

Comparison of genotypic and phenotypic strategies for population screening in hemochromatosis: Assessment of anxiety, depression, and perception of health

Christine Patch, PhD, Senior Research Fellow¹, Paul Roderick, MD, FRCP¹, Professor William Rosenberg, PhD, FRCP²

Purpose: Hemochromatosis is a treatable disorder with a major genetic predisposition. It provides an example in which genotypic and phenotypic strategies for screening may be compared. We previously showed noninferiority of uptake of a genotypic population screening strategy for hemochromatosis compared with a phenotypic strategy. In this article we present the psychologic effects of each strategy. **Methods:** A sample of 3000 individuals from primary care were randomly allocated to a phenotypic or genotypic screening strategy for hemochromatosis, and the 939 individuals who accepted screening provide the sample for this article. Standardized assessments of anxiety, general health, and depression were made at invitation, testing, result-giving, and 6 months. **Results:** Screening did not lead to significant changes in the self-rated assessments of anxiety, depression, and general health over time, and there were no significant differences between the two screening strategies. The unemployed or permanently disabled had lower ratings of health and higher anxiety and depression. **Conclusion:** The two screening strategies appeared to cause little adverse psychologic disturbance in the short term, and there was no difference between the two strategies. This study provides some empiric data to support arguments against “genetic exceptionalism” and suggests that genetic testing when used for population screening for a treatable disease has few adverse effects. *Genet Med* 2005;7(8):550–556.

Key Words: hemochromatosis, screening, psychologic effects

In the future, genetic technologies may be used in well populations for the detection of high-risk individuals to initiate treatment or for the prevention of common diseases.¹ Hemochromatosis, a treatable adult-onset disorder of iron metabolism, is a condition in which population screening could be an appropriate case-detection strategy. The abnormal genotype is a necessary but not sufficient cause of the disease and thus may be considered a predictive risk factor. There are two approaches to screening: phenotypic, using transferrin saturation, and genotypic. Although the current consensus is that population screening is premature because of uncertainties about the natural history of the condition, hemochromatosis provides a situation in which comparisons may be made between the consequences of a phenotypic and a genotypic screening strategy.^{2–5} In population screening programs for conditions of low prevalence, the majority of people will screen

negative. Although there may be the possibility of receiving reassurance from a negative result, they will not receive any benefit from preventive or treatment strategies. There may be harm caused by the offer of screening itself, unnecessary investigations, and anxiety in those who initially screen positive and subsequently are shown not to have the condition (false positives), and false reassurance in those who screen negative but are affected with the condition (false negatives). The program itself therefore has the potential for harming a significant number of people. Evaluation of screening should include evidence relating to the effects on psychologic well-being and quality of life of the participants in addition to assessment of the direct health benefits.^{6,7}

Concern about genetic testing and screening is evidenced by a number of reports that focus on issues of potential stigmatization, discrimination, family implications, and the possible psychologic consequences.^{8–10} In these reports genetic information is accorded special status because it may lead to specific harm. However, there is little empiric evidence of genetic testing programs leading to serious adverse effects, although this is in the context of rigorous test preparation and short- to medium-term follow-up.^{11–13} To date most testing has been performed for highly penetrant single gene disorders such as Huntington’s disease or dominantly inherited cancers. It has also occurred in the context of family testing. In these disorders the

From the ¹Public Health Sciences and Medical Statistics, and ²Liver Research Group, University of Southampton, United Kingdom.

Christine Patch, PhD, Address Mailpoint 805, level C, South Academic Block, Southampton General Hospital, Southampton SO16 6YD, UK.

Submitted for publication April 21, 2005.

Accepted for publication August 4, 2005.

DOI: 10.1097/01.GIM.0000182466.87113.ce

test is being offered against the background of a known high genetic risk, and prevention or treatment is not possible or may be radical (e.g., prophylactic mastectomy in the case of breast cancer genes). The participant is therefore gaining knowledge from the test to be informed on decisions about previous knowledge of the condition from family experience and known high prior risk. Genotypic screening, however, can be used to detect conditions such as hemochromatosis, for which there are simple effective treatments. Genotypic screening can also be used for population screening and extended family testing. In population screening for hemochromatosis, most individuals will have no direct experience of the condition, and the purpose of testing is to initiate effective treatment and prevent disease. The genetic test is therefore being used in a different context than its use in family-based testing in which, as stated before, the individual is already identified as being at risk and has prior knowledge of the condition, and in which the actual risk of disease may be higher because of the aggregation of familial genetic and environmental risk factors. The use of a genetic test in the context of population screening for hemochromatosis is more analogous to other screening tests that are used to identify individuals at risk of disease to offer treatment, such as routine cholesterol testing.

In this article we report the results of a pragmatic trial comparing a genotypic strategy with a phenotypic strategy for population screening for hemochromatosis. Assessments were made of the impact of the two strategies on perceptions of health, anxiety, and depression as part of a randomized controlled equivalence trial comparing the uptake, feasibility, and cost of each strategy offered to an adult primary care population in the United Kingdom.

MATERIALS AND METHODS

The full methods have been reported.¹⁴ In brief the trial randomly selected 3000 participants from two general practice (family doctor) registers; the participants were aged between 30 and 70 years and stratified by two age groups and gender. They were randomized to the phenotypic or the genotypic screening strategy and invited by their general practitioner to participate in the study. Both groups were given written information about hemochromatosis stating that it was a treatable disorder, it was genetic, there were familial implications to a diagnosis, and the benefit of early diagnosis was not clear. The phenotypic group was offered an appointment at the general practitioner's surgery for a blood test to be taken measuring transferrin saturation. They were told that this test would measure the levels of iron in their blood. The participants in the genotypic group were offered a genetic test to be performed on a mouthwash sample at home and posted back to the laboratory. They were told that they would be tested for the genes that predisposed to developing hemochromatosis. Both groups were advised that they would be informed whether the result of the screening test was positive or negative. If it was negative they would be at low risk of developing iron overload because of hemochromatosis, and there would be no further investiga-

tions. If it was positive they were informed that they may be at risk of developing iron overload, and further investigations were offered as detailed next.

Individuals who refused testing were asked to return a short questionnaire indicating their reasons for refusal.

Individuals who screened positive (transferrin saturation > 45% or homozygosity for the C282 years mutation or compound heterozygosity for the C282 years and the H63D mutation) were invited to the hospital for a clinical assessment by a consultant hepatologist and further testing (fasting transferrin saturation, ferritin, liver function tests, and genotyping). A complete history was obtained, and a thorough physical examination was performed. This included particular focus on factors influencing iron status and risk factors for chronic liver disease including alcohol, metabolic syndrome, and viral hepatitis. Particular attention was given to symptoms and signs of hemochromatosis including chronic liver disease and joint disease. They were then assigned to one of three management pathways: discharged, monitored, or venesected (treated with phlebotomy). Full details of the clinical decisions and outcomes have been reported.¹⁴ Individuals who were homozygous for the C292Y mutation were venesected if their ferritin was increased or monitored if their ferritin was normal. Individuals with other genotypes were venesected if their ferritin was increased or at the upper limits of normal in combination with disturbed liver function test results. Compound heterozygotes were monitored if the individuals were male with increased iron indices (based on normal range in the laboratory) or were female but still menstruating. All other individuals were discharged as being at low risk of developing progressive iron overload. For the purpose of analysis, the individuals who were discharged were categorized as false-positive results and the individuals who were monitored or treated were categorized as true positives.

The study was conducted between 2001 and 2003, and ethical approval and signed consent were obtained (Southampton and Southwest Hampshire Local Research Ethics Committee Reference 382/00). Self-completed questionnaires were collected at invitation, testing, result giving, and 6 months follow-up. Psychologic assessments were made at each of these time points, and sociodemographic data were collected at invitation. The process of the study is outlined in Figure 1.

Validated tools were used; however, attention was paid to the length of the tools used and their relevance to a well population to address issues of feasibility, acceptability, and appropriateness.^{15,16} The tools used were the anxiety scale of the short form of the Spielberger State-Trait Anxiety Inventory,¹⁷ the depression subscale of the Hospital Anxiety and Depression scale,¹⁸ and the general health domain of the Short Form-36.¹⁹

Data analysis

Statistical analysis was conducted in SPSS (v11.5) (SPSS Inc., Chicago, IL). A one-way between-groups analysis of covariance was conducted to compare the psychologic assess-

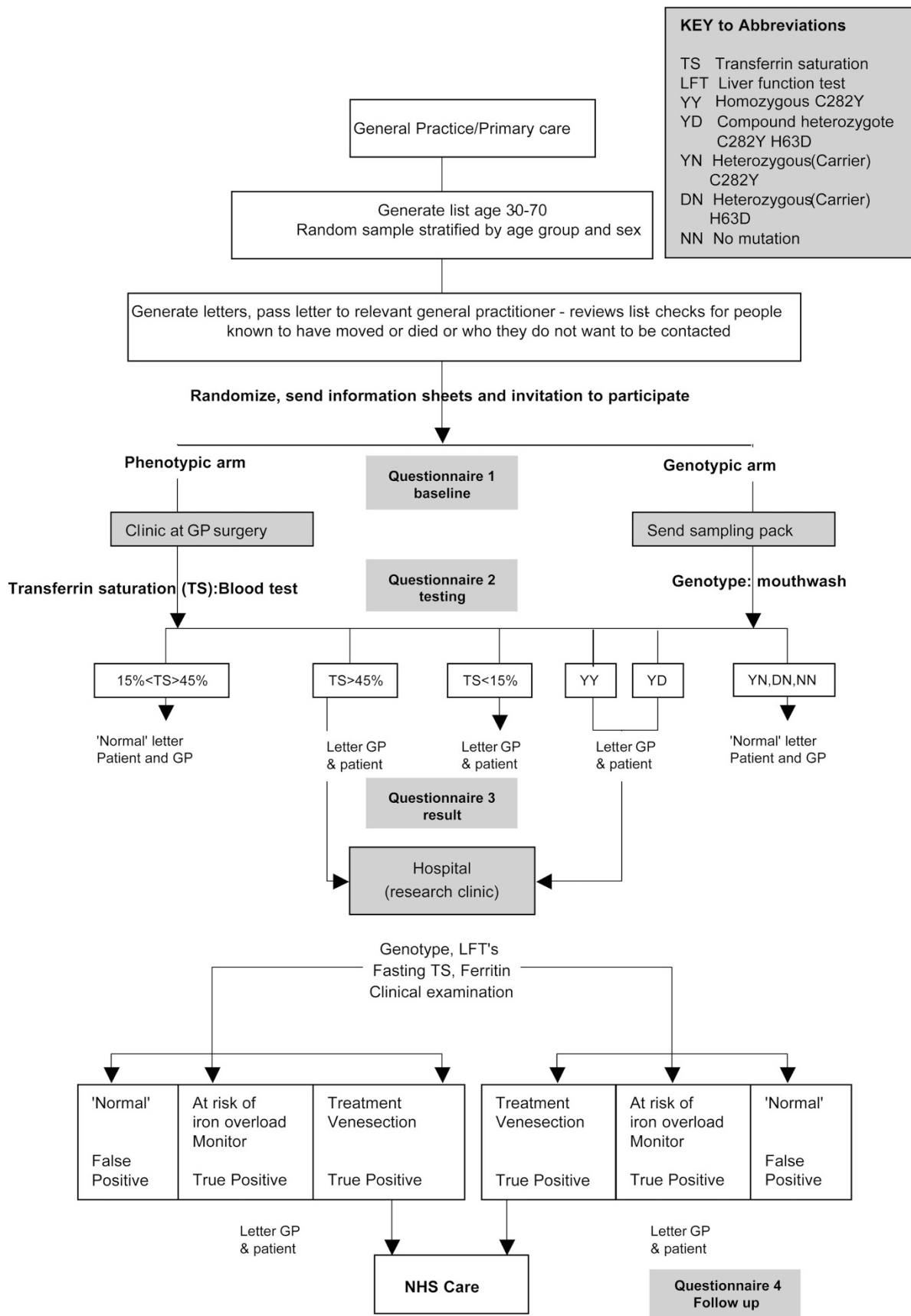


Figure 1. Flowchart of study.

ments between the two screening arms at four time points (invitation, testing, result giving, and follow-up). The independent variable was the screening strategy, and the dependent variable was the particular psychologic measure at testing, result, or follow-up. The psychologic assessment at baseline was used as the covariate in each analysis.²⁰

To investigate further the variables that might be affecting the outcome measures, the data were analyzed using a repeated-measurement, linear, mixed-effects model with time modeled as a fixed effect and participant effects as random effects.²¹ This analysis was performed in STATA (v8.2) (StataCorp, College Station, TX). All additional variables on which we had data were examined as fixed effects. These were age, sex, outcome of screening (screen negative, screen positive without iron overload [false positives], screen positive at risk of iron overload requiring monitoring or treatment), employment category (employed, retired/other, unemployed, disabled, as defined by the UK 2001 census questions),²² deprivation score (Townsend deprivation index derived from area of residence),²³ and screening strategy (phenotypic or genotypic). The Townsend Score is a measure of levels of material deprivation and includes four variables: unemployment, overcrowding (material living conditions), lack of owner-occupied accommodation (a proxy indicator of wealth), and lack of car ownership (a proxy indicator of income). The Townsend Score is a summation of the standardized scores (z scores) for each variable (scores >0 indicate greater levels of material deprivation). By using data from the UK Census, Townsend deprivation scores were mapped to geographic areas and patients' postcodes were linked to these areas.

All variables were included in the model.

RESULTS

The 939 individuals who accepted testing form the sample for this study. The response rates for the subsequent questionnaires were 94% at testing, 84% at result giving, and 77% at follow-up. There were no statistically significant characteristics in the baseline characteristics or measurements between those who returned questionnaires and those who did not, or between the two screening strategies. The baseline characteristics of participants together with the outcomes of the screening are shown in Table 1.

As reported previously, the 220 individuals who gave information as to the reasons for refusing screening had the same characteristics as those who accepted screening; they were more likely to be female and older, and to live in areas of less social deprivation.¹⁴ They were therefore not representative of the total sample. The most frequent reason given was "not interested" or "not enough time," and this was the same for both phenotypic and genotypic strategies. Twenty-seven percent of the genotypic screening arm said they did not want a genetic test. Twenty-four percent in the phenotypic arm did not want a blood test. Only one person mentioned insurance as a specific concern. This person was in the phenotypic arm and had recent difficulties with obtaining life insurance because of an increased cholesterol level. Additional data were collected at baseline relating to previous experience of genetics services, genetic testing, or knowledge of hemochromatosis. There were no significant differences between those accepting and those refusing screening. Approximately 2% of people had been seen by a genetics department, 1% had undergone a genetic test, and 9% had heard of hemochromatosis

Table 1
Baseline characteristics of participants and outcomes of screening n = 939

	Genotypic screening strategy N=497	Phenotypic screening strategy N=442
Sex Men	39%	40%
Mean age (SD)	52.6 (11.0)	53 (10.7)
Townsend deprivation score (median, interquartile range)	-0.29 -0.58 to 3.53	-0.29 -0.36 to 3.53
Employment category		
1 Employed	63%	64%
2 Retired/other	32%	30%
3 Unemployed	2%	2%
4 Disabled	3%	4%
Outcome of screening		
1 Negative	98.6%	88.7%
2 Screen positive; at risk of iron overload requiring monitoring or treatment.	1% n=5	1.1% n=6
3 Screen positive; without iron overload.	0.4% n=2	10.2% n=44

SD, standard deviation.

There were no statistically significant differences in the results of the analyses of covariance between the two screening strategies apart from at testing, with those in the phenotypic screening group having lower depression and anxiety and higher health perception. The differences in the adjusted means, although statistically significant, were small and unlikely to be clinically significant (Table 2).

The study was designed as an equivalence trial with a primary outcome of uptake of screening, and the sample size was determined accordingly. In an equivalence trial sample size is determined by setting a predefined limit of equivalence, in addition to the probability of a type 1 or 2 error; if the 95% confidence interval (CI) around the estimate of the difference is within this limit there is deemed to be no difference.²⁴ A post hoc analysis of equivalence of the outcome measures demonstrated equivalence between the two screening strategies at follow-up. The limit of difference was set at $\pm 10\%$ of each scale. The general health profile subscale of the Short Form-36 is divided into bands of 10 points for the purpose of comparison with other scales. The depression scale of the Hospital Anxiety and Depression scale and the State-Trait Anxiety Inventory are divided into three bands for clinical comparisons. The choice of $\pm 10\%$ of each scale as the limit of equivalence is therefore conservative.

A further one-way between-groups analysis of covariance was conducted to compare the psychologic assessments between the screening outcomes (screen positive and screen negative) at the four time points. The independent variable was the screening outcome (screen positive or screen negative), and the dependent variable was the particular psychologic measure at testing, result, or follow-up. The psychologic assessment at baseline was used as the covariate in each analysis. There were no statistically significant differences. The numbers in the true and false-positive outcome groups were too small to perform further meaningful statistical analyses.

The results of the modeling confirmed no effect of screening strategy or time on any of the measures. In the model there was a small effect of screening outcome on anxiety, although not on depression or general health; those who might be categorized as false positives were less anxious after adjustment for the other variables in the model (regression estimate -3.53 95% CI -6.77 to -0.29). There appeared to be a small improve-

ment in depression and anxiety from invitation over time with no change in health perception. There was no effect of age or deprivation score on the outcomes when adjusted for all the variables. There was no effect of gender or screening outcome on depression or health perception. There was an effect of gender on anxiety with women being slightly more anxious (regression estimate 1.98 95% CI 0.55 – 3.42).

In this study employment category appeared to have an effect on the outcome measures; individuals who were unemployed or disabled were significantly more anxious and depressed, and had a lower health perception adjusted for the other variables in the model. The regression estimate for an unemployed individual compared with an employed individual was -32.6 (95% CI -39.4 to -25.7) on the general health perception scale (maximum score 100 higher score = better health), 9.75 (95% CI 5.71 to 13.8) on the State-Trait Anxiety Inventory (maximum score 80 higher score = more anxious), and 4.96 (95% CI 3.87 to 6.05) on the depression scale (maximum score 21 higher score = more depressed).

DISCUSSION

This study was designed to compare two pragmatic screening strategies, genotypic and phenotypic, for a treatable disease offered to a previously unaware population. Response rates were high in the screened individuals who were the sample for this study, with no differences between those who dropped out and those who did not. In addition, the use of the repeated-measurement, linear, mixed-effects model allowed the use of data from subjects who provided baseline responses and at least one other response. Thus the likelihood and size of bias in the results caused by missing data and dropouts are reduced.

There was no effect of the screening strategy on the outcome measures. Our previous analysis of equivalence/noninferiority showed no difference in uptake between the two strategies.¹⁴ These findings suggest that when a genotypic strategy is used for screening in this context the outcomes are similar to a phenotypic strategy. The method of testing was different between the two groups in that the genotypic testing was performed at home on a mouthwash sample, and the phenotypic testing was performed on a blood sample taken at the doctor's surgery. This is a potential limitation of the study. However, as recommended in guidelines for evaluating screening programs, the study was designed to compare two strategies rather than two tests, and one advantage of genetic testing strategies is that they could be offered in this way. This provides some further support to arguments against "genetic exceptionalism," that is, the claim that genetic information is unique and has specific properties that require special consideration in all situations. Although it is argued that the use of genetic information to predict disorders such as Huntington's disease in asymptomatic individuals does require special consideration, tests should be categorized by their qualities and by reference to the circumstances in which they are used rather than giving genetic tests special status.^{25,26} In this study both the phenotypic and the genotypic tests were being used in the same context, and there

Table 2

Psychologic assessments over time by screening arm adjusted by baseline

Assessment Time	Assessment Type	p-Value	Adjusted Means (95% CI)	
			Genetic	Biochemical
Testing	HAD (Depression scale)	<0.001	3.9 (3.8 to 4.1)	3.4 (3.2 to 3.6)
	STAI (Anxiety scale)	0.005	34.8 (33.9 to 35.7)	32.9 (31.0 to 33.9)
	GHP (General Health Perception scale)	<0.001	69.0 (68.0 to 70.1)	72.0 (70.9 to 73.1)

CI, confidence interval; HAD, Hospital Anxiety and Depression scale; STAI, State-Trait Anxiety Inventory.

appeared to be no difference in the outcomes for the participants.

As found in other studies, there appeared to be an improvement in some of the outcome measures over time compared with the assessments at invitation.²⁷ These improvements were small and unlikely to be of clinical significance, but do suggest that the process of screening does not lead to increases in anxiety or depression or decreasing health perception. Invitation to participate in screening programs or research may cause temporary disturbances to quality of life, and the small improvement seen in this study may reflect a return to the baseline state. The distributions of the scores in the measures used were similar to available population norms.¹⁹ However, there are limited published data from comparative UK populations. In this study the majority of persons screened negative. Therefore, it would be expected that any disturbance would be temporary. There also appeared to be no increase in anxiety or depression or decrease in general health perception in those who screened positive at the first stage of the screening process.

More false positives were detected with the phenotypic screening strategy. In the direct comparison of the outcome groups there did not appear to be an effect of screening outcome (i.e., screen positive or screen negative on the measures). The numbers in the true and false-positive outcome groups were too small to perform further direct comparisons. However, in the model there appeared to be a small decrease in anxiety adjusted for all other factors in the model including those in the screening group categorized as false positives. This is difficult to interpret but may reflect the greater clinical involvement with these individuals, together with the positive aspects of being told one is at low risk of the condition being screened for. Other studies investigating the acceptability of genotyping in the context of testing or screening for hemochromatosis have also found few adverse effects.²⁸ In this study the genetic test was used as a screening test, and participants were aware that they would be told if they were “screen positive” or “screen negative.” They were not informed of their carrier status, and therefore we have no information relating to any possible adverse effects of this information. The participants who were categorized as “true” positives (i.e., were treated by venesection [phlebotomy] or monitored) were seen by the genetic counseling services, and testing of first-degree relatives was performed according to the usual clinical practice.

It is known that there is a link between employment status and health assessed on a variety of measures, and this was also found in this study.²⁹ Deprivation did not seem to be a contributory factor in the analysis. The effect of employment status on outcomes is rarely reported in studies evaluating genetic screening or testing strategies. In a recent review of the methodology of longitudinal studies in genetics, 14 studies in adults were considered to be of adequate methodologic quality, none of which reported data on employment status.²¹ In this current study there was no apparent interaction between employment status and screening strategy. However, the fact that we have demonstrated a large effect on the outcome measures indicates that employment status is a potential powerful confounder

that might introduce bias into future studies if not taken into account in their design and analysis.

In summary this study compared a genotypic strategy with a phenotypic strategy for screening for a treatable disease in a previously unaware population. Neither screening strategy had significant effects on health perception, anxiety, or depression. In addition there appeared to be no differences in the outcome measures between the two strategies. This study provides some empiric data to support arguments against “genetic exceptionalism” and suggests that genetic testing when used for population screening for a treatable disease has few adverse effects

Acknowledgements

The work was funded by the NHS Research and Development directorate award number RDF084 and the European Union through Grant QLK6-1999-02237 from Framework 5 to the PISDAP consortium. Christine Patch is currently funded by The Health Foundation (UK), previously the PPP Foundation.

References

- Collins FS, McKusick VA. Implications of the Human Genome Project for medical science. *JAMA* 2001;285(5):540–544.
- Burke W, Thomson E, Khoury MJ, McDonnell SM, et al. Hereditary hemochromatosis: gene discovery and its implications for population-based screening. *J Am Med Assoc* 1998;280(2):172–178.
- Cogswell ME, Burke W, McDonnell SM, Franks AL. Screening for hemochromatosis. A public health perspective. *Am J Prev Med* 1999;16(2):134–140.
- McCullen MA, Crawford DHG, Hickman PE. Screening for hemochromatosis. *Clin Chim Acta* 2002;315:169–186.
- Witte DL, Crosby WH, Edwards CQ, Fairbanks VF, et al. Practice guideline development task force of the College of American Pathologists. Hereditary hemochromatosis. *Clin Chim Acta* 1996;245(2):139–200.
- Burke W, Atkins D, Gwinn M, Guttmacher A, et al. Genetic test evaluation: information needs of clinicians, policy makers, and the public. *Am J Epidemiol* 2002;156(4):311–318.
- National Screening Committee. *Second Report of the UK National Screening Committee*. London: Department of Health; 2000.
- Ayme S, Bobrow M, Coviello D, Evers-Kiebooms G, et al. Population genetic screening programmes: proposed recommendations of the European Society of Hum Genet. *Eur J Hum Genet* 2000;8(12):998–1000.
- Nuffield Council on Bioethics. *Genetic Screening: Ethical Issues*. London: Nuffield Council on Bioethics; 1993.
- World Health Organisation. *Proposed International Guidelines on Ethical Issues in Medical Genetics and Genetics Services*. Geneva: World Health Organization; 1998.
- Almqvist EW, Bloch M, Brinkman R, Craufurd D, et al. A worldwide assessment of the frequency of suicide, suicide attempts, or psychiatric hospitalization after predictive testing for Huntington disease. *Am J Hum Genet* 1999;64(5):1293–1304.
- Almqvist EW, Brinkman RR, Wiggins S, Hayden MR. Psychological consequences and predictors of adverse events in the first 5 years after predictive testing for Huntington’s disease. *Clin Genet* 2003;64(4):300–309.
- Meiser B, Butow P, Friedlander M, Barratt A, et al. Psychological impact of genetic testing in women from high-risk breast cancer families. *Eur J Cancer* 2002;38(15):2025–2031.
- Patch C, Roderick P, Rosenberg W. Factors affecting the uptake of screening: a randomised controlled non-inferiority trial comparing a genotypic and a phenotypic strategy for screening for haemochromatosis. *J Hepatol* 2005;43:149–155.
- Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321(7262):694–696.
- Fitzpatrick R. Measurement issues in health-related quality of life: challenges for health psychology. *Psychol Health* 2000;15(1):99–108.
- Marteau TM, Bekker H. The development of a 6-item short form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *Br J Clin Psychol* 1992;31:301–306.

18. Zigmond AS, Snaith P. The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica* 1983;67:361–370.
19. McHorney CA, Ware JE, Raczek AE. The MOS 36 item short form health survey (SF-36) psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31:247–263.
20. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *BMJ* 2001;323:1123–1124.
21. Timman R, Stinjen T, Tibben A. Methodology in longitudinal studies on psychological effects of predictive DNA testing: a review. *J Med Genet* 2004;41:e100.
22. Census dissemination unit. Census knowledge base. *Office of National Statistics*. Available at: <http://www.census.ac.uk/cdu/>. Accessed April 21, 2005.
23. Phillimore P, Beattie A, Townsend P. Widening inequality of health in northern England, 1981–91. *BMJ* 1994;308:1125–1128.
24. Jones B, Jarvis P, Lewis JA, Ebbutt AF. (1996). Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996;313:36–39.
25. Burke W, Pinsky LE, Press NA. Categorizing genetic tests to identify their ethical, legal, and social implications. *Am J Med Genet* 2001;106(3):233–240.
26. Green MJ, Botkin JR. “Genetic exceptionalism” in medicine: clarifying the differences between genetic and nongenetic tests. *Ann Intern Med* 2003;138(7):571–575.
27. Van Maarle MC, Stouthard ME, Bonsel GJ. Quality of life in a family based genetic cascade screening programme for familial hypercholesterolaemia: a longitudinal study among participants. *J Med Genet* 2003;40(1):E3.
28. Hicken BL, Calhoun DA, Barton JC, Tucker DC. Attitudes about and psychosocial outcomes of HFE genotyping for hemochromatosis. *Genet Test* 2004;8(2):90–97.
29. Dooley D, Fielding J, Levi L. Health and unemployment. *Ann Rev Public Health* 1996;17:449–465.