RESEARCH ETHICS

Non-reporting and inconsistent reporting of race and ethnicity in articles that claim associations among genotype, outcome, and race or ethnicity

H Shanawani, L Dame, D A Schwartz, R Cook-Deegan

J Med Ethics 2006;**32**:724–728. doi: 10.1136/jme.2005.014456

See end of article for authors' affiliations

Correspondence to: Hasan Shanawani, Department of Biostatistics and Research Epidemiology, Henry Ford Hospitals, 1 Ford Place, 5C-69 Detroit, MI 48202, USA; hshanawani@yahoo.com

Received 16 September 2005
Revised 14 March 2006
Accepted 16 March 2006

Background: The use of race as a category in medical research is the focus of an intense debate, complicated by the inconsistency of presumed independent variables, race and ethnicity, on which analysis depends. Interpretation is made difficult by inconsistent methods for determining the race or ethnicity of a participant. The failure to specify how race or ethnicity was determined is common in the published literature.

Hypothesis: Criteria by which they assign a research participant to racial or ethnic categories are not reported by published articles.

Methods: Methods were reviewed for assigning race and ethnicity of research participants in 268 published reports reporting associations among race (or ethnicity), health outcome and genotype.

Results: Of the 268 published reports reviewed, it was found that 192 (72%) did not explain their methods for assigning race or ethnicity as an independent variable. This was despite the fact that 180 (67%) of those reports reached conclusions about associations among genetics, health outcome and race or ethnicity.

Conclusions: More attention needs to be given to the definition of race and ethnicity in genetic studies, especially in those diseases where health disparities are known to exist.

Different socially identified groups have different rates of diseases,^{1–4} leading to intense study, debate on the possible causes of such health disparities and controversy about the use of racial classification in medicine, biomedical research in general^{5–7} and genetics research in particular.^{8–9} Although some suggest that there may be no biological basis for race and therefore racial categories will be misleading in the search for biological differences underlying health disparities,¹⁰ others believe that race may correlate to some genetic factors associated with disease states, and that too much important information may be missed if race is not considered. Social epidemiologists and other investigators express concern that social factors in health disparity groups may be missed if race is not studied.^{11–14} Although many investigators pursue issues of racism and health inequality as the source of health disparities, they are also interested in genetic and other biological factors in causal networks associated with health disparities. Some investigators of genetics believe that it is crucial to continue to draw categories along racial lines¹⁵ and offer guidelines on how to assign membership of a research participant to a particular social group.

In an effort to understand the genetic and environmental causes of diseases where health disparities exist, an important goal of scientific research findings is generalisability and external validity.^{16–17} An important step to attaining that external validity is the ability of other investigators to replicate studies reported in the scientific literature. We believe that ambiguous definitions of race coupled with generalisations made about race are a principal impediment to successful genetics research.

Investigators studying the use of race in biomedical research reports have found inconsistent reporting practices.¹⁸ Journals have been noted to have inconsistent editorial policies and variable enforcement of those policies.^{19–20} The definitions of terms used to refer to race and

ethnicity have also varied from article to article in biomedical and social science journals.^{21–22} We are, however, unaware of any attempt to characterise the use of race in genetics research articles, where the use of race is, in our opinion, the most challenging and controversial.

We hypothesised that clinicians and research participants use disparate definitions of race and do not report the criteria by which they assigned or were assigned to a cohort. We set out to investigate the reporting practices of investigators who published race, health outcome and genetic polymorphisms.

METHODS

We completed a Medline search to identify articles investigating genetic phenomena and health outcomes that included keywords relating to “race” and “ethnicity”. Our Medline search included articles published from January 2003 through to December 2003 (search completed in April 2004). To improve our search sensitivity, a broad search strategy using Ovid (New York city, New York, USA) was used (fig 1). Articles were then obtained and individually reviewed. Articles were included if they identified a racial or ethnic cohort, a genetic procedure or test and an identifiable health outcome.

We reviewed each article with a standard evaluation form prepared for this investigation. We recorded the terms used by articles that reported the study of any group and used a term commonly understood to refer to a race, ethnicity, tribe, or national or geographical origin. We did not attempt to distinguish whether the group terms used would be considered to refer to a race, ethnicity or other classification, as this was not the focus of our study. We recorded what groups were studied. We then recorded the reporting of any term referring to race, ethnicity or geographical origin in the Methods section of the published paper. If there was such a term, we recorded whether the Methods section reported any method to determine the participant's inclusion or exclusion

#	Search history	#	Search history
1	Exp genetics/	17	15 and 16
2	Exp genetic phenomena/	18	Ep.fs
3	Exp genetics structures/	19	Exp prognosis/
4	Anticipation, genetic/	20	Exp mortality/
5	Ge.fs.	21	Mo.fs.
6	1 or 2 or 3 or 4 or 5	22	Exp morbidity/
7	Exp population groups/	23	Exp survival analysis/
8	Eh.fs.	24	Severity of illness index/
9	(Race or racial\$ or ethnic\$).ti.	25	Exp risk
10	7 or 8 or 9	26	18 or 19 or 20 or 21 or 22 or 23 or 25
11	6 and 10	27	17 and 26
12	Exp *genetics/ or exp *genetic phenomena/or exp *genetic structures/ or *anticipation, genetic/	28	11 and 26
13	Gene\$.ti.	29	Duke. in.
14	12 or 13	30	28 and 29
15	14 and 11	31	From 30 keep 1 to 10
16	Exp *population groups/ or 9	32	Limit 28 to yr = 2003

Figure 1 Ovid search strategy

from the group referred to. We separately recorded if the method used self-identification or self-report as part of it. If the Methods section referred to a previous publication, we obtained that reference and studied it as part of the review of the paper found in the original search. Finally, we determined if any conclusion of the paper explicitly included a “three-way association” between the health outcome studied, the genetic phenomenon measured, and the racial, ethnic or geographical cohort studied.

In articles that reported a three-way association, we reviewed the original hypothesis to determine if the goals of the study specified the three elements that were the basis of the three-way association reported.

We also collected information on the genetic variants studied, type of study conducted, how participants or samples were obtained for the study, the funding source, geographical location of the first author of the paper, whether the journal was a genetics journal and whether the groups studied in the paper were explicitly referred to as “minority”.

We classified the journals found in our search as “genetics” journals or “non-genetics” journals. This determination was based on querying the title of the journal and whether it included a term “gene”, “genetics”, “heredity” or “mutation”.

We examined whether the Methods section explained how membership of a participant to a particular race or ethnicity group was determined as a yes/no variable. If the answer was yes, we additionally examined “how” as a categorical variable. We examined where any three-way association was reported as both a bivariate and a categorical variable (“mention in title”, “abstract”, “other”, “not at all”). The data were stratified by whether the study was published in a genetics or non-genetics journal.

To determine whether our Medline search strategy missed articles that met our inclusion criteria, we hand searched five journals by using methods described previously.^{23–26} We selected *The American Journal of Human Genetics*; *The American Journal of Medical Genetics*; *Cancer Epidemiology, Biomarkers, and Prevention*; *Neurology*; and *The American Journal of Epidemiology*. We chose two genetics journals, two medical specialty journals and one epidemiology journal. These particular journals were selected because they

represented the largest number of citations in our original search strategy.

The five journals selected accounted for 42 of the 268 articles from our original search. Our hand search of all titles and abstracts of articles published in those journals yielded an additional 25 articles, which were not found in our original Ovid search but would have been included in our study based on the title or abstract. To determine why these articles had not been found in our Medline search, we searched those articles in Medline and found that they had no Mesh heading related to race or ethnicity, but did include a Mesh term “Geographic Location”, which was not included in our original search strategy. This term is defined (PubMed) as follows:

All of the continents and every country situated within, the UNITED STATES and each of the constituent states arranged by region, CANADA and each of its provinces, AUSTRALIA and each of its states, the major bodies of water and major islands on both hemispheres, and selected major cities. Although the geographic locations are not printed in INDEX MEDICUS as main headings, in indexing they are significant in epidemiologic studies and historical articles and for locating administrative units in education and the delivery of health care.

We used our Ovid strategy once again, including the exploded term “geographic locations” in our search. This doubled our citations (741 to 1571; search completed in September 2004). These new citations were not included in further analysis, as we believed that prevailing hypotheses of population genetics were not based on diplomatic or political borders between or within nations, which we understood to be the essence of this particular MESH heading. Conversely, we noted whether the articles in our original strategy based their cohort on a municipality, citizenship in a country or political boundary other than tribal membership.

To gain qualitative insight and to supplement the literature review, HS conducted unstructured interviews on a convenience sample of nine authors in the US, who were selected from reviewed papers that did not specify their method of determining race or ethnicity.

This work was completed as part of a fellowship funded by an independent Duke Endowment as well as an institutional T32 grant.

RESULTS

We excluded 300 of the 568 articles identified by our initial Medline search (table 1). We excluded articles that did not report the study of a genetic phenomenon, or were review articles, meta-analyses, letters to the editor, editorials and news reporting in scientific journals, non-human studies, case reports, case series and articles in languages other than

Table 1 Excluded articles (n = 300)

Not a genetics article	112
Review or meta-analysis	58
Letter to the editor	32
Non-human genetics	32
Case report or series	19
Not in English	12
Editorial	11
Duplicate citation in Ovid	10
News article	8
Others	6

English. Ten citations were duplicates and were included only once. Our final review was of the remaining 268 articles.

We sought to determine if the study authors included an explanation on inclusion or exclusion criteria of their research participants, which could be tested for external validity—that is, were we to attempt to duplicate their study, could we establish their population frame based on their scientific report. Only 76 (28%) of the 268 articles in our analysis reported a method by which the research participant's race, ethnicity or membership to another studied cohort was assigned (tables 2, 3). Of the remaining 192 articles, 113 referred to the population term that was the topic of their article in their Methods section, but did not report any method by which participants could be included or excluded from that population they were reporting to have studied; 61 articles had no term in their Methods section, but reported on race or ethnicity in the Results section; 12 articles referred to another paper, but the articles they referenced were reviewed and were also found to lack a described method for assigning race or ethnicity.

Of the 268 articles, 180 (67%) articles reported a three-way association between (1) race or ethnicity or geographical origin, (2) a genetic variant and (3) a health outcome (table 4). Of the 180 articles, 126 articles that were associated with genotype, outcome and race or ethnicity did not describe how race or ethnicity was determined. Of the 180 articles, 137 (76%) reported a three-way association in the title or abstract of the paper (table 2), suggesting that it was an important finding.

Of the 180 articles, 76 (42%) articles did not explicitly frame a hypothesis that was the basis of their reported findings, suggesting that in at least some cases the findings were reported when carrying out subpopulation analysis, but studying racial or ethnic differences may not have been the original purpose.

The 76 articles that explained how race and ethnicity were assigned to participants were studied further (table 3). The most common method was self-identification, used in 25 (33%) of the 76 articles, 20 of which did so as part of initial recruitment or by offering a choice of categories on a questionnaire. The remaining five studies had open recruitment and did not specify the method other than it occurred during an interview.

Of the 76 studies, 5 (7%) used data found during a clinical record review of the research participant. An illustrative example is included:²⁷

For each patient, records were assessed for clinical status, date and place of birth, ethnic background, family history, and other affected family members. DM patients were grouped as followed: the Ashkenazi Jews originated from European countries (excluding the Balkan countries) and North America. The Sephardim/Oriental Jews originated from North Africa, Asia, and the Balkans and the Yemenite Jews who originated from Yemen. Patients, whose origin was different, such as India, were classified as others.

Four studies relied on language to frame their study population, although they did not clearly state “self-reported” language use. For example, one study²⁸ reported that the DNA samples used were extracted from “tumor tissues ... [obtained from] Xhosa-speaking esophageal cancer patients”. How the investigators determined language ability of the subjects was, however, never reported.

Of the 76 articles, 17 (22%) framed their study population by geography. For example:²⁹

Table 2 Articles reporting a three-way association: location of three-way association and relationship to reported hypothesis of study

	Based on reported hypothesis	Not based on reported hypothesis	Total
Reported in title or abstract	72	65	137
Reported elsewhere in paper	32	11	43
Total	104	76	180

Table 3 Method cited in articles reporting a method of identifying race or ethnicity (n=76)

Self-reported inclusion in population frame	29
Questionnaire or specific question	15
Recruitment	9
Interview	5
Eligible study participants determined by geographical frame	19
Presentation to medical centre	10
Record review	5
Eligible study participants determined by language frame	4
Other or variable methods	9

Table 4 Distribution of articles reporting on race or ethnicity

	Did report race or ethnicity	Did not report race or ethnicity	Total
Reported a three-way association	54	126	180
Did not report a three-way association	22	66	88
Total	76	192	268

The subjects in the study came from a non-urban community of the dikgale district in the northern province of South Africa. This district is situated 15 km northeast of the university of the north and 40 km from the province's capital, Polokwane (Pietersberg).

An additional nine studies relied on referral patterns to the medical centre where the study was performed. (“Patients over 16 years old with tuberculosis were identified at the clinic in the southwestern region of Croatia.”³⁰)

Although inclusion was often initially reported in a straightforward fashion, details were often missing—for example, one report noted that “Only indigenous Zulu-speaking Black African women were included in the study and those with known nonindigenous relatives were excluded.”³¹ The method of determination that a relative was “non-indigenous” was, however, never reported.

Methods differed among studies of defined populations. One study on Native Americans, for example, took care to sample “from different areas of the reservation, to obtain a representative and random sample of the Navajo Tribe”,³² whereas another used self-report of tribal membership.³³ In two studies on Ashkenazim in Israel, one collected data on

European origin,³⁴ whereas the other³⁵ reported no further demographic query.

Only five (6.5%) studies stated that their demographic determination was based on more than one generation in the participant's family; of these, four relied on racial or ethnic background of three generations in the participant's family, as recommended previously.³⁶

The nine authors who faced our unstructured interviews were asked how demographic data on research participants were obtained and recorded. We found that investigators intended to use self-identification, and this was usually based on pre-established categories. Many participants were, however, not directly asked nor given the opportunity to suggest their race or ethnicity. In those cases, classifications of participants were made on the basis of clinical chart review, second-hand information from clinicians asked by study recruiters or assumed by the study recruiter without asking the participant. Some study cohorts combined participants from different studies that used different methods of categorisation. In these instances where different methods of categorisation were used, some investigators "inherited" data or a cohort from the second study or from coinvestigators for analysis. Despite publishing an article, some investigators simply did not know how classifications were done.

DISCUSSION

Genomic studies illustrate the complexities and ambiguities of group labels in biomedical research. As most (72%) of the articles in our sample included no descriptions of how race or ethnicity was determined, we cannot draw conclusions on the consistency or applicability of definitions in published investigations. Also, we cannot build empirically based categories to describe how race and ethnicity are "measured". We were unsure, for example, as to whether to classify Chinese as an ethnic group,³⁷ a geographical population³⁸ or a nationality.³⁹ In these three examples, the respective article offered no guidance. The assessment of race and ethnicity in epidemiological research is fundamental to any effort to reduce excesses in poor health among racial or ethnic groups.⁴⁰ The ambiguous use of terms, however, poses a challenge to this assessment.

The National Institutes of Health or other US federal agencies funded 44% of the studies. They would therefore fall under OMB directive 15, which states that investigators should generally rely on "self-reporting or self-identification" to establish the race or ethnicity of a research participant. Although we found that studies funded by the US were more likely (33%) to report methods in our study, this difference was not statistically significant. This is consistent with findings in other, non-genetic studies⁴¹ in which race and ethnicity were not being reported at all in clinical studies of diseases with known health disparities.

Nearly two thirds of the articles reported a three-way association among the genetic phenomenon, health outcome, and racial or ethnic cohort studied, and more than three quarters did so in the title or abstract of their article, suggesting that it was a central feature of their report. We are concerned that uncritical acceptance of such claims, especially with ambiguous definitions of the cohort studied, may lead to unnecessary and potentially harmful social or biological exaggerations of population differences. This concern has been expressed previously.¹⁷ The opposite is likewise true: lack of methodological consistency can also lead to understating of genetic or social risks, and missed opportunities in genetics research of complex diseases.⁴²

Conflating race or ethnicity as a social concept and as a label for regional populations sows seeds of confusion and breeds controversy. As terms of race and ethnicity are used

inconsistently to define social cohorts and distinct populations exhibiting possible genetic differences (whether differences in alleles or differences in allele frequency), questions of the relationship among race, genes and population health become harder to characterise, more complex to study and increasingly controversial. An important first step is to define terms and populations of study as precisely as possible. This should occur during study design and recruitment, as suggested by leading investigators.

Tolerance of ambiguous definitions may also lead to more ominous threats to health. It has been argued that "conservative foundations are seeking to frame debates over determinants of racial/ethnic health disparities as a matter of 'politically correct' unscientific ideology ... vs scientific yet 'politically incorrect' expertise rooted in biological facts".⁴³ The argument over the existence of biological categories threatens to distract societies from well-described social causes of health disparities.⁴⁴⁻⁴⁷ We and others⁴⁸ believe that a common ground exists between geneticists and social scientists. We argue that achieving this common ground will depend on methodological consistency between the two research communities, which starts with the consistent measuring of the populations and studying of ecological settings.

Growing sophistication in descriptive epidemiology has allowed for more precise modelling of risk factor exposures, race and ethnicity, and disease rates. Genetics investigators have an important assignment: recognising the degree of genetic variation within and between racial and ethnic groups, linking it to variations in health outcome, and tying differences to social risk factors, biological risk factors, environmental risk factors and other nodes in a complex causal network. Studies that do not clearly specify the independent variable of race or ethnicity make it difficult, if not impossible, to know exactly what factors are being studied and how to interpret results.

ACKNOWLEDGEMENTS

We thank G Corbie-Smith, J Reardon, E Hauser and W Kraus for their contribution to the important background work for this project and A Powers and C Whitener for assistance with the Medline search strategy.

Authors' affiliations

H Shanawani, Department of Biostatistics and Research Epidemiology, Henry Ford Hospitals, Detroit, MI, USA

L Dame, R Cook-Deegan, Duke University Center for Genome Ethics, Law, and Policy, Durham, North Carolina, USA

D A Schwartz, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Competing interests: None.

Contributors: HS generated the idea for this project and wrote the protocol that was further elaborated by the other two authors. He also obtained the data from the articles. All authors interpreted the data and analyses. HS wrote the final draft. LD, DAS and RCD commented on it critically.

REFERENCES

- 1 **Cooper R**. The role of genetic and environmental factors in cardiovascular disease in African Americans. *Am J Med Sci* 1999;**317**:208-13.
- 2 **Sorlie PD**, Sharrett AR, Patsch W, *et al*. The relationship between lipids/lipoproteins and atherosclerosis in African Americans and whites: the atherosclerosis risk in communities study. *Ann Epidemiol* 1999;**9**:149-58.
- 3 **Cooper RS**, Zhu X. Racial differences and the genetics of hypertension. *Curr Hypertens Rep* 2001;**3**:19-24.
- 4 **Williams MD**, Sandler AB. The epidemiology of lung cancer. *Cancer Treat Res* 2001;**105**:31-52.
- 5 **Burchard EG**, Ziv E, Coyle N, *et al*. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med* 2003;**348**:1170-5.
- 6 **Phimister EG**. Medicine and the racial divide. *N Engl J Med* 2003;**348**:1081-2.

- 7 **Osborne NG**, Feit MD. The use of race in medical research. *JAMA* 1992;**267**:275-9.
- 8 **Freedman ML**, Reich D, Penney KL, et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004;**36**:388-93.
- 9 **Lander ES**, Schork NJ. Genetic dissection of complex traits. *Science* 1994;**265**:2037-48.
- 10 **Schwartz RS**. Racial profiling in medical research. *N Engl J Med* 2001;**344**:1392-3.
- 11 **Goodman AH**. Why genes don't count (for racial differences in health). *Am J Public Health* 2000;**90**:1699-702.
- 12 **Aldhous P**. Geneticist fears 'race-neutral' studies will fail ethnic groups. *Nature* 2002;**418**:355-6.
- 13 **LaVeist TA**. Segregation, poverty, and empowerment: health consequences of African Americans. In: LaVeist TA, ed. *Race, ethnicity, and health: a public health reader*. San Francisco: Jossey-Bass, 2002.
- 14 **Williams DR**, Collins C. U.S. socioeconomic and racial differences in health: patterns and explanations. In: LaVeist TA, ed. *Race, ethnicity, and health: a public health reader*. San Francisco: Jossey-Bass, 2002.
- 15 **Risch N**, Burchard E, Ziv E, et al. Categorization of humans in biomedical research: genes, race and disease [comment]. *Genome Biol* 2002;**3**:2007.
- 16 **Foster MW**, Sharp RR. Beyond race: towards a whole-genome perspective on human populations and genetic variation. *Nat Rev Genet* 2004;**5**:790-6.
- 17 **Wang VO**, Sue S. In the eye of the storm: race and genomics in research and practice. *Am Psychol* 2005;**60**:37-45.
- 18 **Bhopal R**, Donaldson L. White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. *Am J Public Health* 1998;**88**:1303-7.
- 19 **Bennett T**, Bhopal R. US health journal editors' opinions and policies on research in race, ethnicity, and health. *J Natl Med Assoc* 1998;**90**:401-8.
- 20 **Bhopal R**, Kohli H, Rankin J. Editors' practice and views on terminology in ethnicity and health research. *Ethn Health* 1997;**2**:223-7.
- 21 **Jones CP**, LaVeist TA, Lillie-Blanton M. "Race" in the epidemiologic literature: an examination of the American Journal of Epidemiology, 1921-1990. *Am J Epidemiol* 1991;**134**:1079-84.
- 22 **Williams DR**, Lavizzo-Mourey R, Warren RC. The concept of race and health status in America. *Public Health Rep* 1994;**109**:26-41.
- 23 **Patrick TB**, Demiris G, Folk LC, et al. Evidence-based retrieval in evidence-based medicine. *J Med Libr Assoc* 2004;**92**:196-9.
- 24 **Alderson P**, Green S, Higgins J, eds. *Cochrane Reviewers' Handbook 4.2.2 [updated march 2004]*. In: *The Cochrane Library*, Issue 1. Chichester: Wiley, 2004.
- 25 **Avenell A**, Handoll HH, Grant AM. Lessons for search strategies from a systematic review, in The Cochrane Library, of nutritional supplementation trials in patients after hip fracture. *Am J Clin Nutr* 2001;**73**:505-10.
- 26 **Helmer D**, Savoie I, Green C, et al. Evidence-based practice: extending the search to find material for the systematic review. *Bull Med Libr Assoc* 2001;**89**:346-52.
- 27 **Segel R**, Silverstein S, Lerer I, et al. Prevalence of myotonic dystrophy in Israeli Jewish communities: inter-community variation and founder premutations. *Am J Med Genet* 2003;**119A**:273-8.
- 28 **Vos M**, Adams CH, Victor TC, et al. Polymorphisms and mutations found in the regions flanking exons 5 to 8 of the TP53 gene in a population at high risk for esophageal cancer in South Africa. *Cancer Genet Cytogenet* 2003;**140**:23-30.
- 29 **Choma SR**, Alberts M, Potgieter HC, et al. The effect of different apo(a) isoforms on plasminogen activation in black South African subjects. *Cardiovasc J South Africa* 2003;**14**:120-4.
- 30 **Fraser DA**, Bulat-Kardum L, Knezevic J, et al. Interferon-gamma receptor-1 gene polymorphism in tuberculosis patients from Croatia [see comment]. *Scand J Immunol* 2003;**57**:480-4.
- 31 **Pegoraro RJ**, Hira B, Rom L, et al. Plasminogen activator inhibitor type 1 (PAI1) and platelet glycoprotein IIIa (PGIIa) polymorphisms in Black South Africans with pre-eclampsia. *Acta Obstet Gynecol Scand* 2003;**82**:313-7.
- 32 **Wall TL**, Carr LG, Ehlers CL. Protective association of genetic variation in alcohol dehydrogenase with alcohol dependence in Native American mission Indians [see comment]. *Am J Psychiatry* 2003;**160**:41-6.
- 33 **Yi Z**, Garrison N, Cohen-Barak O, et al. A 122.5-kilobase deletion of the P gene underlies the high prevalence of oculocutaneous albinism type 2 in the Navajo population. *Am J Hum Genet* 2003;**72**:62-72.
- 34 **Offit K**, Pierce H, Kirchhoff T, et al. Frequency of CHEK2*1100delC in New York breast cancer cases and controls. *BMC Med Genet* 2003;**4**:1.
- 35 **Goffin JR**, Chappuis PO, Begin LR, et al. Impact of germline BRCA1 mutations and overexpression of p53 on prognosis and response to treatment following breast carcinoma: 10-year follow up data. *Cancer* 2003;**97**:527-36.
- 36 **Tang H**, Quattermous T, Rodriguez B, et al. Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *Am J Hum Genet* 2005;**76**:268-75.
- 37 **Kadam-Pai P**, Su XY, Miranda JJ, et al. Ethnic variations of a retinoblastoma susceptibility gene (RB1) polymorphism in eight Asian populations. *J Genet* 2003;**82**:33-7.
- 38 **Yang JD**, Feng GY, Zhang J, et al. Apolipoprotein E -491 promoter polymorphism is an independent risk factor for Alzheimer's disease in the Chinese population [correction appears in *Neurosci Lett* 2003;30:215]. *Neurosci Lett* 2003;**350**:25-8.
- 39 **Chang YT**, Tsai SF, Lin MW, et al. SPR1 gene near HLA-C is unlikely to be a psoriasis susceptibility gene. *Exp Dermatol* 2003;**12**:307-14.
- 40 **Hahn RA**, Stroup DF. Race and ethnicity in public health surveillance: criteria for the scientific use of social categories. *Public Health Rep* 1994;**109**:7-15.
- 41 **Corbie-Smith G**, St George DM, Moody-Ayers S, et al. Adequacy of reporting race/ethnicity in clinical trials in areas of health disparities. *J Clin Epidemiol* 2003;**56**:416-20.
- 42 **Reiner AP**, Ziv E, Lind DL, et al. Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am J Hum Genet* 2005;**76**:463-77.
- 43 **Krieger N**. Stormy weather: race, gene expression, and the science of health disparities. *Am J Public Health* 2005;**95**:2155-60.
- 44 **Wilkinson DY**, King G. Conceptual and methodological issues in the use of race as a variable: policy implications. *Milbank Q* 1987;**65**(Suppl 1):56-71.
- 45 **Phelan JC**, Link BG. Controlling disease and creating disparities: a fundamental cause perspective. *J Gerontol B Psychol Sci Soc Sci* 2005;**60**(Spec No 2):27-33.
- 46 **Smedley A**, Smedley BD. Race as biology is fiction, racism as a social problem is real: anthropological and historical perspectives on the social construction of race. *Am Psychol* 2005;**60**:16-26.
- 47 **Bloche MG**. American medicine and the politics of race. *Perspect Biol Med* 2005;**48**(Suppl 1):S54-67.
- 48 **Rivera JH**, Rivera EY. Racial and ethnic disparities in health status and health care: political common ground. *Del Med J* 2005;**77**:361-5.