Outcomes of an Insurance Company-Sponsored Multichannel Chemistry Screening Initiative

James W. Mold, MD; Cheryl B. Aspy, PhD; and Frank H. Lawler, MD, MPSH
Oklahoma City, Oklahoma

BACKGROUND. The use of serum chemistry panels as screening tests has been studied in a variety of clinical and nonclinical settings. None of the studies, however, has attempted to carefully examine any potential harm done to participants, and none has measured the impact on health–related quality of life.

METHODS. Consenting participants in an insurance company–sponsored screening initiative completed a questionnaire before and 6 months after having blood drawn for a 25-item chemistry panel and a lipid profile; for men older than 50, a prostate-specific antigen (PSA) test was also done. The prescreening questionnaire included demographic and health information. The postscreening questionnaire included questions about specific outcomes. Included in both questionnaires were single-item measures of self-rated health and self-rated worry about health, and the 17-item Duke Health Profile (DUKE), a measure of health-related quality of life. Various outcomes were examined, including the numbers of new diagnoses, numbers and types of new treatment recommendations, change in self-reported health and worry, and change in DUKE subscale scores. Participants who were potentially helped and those who were potentially harmed by the serum chemistry panels screening program were identified and further characterized.

RESULTS. Of the 2249 subjects who satisfactorily completed both questionnaires, 2012 (89%) had at least one abnormal test result, but only 985 of these (49%) remembered having discussed their test results with a physician. A total of 342 individuals received new treatment advice. However, 29 (10%) of them indicated that they would be "somewhat unlikely" to "very unlikely" to follow it. Following the intervention questionnaire, there were statistically significant average decrements in the General Health, Physical Health, and Pain subscales of the DUKE for participants with abnormal results. Self-rated health status did not change, but level of worry about health increased significantly. At least 250 (11%) subjects were potentially helped by the screening initiative, but at least 574 (26%) were potentially harmed by it.

CONCLUSIONS. The use of serum chemistry panels as screening tests in nonclinical settings should probably be discouraged, since health-related quality of life is not improved and the intervention may harm more individuals than it benefits.

KEY WORDS. Clinical chemistry; outcome assessment; laboratory; data analysis. *(J Fam Pract 1998; 47:110-117)*

Serum chemistry panels, first made possible in the 1960s by automated multichannel analyzers, dramatically changed the way physicians order common blood tests. Rarely are all of the tests on a panel indicated even in an illness situation. These panels are often used as screening tests by health departments at health fairs and by physicians as part of annual physical examinations. Proponents argue that conditions such as diabetes and hyperparathyroidism are occasionally discovered, and early treatment may improve health outcomes. Critics mention that the high rates of false-positive test results lead to additional testing, and the discovery of new diagnoses may only increase anxiety without improving health.

Submitted, revised, February 19, 1998.
From the Department of Family and Preventive Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma. Requests for reprints should be addressed to James W. Mold, MD, University of Oklahoma Health Sciences Center, 900 NE 10th Street, Oklahoma City, OK 73104.

During the last 30 years, a number of studies have been conducted to determine the impact of the use of serum chemistry panels on cost and outcomes in both clinical[1-13] and nonclinical[14-16] settings. Representative of the clinical studies are those of Romm[4] in the United States and Ruttiman and colleagues[5] in Switzerland that demonstrated that relatively few new diagnoses (mainly, elevated lipid concentrations) and few new treatment decisions resulted from their use.

Studies conducted in nonclinical settings report similar results, with the additional finding that physicians respond to abnormal results discovered through serum chemistry tests in only a minority of cases (15% to 30%). Reasons given include: the result lacked significance; the result was previously known or suspected; the physician questioned the testing method; and the patient was no longer under the physician's care or was uninterested in further evaluation. The largest percentage of new diagnoses and management changes were again associated with elevated blood glucose and lipid levels.[14]

On the basis of these same findings, proponents argue that chemistry panels lead to the discovery of an occasional patient with an unsuspected, treatable condition, and the additional costs are not unreasonable. However, this argument assumes that the additional costs are entirely monetary, and that no actual harm is done to patients who have false-positive or positive but clinically unimportant test results, or to those who have results that lead to the diagnosis of untreatable conditions or conditions the patient chooses not to treat.

Unfortunately, when adverse effects of chemistry panels as screening instruments have been looked for, they have often been found.[17-24] In one study to evaluate adverse psychological effects associated with screening, Witte et al[16] found that 41% of individuals ignored abnormal results, 9% admitted to increased worry, and only 4% actually benefitted.

The purpose of our study of multichannel chemistries, offered by an insurance company as a promotion, was to measure both the positive and negative health effects of multichannel serum chemistry screening in a relatively healthy population in a nonclinical setting, including a measurement of health-related quality of life.

METHODS

Between February 1 and March 15, 1994, the Oklahoma State and Educational Employees Insurance Group offered its approximately 100,000 adult members older than 25 the opportunity to have a Chem 25 and lipid profile for $15. Men 50 years of age and older were also offered a prostate-specific antigen (PSA) test at no additional cost. Phlebotomy centers were established in 300 locations throughout the state. A total of 8818 enrollees chose to have the blood tests done. Included in the Chem 25 panel were measurements of serum glucose, blood urea nitrogen (BUN), creatinine, BUN to creatinine ratio, uric acid, sodium, sodium balance, potassium, chloride, carbon dioxide, calcium, phosphate, total protein, albumin, globulins, albumin to globulin ratio, total cholesterol, triglycerides, total bilirubin, alkaline phosphatase, aspartate aminotransferase (SGOT, AST), creatine phosphokinase (CK, CPK), lactate dehydrogenase (LD), iron, alanine aminotransferase (SGPT, ALT), gammaglutamyl transpeptidase (GGT, GGTP), and iron. The lipid panel included high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), very low-density lipoprotein cholesterol (VLDL), total cholesterol to HDL ratio, and LDL plus VLDL. Height, weight, and seated blood pressure were also measured, and participants were queried regarding their personal and family history of heart attack. Prostate-specific antigen levels were done on 2268 of the 2342 eligible men. For unknown reasons an additional 285 men younger than 50 also had PSA levels measured.

Participants were asked to fast for 8 to 12 hours before testing. Tests were performed using a DAX multichannel analyzer (Miles Inc, Tarrytown, NY). Results were sent to participants and to their primary care physicians when requested. In addition to the actual test results, participants were sent information by the laboratory about their cardiac risk status, which included lipid levels, age, body mass index, blood pressure, and family history of heart attack.

Before having blood drawn, each individual was invited to participate in a research project involving completion of questionnaires before and again several months after the blood chemistry testing. The first questionnaire was completed by 4121 people. Follow-up questionnaires were mailed on July 22, 1994. Subjects who had not returned their follow-up questionnaire by October 4, 1994, were sent a second one.

The questionnaire contained 5-point Likert scale response options regarding self-rated health (poor to excellent) and level of worry about health (very worried to not at all worried). It also included the Duke Health Profile (DUKE), a 17-item health-related quality-of-life measure from which six health measures and four dysfunction measures were generated. The health subscales of the DUKE include physical, mental, social, general, perceived health, and self-esteem. The dysfunction subscales include anxiety, depression, pain, and disability. Subscale scores range from 0 to 100, reflecting low to high quantities of the construct measured. Since some items are shared, subscales cannot be considered independent.

Included in the second questionnaire were questions about outcomes of the screening initiative, such as whether abnormal results were found, whether these had been discussed with a doctor, whether new diagnoses were made or new treatments started, the likelihood that these new treatment recommendations would be followed (measured by a 5-point Likert scale ranging from very unlikely to very likely), the number of additional physician visits resulting from the testing, perceived effects of the screening program, and perceived effects of the program on individual and family concerns about health.

ANALYSIS

To evaluate the potential impact of loss to follow-up, descriptive statistics were tabulated for individuals who declined to participate in the study, those who completed the first questionnaire only, and those who completed both questionnaires. Information collected by the laboratory (sex, age, history of heart attack, blood pressure, height and weight, the percentage who had the PSA test done, and the percentage with at least one laboratory test abnormality) were compared using the chi-square statistic for categorical variables and analysis of variance for continuous variables.

Within the final study group (those who completed both questionnaires), subjects with at least one abnormal laboratory test result were identified and compared with those with no abnormal results. Results considered to be

of no clinical relevance were excluded. Thus, we did not count as abnormal values below the lower reference limit for uric acid, creatinine, PSA, ALT, total bilirubin, alkaline phosphatase, AST, or CK.

A "healthy" subgroup was defined by negative responses to three multi-item questions about current medical problems, past abnormal laboratory test results, and current use of medications; alcohol consumption of less than five drinks per week; and a score ≥ 90 on the general health subscale of the DUKE.

A subset of subjects was defined as "potentially helped" by the intervention by having received a new treatment and being likely or very likely to adhere to it. A subset of subjects was presumed to have been "potentially harmed" by the intervention if they had one or more unanticipated laboratory test abnormality and either (1) no new treatments prescribed, or (2) new treatments prescribed to which they were neutral, unlikely, or very unlikely to adhere. This definition was based on studies by Haynes et al,[17] Charlson and coworkers,[18] and Lerman and colleagues[19] who showed that adverse labeling effects were most often seen in those with a new diagnosis who did not comply with recommended treatment. Those indicating neutrality regarding adherence were included with those unlikely to adhere based on the assumption that even those who plan to adhere, often do not. Logistic regression was used to model the relationship between either "harmed" or "not harmed" and "helped" or "not helped" with demographic and the DUKE variables.

Differences between the DUKE subscale scores were analyzed separately for those participants with at least one abnormal test result and for those with no abnormalities, using *t* tests for dependent samples. Differences between levels of self-perceived health and worry for the two groups were tested for significance using the Wilcoxon signed rank test. Standardized response means (mean change scores divided by the standard deviation of change scores) for the DUKE subscales and change scores for levels of self-perceived health and worry were calculated. Change scores for the two groups were compared using independent *t* tests for the DUKE and the Wilcoxon rank sum test for self-perceived health and worry.

RESULTS

One thousand seven hundred forty-four people responded to the first mailing of the follow-up questionnaire. An addi-

TABLE 1

Comparison of Subjects Who Had the Blood Test Only, Those Who Also Completed the First Questionnaire, and Those Who Completed Both Questionnaires

	Blood Test Only (N=4745)*	Blood Test and Prescreening Questionnaire (N=1716)*	Complete Data (N=2249)*	Total (N=8710)*
Sex,† no. (%)				
Male	2009 (42)	746 (43)	885 (39)	3640 (41)
Female	2736 (58)	970 (58)	1363 (61)	5069 (58)
Patient age, years,‡ no. (%)				
16 to 30	88 (2)	85 (5)	36 (2)	209 (2)
31 to 45	1138 (24)	518 (31)	470 (21)	2126 (25)
46 to 65	2591 (56)	942 (56)	1286 (58)	4819 (56)
>65	860 (18)	143 (8)	416 (20)	1449 (17)
MI, no. (%)				
Yes	169 (4)	51 (3)	81 (4)	301 (4)
No	4543 (96)	1643 (97)	2164 (96)	8350 (96)
PSA done,‡ no. (%)				
Yes	1436 (30)	427 (25)	665 (30)	2528 (29)
No	3276 (70)	1267 (75)	1580 (70)	6123 (71)
Systolic blood pressure,‡ mean (SD)	133 (19)	131 (18)	131 (18)	132 (19
Height, mean (SD)	67 (3.9)	67 (4)	67 (4)	67 (4)
Weight,‡ mean (SD)	174 (41)	178 (43)	170 (37)	174 (40)
No. abnormal tests,‡ mean (SD)	3.8 (2.2)	3.8 (2.2)	3.6 (2.2)	3.74 (2.2)

MI denotes myocardial infarction; PSA, prostate-specific antigen; SD, standard deviation.
*Variable totals differ because of incomplete data.
† $P < .01$.
‡ $P < .001$.

TABLE 2

Results from the Postscreening Questionnaire (N=2249)

Item	No.	(%)
Your overall health		
Fair to poor	183	(8)
Average	421	(19)
Good to excellent	1645	(73)
Worry level in the past month		
Not at all to a little worried	1907	(85)
Somewhat worried	250	(11)
Worried to very worried	84	(4)
Discussed screening results with a physician	985	(45)
Screening resulted in a new diagnosis	278	(13)
New treatment was prescribed	342	(16)
New prescription	113	(33)
Stopped a prescription	12	(4)
Surgery	11	(3)
Hospitalization	5	(0.2)
Change in diet	197	(58)
Change in exercise	180	(53)
Other	48	(2)
Likelihood of following recommendations		
Somewhat to very unlikely	29	(10)
Neutral	9	(3)
Somewhat to very likely	256	(87)
Effect of screening on level of health concern		
Somewhat more to much more concerned	917	(42)
No effect	739	(34)
Somewhat less to much less concerned	544	(25)
Screening was helpful		
Yes	2069	(94)
No	132	(6)

tional 505 responded to the second mailing, for a total of 2249 people with useable initial and follow-up questionnaire data (response rate = 55%). Table 1 is a comparison of known variables for the 4745 people who had blood drawn but declined to complete the first questionnaire, the 1716 who completed the first questionnaire only, and the 2249 who provided complete information on questionnaires. On the basis of these variables, the groups appear comparable. The differences that are statistically significant are unlikely to be of any great importance to the subsequent analyses or conclusions. Table 2 displays outcomes reported by all 2249 subjects on the postscreening questionnaire.

Table 3 lists specific laboratory test abnormalities and the numbers of participants with these abnormalities who received new diagnoses. Lipid levels were the most likely to be abnormal. However, these abnormal results were associated with a new diagnosis only 15% of the time. Serum CO_2 was the next most likely to be abnormal, but only 12% of those with abnormal CO_2 values received a new diagnosis. Participants with abnormalities of lactate dehydrongenase, BUN, sodium, PSA, glucose, iron, cholesterol, triglycerides, or calcium were more likely to remember discussing their test results with a physician.

Comparisons of the DUKE subscale scores pre- and postintervention for subjects with and without abnormal test results are shown in Table 4. Significant decrements in physical health, general health, perceived health, and pain were seen in the group with one or more abnormal results.

TABLE 3

Number of Abnormal Test Results for Each Test and the Resulting New Diagnoses (N=2249)

Test	Total No. Abnormal Test Results (%)	New Diagnoses
Cholesterol	1372 (62)	196
Triglycerides	1242 (55)	192
LDL cholesterol	964 (43.5)	136
HDL cholesterol	944 (43)	134
CO_2	501 (23)	59
Glucose	176 (8)	30
GGT	114 (5)	24
Uric acid	134 (6)	20
Iron	116 (5)	20
Bilirubin	82 (3.7)	17
Potassium	103 (5)	16
PSA	44 (2)	12
SGPT	99 (4.5)	12
Alkaline phosphatase	45 (2)	11
CK	62 (2.8)	11
SGOT	54 (2.4)	9
LD	72 (3)	9
Calcium	24 (1)	6
Sodium	22 (1)	6
Phosphate	41 (2)	5
Creatinine	33 (1.5)	4
BUN	16 (1)	4
Protein	17 (1)	2
Chloride	19 (1)	1
Albumin	1 (.05)	0
Globulin	1 (.05)	0

Note: Each respondent was asked, "Did the recent screening blood tests result in the diagnosis of any new health problems?" Responses to this item for those with each abnormal test value are included in this table. Therefore, diagnoses do not necessarily relate directly to specific test abnormalities.

GGT denotes gammaglutamyl transpeptidase; SGPT, alanine aminotransferase; PSA, prostate-specific antigen; CK, creatinine phosphokinase; SGOT, aspartate aminotransferase; LD, lactate dehydrongenase; BUN, blood urea nitrogen.

The group with all normal laboratory test results experienced statistically insignificant decrements in most of the subscales. No significant differences were found between the two subgroups with respect to the DUKE subscale scores.

Levels of self-reported health did not change for either group. Subjects with abnormal results reported significantly more worry regarding their health after the screening program (P = .01). However, subjects with normal test results also reported an increase in worry that approached statistical significance (P=.053).

Based on our definitions of "potentially helped" and "potentially harmed," 250 (11%) subjects may have been helped and 574 (26%) may have been harmed by the intervention. Factors associated with being potentially helped, after controlling for new treatment, were age (odds ratio [OR] = 1.02; 95% confidence interval [CI], 1.01 -1.03; P = .03), self-rated health (OR = 0.97; CI, 0.95 - 0.99; P =.03), and normality of all test results (OR = 0.29, CI, 0.10 - 0.67; P = .01). Thus, individuals who were older, rated their health lower, and had one or more abnormal results were more likely to be "helped." Of those potentially helped, 31% reported more worry, 51% the same level of worry, and 18% less worry following the intervention.

Factors associated with being potentially harmed included age (OR = 0.97; CI, 0.96 - 0.98; P =.0001), race (OR = 1.83; CI, 1.09 - 3.22; P =.026), sex (OR = 0.64; CI, 0.5 - 0.81; P =.0002), physician consultation about results (OR = 0.65; CI, 0.50 -0.85; P =.002), self-rated health (OR=1.02; CI, 1.00 - 1.04; P =.03), "healthy" (OR=11.77; CI, 7.55 - 18.78; P =.0001), DUKE social health (OR=0.98; CI, 0.97 - 0.99; P =.0002), DUKE perceived health (OR=1.01; CI, 1.00 - 1.01; P =.037), and DUKE self-esteem (OR=1.01; CI, 1.00 - 1.03; P =.04). Thus individuals were more likely to be harmed if they were younger, white, or male; had not seen a physician during 2 years prior to the screening program; rated their health as good or better; were "healthy" (by our definition); and had a lower social health score, a higher perceived health score, and a higher self-esteem score on the DUKE. Of those potentially harmed, 18% reported more worry, 60% the same level of worry, and 22% less worry following the intervention. Standardized response means for the DUKE subscales did not differ significantly between those potentially helped and those potentially harmed.

In response to two follow-up questions, 2069 (94%) subjects said that the screening initiative was helpful, while 132 (6%) said that it was not. An overlapping 2064 (94%) said that they would participate again if it were offered to them, whereas 139 (6%) would not.

TABLE 4

Standard Response Means (SRM) Before and After Serum Chemistry Screening Panel Intervention for Subjects with and Without Abnormal Test Results

The DUKE Subscales*	Within Groups Comparisons: All Normal Results N	SRM	P	One or More Abnormal Result N	SRM	P	Between Groups Comparison, P
Physical health	151	-.12	.14	1920	-.11	.0001	.99
Mental health	150	+.04	.65	1883	-.03	.13	.39
Social health	153	-.12	.15.	1910	-.01	.59	.20
General health	141	-.11	.23	1786	-.09	.0002	.77
Perceived health	156	-.03	.70	1961	-.06	.006	.55
Self-esteem	151	-.13	.10	1888	-.01	.65	.12
Anxiety	150	+.10	.22	1864	+.03	.17	.41
Depression	150	+.02	.78	1889	+.04	.07	.82
Pain	158	-.09	.27	1985	-.08	.0002	.98
Disability	157	-.09	.29	1992	-.03	.14	.54

*The DUKE subscales range from 0 to 100, with 100 representing the highest level of the construct (ie, higher levels of anxiety, depression, pain, and disability are undesirable).

DISCUSSION

Our study provides more evidence suggesting that caution should be exercised by those who initiate mass screening initiatives of any kind, particularly those disconnected from the site and source of clinical care.[25-27] Unfortunately, it cannot put the arguments regarding multiphasic chemistry screening firmly to rest, since it was impossible to attach values to the various outcomes.

It is important to point out that the serum chemistry panel was not the only intervention to which our subjects were exposed. In addition to blood tests, they had their height, weight, and blood pressure checked, their cardiac risk factors determined, and an estimate of their overall risk for heart attack calculated and reported to them. Some had a PSA determination. These may have account-

ed for some of the effects observed.

While 89% of the participants had at least one abnormal test result, new treatments were prescribed for only 16%, the majority of which were recommendations regarding diet or exercise. Lipid level abnormalities were associated with the largest numbers of both new diagnoses and new treatments, mirroring the findings of Romm[4] and Ruttimann.[5] The next most frequent abnormality was CO_2, which was associated with, but probably not the reason for, 59 new diagnoses (Table 3). Abnormal test results most likely to be associated with new diagnoses were alkaline phosphatase (24%), uric acid (19%), calcium (25%), sodium (27%), BUN (25%), and PSA (36%). These results confirm the potential value of screening programs such as this one, since new diagnoses and treatment recommendations are made in a number of cases.

A previously demonstrated principle well illustrated by our study is that when screening programs are separated from the source of medical advice and care, a large number of people "fall through the cracks." Many of the laboratory test results never reached physicians' offices, and many people never discussed their abnormal test results with a physician. Only 43% of all subjects and 49% of those with abnormal results remembered having discussed their results with a physician. This was a somewhat lower percentage than was found by Witte et al[16] (59%), but a somewhat higher percentage than Hyman and colleagues[28] reported for cholesterol screening (33%).

The small, though statistically significant, drops in several of the DUKE subscale scores observed may or may not have clinical meaning. General health-related quality-of-life (HRQL) measures have sometimes been relatively insensitive to change even following significant interventions such as radical prostatectomy[29] and treatment for sinusitis.[30] Standardized response means (SRMs) of 0.3 to 1.13 have generally been reported from trials in which clinically significant changes in the components of HRQL were felt to have occurred.[31,32] For instance, in a recent study using the DUKE, successful treatment of men with impotence resulted in mental health SRMs of +0.36 (G. Parkerson. Personal communication, November 1997.) The greatest SRM seen in our study was -0.17 for perceived health and -0.16 for physical health in those potentially helped. This may reflect an appropriate response to the recognition of health problems. It is also conceivable that a number of those who chose to be screened did so because they suspected they were becoming ill and subsequently did. It is probably reasonable to conclude, however, on the basis of the small, but generally negative, changes observed in our study that HRQL did not improve as a result of the intervention.

Worry increased significantly among those who had at least one abnormal test result. That worry increased to some degree in the group without abnormal test results is more interesting. This may be attributable, in part, to the other interventions (height, weight, blood pressure, and cardiac risk assessment); to concern about the clinically nonsignificant abnormal results not included in our analyses (low uric acid, creatinine, PSA, ALT, and so forth) among those who never discussed the results with a physician; or to questions included in our questionnaires that may have increased their focus on ongoing health concerns. It is unlikely to be a secular (due to the passage of time) or seasonal (more health problems in certain seasons) trend, since the duration of the study was only 6 months and it was conducted from spring to early fall.

If the chemistry panel intervention was helpful, it was primarily because it detected abnormalities that led to treatment advice that resulted in better health outcomes. Our relatively narrow definition of "potentially helped," therefore, did not include those who were reassured by normal results. However, it no doubt included a number of individuals who were given new treatment advice that did not ultimately prove to be effective or that they, despite their best intentions, did not follow. These individuals might either not have been helped or might have been harmed by the intervention.

Our definition of "potentially harmed" was chosen according to the results of the Haynes hypertension screening study and the Lerman Pap smear study, both of which suggest that those patients who are given a new diagnosis (in our study, an unanticipated laboratory test abnormality) and who are unlikely to follow treatment advice are the ones most likely to be harmed.[17-19] Also included in our definition were those with an abnormal result for which no treatment advice was given. This is supported by studies documenting the adverse impact of the discovery of conditions for which treatment is unavailable or unnecessary.[20-23] Clearly not all of the individuals classified as potentially harmed were actually harmed by the intervention. In many cases the abnormality was dismissed by their physician as unimportant, and they thought no more about it. Some may actually have been helped, for example, by this opportunity to discuss other concerns with their physician.

Using the outcome measures chosen, we were unable to document an adverse impact associated with our definition of potential harm. In retrospect, this could probably have been anticipated. The labeling effects seen in most other screening studies were documented using measures that did not depend on self-reported discomfort or ill health (eg, absenteeism from work). The potential harm associated with diagnosis or treatment may be more likely to occur in individuals who choose to minimize or deny the importance of their condition. This possibility is supported by the findings of Haynes[17] and Charlson[18] that those most likely to demonstrate increased absenteeism were less likely to seek or adhere to treatment. The subgroup we identified as potentially harmed were less likely to have seen a physician during the 2 years previous to the screening program and reported better health but had lower social health scores on the DUKE than other partic-

ipants. That we were unable to document harm should in no way lead to the conclusion that no harm was done.

STUDY LIMITATIONS

Several potential limitations of the study deserve attention. As with any study in which incomplete participation and loss to follow-up occur, there is the potential for selection bias. Individuals who chose to participate in the screening initiative and those who completed the follow-up questionnaire were certainly different in some ways from those who did not. It is reassuring that the differences found in known variables did not appear to be clinically important even though some were statistically significant. No significant differences were found between early and late responders to the second questionnaire with regard to sex, race, worry, or self-rated health. We did find that the early responders were younger ($\chi^2 = 35.8$, $df = 3$, $P = .0001$). We suspect, but cannot prove, that those who did not complete the second questionnaire were likely to have been less pleased with some aspect of the intervention than those who did. This may have resulted in an overestimate of the benefits and an underestimate of the harm caused by the intervention.

Much of the data regarding outcomes is from self-report and is therefore subject to the errors inherent in that method. The DUKE and the question regarding self-reported health are well-validated measures. However, other questions included in the questionnaires were designed for use in this study and may or may not suffer from problems of reliability and validity. Since the second questionnaire was sent out as much as 8 months after the initial screening, some participants may have had difficulty accurately recalling certain information, such as number of additional physician visits or whether a new diagnosis or treatment was the result of the blood test results. Any analysis of health outcomes should take into account the possibility that short- and long-term outcomes may be different. That is, effective diagnosis and treatment may initially cause discomfort and disability but eventually result in an increase in quality or quantity of life. It is therefore difficult to know at which point to measure outcomes. We chose 6 months, hoping that most of the acute adverse effects of the intervention would have faded and that long-term effects would have begun to be appreciated. A longer time before follow-up would have been problematic because of recall bias and greater loss to follow-up.

CONCLUSIONS

The use of clinical chemistry panels for screening in nonclinical settings should probably be discouraged. Only 49% of the participants with abnormal test results ever discussed them with a physician. Of those who did, approximately one third received new treatment advice, and of these, 10% admitted to being unlikely to follow the advice. Based on previous research, the kinds of advice given are unlikely to have a major positive long-term health impact. The intervention appears to increase worry about health and may result in either no change in or a lower health-related quality of life. Finally, it appears likely that more individuals are potentially harmed than are potentially helped by such an intervention.

ACKNOWLEDGMENTS

The authors wish to thank the Oklahoma State and Educational Employees Insurance Program for contributions in data collection, Debra Bemben for help with data analysis, and Jan Wilmoth for preparation of this manuscript.

REFERENCES

1. Carmalt MH, Freeman P, Stephens AJ, Whitehead TP. Value of routine multiple blood tests in patients attending the general practitioner. BMJ 1970: 7:620-3.
2. Friedman GD, Goldberg M, Ahuja JN, Siegelaub AB, Bassis ML, Collen MI. Biochemical screening tests: effect of panel size on medical care. Arch Intern Med 1972; 129:91-7.
3. Schneiderman LJ, DeSalvo L, Baylor S, Wolf PL. The "abnormal" screening laboratory results: its effect on physician and patient. Arch Intern Med 1972; 129:88-90.
4. Romm FJ. Routine chemistry testing. Fam Med 1986;18:230-2.
5. Ruttimann S, Dreifuss M, Clemencon D, DiGallo A, Dubach UC. Multiple biochemical blood testing as a case-finding tool in ambulatory medical patients. Am J Med 1993; 94:141-8.
6. Cebul RD, Beck JR. Biochemical profiles: applications in ambulatory screening and preadmission testing of adults. Ann Intern Med 1987; 106:403-13.
7. Cullen SA. Place of biochemical tests in aircrew medical examinations. Aviat Space Environ Med 1989; A48-9.
8. Durbridge TC, Edwards F, Edwards RG, Atkinson M. Evaluation of benefits of screening tests done immediately on admission to hospital. Clin Chem 1976; 22:968-71.
9. Fletcher SW, Sourkes M, Rabzel M, Fletcher RH. Multiphasic screening: case-finding tool in a teaching hospital medical clinic. JAMA 1977; 237:887-91.
10. Kaplan EB, Sheiner LB, Boeckmann AJ, Roizen MF, Beal SL. The usefulness of preoperative laboratory screening. JAMA 1985; 253:3576-81.
11. Rubenstein LV, Ward NC, Greenfield S. In pursuit of the abnormal serum alkaline phosphatase: a clinical dilemma. J Gen Intern Med 1986; 1:38-43.
12. Sheard SC. Haematological and biochemical screening of potential Royal Naval aircrew. J Roy Nav Med Serv 1991; 77:11-6.
13. Williamson E, VanPeenen HJ. Patient benefit in discovering occult hyperparathyroidism. Arch Intern Med 1974; 133: 430-1.
14. Bates B, Yellin JA. The yield of multiphasic screening. JAMA 1972; 222:74-8.
15. Olsen DM, Kane R.L, Proctor PH. A controlled trial of multiphasic screening. N Engl J Med 1976; 294:925-30.
16. Witte DL, Angstadt DS, Schweitzer JK. Chemistry profiles in "wellness programs": test selection and participant outcomes. Clin Chem 1988; 34:1447-50.
17. Haynes RB, Sackett DL, Taylor DW, Gibson ES, Johnson AL. Increased absenteeism from work after detection and labeling of hypertensive patients. N Engl J Med 1978; 299:741-4.
18. Charlson ME, Alderman M, Melcher L. Absenteeism and labelling in hypertensive subjects. Am J Med 1982; 73:165-70.
19. Lerman C, Miller SM, Scarborough R, et al. Adverse psychologic consequences of positive cytologic cervical

screening. Am J Obstet Gynecol 1991; 165:658-62.

20. Cadman D, Chambers LIW, Walter SD, et al. Evaluation of public health preschool child development screening: the process and outcomes of a community program. Am J Public Health 1987; 77:45-51.
21. Hampton MI, Anderson J, Lavizzo BS, Bergman AB. Sickle cell nondisease. Am J Dis Child 1974; 128:58-61.
22. Berman AB, Stamm SJ. The morbidity of cardiac nondisease in schoolchildren. New Eng J Med 1967; 276:1008-13.
23. Sorenson JR, Levy HL, Mangione TW, Sepe SJ. Parental response to repeat testing of infants with "false positive" results in a newborn screening program. Pediatrics 1986; 73:183-7.
24. Fyro K, Bodegard G. Four year follow-up of psychological reactions to false positive screening tests for congenital hypothyroidism. Acta Paediatr Scand 1987; 76:107-14.
25. Berwick DM. Screening in health fairs. JAMA 1985; 254:1492-8.
26. Donaldson RM. The trouble with screening. Cortlandt Forum 1988: 11.
27. Lentsch PM. Screening of older adults — an ethical issue [editorial]. J Gerontol Nurs 1981; 7:215-6.
28. Hyman DJ, Flora JA, Reynolds KD, et al. The impact of public cholesterol screening on diet, general well-being, and physician referral. Am J Prev Med 1991; 7:268-72.
29. Litwin MS, Hays RD, Fink A, et al. Quality-of-life outcomes in men treated for localized prostate cancer. JAMA 1975; 273:129-35.
30. Gliklich RE, Hilinski JM. Longitudinal sensitivity of generic and specific health measures in chronic sinusitis. Qual Life Res 1995; 4:27-32.
31. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. J Clin Epidemiol 1997; 50:79-93.
32. Pouchot J, Buillemin F, Coste J, Bregeon C, Sany J. Validity, reliability, and sensitivity to change of a French version of the arthritis impact measurement scales 2 (AIMS2) in patients with rheumatoid arthritis treated with methotrexate. J Rheumatol 1996; 23:52-60.