	N/R	X	✓	No
Van Vanukis et al. (1989)[79]	Smoking	✓	N/R	✓	✓	No
Wagenknecht et al. (1992)[73]	Smoking	✓	✓	✓	✓	Yes
Suadicani et al. (1994)[77]	Smoking	✓	✓	✓	✓	Yes
Wills et al. (1997)[78]	Smoking	✓	✓	X	✓	No

Study	Risk factor(s) investigated	Adequate gold standard?	Adequate consent to validation?	Adequate sample size?	≥1 month self-report to validation?	Study included?
Bowlin et al. (1993)[4]	Smoking, hypertension, blood pressure, diabetes, cholesterol level, hypercholesterolemia, weight, height & body mass index (BMI)	✓	✓	✓	✓	Yes
Bowlin et al. (1996)[60]	Smoking, BMI, hypertension, diabetes,hypercholesterolemia, uncontrolled hypertension	✓	✓	✓	✓	Yes
Robinson et al. (1997)[88]	Hypertension, diabetes, cholesterolemia	✓	✓	✓	✓	Yes
Ford et al. (1990)[91]	Hypertension	✓	N/R	✓	N/R	No
Giles et al. (1995)[87]	Hypertension	✓	X	✓	✓	No
Horwitz et al. (1996)[90]	Hypertension	X	✓	X	N/A	No
Vargas et al. (1997)[89]	Hypertension	✓	✓	✓	✓	Yes
Jalkanen et al. (1987)[54]	Weight	✓	✓	✓	✓	Yes
Stevens et al (1990)[65]	Weight	X	✓	✓	✓	No
Schmidt et al. (1993)[59]	Weight	✓	✓	✓	✓	Yes

Study	Risk factor(s) investigated	Adequate gold standard?	Adequate consent to validation?	Adequate sample size?	≥1 month self-report to validation?	Study included?
Le Marchand et al. (1988)[57]	Weight & height	✓	✓	X	X	No
Stewart et al. (1987)[50]	Weight, height & BMI	✓	✓	✓	✓	Yes
Kuskowska-Wolk et al. (1989)[58]	Weight, height & BMI	✓	✓	✓	✓	Yes
Rowland (1990)[55]	Weight, height & BMI	✓	✓	✓	✓	Yes
Lackland et al. (1990)[52]	Weight, height & BMI	✓	N/R	✓	N/R	No
Kuskowska-Wolk et al. (1991)[56]	Weight, height & BMI	✓	✓	✓	X	No
Tienboon et al. (1992)[53]	Weight, height & BMI	✓	N/R	✓	✓	No
Waters (1993)[51]	Weight, height & BMI	✓	✓	✓	✓	Yes
Roberts (1995)[61]	Weight, height & BMI	✓	X	✓	X	No
Weaver et al. (1996)[62]	Weight , height, body circumferences of waist, hip & bust	X	✓	X	N/R	No
Kushi et al. (1988)[66]	Body fat distribution	✓	N/R	X	X	No
Hall et al. (1989)[63]	Body fat distribution, prescence of central fat pattern (waist :hip ratio)	✓	✓	✓	✓	Yes

Study	Risk factor(s) investigated	Adequate gold standard?	Adequate consent to validation?	Adequate sample size?	≥1 month self-report to validation?	Study included?
Freudenheim et al. (1991)[64]	Body fat distribution			X		No
Meier et al. (1987)[67]	Alcohol intake			X		No
Cutler et al. (1988)[68]	Alcohol intake	X			N/R	No
Duckert et al. (1992)[71]	Alcohol intake	X	N/R	X		No
Hoyer et al. (1995)[69]	Alcohol intake	X			N/A	No
Gronbaek et al. (1996)[70]	Alcohol intake	X	N/R			No
Chase et al. (1984)[84]	Physical activity	X			N/A	No
Washburn et al. (1990)[2]	Physical activity	X				No
Smith et al. (1991)[82]	Physical activity	X				No
Arroll et al. (1991)[83]	Physical activity	X	X	X	X	No
Cousins et al. (1997)[85]	Physical activity	X	N/R		N/A	No
De Vries et al. (1994)[94]	Diet, energy intake	X			X	No

Study	Risk factor(s) investigated	Adequate gold standard?	Adequate consent to validation?	Adequate sample size?	≥1 month self-report to validation?	Study included?
Tsubono et al. (1994)[106]	Gastric photofluorographic screening (for stomach cancer)				N/A	Yes
Coulter et al. (1987)[95]	Pap screening	X		X	N/A	No
Walter et al. (1988)[97]	Pap screening	X			N/A	No
Sawyer et al. (1989)[96]	Pap screening	X	N/R	X	N/A	No
Bowman et al.(1991)[39]	Pap screening			X	N/A	No
Fruchter et al. (1992)[40]	Pap screening				N/A	Yes
Bowman et al. (1997)[99]	Pap screening				N/A	Yes
Degnan et al. (1992)[102]	Mammography screening				N/A	Yes
Fulton-Kehoe et al. (1992)[105]	Mammography screening	X		X	N/A	No
Crane et al. (1996)[104]	Mammography screening				N/A	Yes
Zapka et al. (1996) [103]	Mammography screening				N/A	Yes
Suarez et al. (1995)[100]	Pap and mammography screening	X			N/A	No

Study	Risk factor(s) investigated	Adequate gold standard?	Adequate consent to validation?	Adequate sample size?	≥1 month self-report to validation?	Study included?
Paskett et al. (1996)[101]	Pap and mammography screening	X			N/A	No
Gordon et al. (1993)[3]	Pap, mammography, clinical breast exam, sigmoidoscopy, faecal occult blood test & digital rectal exam screening	(not breast & rectal exam)			N/A	Yes
Hiatt et al. (1995)[98]	Mammography, clinical breast exam, pap smear, sigmoidoscopy, fecal occult blood test,digital rectal exam				N/A	Yes
Brown et al. (1992)[86]	Blood pressure, mammography and digital rectal exam screening	(mamm only)			N/A	Yes
Little et al. (1995)[93]	Body mole counts			X		No
Titus-Ernstoff et al. (1996)[92]	Nevi counts				X	No
Silberberg et al. (1994)[108]	Family history of heart disease and cancer	X	N/R	X	N/A	No
Aitken et al. (1995)[109]	Family history of colorectal cancer	X		X	N/A	No
Parent (1995)[107]	Family history of breast cancer	X	N/R	X	N/A	No

Study	Risk factor(s) investigated	Adequate gold standard?	Adequate consent to validation?	Adequate sample size?	≥1 month self-report to validation?	Study included?
Kerber et al. (1997)[110]	Family history of cancer	X	N/R	X	N/A	No

N/R This information was not reported & the reader was not referred to another source to find it.
N/A This criteria was not applicable to the behaviour being investigated.

Table 3: The accuracy of self-reported health risk status.

Study, by risk factor	Sample and survey method*	Self-report measure	Gold standard measure**	SE[1]	SP[1]	PPV[1]	NPV[1]	% mis-class	% at risk missed
Smoking status									
Fortmann et al. (1984)[76]	1,279 people aged 18-74: FS + PE	Smoke ≥ 9 cigarettes a day & had a cigarette in last 48 hours	Exhaled CO ≥ 8ppm	91%	99%	97%	96%	4%	9%
Pierce et al. (1987)[75]	1,172 people aged 15+: DS	Regularly smoke & smoked in last 24 hours	SC ≥ 250 nmol/L	93%	94%	88%	96%	7%	7%
Wagenknecht et al. (1992)[73]	4,984 people aged 18-30:TS +PE	Current smoker of >5 cigarettes per week	SeC >14ng/ml	91%	98%	95%	96%	4%	9%
Bowlin et al. (1993)[4]	596 people aged 20-69: TS+PE	Current smoker & ≥ 100 cigarettes ever	Exhaled CO ≥ 8ppm	♀ 78% ♂ 86%	97% 96%	89% 90%	92% 95%	- -	22% 14%
Bowlin et al. (1996)[60]	628 people aged 20-69: TS+FS+PE.	Current smoker	Exhaled CO ≥ 8ppm	TS 82% FS 87%	96% 95%	89% 86%	94% 95%	- -	18% 13%
Suadicani et al. (1994)[77]	2,833 ♀ aged 53-75: FS+PE	Current smoker	SeC >100ng/ml	95%	100%	-	-	-	5%

Study, by risk factor	Sample and survey method*	Self-report measure	Gold standard measure**	SE[1]	SP[1]	PPV[1]	NPV[1]	% mis-class	% at risk missed
Weight									
Jalkanen et al. (1987)[54]	11,880 people aged 30-64: MS+PE	Weight in light clothing	Measured weight in light clothing	↵ underestimated by an average 0.4kg ↓ underestimated by an average 0.6kg					-
Stewart et al. (1987)[50]	1,598 people aged 35-65: DS	Weight in light clothes	Measured weight in light clothes	↵ & ↓ underestimated by an average 0.6kg					-
Kuskowska-Wolk et al. (1989)[58]	301 people aged 16+: FS+PE	Weight	Measured weight in light underwear	↵ underestimated by an average 0.5kg ↓ underestimated by an average 0.6kg					-
Rowland et al. (1990)[55]	11,284 people aged 20-74: DS+PE	Weight without clothes & shoes	Measured weight in paper suit & foam slippers	↵ overestimated by an average 0.4kg ↓ underestimated by an average 1.0kg					-
Bowlin et al. (1993)[4]	See above	Weight without shoes	Measured weight in street clothes, without shoes	↵ underestimated by an average 1.4kg ↓ underestimated by an average 2.9kg					-
Waters (1993)[51]	6,898 people aged 20-69: PE	Weight without clothes & shoes	Measured weight in light clothes (- 1kg for clothing)	↵ underestimated by an average 0.2kg ↓ underestimated by an average 0.4kg					-

Study, by risk factor	Sample and survey method*	Self-report measure	Gold standard measure**	SE[1]	SP[1]	PPV[1]	NPV[1]	% mis-class	% at risk missed
Schmidt et al. (1993)[59]	659 people aged 15-64: DS + PE	Weight	Measured weight in light clothing & without shoes	↵ overestimated by an average 0.3kg ↛ underestimated by an average 0.3kg					-
Height									
Stewart et al. (1987)[50]	See above	Height without shoes	Measured height without shoes	↵ overestimated by an average 2.1cm ↛ overestimated by an average 1.6cm					-
Kuskowska-Wolk et al. (1989)[58]	See above	Height	Measured height without shoes	↵ overestimated by an average 0.9cm ↛ overestimated by an average 1.9cm					-
Rowland (1990)[55]	See above	Height without shoes	Measured height in foam slippers	↵ overestimated by an average 1.4cm ↛ overestimated by an average 0.6cm					-
Bowlin et al. (1993)[4]	See above	Height without shoes	Measured height without shoes	↵ overestimated by an average 2.7cm ↛ overestimated by an average 1.7cm					-
Waters (1993)[51]	See above	Height without shoes	Measured height without shoes	↵ overestimated by an average 1.1cm ↛ overestimated by an average 0.5cm					-

Study, by risk factor	Sample and survey method*	Self-report measure	Gold standard measure**	SE[1]	SP[1]	PPV[1]	NPV[1]	% mis-class	% at risk missed
Body Mass Index (BMI)									
Stewart et al. (1987)[50]	See above	BMI ≥25 from S/R weight & height	Measured BMI ≥25	76%	95%	95%	78%	15%	24%
Kuskowska-Wolk et al. (1989)[58]	See above	BMI >25.0(♩), >23.8(♀)from weight & height	Measured BMI >25.0(♩), >23.8(♀)	80%	96%	94%	83%	12%	20%
Rowland et al.(1990)[55]	See above	BMI ≥27.8(♩), ≥27.3(♀) from S/R weight & height	Measured BMI ≥27.8(♩), ≥27.3(♀)	29% of O/W ♩ underestimated weight by ≥2.3kg, 10% by ≥4.5kg; 43% of O/W ♀ underestimated by ≥2.3kg, 23% by ≥4.5kg					-
Bowlin et al. (1993)[4]	See above	BMI ≥27.5 from S/R weight & height	Measured BMI ≥27.5	♩ 77% ♀ 72%	99% 99%	99% 99%	85% 86%	- -	23% 28%
Waters et al. (1993)[51]	See above	BMI >25 from S/R weight & height	Measured BMI >25	♩ 83% ♀ 84%	94% 97%	94% 94%	83% 92%	12% 7%	17% 16%
Bowlin et al. (1996)[60]	See above	BMI ≥27.5 from S/R weight & height	Measured BMI ≥ 27.5	TS 74% FS 84%	99% 99%	99% 99%	94% 95%	- -	26% 16%

Study, by risk factor	Sample and survey method*	Self-report measure	Gold standard measure**	SE[1]	SP[1]	PPV[1]	NPV[1]	% mis-class	% at risk missed
Body Fat Distribution									
Hall et al. (1989)[63]	200 people aged 30-85: FS + PE	S/R hip & waist measurement	Measured hip & waist.	♩ 64%	87%	-	-	-	36%
				♪ 63%	87%	-	-	-	37%
Blood Pressure									
Bowlin et al. (1993)[4]	See above	Most recent BP reading	Measured BP	♩ & ♪ underestimated systolic & diastolic BP by average 1 mmHg , younger people overestimated and older ones underestimated					-
Hypertension									
Bowlin et al. (1993)[4]	See above	Been told had high BP once or currently on medication for high BP	Measured BP >140/90 mmHg	♩ 40%	87%	53%	79%	-	60%
				♪ 46%	87%	56%	82%	-	54%
Bowlin et al. (1996) [60]	See above	S/R of controlled hypertension	Measured BP ≥140/90 mmHg	TS 57%	82%	54%	84%	-	43%
				FS 60%	83%	56%	5%	-	40%
Bowlin et al. (1996) [60]	See above	S/R of uncontrolled hypertension	Measured BP ≥140/90 mmHg	TS 91%	11%	42%	62%	-	9%
				FS 88%	13%	44%	58%	-	12%
Vargas et al. (1997)[89]	8409 people: FS + PE	Ever diagnosed with hypertension	Measured BP ≥140/90 mmHg	71%	90%	72%	89%	-	29%

Study, by risk factor	Sample and survey method*	Self-report measure	Gold standard measure**	SE[1]	SP[1]	PPV[1]	NPV[1]	% mis-class	% at risk missed
Robinson et al. (1997)[88]	2728 people aged 18-74: interviewed	Been told had high BP	HIR - had in last 3 years	78%	86%	63%	93%	-	22%
Diabetes									
Bowlin et al. (1993)[4]	See above	Ever told have diabetes	Measured fasting serum glucose ≥ 140 mg/dl	↵ 68%	98%	44%	99%	-	32%
				↞ 80%	98%	50%	99%	-	20%
Bowlin et al. (1996)[60]	See above	S/R diabetes status	Fasting serum glucose ≥140 mg/dl	TS 75%	98%	48%	99%	-	25%
				FS 75%	98%	44%	99%	-	25%
Robinson et al. (1997)[88]	See above	Been told have diabetes	HIR – in last 3 years	72%	98%	76%	98%	-	28%
Cholesterol level									
Bowlin et al. (1993)[4]	See above	S/R blood cholesterol level	Measured fasting serum cholesterol	↵ underestimated by an average 3 mg/dl ↞ underestimated by an average 1 mg/dl					
Robinson et al. (1997)[88]	See above	Been told have high cholesterol	HIR - in last 3 years	79%	90%	32%	99%	-	21%
Hypercholesterolemia									
Bowlin et al. (1993)[4]	See above	Ever told have high cholesterol	Measured fasting serum cholesterol ≥ 200 mg/dl	↵ 41%	75%	65%	53%	-	59%
				↞ 47%	88%	85%	52%	-	53%
Bowlin et al. (1996)[60]	See above	S/R cholesterol level	Measured fasting serum cholesterol ≥200 mg/dl	TS 47%	80%	77%	52%	-	53%
				FS 51%	83%	81%	52%		49%

Study, by risk factor	Sample and survey method*	Self-report measure	Gold standard measure**	SE[1]	SP[1]	PPV[1]	NPV[1]	% mis-class	% at risk missed
Pap testing									
Fruchter et al. (1992)[40]	306 ♀: FS	No. yrs since & approx date of last	PLR - ever had	98%	39%	79%	92%	19%	61%
Gordon et al. (1993)[3]	431 ♀ aged 40-74: MS or TS	If had & no. yrs since last	HIR - had in last 2 years	97%	35%	78%	84%	23%	65%
Hiatt et al. (1995)[98]	401 Hispanic ♀: TS	If had & no. had in last 2 yrs	HIR - had in last 2 years	85%	40%	-	-	-	60%
Hiatt et al. (1995)[98]	290 non-Hispanic white ♀: TS	If had & no. had in last 2 yrs	HIR - had in last 2 years	90%	37%	-	-	-	63%
Bowman et al.(1997)[99]	455 ♀ aged 18-70: DS	If had & no. had in last 1-4 yrs or ever	PLR - had in last 1,2,3 or 4 years	1yr 89%	64%	40%	95%	-	36%
				2yr 97%	49%	52%	97%	-	51%
				3yr 96%	42%	61%	92%	23%	58%
				4yr 96%	38%	63%	89%	-	62%
Mammographic Screening									
Brown et al. (1992)[86]	189 ♀ aged 17-79: TS	If had at initial visit to health centre	RCR - had at initial visit (1-3 months prior)	92%	85%	79%	95%	12%	15%
Degnan et al. (1992)[102]	456 ♀ aged 50-74: TS	No. yrs since last	RCR - had in last 2 years	99%	80%	86%	99%	9%	20%
Gordon et al. (1993)[3]	See above	If had & no. yrs since last	HIR - had in last 2 years	98%	51%	81%	94%	16%	49%
Hiatt et al. (1995)[98]	401 Hispanic ♀: TS	If had in last 2 yrs	HIR - had in last 2 years	81%	63%	-	-	-	37%
Hiatt et al. (1995)[98]	290 non-Hispanic white ♀: TS	If had in last 2 yrs	HIR - had in last 2 years	92%	50%	-	-	-	50%

Study, by risk factor	Sample and survey method*	Self-report measure	Gold standard measure**	SE[1]	SP[1]	PPV[1]	NPV[1]	% mis-class	% at risk missed
Zapka et al. (1996)[103]	392 ♀ aged 50-74: TS or MS	Month & year of most recent screen	RCR - month & year of most recent screen	SR to V : exact 31%; +/-3 months 54%; : +/-12 months 83%; incomplete date 15%					-
Crane et al. (1996)[104]	576 ♀ aged 50+: FS	S/R date of last mammogram	RCR – had in last 19 months	82%	-	-	-	-	26%
Gastric photofluorographic screening (for stomach cancer)									
Tsubono et al. (1994)[106]	307 people aged 45-64: MS	Frequency of attendence at annual screens	Attendance & diagnosis records at screening centre	100%	76%	88%	100%	9%	24%
Sigmoidoscopy									
Gordon et al. (1993)[3]	779 people aged 40-74: MS/TS	If had & no. years since last	HIR - had in last 2 years	79%	87%	40%	98%	13%	13%
Hiatt et al. (1995)[98]	815 Hispanic people aged 35-74: TS	If & no. in last 2 yrs	HIR - had in last 2 years	♀ 100% ♂ 40%	96% 92%	- -	- -	- -	4% 8%
Hiatt et al. (1995)[98]	483 non-Hispanic whites aged 35-74: TS	If & no. in last 2 yrs	HIR - had in last 2 years	♀ 33% ♂ 50%	94% 94%	- -	- -	- -	6% 6%
Fecal Occult Blood Test									
Gordon et al. (1993)[3]	See above	If had & no. yrs since last	HIR - had in last 2 years	92%	71%	66%	93%	21%	29%
Hiatt et al. (1995)[98]	815 Hispanic people aged 35-74: TS	If & no. in last 2 yrs	HIR - had in last 2 years	♀ 53% ♂ 62%	82% 81%	- -	- -	- -	18% 19%

Study, by risk factor	Sample and survey method*	Self-report measure	Gold standard measure**	SE[1]	SP[1]	PPV[1]	NPV[1]	% mis-class	% at risk missed
Hiatt et al. (1995)[98]	483 non-Hispanic whites aged 35-74: TS	If & no. in last 2 yrs	HIR - had in last 2 years	♂ 68%	83%	-	-	-	17%
				♀ 63%	76%	-	-	-	24%
Digital Rectal Examination									
Hiatt et al. (1995)[98]	815 Hispanic people aged 35-74: TS	If & no. in last 2 yrs	HIR - had in last 2 years	♂ 48%	66%	-	-	-	34%
				♀ 63%	72%	-	-	-	28%
Hiatt et al. (1995)[98]	483 non-Hispanic whites aged 35-74: TS	If & no. in last 2 yrs	HIR - had in last 2 years	♂ 66%	69%	-	-	-	31%
				♀ 69%	66%	-	-	-	34%
Clinical Breast Examination									
Hiatt et al. (1995)[98]	401 Hispanic ♂: TS	If had & no. in last 2 yrs	HIR - had in last 2 years	87%	34%	-	-	-	66%
Hiatt et al. (1995)[98]	290 non-Hispanic white ♂: TS	If had & no. in last 2 yrs	HIR - had in last 2 years	92%	24%	-	-	-	76%

* Survey method: DS = door-to-door survey; TS = telephone survey; MS = mailout survey; FS = field survey; PE = physical examination.

** Gold standard: SC = salivary cotinine; SeC = serum cotinine; CO = carbon monoxide; BMI = body mass index; HIR = health insurance records; PLR = pathology laboratory records; RCR = radiology centre records.

1 SE = sensitivity; SP = specificity; PPV = positive predictive value; NPV = negative predictive value.

Table 4: Estimated risk factor prevalence by self-report (SR) and gold standard (GS)

Study, by Risk Factor	Sample	N	Prevalence		Z test
			SR	GS	
Smoking					
Fortmann et al. (1984)[76]	↵ & א	1,279	30%	33%	NS
Pierce et al. (1987)[75]	↵ & א	975	36%	34%	NS
Wagenknecht et al. (1992)[73]	↵ & א	4,984	31%	32%	NS
Bowlin et al. (1993)[4]	↵	254	22%	28%	NS
Bowlin et al. (1993)[4]	א	342	26%	30%	NS
Bowlin et al. (1996)[60]	↵ & א - phone survey	628	25%	29%	NS
Bowlin et al. (1996)[60]	↵ & א - field survey	628	27%	30%	NS
Overweight, by body mass index (BMI)					
Stewart et al. (1987)[50]	↵ & א - BMI>25	1,474	42%	53%	p<0.0001
Kuskowska-	א - BMI>23.9	182	43%	52%	p<0.05
Wolk et al. (1989)[58]	↵ - BMI>25	119	39%	45%	NS
Waters (1993)[51]	↵ - BMI>25	3,587	47%	53%	p<0.0001
	א - BMI>25	3,311	30%	34%	p<0.0005
Schmidt et al. (1993)[59]	↵ & א - BMI>25	659	37%	39%	NS
Bowlin et al. (1993)[4]	↵ - BMI≥27.5	270	34%	44%	p<0.01
	א - BMI≥27.5	333	26%	36%	p<0.005
Bowlin et al. (1996)[60]	↵ & א - BMI≥27.5 – phone survey	628	30%	40%	p<0.0005
Bowlin et al. (1996)[60]	↵ & א - BMI≥27.5 – field survey	628	34%	40%	p<0.0005
High abdominal fat ratio					
Hall et al. (1989)[63]	↵ & א	200	17%	26%	p<0.05
Hypertension (H)					

Study, by Risk Factor	Sample	N	Prevalence		Z test
			SR	GS	
Bowlin et al. (1993)[4]	↵	280	21%	37%	p<0.0001
	א	342	22%	36%	p<0.0001
Bowlin et al. (1996)[60]	↵ & א - controlled H – phone survey	628	29%	40%	p<0.0001
	↵ & א - controlled H – field survey	628	29%	40%	p<0.0001
	↵ & א - uncontrolled H – phone survey	628	10%	58%	p<0.0001
	↵ & א - uncontrolled H – field survey	628	13%	56%	p<0.0001
Robinson et al. (1997)[88]	↵ & א	2651	29%	24%	p<0.0001
Vargas et al. (1997)[89]	א	4415	28%	25%	p<0.0001
	↵	3994	25%	29%	p<0.0001
Diabetes					
Bowlin et al. (1993)[4]	↵	280	3%	4%	NS
	א	344	5%	5%	NS
Bowlin et al. (1996)[60]	↵ & א - phone survey	628	4%	5%	NS
	↵ & א - field survey	628	4%	5%	NS
Robinson et al. (1997)[88]	↵ & א	2651	7%	7%	NS
Hypercholesterolemia					
Bowlin et al. (1993)[4]	↵	153	33%	65%	p<0.0001
	א	183	33%	66%	p<0.0001
Bowlin et al. (1996)[60]	↵ & א - phone survey	628	36%	66%	p<0.0001
	↵ & א - field survey	628	37%	65%	p<0.0001
Robinson et al. (1997)[88]	↵ & א	2651	14%	6%	p<0.0001
Inadequate Pap testing					
Fruchter et al. (1992)[40]	א - never had one	189	13%	30%	p<0.0001
Gordon et al. (1993)[3]	א - not in last 2 years	352	12%	30%	p<0.0001

Study, by Risk Factor	Sample	N	Prevalence		Z test
			SR	GS	
Hiatt et al. (1995)[98]	Hispanic ♀ aged 35-74 - not in last 2 years	398	26%	46%	p<0.0001
Hiatt et al. (1995)[98]	non-Hispanic white ♀ aged 35-74 - not in last 2 years	288	22%	41%	p<0.0001
Bowman et al. (1997)[99]	♀ - not in last 3 yrs	455	51%	66%	p<0.0001
Inadequate mammographic screening					
Brown et al. (1992)[86]	♀ - not in last 3 months	189	57%	63%	NS
Degnan et al. (1992)[102]	♀ - not in last year	456	46%	64%	p<0.0001
	♀ - not in last 2 years	456	37%	46%	p<0.005
Gordon et al. (1993)[3]	♀ - not in last 2 years	386	22%	36%	p<0.0001
Hiatt et al. (1995)[98]	Hispanic ♀ aged 40-49 - not in last 2 years	120	43%	55%	NS
	Hispanic ♀ aged 50-74 - not in last 2 years	180	30%	46%	p<0.005
	non-Hispanic white ♀ aged 40-49 - not in last 2 years	108	28%	57%	p<0.0001
	non-Hispanic white ♀ aged 50-74 - not in last 2 years	129	18%	30%	p<0.05
Crane et al. (1996)[104]	♀ - not in last 19 months	576	56%	64%	p<0.01
No gastric photofluorographic screening in last 5 years (for stomach cancer)					
Tsubono et al. (1994)[106]	♂ & ♀	337	28%	36%	p<0.05
No sigmoidoscopic screening in last 2 years					
Gordon et al. (1993)[3]	♂ & ♀	672	81%	91%	p<0.0001
Hiatt et al. (1995)[98]	non-Hispanic white ♀ aged 50-74	128	88%	97%	p<0.01
	Hispanic ♀ aged 50-74	177	93%	99%	p<0.01
	non-Hispanic white ♂ aged 50-74	99	87%	93%	NS
	Hispanic ♂ aged 50-74	215	85%	98%	p<0.0001

Study, by Risk Factor	Sample	N	Prevalence		Z test
			SR	GS	
No fecal occult blood test in last 2 years					
Gordon et al. (1993)[3]	♂ & ♀	601	46%	61%	p<0.0001
Hiatt et al. (1995)[98]	non-Hispanic white ♀ aged 50-74	126	53%	57%	NS
	Hispanic ♀ aged 50-74	179	58%	65%	NS
	non-Hispanic white ♂ aged 50-74	106	55%	65%	NS
	Hispanic ♂ aged 50-74	218	57%	71%	p<0.005
No digital rectal examination in last 2 years					
Hiatt et al. (1995)[98]	non-Hispanic white ♀ aged 40-74	232	49%	46%	NS
	Hispanic ♀ aged 40-74	298	56%	49%	NS
	non-Hispanic white ♂ aged 40-74	158	53%	72%	p<0.0005
	Hispanic ♂ aged 40-74	330	60%	77%	p<0.0001
No clinical breast examination in last 2 years					
Hiatt et al. (1995)[98]	non-Hispanic white ♀ aged 40-74	233	13%	2%	p<0.0001
	Hispanic ♀ aged 40-74	297	22%	37%	p<0.0001

Notes: ♂ = male, ♀ = female

Table 5: Predictors of inaccurate self-report.

Study, by risk factor	Predictors of inaccurate self-report	Factors not affecting accuracy
Smoking status		
Wagenknecht et al. (1992)[73]	Being black, less educated or a self-reported ex-smoker	Gender
Bowlin et al. (1993)[4]	Being male, aged 30-39 (↵ only) or 20-29 (א only)	-
Weight		
Jalkanen et al. (1987)[54]	Being older, heavier, overweight, male, lower annual family income (↵ only), higher annual family income (א only) or more educated (א only)	Urban or rural residence, weight loss attempts in last 6 months
Stewart et al. (1987)[50]	Being heavier	Age
Kuskowska-Wolk et al. (1989)[58]	Being younger (א only), underweight or overweight	-
Rowland (1990)[55]	Being א, overweight, aged 20-34, more educated (א only) or rounding off (א only)	-
Bowlin et al. (1993)[4]	Being female or younger	-
Waters (1993)[51]	Being male or aged 35-39 (א only)	-
Schmidt et al. (1993)[59]	Being taller or older. ↵ more likely to overestimate, א more likely to underestimate	Race, education
Height		
Stewart et al. (1987)[50]	Being male or older	-
Kuskowska-Wolk et al. (1989)[58]	Being female, older (א only) or overweight (↵ only)	-
Rowland (1990)[55]	Being older, ↵, underweight (↵ only), or shorter & overweight	-
Waters (1993)[51]	Being male or older	-

Study, by risk factor	Predictors of inaccurate self-report	Factors not affecting accuracy
Bowlin et al. (1993)[4]	Being male or older	-
Body fat distribution		
Hall et al. (1989)[63]	Being ♀ and larger	-
Blood pressure level		
Bowlin et al. (1993)[4]	Younger people tend to underestimate & older ones to overestimate	-
Hypertension status		
Bowlin et al. (1993)[4]	Being older	Gender
Vargas et al. (1997)[89]	No recent doctor's visit, being male or not being overweight	Education
Diabetes status		
Bowlin et al. (1993)[4]	-	Gender, age
Cholesterol level		
Bowlin et al. (1993)[4]	Some age groups over-estimated, others underestimated	Gender
Hypercholesterolemia status		
Bowlin et al. (1993)[4]	Being older	Gender
Pap testing recency		
Fruchter et al. (1992)[40]	Longer time since last pap test	Age, birthplace
Gordon et al. (1993)[3]	Being older or less educated	-
Hiatt et al. (1995)[98]	Being non-Hispanic white	Gender
Bowman et al. (1997)[99]	Uncertainty about recency of last test	Perceived severity of cervical cancer, many demographic variables

Study, by risk factor	Predictors of inaccurate self-report	Factors not affecting accuracy
Mammographic screening recency		
Brown et al. (1992)[86]	Being older	Gender, time since visit
Degnan et al. (1992)[102]	-	Age, education, employment, income
Gordon et al. (1993)[3]	-	Age, education
Hiatt et al. (1995)[98]	-	Gender, ethnicity
Zapka et al. (1996)[103]	Longer time since last mammogram, being Hispanic or black	Age, education, income
Gastric photofluorographic screening recency		
Tsubono et al. (1994)[106]	Having a history of stomach cancer	Age, gender, education, occupation
Sigmoidoscopic screening recency		
Gordon et al. (1993)[3]	Being older	Gender, education
Hiatt et al. (1995)[98]	-	Gender, ethnicity
Fecal occult blood testing recency		
Gordon et al. (1993)[3]	Being older	Gender, education
Hiatt et al. (1995)[98]	-	Gender, ethnicity
Digital rectal examination recency		
Hiatt et al. (1995)[98]	-	Gender, ethnicity

Notes: ♂ = male, ♀ = female

Table 6: Characteristics predictive of non-consent to validation.

Study	Risk factor	Characteristics predictive of non-consent
Pierce et al. (1987)[75]	Smoking	Being female, aged over 65 or being a self-reported non-smoker
Bowlin et al.(1993)[4]	Smoking, hypertension, diabetes, cholesterol level, weight, height, BMI	Being less educated, not married or having no self-reported history of cardiovascular problems
Tsubono et al.(1994)[106]	Stomach cancer screen	Being male or younger
Bowlin et al. (1996)[60]	Smoking, BMI, hypertension, diabetes, hypercholesterolemia	Being less educated, not married or being a self-reported current smoker.
Crane et al. (1996)[104]	Mammography screening	Not being Hispanic, being born outside the United States and self-reported non-compliance with a mammogram referral.
Bowman et al. (1997)[99]	Pap testing	Being less educated or of lower occupational status

Figure 1: **A hypothetical data set for the calculation of accuracy data**

Had a pap test in the last 2 years?		**Gold Standard (Pathology Laboratory Records)**		
		Yes	**No**	**Total**
Self-report data	**Yes**	168 **(a)**	66 **(b)**	234 **(a+b)**
	No	4 **(c)**	54 **(d)**	58 **(c+d)**
	Total	172 **(a+c)**	120 **(b+d)**	292 **(a+b+c+d)**