obesity (BMI 30+ kg/m^2) of 1% in a rural area and 3.4% in a semi-urban area, which compares with a Global Infobase estimate of 0.1%.[3] We think it is perfectly reasonable to draw the attention of end-users to such marked variability in estimation as it is only through use and revision that information is improved.

We welcome and agree with WHO's strategy for the prevention of chronic diseases and did cite it positively in our editorial.[4] While Strong and colleagues think the time for debate is over, it is perfectly clear that in most low-income and middle-income countries the issues we discuss are far from clarified—hence the need for WHO's *Preventing Chronic Diseases* document. The most important tool for action—resources focused on chronic disease control and prevention—is lacking in virtually all low-income and most middle-income countries. Getting the resources in place involves debate with those who have competing priorities and are selling different visions of the future. As scientists, we would emphasize the importance of relevant and robust evidence of the effectiveness, and where possible, the cost-effectiveness, of population and high-risk strategies for the prevention of chronic diseases.

References

[1] Strong K, Mathers C, Epping-Jordan J, Beaglehole R. Preventing chronic diseases. A priority for global health. *Int J Epidemiol* 2006;**35**:492–94.

[2] Ebrahim S, Smeeth L. Non-communicable diseases in low and middle income countries: a priority or a distraction? *Int J Epidemiol* 2005;**34:**961–66.

[3] King H, Keuky, Seng S, Khun T, Roglic G, Pinget M. Diabetes and associated disorders in Cambodia: two epidemiological surveys. *Lancet* 2005;**366:**1633–39.

[4] World Health Organisation. *Preventing Chronic Disease: A Vital Investment*. Geneva: WHO, 2005.

doi:10.1093/ije/dyi316
Advance Access publication 30 January 2006

Comparison between cancers identified by state cancer registry, self-report, and death certificate in a prospective cohort study of US radiologic technologists

From D MICHAL FREEDMAN,[1]* ALICE J SIGURDSON,[1] MICHELE M DOODY,[1] SHARIFA LOVE-SCHNUR[1,2] and MARTHA S LINET[1]

When medical records or pathology reports to confirm diagnoses are unavailable, epidemiological studies often rely on self-reported diagnoses or death certificates to identify cancer outcomes. In assessing the validity of these sources, some studies focus on the positive predictive value, the percentage of reported cases in which cancer is correctly identified.[1] A second important measure of validity is false negative response, i.e. the percentage of cases, according to a gold standard, that are not identified in the self-report or death certificate or because of survey non-response. This is the complement of sensitivity, which is the percentage of cancers that are captured by self-reports and other sources. Here we present the extent of under-ascertainment due to self-report, death certificate, and survey non-response for cancers overall and six major cancer sites in a prospective, largely female cohort, the US Radiologic Technologists (USRT) Study, based on state cancer registries as the gold standard. A previous analysis of the USRT study presented the positive predictive values for cancers reported by cohort participants and death certificates.[2]

Self-administered questionnaires were sent to members of the USRT Cohort in three surveys carried out during 1983–89, 1994–98, and 2003 to the present. More than 110 000 study participants completed at least one of the first two questionnaires. Previous published reports analysing USRT data have relied on several sources to ascertain cancer outcomes: self-report, with cancers validated by medical records to the extent feasible; death certificates or the National Death Index, which include cancers identified as underlying or contributing causes of death; and additional subject contacts for ongoing studies.

To estimate the extent of under-reported cancers, we matched the 110 000 study participants with state-wide cancer registries in four states, California, Pennsylvania, Michigan, and Florida; about one-fourth of participating members reside in these four states. In 1990 the self-reported completeness of these state cancer registries ranged from 88 to 98%, although no completeness summary was reported for Michigan.[3] Moreover, limited participant mobility contributed to the comprehensiveness of registry ascertainment, with almost 90% of participants living in the same state during the first survey (1983–89) as at the time of diagnosis in the registry state. Matches between persons with cancer diagnoses listed in registries and cohort members relied on a combination of probabilistic[4] and deterministic algorithms and were based on participant social security numbers (97% available), names (100%), birthdates (100%), and gender (100%). We assessed under-reporting for cancers diagnosed between 1990 and the completion of the second questionnaire (1994–98), the date of death, or August 31, 1998 (for living participants who did not complete the second questionnaire), whichever occurred first.

There were a total of 761 invasive cancers reported in the registries for study participants during this period; 42% were diagnosed in California, 26% in Florida, 13% in Michigan, and 19% in Pennsylvania. The majority was diagnosed in women (69%) and those under age 60 in 1990 (65%).

[1] National Cancer Institute, Division of Epidemiology and Genetics, NIH, DHHS, Bethesda, MD, USA.

[2] Yale University, USA.

* Corresponding author. National Cancer Institute, Division of Cancer Epidemiology and Genetics, Radiation Epidemiology Branch, Executive Plaza South, Room 7036, 6120 Executive Boulevard, Bethesda, MD 20892, USA. E-mail: mf101e@nih.gov.

Table 1 Under-ascertainment of cancers in the US Radiologic Technologists Study based on linkage with four State Cancer Registries (1990–98)[a]

Type of cancer	Cancer registry-identified cancers by subject status at second survey	Cancer registry-identified cancers not identified in second survey by subject status
All malignancies	761	
Questionnaire responder	508	62 (12.2)[b]
Deceased	165	58 (35.2)
Non-responder/lost to follow-up	88	76 ()[c]
Female breast	243	
Questionnaire responder	189	5 (2.6)
Deceased	21	2(9.5)
Non-responder/lost to follow-up	33	31 ()
Prostate	84	
Questionnaire responder	62	3 (4.8)
Deceased	14	8 (57.1)
Non-responder/lost to follow-up	8	7 ()
Colon	33	
Questionnaire responder	22	1 (4.5)
Deceased	9	4 (44.4)
Non-responder/lost to follow-up	2	1 ()
Lung	67	
Questionnaire responder	25	4 (16.0)
Deceased	38	12 (31.6)
Non-responder/lost to follow-up	4	3 ()
Melanoma	31	
Questionnaire responder	18	4 (22.2)
Deceased	5	1 (20.0)
Non-responder/lost to follow-up	8	8 ()
Endometrial	38	
Questionnaire responder	27	5 (18.5)
Deceased	5	4 (80.0)
Non-responder/lost to follow-up	6	5 ()

[a] The four states are California, Florida, Pennsylvania, and Michigan.
[b] Numbers in parentheses are percentages.
[c] Some cancers were identified through follow-up related to studies within the cohort, even though the participant did not respond to the second questionnaire.

A total of 196 of the 761 (25.8%) cancers identified by the registries were not captured by self-report, death certificate, or other follow-up (Table 1). Self-reports (from the second questionnaire) failed to identify 12.2% of total cancer diagnoses, although respondents failed to report only 2.6% of female breast cancers and <5% each of prostate and colon cancers. Ascertainment was less complete for death certificates, which did not identify 35.2% of cancers identified by the registries in those participants who had died before completing the second questionnaire.

Estimates of the under-reporting of self-reported cancer in other studies vary widely, with high rates, 39%, for example, in a community-based study[5]; and more modest rates, 21%, in an American Cancer Society cohort.[6] The extent of false-negative reporting has also varied substantially by cancer site across studies.[5,7–9] Yet several studies, including this one, reflect a similar pattern of ascertainment, i.e. higher ascertainment, for breast, colon, and prostate cancers, and lower ascertainment for melanoma and uterine/endometrial cancers. Levels of under-reporting by site are relatively low in the USRT cohort, with no more than ~23% of the cancers under-reported by respondents in any of the six sites presented (Table 1). Because USRT study members work in health care settings, they may be more knowledgeable about their history of serious diseases, and this may at least partially account for their low rates of under-reported cancers.

Acknowledgements

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute.

References

[1] Colditz GA, Martin P, Stampfer MJ *et al*. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol* 1986;**123**:894–900.

[2] Sigurdson AJ, Doody MM, Rao RS *et al*. Cancer incidence in the US radiologic technologists health study, 1983–1998. *Cancer* 2003;**97**:3080–89.

[3] *Cancer Incidence in North America, 1988–1991*. Sacramento, CA: North American Association of Central Cancer Registries, 1995.

[4] Jaro MA. Probabilistic linkage of large public health data files. *Stat Med* 1995;**14**:491–98.

[5] Desai MM, Bruce ML, Desai RA, Druss BG. Validity of self-reported cancer history: a comparison of health interview data and cancer registry records. *Am J Epidemiol* 2001;**153**:299–306.

[6] Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol* 1998;**147**:556–62.

[7] Berthier F, Grosclaude P, Bocquet H, Faliu B, Cayla F, Machelard-Roumagnac M. Prevalence of cancer in the elderly: discrepancies between self-reported and registry data. *Br J Cancer* 1997;**75**:445–47.

[8] Parikh-Patel A, Allen M, Wright WE. Validation of self-reported cancers in the California Teachers Study. *Am J Epidemiol* 2003;**157**:539–45.

[9] Schrijvers CT, Stronks K, van de Mheen DH, Coebergh JW, Mackenbach JP. Validation of cancer prevalence data from a postal survey by comparison with cancer registry records. *Am J Epidemiol* 1994;**139**:408–14.

doi:10.1093/ije/dyi286
Advance Access publication 5 December 2005

A refinement to 'how many genes underlie the occurrence of common complex diseases in the population?'

From RAMAL MOONESINGHE

Sirs—In the paper entitled 'How many genes underlie the occurrence of common complex diseases in the population?',[1] the authors indicated that they had not found closed forms for the number of genes needed to achieve a particular population attributable fraction (PAF) with varying genotype prevalence G and risk ratios R_g for the additive models and multiplicative models, Equations (1) and (2), respectively. They used a computing algorithm to estimate N for any given PAF. However, closed forms for N do exist and the derivations of these formulas are given below.

Additive effects model

In the paper, Equation (1) was as follows:

$$D = I\sum_{j=0}^{N}\frac{N!}{j!(N-j)!}G^j(1-G)^{(N-j)}\big(jR_g-(j-1)\big), \qquad (1)$$

where I is the background risk of disease in the absence of the N susceptibility genotypes and j ($j = 0, 1, 2, \ldots, N$) indicates the number of disease susceptibility genotypes.

From Equation (1), $D = IE[jR_g - (j-1)]$, where $E(.)$ denotes the expected value and for a Binomial distribution with parameter G, $E(j) = NG$. Hence,

$$D = I\big(NGR_g - NG + 1\big)$$

and

$$\mathrm{PAF} = \frac{D-I}{D} = \frac{NG\big(R_g-1\big)}{NG\big(R_g-1\big)+1}.$$

Therefore,

$$N = \frac{\mathrm{PAF}}{1-\mathrm{PAF}}\frac{1}{G\big(R_g-1\big)}.$$

Multiplicative effects model

In the paper, Equation (2) was as follows:

$$D = I\sum_{j=0}^{N}\frac{N!}{j!(N-j)!}G^j(1-G)^{(N-j)}R_g^j, \qquad (2)$$

D is the product of I and the Binomial expansion of $(GR_g + (1-G))^N$.

Hence,

$$\mathrm{PAF} = \frac{\big(GR_g+(1-G)\big)^N-1}{\big(GR_g+(1-G)\big)^N}.$$

And finally,

$$N = \frac{-\mathrm{Log}(1-\mathrm{PAF})}{\mathrm{Log}\big(GR_g+(1-G)\big)}.$$

Reference

[1] Yang Q, Khoury MJ, Friedman JM, Little J, Flanders WD. How many genes underlie the occurrence of common complex diseases in the population? *Int J Epidemiol* 2005;**34:**1129–37.

doi:10.1093/ije/dyi288
Advance Access publication 14 December 2005

Office of Genomics and Disease Prevention, Coordinating Center for Health Promotion, Centers for Disease Control and Prevention, MS-K89, 4770 Buford Highway, NE, Atlanta, GA 30341-3717, USA. E-mail: Rmoonesinghe@cdc.gov