Papers

Cohort study of association of risk of breast cancer with cyst type in women with gross cystic disease of the breast

Paolo Bruzzi, Luigi Dogliotti, Carlo Naldoni, Lauro Bucchi, Massimo Costantini, Alessandra Cicognani, Mirella Torta, Gian Franco Buzzi, Alberto Angeli

Abstract

Objective: To assess correlation between type of breast cyst and risk of breast cancer in women with gross cystic disease of the breast.
Design: Cohort study of women with breast cysts aspirated between 1983 and 1993 who were followed up until December 1994 for occurrence of breast cancer.
Setting: Major cancer prevention centre.
Subjects: 802 women with aspirated breast cysts.
Main outcome measures: Type of breast cyst based on cationic content of cyst fluid: type I (potassium:sodium ratio >1.5), type II (potassium:sodium ratio <1.5), or mixed (both types). Subsequent occurrence and type of breast cancer.
Results: After median follow up of six years (range 2-12 years) 15 cases of invasive breast cancer and two ductal carcinomas in situ were diagnosed in the cohort: 12 invasive cancers (and two carcinomas in situ) among the 417 women with type I cysts, two cancers among the 325 women with type II cysts, and one among the 60 women with mixed cysts. The incidence of breast cancer in women with type I cysts was significantly higher than that in women with type II cysts (relative risk 4.62 (95% confidence interval 1.26 to 29.7)). These results were confirmed after adjustment for several risk factors for breast cancer (relative risk 4.24 (1.12 to 27.5)).
Conclusions: The increased risk of breast cancer of women with breast cysts seems to be concentrated among women with type I breast cysts.

Introduction

Gross cystic disease of the breast, defined on the basis of the presence of palpable cysts, is reported to occur in 7% of women in the Western world.[1] Overall, the available evidence suggests that women with breast cysts have a moderately increased risk of breast cancer.[1-11] It should be noted that breast cysts are not considered to be premalignant lesions but are simply markers of an increased risk affecting the whole organ.[1]

It has been observed that the fluid in breast cysts consistently shows a bimodal distribution in the concentration of many substances, including cations,[12-17] hormones,[12-19] and growth factors.[15 19 20] Thus, two major types of cysts can be identified[21]: type I cysts have high concentrations of potassium and low concentrations of sodium and chloride, high concentrations of androgen and oestrogen conjugates, and high concentrations of epidermal growth factor; type II cysts have an electrolyte composition more similar to that found in plasma (high concentrations of sodium and chloride and low concentrations of potassium) and lower concentrations of sex hormones and epidermal growth factor. It has been suggested that women with type I breast cysts have an increased risk of developing breast cancer.[10 21]

In 1983 we started a cohort study to evaluate the relation between the risk of breast cancer and the cationic composition of breast cysts in women with gross cystic disease. We previously reported the distribution of epidemiological and mammographic risk factors for breast cancer and the rates of cyst recurrence among the women of this cohort according to their type of cyst(s).[22-25] We present here the incidence of breast cancer by cyst type.

Methods

Subjects

Between 25 February 1983 and 15 February 1993, 1323 women aged 30-69 underwent aspiration of one or more breast cysts at the Cancer Prevention Center of Ravenna, Italy. Of these, we excluded 229 women who did not meet the eligibility criteria for our study—38 had had a previous diagnosis of cancer of any site including breast, eight had breast cancer diagnosed at the initial examination, 22 were not living in the local area, and 161 had provided samples of cyst fluid of less than 1 ml. For a further 292 women, the fluid of their breast cyst was not assayed for cationic content—either because, in accordance with the original study protocol, electrolyte concentrations were not measured for women enrolled before 1990 if less than 3 ml of fluid was aspirated from their cyst or because of accident. As a consequence, our present report concerns 802 women.

Classification of cysts

Sodium and potassium concentrations in cysts fluid were determined by flame photometry after 15 minutes of centrifugation and appropriate dilution. We

Unit of Clinical Epidemiology and Trials, National Institute for Cancer Research, Largo Rosanna Benzi 10, 16132 Genova, Italy
Paolo Bruzzi, *head*
Massimo Costantini, *epidemiologist*
Department of Clinical and Biological Sciences, Medical Oncology, University of Turin at St Luigi Hospital, 10043 Orbassano (Torino), Italy
Luigi Dogliotti, *professor of oncology*
Mirella Torta, *research assistant*
Cancer Prevention Center, St. Maria delle Croci Hospital, Via Missiroli 10, 48100 Ravenna, Italy
Carlo Naldoni, *deputy head*
Lauro Bucchi, *epidemiologist*
Alessandra Cicognani, *research fellow*
Gian Franco Buzzi, *director*
Department of Clinical and Biological Sciences, Internal Medicine, University of Turin at St. Luigi Hospital, 10043 Orbassano (Torino), Italy
Alberto Angeli, *professor of internal medicine*
Correspondence to: Dr Bruzzi.

BMJ 1997;314:925–8

Table 1 Baseline characteristics of 802 women with gross cystic disease by type of breast cyst aspirated (values are numbers (percentages))

Characteristic	Type of breast cyst*			P value†
	Type I (n=417)	Type II (n=325)	Mixed (n=60)	
Age at entry to study (years):				
30-39	60 (14)	46 (14)	7 (12)	0.047
40-44	132 (32)	67 (21)	16 (27)	
45-49	140 (34)	129 (40)	22 (37)	
50-68	85 (20)	83 (26)	15 (25)	
Menopausal status:				
Premenopausal or perimenopausal	383 (92)	288 (89)	56 (93)	0.246
Postmenopausal	34 (8)	37 (11)	4 (7)	
Age at menarche‡:				
<13	184 (44)	145 (45)	21 (35)	0.411
13	114 (27)	80 (25)	15 (25)	
>13	119 (29)	99 (30)	24 (40)	
No of births:				
0	56 (13)	39 (12)	6 (10)	0.005
1	153 (37)	82 (25)	16 (27)	
>1	208 (50)	204 (63)	38 (63)	
Age at first birth (years):				
Nulliparous	56 (13)	39 (12)	6 (10)	0.390
14-25	232 (56)	189 (58)	35 (58)	
26-30	109 (26)	70 (22)	14 (23)	
31-49	20 (5)	27 (8)	5 (8)	
Family history of breast cancer:				
No	393 (94)	305 (94)	57 (95)	0.933
Yes	24 (6)	20 (6)	3 (5)	
No of cysts aspirated:				
Solitary	368 (88)	314 (97)	0	<0.001§
Multiple	49 (12)	11 (3)	60 (100)	

*See text for details of cyst types.
†Calculated by χ^2 test for heterogeneity.
‡Unknown in one woman.
§Type I cyst *v* type II.

calculated the ratio of potassium to sodium concentrations and, according to the original protocol,[23] used a cut off value of 1.5 in the ratio to divide breast cysts into two groups—type I cysts with a potassium:sodium ratio of ≥ 1.5 and type II cysts with a ratio of < 1.5.

Detection of breast cancer

All the women in the cohort were included in a follow up programme for detecting breast cancer based on yearly clinical and ultrasound examination and biennial mammography. We also searched for the women's names in the files of the Romagna Cancer Registry.[26] After 31 December 1994, we questioned all the women by telephone about their medical history. Thus far, only seven women have been lost to follow up: one emigrated abroad after 122 months; five could not be traced after 24, 31, 49, 104, and 113 months of follow up respectively; and one could not be traced after the initial aspiration of her cyst.

Statistical analysis

We classified the women according to the type of their first cyst that had been assayed for cationic composition. Those presenting with multiple cysts of the same type were assigned to the corresponding group, but if they had cysts of both types they were assigned to a third group (mixed). We calculated the time of follow up for each woman from the date of aspiration of the index cyst to 31 December 1994 or to the date of diagnosis of breast cancer or death from any cause, whichever occurred first. The follow up times for the seven women who were lost to follow up and for one woman who underwent prophylactic bilateral mastectomy were censored at the time of the last follow up examination.

We calculated age standardised ratios of incidence of invasive breast cancer to compare the rates in the three groups of women in our cohort with those of the general population living in the local area (derived from the age specific rates reported by the Romagna Cancer Registry from 1986 to 1988[26]). We calculated exact 95% confidence intervals by direct exploration of the profile likelihood function in a Poisson regression model,[27] and, similarly, used a multivariate Poisson regression model to estimate relative risks and exact 95% confidence intervals when we compared the three groups directly, both including and excluding cases of ductal carcinoma in situ.

Our study was designed to detect a fourfold increase in risk of breast cancer in women with type I breast cysts relative to women with other types of cysts, assuming that the incidence in the whole cohort was twice that observed in the general population of the Romagna Cancer Registry.[23] The analysis was planned after 4400 person years of observation. By 31 December 1994, the 802 women in the cohort had contributed 5112 person years of observation, and the median follow up time was 2219 days (range 684-4358).

Results

Table 1 shows the baseline characteristics of the 802 women in the study cohort: 417 (52%) had type I breast cysts, 325 (41%) had type II, and 60 (7%) had mixed cysts. Compared with those with type II and mixed cysts, the women with type I cysts were younger ($P = 0.047$) and reported fewer births ($P = 0.005$).

By 31 December 1994, 17 cases of breast cancer had been diagnosed in the cohort—15 cases of invasive breast cancer (14 epithelial cancers and one malignant phyllodes tumour) and two cases of ductal carcinoma in situ. Fourteen of the cancers (including the two ductal carcinomas in situ) occurred in the women with type I cysts, two occurred in those with type II cysts, and one in those with mixed cysts. The median time from aspiration of the index cyst to diagnosis of cancer was 38.1 months (range 2.3-129.0). In nine cases the cancer was found in the same breast as the index cyst had been, in five cases it was found in the other breast, and in three cases bilateral cysts had been aspirated.

The incidence of invasive breast cancer in the whole cohort was higher than that expected from the rates reported by the Romagna Cancer Registry (age standardised incidence ratio 1.69 (95% confidence interval 0.97 to 2.70)) (table 2). The increased risk was concentrated in the women with type I cysts (12 cases observed *v* 4.58 expected, age standardised incidence ratio 2.62 (1.40 to 4.40)), while no excess was seen among the women with type II and mixed cysts.

Table 3 shows the results of direct comparison among the three groups of women: the risk of invasive breast cancer in the women with type I cysts was significantly higher than in the women with type II cysts both in univariate analysis (relative risk 4.62, (95% confidence interval 1.26 to 29.7)) and in multivariate analysis (4.24 (1.12 to 27.5)). The association was stronger when the two cases of ductal carcinoma in situ

were included in the analyses (relative risks 5.41 (1.51 to 34.4) and 5.06 (1.38 to 32.5) respectively). The women with multiple cysts at enrolment were not at increased risk of breast cancer compared with women with solitary cysts.

Discussion

Our study shows that among 802 women with gross cystic disease the risk of breast cancer was associated with the cationic content of fluid from their breast cysts. Despite the small number of cases of cancer, the association was significant, and various considerations suggest that it was not due to chance, bias, or confounding from known risk factors for breast cancer.

Validity of results

The primary aim of our study was to compare the incidence of breast cancer in women with type I cysts with that in women with type II cysts, and the analysis was conducted at a time specified in advance. The choice of a cut off value of 1.5 in the ratio of potassium to sodium concentrations for classifying the cysts was already indicated in the original study protocol.[23] The appropriateness of this cut off value was supported by our subsequent analysis of a subset of women from the same cohort in which we used an independent criterion—the rate of recurrence of cysts.[22] Indeed, the association between type of cyst and risk of breast cancer would have been even stronger if we had used the cut off value of 0.33 for the potassium:sodium ratio that was proposed by Dixon *et al*[13] and used in several studies since one of the women with invasive breast cancer had had an index cyst with a potassium:sodium ratio of 1.45 (see fig 1).

In our series of 17 cancers two (12%) were ductal carcinomas in situ, a higher proportion than that reported in the same age groups by the Romagna Cancer Registry (5.3%). This suggests that our comparison with the registry rates might be biased due to overdiagnosis. However, this problem does not apply to the internal comparisons among the different groups of women within the cohort since compliance to follow up was virtually complete and was independent of cyst type. Incidentally, the age standardised incidence ratio of breast cancer in our cohort was only 1.69 compared with the general population of the area, close to the relative risk reported in another Italian study,[4] which was the lowest of those reported from studies of similar cohorts.[28]

Finally, the relative risk of breast cancer remained virtually unchanged after we had adjusted for several known risk factors for breast cancer. Therefore, our observed association seems to have been a real one, which supports the view that breast cysts are not a homogeneous entity but should be separated into two major populations.

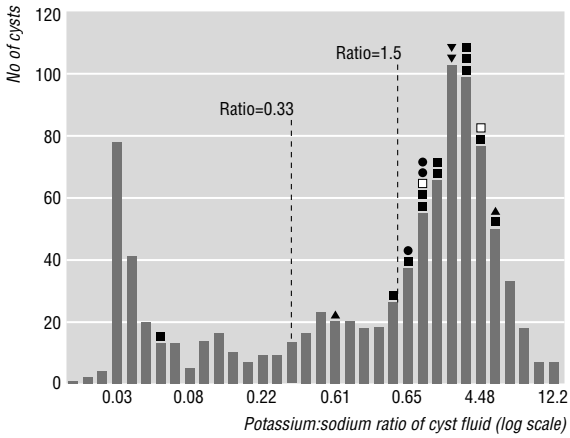

Fig 1 Distribution of potassium:sodium ratio among the 948 breast cysts aspirated from 802 women with gross cystic disease. Twenty one cysts were aspirated from the 17 women who developed breast cancer: 12 of the 15 women who developed invasive breast cancer had a single cyst aspirated (●), two women had two cysts each (▲, ▼), and one woman had three cysts (■), while both women who developed ductal carcinoma in situ had a single cyst (□). Only two of these 21 cysts had potassium:sodium ratios below cut off value of 1.5 used to separate type I and type II cysts, and only one cyst showed a ratio below cut off value of 0.33 that was proposed by Dixon *et al*[13]

Table 2 Incidence of breast cancer among 802 women with gross cystic disease by type of breast cyst aspirated and compared with incidence among general female population covered by Romagna Cancer Registry[26]

	Type of breast cyst*			
	Type I (n=417)	Type II (n=325)	Mixed (n=60)	Total (n=802)
No of cases of invasive breast cancer	12	2	1	15
Person years of observation	2658.5	2054.9	398.2	5111.6
Expected No of cases†	4.58	3.59	0.69	8.86
Age standardised incidence ratio (95% CI)	2.62 (1.40 to 4.40)	0.56 (0.09 to 1.72)	1.44 (0.08 to 6.33)	1.69 (0.97 to 2.70)

*See text for details of cyst types.
†Based on the age specific incidence rates reported by Romagna Cancer registry (1986-1988)[26]

Table 3 Relative risk of invasive breast cancer among 802 women with gross cystic disease by type of breast cyst and number of cysts aspirated at enrolment

	No of subjects	No of cases of cancer	Relative risk (95% CI)	
			Univariate analysis	Multivariate analysis†
Type of breast cyst*:				
Type II	325	2	1.00	1.00
Type I	417	12	4.62 (1.26 to 29.7)	4.24 (1.12 to 27.5)
Mixed	60	1	2.57 (0.12 to 26.9)	1.98 (0.07 to 34.4)
No of cysts aspirated:				
Solitary	682	12	1.00	1.00
Multiple	120	3	1.35 (0.31 to 4.24)	1.26 (0.19 to 4.75)

*See text for details of cyst types.
†Adjusted for age, age at menarche, No of births, and family history of breast cancer.

Comparison with other studies

Our observation of an association between type I breast cysts and risk of breast cancer is supported by other studies showing an increased risk of breast cancer among women with multiple aspirations of cysts[2 9] and among women with apocrine cysts,[11] since an increased rate of cyst recurrence has been observed among women with type I cysts[22 29] and these are more often lined by apocrine epithelium than type II cysts.[13]

Miller *et al* reported that, among 18 women who subsequently developed breast cancer, the distribution of cyst type was clearly different from that observed in a group of women who did not develop breast cancer.[21] In contrast, however, Ebbs and Bates failed to find any relation between the potassium:sodium ratio in fluid from breast cysts and breast cancer in a consecutive series of 101 women with breast cysts,[30] but methodological problems possibly invalidate their report.[29]

Key messages

- Several studies have shown that women with palpable cysts in their breasts are at increased risk of breast cancer
- Two types of breast cyst can be identified—type I cysts, with low concentrations of sodium and high concentrations of potassium ions, and type II cysts, with opposite characteristics
- We investigated the correlation between cyst type and risk of breast cancer in 802 women with aspirated breast cysts
- After median follow up of six years, the women had a relative risk of breast cancer of 1.69 compared with the general population, and those with type I cysts had a risk four times higher than those with type II cysts
- The excess risk of breast cancer of women with breast cysts seems to be concentrated among women with type I cysts, but the size and duration of this increased risk are still to be assessed

Explanation of results

There is no clear explanation for the increased risk of breast cancer among women with type I breast cysts. Breast cysts are not precursor lesions of cancer, and the occurrence of a cancer within a cyst is quite rare. In our study, in agreement with previous reports,[1] the risk of cancer was not limited to the breast from which a cyst had been aspirated. Thus, a high potassium:sodium ratio in cyst fluid seems to be a marker of an increased risk affecting the whole breast tissue.

The bimodal distribution of cations in breast cysts' fluid is mirrored by the distribution of other substances, such as steroid sex hormones[12-14 18 19] and growth factors.[15 19 20] The concentration of mitogenic growth factors such as epidermal growth factor is substantially higher in type I cysts,[15] whereas transforming growth factor β—which is claimed to have an inhibitory effect on growth of tumour cells of epithelial origin—is preferably accumulated in type II cysts.[31]

Conclusion

Our study suggests that the increased risk of breast cancer among women with gross cystic disease, previously observed in several cohort studies, is limited to a subgroup of these women. Such women can be identified on the basis of a high ratio of potassium to sodium ions in the fluid aspirated from breast cysts. The potential clinical implications of this finding are considerable in view of the simplicity and worldwide availability of these measurements. From on our data, however, it is not possible to estimate with acceptable precision the size of the increased risk among women with type I cysts, nor to assess its duration. These issues need to be further explored by longer follow up of our cohort and in larger, independent cohort studies.

Presented in part at the 31st Annual Meeting of the American Society of Clinical Oncology, Los Angeles CA, USA, 20-24 May 1995, and at the 4th Cherasco Meeting on Gross Cystic Disease of the Breast, Cherasco, Italy, 23 June 1995. We thank Licia Tavolazzi (Chemical Pathology Laboratory, Santa Maria delle Croci Hospital, Ravenna), Antonio Mazzotti (RIA Laboratory, Santa Maria delle Croci Hospital, Ravenna), and Luca Boni (Unit of Clinical Epidemiology and Trials, National Institute for Cancer Research, Genova) for their participation in the study.

Funding: This study was supported by a grant from the Istituto Oncologico Romagnolo, Forlì, Italy (contract No 90133.1); a grant from the Associazione Italiana per la Ricerca sul Cancro (AIRC), Italy (Family history of cancer, inheritance, and molecular mechanisms in benign and malignant breast diseases); and a 60% grant from MURST, Italy.

Conflict of interest: None.

1 Haagensen CD, Bodian C, Haagensen DE Jr, eds. *Breast carcinoma. Risk and detection.* London: WB Saunders, 1981.
2 Bodian CA, Lattes R, Perzin KH. The epidemiology of gross cystic disease of the breast confirmed by biopsy or by aspiration of cyst fluid. *Cancer Detect Prev* 1992;16(suppl):95-103.
3 Harrington E, Lesnick G. The association between gross cysts of the breast and breast cancer. *Breast* 1980;7:13-7.
4 Ciatto S, Biggeri A, Rosselli del Turco M, Bartoli D, Iossa A. Risk of breast cancer subsequent to proven gross cystic disease. *Eur J Cancer* 1990;26: 555-7.
5 Dupont WD, Page DL. Risk factors for breast cancer in woman with proliferative breast disease. *N Engl J Med* 1985;312:146-51.
6 Jones BM, Bradbeer JW. The presentation and progress of macroscopic breast cysts. *Br J Surg* 1980;67:669-71.
7 Roberts MM, Jones V, Elton RA, Fortt RW, Williams W, Gravelle IH. Risk of breast cancer in women with history of benign breast disease of the breast. *BMJ* 1984;288:275-8.
8 Hutchinson WB, Thomas DB, Hamlin WB, Roth GJ, Peterson AV, Williams B. Risk of breast cancer in women with benign breast disease. *J Natl Cancer Inst* 1980;65:13-20.
9 Bundred NJ, West RR, Dowd JO, Mansel RE, Hughes LE. Is there an increased risk of breast cancer in women who have had a breast cyst aspirated? *Br J Cancer* 1991;64:935-55.
10 Dixon JM, Lumsden AB, Miller WR. The relationship of cyst type to risk factors for breast cancer and the subsequent development of breast cancer in patients with breast cystic disease. *Eur J Cancer Clin Oncol* 1985; 21:1047-50.
11 Bodian CA, Perzin KH, Lattes R, Hoffman P, Abernathy TG. Prognostic significance of benign proliferative breast disease. *Cancer* 1993;71:3896-907.
12 Miller WR, Dixon JM, Scott WN, Forrest AP. Classification of human breast cysts according to electrolyte and androgen conjugate composition. *Clin Oncol* 1983;9:227-32.
13 Dixon JM, Miller WR, Scott WN, Forrest APM. The morphological basis of human breast cyst population. *Br J Surg* 1983;70:604-6.
14 Dogliotti L, Orlandi F, Torta M, Buzzi G, Naldoni C, Mazzotti A. Cations and dehydroepiandrosterone-sulfate in cyst fluid of pre and menopausal patients with gross cystic disease of the breast. Evidence for the existence of subpopulations of cysts. *Eur J Cancer Clin Oncol* 1986;22:1301-7.
15 Boccardo F, Valenti G, Zanardi S, Cerruti G, Fassio T, Bruzzi P. Epidermal growth factor in breast cyst fluid: relationship with intracystic cation and androgen conjugate content. *Cancer Res* 1988;45:5860-3.
16 Angeli A, Bradlow HL, Bodian CA, Chasalow FI, Dogliotti L, Hagensen DE Jr. Criteria for classifying breast cyst fluids. *Ann NY Acad Sci* 1990; 586:49-52.
17 Bradlow HL, Skidmore FD, Schwartz MK, Fleisher M, Schwartz D. Cations in breast cyst fluid. In: Angeli A, Bradlow HL, Dogliotti L, eds. *Endocrinology of cystic breast disease.* New York: Raven Press, 1983: 197-201.
18 Belanger A, Caron S, Labrie F, Naldoni C, Dogliotti L, Angeli A. Levels of eighteen non-conjugated and conjugated steroids in human breast cysts fluid: relationship with cyst type. *Eur J Cancer* 1990;26:277-81.
19 Lai LC, Cornell C, Lennard TWJ. Relationship between oestrogen-inducible proteins, oestradiol and electrolyte ratio in breast cyst fluid. *Cancer* 1993;24:668-75.
20 Ness JC, Sedghinasab M, Moe RE, Tapper D. Identification of multiple proliferative growth factors in breast cyst fluid. *Am J Surg* 1993;166:237-43.
21 Miller WR, Scott WN, Phil M, Harris WH, Wang D. Using biological measurements, can patients with benign breast disease who are at high risk for breast cancer be identified? *Cancer Detect Prev* 1992;16(suppl):13-20.
22 Naldoni C, Costantini M, Dogliotti L, Bruzzi P, Bucchi L, Buzzi G. Association of cyst type with risk factors for breast cancer and relapse rate in women with gross cystic disease of the breast. *Cancer Res* 1992;52:1791-5.
23 Naldoni C, Bruzzi P, Bucchi L, Gordini S, Torta M, Caraci P. Cohort study of women affected by Gross Cystic Disease: correlations between cationic and hormonal composition of breast cyst fluid and the risk of breast cancer. *Ann NY Acad Sci* 1990;586:272-5.
24 Angeli A, Dogliotti L, Naldoni C, Orlandi F, Puligheddu B, Caraci P. Steroid biochemistry and categorization of breast cyst fluid: relation to breast cancer risk. *J Steroid Biochem Mol Biol* 1994;49:333-9.
25 Bucchi L, Costantini M, Buzzi G, Bravetti P, Cicognani A, Torta M. Wolfe's mammographic patterns in women with gross cystic disease of the breast. *J Clin Epidemiol* 1995;48:969-76.
26 Amadori D, Falcini F, Buzzi GF, Gambi A, Vignutelli P, Callea A. *Registro tumori della Romagna (1986-1988).* Forlì: Istituto Oncologico Romagnolo, 1992.
27 Clayton D, Hills M. *Statistical models in epidemiology.* Oxford: Oxford Science Publications, 1993.
28 Bodian CA. Benign breast diseases, carcinoma in situ, and breast cancer risk. *Epidemiol Rev* 1993;15:177-87.
29 Bundred NJ, Walker RA, Jenkins A, Shaw D, Hall AV, Winder AF. Biochemical analysis of breast cyst fluid predicts the natural history of breast cysts. *Breast Dis* 1992;5:21-7.
30 Ebbs SR, Bates T. Breast cyst type does not predict the natural history of cystic disease or breast cancer risk. *Br J Surg* 1988;75:702-4.
31 Lai LC, Siraj AK, Erbas H, Lennard TWJ. Relationship between basic fibroblast growth factor and transforming growth factor beta 2 in breast cyst fluid. *J Clin Endocrinol Metab* 1995;80(2):711-5.

(Accepted 16 January 1997)

Retirement on grounds of ill health: cross sectional survey in six organisations in United Kingdom

C J M Poole

Abstract

Objective: To assess the process and outcome of retirement due to ill health in six large organisations.
Design: Cross sectional study of the rate of retirement due to ill health by age, sex, and length of service. Principal diagnoses by age and length of service were also compared.
Setting: Four public and two private large employers in the United Kingdom.
Main outcome measures: Rates of retirement on the grounds of ill health by age, sex, and length of service of employees contributing to pension schemes.
Results: Rates of ill health retirement varied from 20 to 250 per 10 000 contributing members, and in two organisations the rate varied geographically within the same organisation. In the two organisations that provided data by sex, women retired at a greater rate than men under age 40 and over age 50. In four organisations the modal age or length of service coincided with enhancements in benefits. In the four that provided information on diagnoses, musculoskeletal and minor psychiatric illnesses were the most common reasons for retirement.
Conclusion: The granting of ill health retirement benefits may not be determined by illness. There is a need for some employers and pension schemes to improve their processes for granting benefits. Doctors should be wary of conflicts of interest and work to guidelines when they advise pension schemes about the merits of an application for benefits.

Dudley Priority Health NHS Trust, Central Clinic, Dudley, West Midlands DY2 7BX
C J M Poole, *consultant occupational physician*

BMJ 1997;314:929–32

Introduction

Applications to retire from occupational pension schemes before the normal retirement age have increased over the past 15 years.[1] Of the 45% of people who retire early, about a third will do so because of ill health.[1] The criteria for awarding a benefit and the process and size of the benefit vary between pension schemes. The criterion may be as stringent as "permanent incapacity which is likely to prevent any gainful employment" or as loose as "incapacity which prevents the employee undertaking regular and effective duties." The process may entail a report from the applicant's general practitioner or the company's medical officer only or from a doctor who is less likely to have a conflict of interest.

Benefits are usually related to the length of service of the applicant and take the form of an increase in the number of pensionable years of service once a minimum number of years has been spent in the organisation (less generous) or an increase in pensionable service as if the applicant had worked to normal retirement age (more generous). During periods of company rationalisation employees may be given financial incentives to leave by taking premature retirement or voluntary redundancy or even by being compulsorily made redundant, which will affect the number of applications for ill health retirement.

Doctors are often asked to advise on whether an applicant fulfils the criteria for ill health retirement benefits, but little is published to assist with these judgments.[2,3] There is anecdotal evidence from employees and employers that decisions about ill health retirement may not always be fair, and there is evidence of poor correlation between doctors when they assess case scenarios for retirement benefits.[4]

I undertook a cross sectional study of six large organisations in the United Kingdom to determine the effects of age, sex, length of service, and diagnosis on retirement due to ill health.

Methods

Data were requested on numbers, age, sex, length of service, and principal diagnosis for employees retiring with ill health from Rover (a car manufacturer), the fire, police, and ambulance services, the Post Office, and the Teachers Pensions Agency for the period 1990-5. Details on age and length of service by sex of members contributing to the pension schemes were also requested. The criteria and number of doctors involved in the process were compared. It was agreed that because of the sensitivity of the data the results would be anonymised so that individual organisations could not be separately identified. Letters were used to represent each organisation.

Data analysis was by frequency distribution, χ^2 and Mann-Whitney tests, and logistic regression as appropriate.

Results

Rates of retirement due to ill health varied more than 10-fold between organisations (table 1). In two organisations, as a proportion of all retirements (1994-5) it varied geographically within the same organisation —from 37% (7/19) to 100% (11/11) in B and from 13% (8/60) to 69% (124/179) in C. Table 1 also shows modal and median ages and lengths of service at retirement. Rates of retirement generally increased with age. In four of the organisations—B, C, D, and E—the mode for age or length of service coincided with enhancements in benefits (table 1). In B, C, and D the modes coincided with maximum enhancement in benefits. By contrast, maximum benefit was paid to only three employees in organisation F (1994-5).

In organisation C the median (interquartile range) age for retirement was 47 (9.0) years for men and 31 (6.0) years for women ($P<0.001$). The respective median lengths of service were 25 (10.0) years and 12 (6.0) years ($P<0.001$; fig 1). Median ages of men and women in organisation C were 37 (12.0) and 30 (8.0) years, respectively. When the data were analysed in five year age bands by sex, the proportion of women retiring on grounds of health was greater than the

Table 1 Retirement on grounds of ill health in six organisations (A, B, C, D, E, and F) for 1994-5

Detail	A	B	C	D	E	F
Normal retirement age (years)	65	55	55	60	60	65
Ill health retirement per 10 000 contributing members	20 (63/31 107)	250 (861/34 516)	177 (2259/127 448)	111 (6075/550 000†)	110 (202/18 423)	180 (56/31 077)
Mode age (years)	62	46	48	54*	56	56
Median (interquartile range) age (years)	56 (10)	46 (9)	46 (12)	51 (8)	48 (16)	52 (11)
Mode length of service (years)	14	27*	27*	32	5*	18
Median (interquartile) length of service (years)	20 (14)	22 (11)	23 (12)	28 (10)	13 (13)	19 (12)

*Coincides with enhancement in benefits.
†Estimated number of contributing members.

proportion of men between ages 21 and 40 years (table 2). Table 3 shows the same data for organisation D, with women retiring at a greater rate than men at ages 26-30 and 51-60.

The criterion for retirement on the grounds of ill health varied between organisations. In A, B, and F ill health had to be permanent and prevent applicants from doing their job; in C ill health had to be permanent and prevent the applicants from doing any job within the organisation (not just the one they were doing at the time); in D ill health needed to last only a year or two; and in E ill health needed only to make applicants incapable of regularly and effectively undertaking the duties of their grade.

In B, C, and E retirement on the grounds of ill health could be granted with supporting evidence from just one doctor, either the applicant's general practitioner or the organisation's medical officer. This compared with two doctors in D and F (usually the general practitioner and a benefits agency doctor) and three doctors in A (the general practitioner, a company occupational physician or medical officer, and an independent occupational physician).

In the four organisations that provided the principal medical diagnoses (A, B, E, and F) musculoskeletal problems affecting the back or joints and minor mental ill health such as stress, anxiety, and depression were the most common reasons for granting ill health retirement benefits. When data for B and E (both predominantly manual workers) were analysed in five year age bands the ratio of diagnoses of musculoskeletal conditions and mental ill health to those of cardiovascular and respiratory conditions by age or length of service ranged from 4:1 to 10:1 (tables 4 and 5, respectively), whereas in A (also predominantly manual workers) the ratio was 1 or less (table 6).

Discussion

The greater rate of ill health retirement before age 40 and after age 50 in women compared with men has not, to my knowledge, previously been reported. The relative risk was almost fivefold in organisation C between ages 26 and 35 years. I cannot identify the reason for this sex difference from this study, but it needs to be examined if expensive training and

Table 2 Rates of retirement due to ill health in active members of pension scheme by sex and age in organisation C, 1994-5

	Men			Women			Relative risk (95% CI) of women retiring compared with men
Age group*	No of members	No who retired	Rate ($\times 10^{-3}$)	No of members	No who retired	Rate ($\times 10^{-3}$)	
21-25	6 258	13	2.1	3566	27	7.6	3.65 (1.89 to 7.06)
26-30	18 636	76	4.1	6221	118	19.0	4.65 (3.49 to 6.20)
31-35	24 481	144	5.9	4434	122	27.5	4.67 (3.68 to 5.94)
36-40	22 392	215	9.6	2230	38	17.0	1.78 (1.26 to 2.50)
41-45	17 017	368	21.6	778	15	19.3	0.89 (0.53 to 1.50)
46-50	15 461	720	46.6	372	15	40.3	0.86 (0.51 to 1.45)
51-55	4 607	361	78.4	74	14	189.2	2.42 (1.49 to 3.91)

*Data for each age are available from author.

Table 3 Rates of retirement due to ill health in active members of pension scheme by sex and age in organisation D, 1994-5

	Men			Women			Relative risk (95% CI) of women retiring compared with men
Age group*	No of members	No who retired	Rate ($\times 10^{-3}$)	No of members	No who retired	Rate ($\times 10^{-3}$)	
21-25	3 856	0	—	16 038	1	—	1.00 (0.99 to 1.00)
26-30	12 663	4	0.3	31 931	31	1.0	3.01 (1.06 to 8.51)
31-35	20 485	30	1.5	32 601	42	1.3	0.88 (0.55 to 1.41)
36-40	36 002	108	3.0	56 841	155	2.7	0.91 (0.71 to 1.16)
41-45	52 448	317	6.0	75 775	399	5.3	0.87 (0.75 to 1.01)
46-50	41 946	788	18.8	58 410	883	15.1	0.81 (0.73 to 0.89)
51-55	23 419	819	35.0	34 495	1320	38.3	1.09 (1.01 to 1.19)
56-60	11 909	422	35.4	16 205	765	47.2	1.34 (1.19 to 1.50)

*Data for each age are available from author.

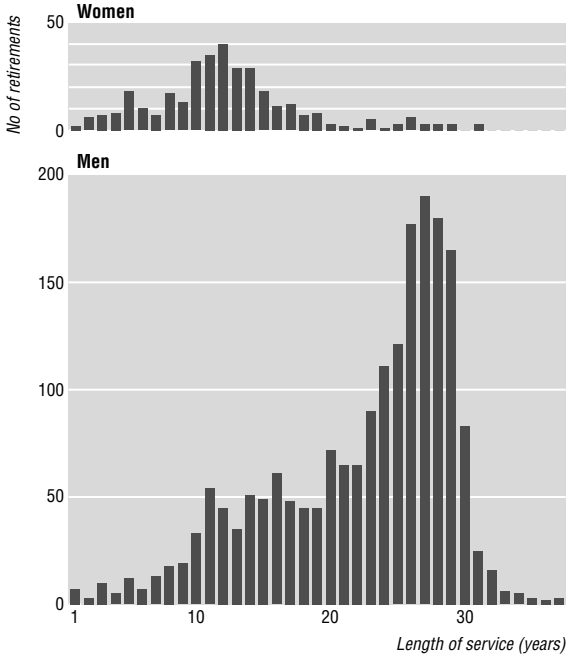

Fig 1 Numbers of employees retiring on grounds of ill health by length of service and sex during 1994-5 in organisation C. Enhancements in benefits are payable after 10 and 27 years of service

Table 4 Numbers of employees in organisation B retiring on grounds of ill health by main diagnostic group and length of service, 1994-5

	Length of service (years)																				
	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Diagnoses																					
Musculoskeletal and psychiatric	4	4	5	12	6	7	3	5	12	9	6	9	5	7	10	14	11	15	5	1	2
Cardiac and pulmonary	1	0	2	0	0	1	2	0	3	3	0	1	3	0	2	4	4	2	1	1	0

Table 5 Numbers of employees in organisation E retiring on grounds of ill health by main diagnostic group and length of service, 1994-5

	Length of service (years)																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Diagnoses																				
Musculoskeletal and psychiatric	1	1	3	6	15	11	8	5	5	4	9	4	4	4	7	6	6	1	5	7
Cardiac and pulmonary	0	0	0	0	1	4	2	1	0	2	1	0	0	2	1	1	0	1	0	1

pension costs are not to be wasted. Women have been recruited increasingly into this organisation over the past 20 years and now form 14% of the workforce. The lower median age of the women in C is unlikely to be the reason for their higher rate of retirement as even those with short service and under age 30 were retiring at four times the rate of men. D was the only other organisation that provided sex specific data. Women retired at a greater rate than men between ages 26 and 30, which coincides with a doubling of reckonable service—for example, five to 10 years—for pension benefits. Women also retired at a greater rate over age 50, but men retired at a greater rate between ages 46 and 50, when ischaemic heart disease is particularly prevalent in men.[5]

The concurrence of modes of ill health retirement by age or length of service with enhancements in benefits in four of the six organisations (B, C, D, and E) probably reflects an understandable desire by retiring employees to secure the optimum pension possible.

Peaks of ill health

There is no medical reason why ill health should peak at these times, though it is possible that employees may carry their ailments for some time before presenting them to a doctor for the purposes of securing ill health retirement. The small peaks at ages 45, 50, and 55 that were seen in A may have occurred for similar reasons at these psychological milestones. A gradual increase in the rate of ill health retirement similar to that of organisation A would be expected if medical reasons were the main factors that determined applications for benefits. The carrying of ill health is probably easier with musculoskeletal problems (such as joint or back pain) and psychiatric problems (such as stress, anxiety, or depression) than with cardiovascular or respiratory illnesses. The comparatively high proportion of musculoskeletal and psychiatric diagnoses around the modes in some of the organisations suggests that this is the case. By contrast, it is unfair that so few employees in organisation F achieved maximum pension benefits.

Influence of non-medical factors

Support for the notion of non-medical factors influencing applications for benefits comes from studies of patients who have undergone coronary artery bypass grafting. The proportion of abnormal electrocardiograms was found to be no different in those who received benefits from those who returned to work,[6] and social, economic, or psychological rather than medical reasons have been reported as the principal reasons for not returning to work after coronary artery bypass grafting.[7]

Table 6 Numbers of employees in organisation A retiring on grounds of ill health by main diagnostic group and age, 1994-5

	Age (years)														
	63	62	61	60	59	58	57	56	55	54	53	52	51	50	49
Diagnoses															
Musculoskeletal and psychiatric	3	8	5	9	4	2	1	1	1	0	2	1	0	1	1
Cardiac and pulmonary	3	4	6	5	2	1	5	1	4	3	1	0	0	2	3

Variations between organisations in the proportions of employees leaving by retiring on the grounds of ill health may simply be a reflection of the different ways in which employment contracts are terminated. A company with a low rate of retirement due to ill health may use other methods of dismissing staff, such as redundancy, premature retirement, or frustration of contract, as is the case with organisation A. By comparison, organisation C rarely makes use of these other methods, preferring the ill health retirement route. This has the effect of "medicalising" dissatisfaction, generating consultations with doctors and expensive investigations for what are comparatively minor illnesses.

Conflicts of interest

Wide variations in the proportions of employees leaving by retiring on the grounds of ill health within the

Key messages

- The rate of retirement due to ill health varies greatly between organisations and may even vary within the same organisation
- Applicants for ill health retirement may be motivated more by financial benefits than by ill health
- Women may retire at a greater rate than men before age 40 and after age 50
- Some pension funds need to improve their processes for granting ill health retirement benefits
- Doctors should beware of conflicts of interest and work to guidelines when advising pension funds about the merits of an application for benefits

same organisation (B and C) suggest that the process is out of control and in need of audit. Possible factors are pressure from management or unions for employees who leave to opt for ill health retirement and inconsistent judgments by doctors. The latter could be dealt with by doctors working to guidelines[3] in which they had had some training. The decision to award ill health retirement is best not left to one doctor (whether the applicant's general practitioner or the organisation's occupational physician), who could be placed in a position of conflict of interest, but to a second doctor who is external to the organisation, trained in occupational health, and working to guidelines.

I thank the many people from the six organisations studied who supplied me with data; Ian Calvert at the Institute of Occupational Health, Birmingham, for statistical advice; and Paul Sims, medical illustration department, Dudley Group of Hospitals NHS Trust.

Funding: No external funding.

Conflict of interest: None.

1 Office of Population Censuses and Surveys. *General household survey (1990-1991).* London: HMSO, 1992.
2 Cox RAF, Edwards FC, McCallum RI, eds. *Fitness for work.* Oxford: Oxford Medical Publications, 1995.
3 Poole CJM, Baron CE, Gunnyeon WJ, O'Hanlon M, Raoof A, Robson S, *et al.* Ill health retirement—guidelines for occupational physicians. *Occup Med* 1996;46:402-6.
4 Elder AG, Symington IS, Symington EH. Do occupational physicians agree about ill health retiral? A study of simulated retirement assessments. *Occup Med* 1994;44:231-5.
5 Institute of Public Health. *Average age-specific death rates for selected causes, common data set, 1995.* London: Department of Health, National Institute of Epidemiology, University of Surrey, 1995.
6 Lundbom J, Myhre HO, Ystgaard B, Aakhus S, Tromsdal A, Sudbo R, *et al.* Exercise tolerance and work ability following aorto-coronary bypass surgery. *Scand J Soc Med* 1994;22:303-8.
7 Monpere C, Francois G, Rondeau du Noyer C, Phan Van J. Return to work after rehabilitation in coronary bypass patients. Role of the occupational medicine specialist during rehabilitation. *Eur Heart J* 1988;9 (suppl L):48-53.

(Accepted 24 January 1997)

A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression

Louis Appleby, Rachel Warner, Anna Whitton, Brian Faragher

School of Psychiatry and Behavioural Sciences, University of Manchester, Withington Hospital, Manchester M20 8LR

Louis Appleby, *professor*

Rachel Warner, *clinical research fellow*

Anna Whitton, *research psychologist*

Research Support Unit, Medical School, Manchester M13 9PT

Brian Faragher, *head of medical statistics*

Correspondence to: Professor Appleby.

BMJ 1997;314:932–6

Abstract

Objective: To study the effectiveness of fluoxetine and cognitive-behavioural counselling in depressive illness in postnatal women: to compare fluoxetine and placebo, six sessions and one session of counselling, and combinations of drugs and counselling.

Design: Randomised, controlled treatment trial, double blind in relation to drug treatment, with four treatment cells: fluoxetine or placebo plus one or six sessions of counselling.

Subjects: 87 women satisfying criteria for depressive illness 6-8 weeks after childbirth, 61 (70%) of whom completed 12 weeks of treatment.

Setting: Community based study in south Manchester.

Main outcome measures: Psychiatric morbidity after 1, 4, and 12 weeks, measured as mean scores and 95% confidence limits on the revised clinical interview schedule, the Edinburgh postnatal depression scale and the Hamilton depression scale.

Results: Highly significant improvement was seen in all four treatment groups. The improvement in subjects receiving fluoxetine was significantly greater than in those receiving placebo. The improvement after six sessions of counselling was significantly greater than after a single session. Interaction between counselling and fluoxetine was not statistically significant. These differences were evident after one week, and improvement in all groups was complete after four weeks.

Conclusions: Both fluoxetine and cognitive-behavioural counselling given as a course of therapy are effective treatments for non-psychotic depression in postnatal women. After an initial session of counselling, additional benefit results from either fluoxetine or further counselling but there seems to be no advantage in receiving both. The choice of treatment may therefore be made by the women themselves.

Introduction

Non-psychotic depressive illness, affecting 8-15% of women in the first few months after childbirth,[1-5] can lead to chronic or recurrent mood disorder in the mother,[2] and disturbances in behaviour and cognitive development in the infant.[6 7] Despite its prevalence and potential impact, and the increasing practice of identifying cases through screening by health visitors, its treatment has rarely been studied in a controlled trial.

Only one intervention study has been published in which cases were identified by screening a community based population of newly delivered mothers.[8] This showed a significant improvement in mood in women receiving non-directive counselling from trained health visitors; the counselling was given in an average of 8.8 sessions over three months. An untreated control group showed no improvement. In another placebo controlled study, treatment with oestrogen was found to have improved mood after two months,[9] although the subjects, being medical referrals, are likely to have been more severely depressed than the majority of community cases.

Anecdotally, women who attend general practitioners with postnatal depression are often treated with a conventional antidepressant, although the anxiety present in many of these patients, the need to avoid oversedation in nursing mothers, the difficulties of prescribing to breastfeeding women, the psychosocial adversity faced by women with postnatal depression,[2 3 10 11] and the good prognosis of non-psychotic depression in the community[12] make it

unclear which, if any, antidepressants are likely to be beneficial. We are not aware of any published trial of antidepressants in this condition, nor of any study that has examined the relative effects of pharmacological and psychological intervention.

We aimed to determine the optimal treatment for non-psychotic depression in childbearing women. The drug used in the study was fluoxetine, a serotonin specific reuptake inhibitor, a class of drugs that is anxiolytic and non-sedating. The psychological treatment was a simple form of counselling based on cognitive-behavioural therapy. All trial subjects received at least one session of counselling, as described below. The hypotheses under test were that six sessions of counselling would be more effective than one; that fluoxetine would be more effective than placebo; and that after one session of counselling, fluoxetine and additional sessions of counselling would be equally effective.

Method

Subjects were women found by screening in an urban health district to be depressed 6-8 weeks after childbirth. From May 1993 to February 1995 women on the maternity wards of two large hospitals in south Manchester were asked to allow assessment of their mood in their homes 6-8 weeks later. This initial approach took place on alternate weekdays; exclusion criteria were inadequate English and living outside the district. The population screened therefore represented a largely unselected systematic sample of newly delivered mothers.

At the screening visit subjects completed the Edinburgh postnatal depression scale,[13] and those who scored ≥ 10 (a threshold at which sensitivity is 89%[14]) were interviewed with the revised clinical interview schedule.[15] Women who scored ≥ 12 on the revised clinical interview schedule, the threshold for significant psychiatric morbidity, and who satisfied research diagnostic criteria[16] for major or minor depressive disorder, were invited to take part in the treatment trial. The main exclusion criteria were chronic (>2 years) or resistant depression, current drug or alcohol misuse, severe illness requiring close monitoring or hospital admission, and breast feeding. Demographic and obstetric data for all subjects were recorded, and additional clinical information was noted on those who agreed to enter the trial.

Subjects were allocated to one of four treatment groups by using computer generated random numbers, to receive a combination of fluoxetine or placebo, plus either one session or six sessions of counselling. The counselling was derived from cognitive-behavioural therapy and was designed to be delivered by non-specialists in mental health—for example, health visitors—after brief training. Each session was structured to offer reassurance and practical advice on four areas of concern to depressed mothers: feelings of not coping, lack of enjoyable activities, lack of practical support, and caring for any older children. In addition, the first session (in week 1 of the trial), which lasted one hour, allowed women time to describe their current circumstances and emotional state. Subsequent sessions lasted 30 minutes; previously agreed tasks, such as taking the baby to a park, or going out socially, were reviewed. These sessions took place in weeks 2, 3, 5, 7, and 11 of the trial.

There was no "no treatment" group. This was for ethical reasons, and because of previous evidence that non-directive counselling was of benefit.[8] Our study was intended to develop this finding by examining the relative impact of drug and counselling interventions, and by comparing the effect of counselling given as a single session and as a course of treatment.

The duration of treatment was three months. Follow up assessments of mood took place using the revised clinical interview schedule as the principal outcome measure after 1, 4 and 12 weeks of treatment. Subjects also completed the Edinburgh postnatal depression scale at these times, and were assessed with the Hamilton depression scale,[17] a standard assessment instrument in antidepressant trials, at entry to the trial and at 12 weeks. The assessment interviews were conducted by a psychiatrist blind to subject treatment group. The counselling was delivered by a psychologist with no previous clinical training, supervised by a second psychiatrist; both were blind to drug treatment, as were trial subjects.

Statistical analysis

The three psychiatric measures were analysed separately by analysis of variance with repeated measures over time. The non-repeated factors evaluated were type of drug treatment and number of counselling sessions; the interaction between these was also examined. All three measures followed approximate log normal distributions so were converted to natural logarithms for analysis; the results were detransformed back into their original units for presentation (geometric means with their 95% confidence limits). Two complete analyses were performed; in the first, only those subjects who completed treatment were included, while in the second ("intention to treat") all subjects who were randomised were included, using last observation carried forward for non-completers.

Results

Subjects

In the 20 month recruitment period 2978 women were eligible to take part in the screening, of whom 2395 (80%) agreed to complete the Edinburgh postnatal depression scale 6-8 weeks later. A total of 503 (21%) scored ≥ 10 on the Edinburgh postnatal depression scale, of whom 406 (81%) agreed to further assessment; 218 (54%) of these were found not to satisfy entry criteria for depression at interview, indicating a low specificity of the Edinburgh postnatal depression scale at this threshold.

We therefore identified 188 confirmed cases of depression (9.7% of initial sample, adjusted for later refusals), of whom 87 agreed to enter the treatment trial. The commonest reason for women to refuse was reluctance to take medication, most commonly because they expected to improve without treatment. The characteristics of the women who entered the trial and those who did not are shown in table 1. Those entering the trial were significantly younger ($P < 0.001$). The characteristics of the four treatment groups are shown in table 2.

Table 1 Characteristics of depressed women who agreed or refused to enter trial of treatment

Characteristic	No agreeing (n=87)	No refusing (n=101)
Single	26	21
Unemployment:		
Subject*	66	67
Partner	22	17
History of subfertility†	3	9
Primiparity	28	36
Unplanned pregnancy	55	58
Complicated pregnancy	26	21
Complicated delivery	18	34
Caesarean section	15	19
Prematurity	15	10
Baby in special care baby unit	11	8

*No job to return to after maternity leave.
†>1 Year failing to conceive before this pregnancy.

Table 3 shows the number of drop outs, the duration of treatment received, and the reasons for dropping out in each treatment group. Drop out rates were similar in the four groups. Drop outs were younger than subjects who completed the study (23.7 (SD 6.2) years *v* 26.3 (5.1) years; $t_{85} = 2.06$, $P = 0.04$) and more likely to have an unemployed partner ($\chi^2 = 3.8$, $df = 1$, $P = 0.05$) and to have had a planned pregnancy ($\chi^2 = 4.6$, $df = 1$, $P = 0.03$), but the groups did not differ on initial psychiatric morbidity scores, employment, obstetric complications, parity, family history, or personal history of depression, including postnatal depression.

Table 2 Characteristics of women in treatment groups

	Fluoxetine plus counselling		Placebo plus counselling	
Characteristics	1 Session (n=22)	6 Sessions (n=21)	1 Session (n=23)	6 Sessions (n=21)
Mean age (years)	25.7	26.6	23.1	26.0
Single	6	7	8	5
Unemployment:				
Subject*	14	14	20	18
Partner	5	4	5	8
Primiparity	9	4	11	4
Unplanned pregnancy	17	9	13	16
Complicated pregnancy	3	10	6	7
Complicated delivery	2	3	8	5
Caesarean section	2	3	6	4
Prematurity	3	5	3	4
Baby in special care baby unit	4	2	2	3
Major depressive disorder	15	11	13	12
Minor depressive disorder	7	10	10	9
History of postnatal depression	9	8	6	7
History of other depression	6	4	2	1
Family history of postnatal depression	8	2	2	4
Family history of other depression	11	4	8	5

*No job to return to after maternity leave.

Table 3 Number of dropouts, timing of dropping out, and reasons for dropping out in treatment groups

	Fluoxetine plus counselling		Placebo plus counselling		
Characteristics	1 Session	6 Sessions	1 Session	6 Sessions	Total
No entering trial	22	21	23	21	87
No of dropouts	6	8	6	6	26
Stage of dropout:					
Before 1 week assessment	3	4	3	6	16
Before 4 week assessment	3	5	3	6	17
Before 12 week assessment	6	8	6	6	26
Reason for dropping out:					
No reason given	3	5	2	4	14
Disliked drug	1	1	2	1	5
Side effects	0	1	2	1	4
Lack of improvement	2	1	0	0	3

Effects of treatment

Tables 4, 5, and 6 show mean scores and their 95% confidence intervals on the three assessment instruments. The results were first analysed for subjects completing 12 weeks of treatment. Highly significant improvements were seen in all four treatment groups. fluoxetine was superior to placebo on all outcome measures. Six sessions of counselling were superior to one session on the revised clinical interview schedule and Hamilton depression scale but not on the Edinburgh postnatal depression scale.

When the revised clinical interview schedule was taken as the main outcome measure, the response to fluoxetine was evident within one week. This was also true of the response to six sessions of counselling, even though only one session had been delivered by this stage. Improvement in all four groups was largely complete after four weeks.

Percentage differences in (geometric) mean scores on the revised clinical interview schedule scores were calculated. The difference between fluoxetine and placebo was 37.1% at 4 weeks (95% confidence interval 5.7% to 58.0%) and 40.7% at 12 weeks (10.9% to 60.6%); the difference between six sessions and one session of counselling was 53.9% at 4 weeks (2.3% to 131.2%) and 38.7% at 12 weeks (– 9.2% to 111.7%).

The women who received a single counselling session with placebo had smaller changes than all other groups in scores on the revised clinical interview schedule and Hamilton depression scale. However, because of the high variability in the responses over time within each of the study groups, the interaction between drug and counselling treatment was not significant.

An "intention to treat" analysis was also performed, in which last observations were carried forward for dropouts (tables 4-6). As table 3 shows, this meant carrying forward baseline scores in most cases. Despite this, the earlier findings were broadly confirmed, although the differences at 12 weeks between women who had received six sessions and one session of counselling were of borderline significance.

Discussion

This study shows the effectiveness of both fluoxetine and cognitive-behavioural counselling in the treatment of women found by community based screening to be depressed 6-8 weeks after childbirth. Combining fluoxetine and six sessions of counselling did not produce additional improvement.

This was not a trial of the effectiveness of counselling itself. All subjects received one session of counselling, and our trial design allowed us to study the additional benefits of an antidepressant and of further counselling sessions. This design was in part intended to reflect actual clinical practice in primary care. A general practitioner or a health visitor who finds that a woman is depressed postnatally is likely to spend some time listening and advising; this study aimed to show

Table 4 Geometric mean scores on revised clinical interview schedule scores (95% confidence intervals) for patients who completed the study [all patients randomised into the study, assessed by intention to treat analysis]

Treatment (drug plus sessions of counselling)	No of patients	Assessment time			
		Baseline	1 Week	4 Weeks	12 Weeks
Fluoxetine:					
Plus 1 session	16 [22]	28.8 (26.4 to 31.4) [29.6 (27.5 to 31.8)]	19.1 (16.0 to 26.1) [21.4 (18.4 to 24.9)]	10.3 (6.5 to 16.1) [13.3 (9.1 to 19.1)]	8.0 (4.4 to 14.1) [11.1 (6.9 to 17.6)]
Plus 6 sessions	13 [21]	26.7 (23.5 to 30.4) [26.8 (23.9 to 30.1)]	13.6 (7.5 to 24.0) [16.5 (11.4 to 23.7)]	6.3 (2.5 to 14.4) [9.9 (5.6 to 17.1)]	7.0 (3.4 to 13.5) [10.5 (6.6 to 16.6)]
Placebo:					
Plus 1 session	17 [23]	30.0 (27.7 to 32.5) [29.3 (27.0 to 31.9)]	23.5 (19.0 to 29.1) [24.0 (20.4 to 28.3)]	17.3 (13.6 to 21.8) [18.9 (15.5 to 23.0)]	17.5 (13.5 to 22.7) [19.1 (15.4 to 23.5)]
Plus 6 sessions	15 [21]	27.1 (24.0 to 30.6) [27.2 (24.7 to 29.9)]	20.5 (16.4 to 25.7) [21.1 (17.7 to 25.1)]	10.7 (7.8 to 14.5) [13.5 (10.2 to 17.7)]	9.9 (6.6 to 14.7) [13.0 (9.2 to 18.1)]
Total fluoxetine	29 [43]	27.8 (25.9 to 29.9) [28.2 (26.4 to 30.1)]	16.4 (12.6 to 21.3) [18.9 (15.6 to 22.8)]	8.3 (5.4 to 12.7) [11.5 (8.4 to 15.8)]	7.6 (5.0 to 11.3) [10.8 (7.9 to 14.8)]
Total placebo	32 [44]	28.6 (26.7 to 30.7) [28.3 (26.6 to 30.1)]	22.1 (19.0 to 25.6) [22.6 (20.1 to 25.4)]	13.8 (11.3 to 16.8) [16.1 (13.6 to 19.0)]	13.4 (10.5 to 17.1) [15.9 (13.1 to 19.3)]
Total 1 session counselling	33 [45]	29.4 (27.8 to 31.1) [29.4 (27.9 to 31.1)]	21.3 (18.5 to 24.4) [23.7 (20.4 to 25.3)]	13.5 (10.5 to 17.3) [15.9 (13.0 to 19.5)]	12.1 (8.8 to 16.5) [14.7 (11.4 to 18.9)]
Total 6 sessions counselling	28 [42]	27.0 (24.8 to 29.3) [27.0 (25.1 to 29.0)]	17.0 (12.8 to 22.5) [18.7 (15.3 to 22.7)]	8.4 (5.6 to 12.5) [11.6 (8.6 to 15.5)]	8.4 (5.9 to 12.0) [11.7 (8.9 to 15.3)]

Scores ⩾12 indicate clinically important morbidity.

Table 5 Geometric mean scores on Edinburgh postnatal depression scale (95% confidence intervals) for patients who completed the study [all patients randomised into the study, assessed by intention to treat analysis]

Treatment (drug plus sessions of counselling)	No of patients	Assessment time			
		Baseline	1 Week	4 Weeks	12 Weeks
Fluoxetine:					
Plus 1 session	16 [22]	16.4 (14.9:18.0) [16.6 (15.4:18.0)]	12.8 (10.7:15.1) [13.2 (11.3:15.4)]	8.1 (6.5:10.0) [9.5 (7.7:11.6)]	5.4 (3.5:8.0) [7.1 (5.0:10.1)]
Plus 6 sessions	13 [21]	16.9 (14.7:19.4) [17.7 (16.1:19.5)]	10.9 (7.2:16.1) [12.1 (9.4:15.7)]	5.9 (3.2:10.3) [8.0 (5.2:11.9)]	5.3 (2.5:10.1) [7.5 (4.6:11.8)]
Placebo:					
Plus 1 session	17 [23]	17.4 (15.2:19.9) [17.5 (15.8:19.3)]	14.2 (11.5:17.5) [14.4 (12.3:16.9)]	8.7 (5.6:13.2) [9.4 (6.8:13.0)]	9.8 (7.2:13.3) [10.3 (8.1:13.2)]
Plus 6 sessions	15 [21]	16.8 (14.6:19.3) [16.4 (14.8:18.1)]	14.2 (11.8:17.1) [13.1 (10.1:16.9)]	9.0 (7.0:11.4) [9.8 (7.8:12.3)]	9.9 (5.5:11.2) [9.5 (7.2:12.5)]
Total fluoxetine	29 [43]	16.6 (15.4:17.9) [17.2 (16.2:18.2)]	11.9 (9.8:14.3) [12.7 (11.0:14.6)]	7.0 (5.3:9.1) [8.7 (7.0:10.8)]	5.3 (3.7:7.5) [7.3 (5.5:9.6)]
Total placebo	32 [44]	17.1 (15.6:18.8) [16.9 (15.8:18.1)]	14.2 (12.5:16.2) [13.8 (12.0:15.9)]	8.8 (6.9:11.2) [9.6 (7.9:11.7)]	8.9 (7.1:11.0) [9.9 (8.3:11.8)]
Total 1 session counselling	33 [45]	16.9 (15.6:18.3) [17.0 (16.0:18.1)]	13.5 (11.9:15.4) [13.8 (12.4:15.4)]	8.4 (6.6:10.5) [9.5 (7.8:11.4)]	7.4 (5.7:9.5) [8.6 (7.0:10.7)]
Total 6 sessions counselling	28 [42]	16.9 (15.4:18.5) [17.0 (15.9:18.2)]	12.6 (10.3:15.3) [12.6 (10.6:15.0)]	7.4 (5.5:9.8) [8.8 (7.0:11.1)]	6.6 (4.6:9.2) [8.4 (6.5:10.9)]

Scores over 9 and over 12 can be used as screening thresholds.

whether it was clinically justified to prescribe an antidepressant as well or provide additional time for further counselling, or both.

An unexpected finding was the great improvement in mood within one week of entering the trial. Cases of "depression" found by community survey are frequently transient and may be more accurately viewed as distress,[12] but there are two reasons for believing that subjects in our trial had true depression. Firstly, the entry criteria (research diagnostic criteria) required subjects to have been depressed for at least two weeks. Secondly, many subjects with symptoms identified by screening were excluded at interview because of insufficient evidence of depressive illness. The improvement after one week was greater in those taking fluoxetine, suggesting that its antidepressant effect begins earlier than is often suggested. The improvement after one week was also greater in those who were going to receive (but had not yet received) further counselling, suggesting that perceived as well as actual support was beneficial.

Table 6 Geometric mean Hamilton scores (95% confidence intervals) for patients who completed the study [all patients randomised into the study, assessed by intention to treat analysis]

Treatment (drug plus sessions of counselling)	No of patients	Assessment time	
		Baseline	12 Weeks
Fluoxetine:			
Plus 1 session	16 [22]	13.3 (11.8 to 15.0) [14.4 (12.8 to 16.2)]	2.9 (1.6 to 4.9) [4.4 (2.4 to 7.4)]
Plus 6 sessions	13 [21]	13.2 (11.3 to 15.4) [14.0 (12.1 to 16.1)]	2.8 (1.1 to 5.8) [5.1 (2.6 to 9.2)]
Placebo:			
Plus 1 session	17 [23]	14.7 (12.7 to 17.1) [14.0 (12.1 to 16.3)]	7.5 (5.3 to 10.4) [8.1 (6.1 to 10.7)]
Plus 6 sessions	15 [21]	13.3 (11.0 to 16.0) [13.8 (11.7 to 16.2)]	3.7 (2.1 to 6.1) [4.9 (3.0 to 8.9)]
Total fluoxetine	29 [43]	13.3 (12.2 to 14.5) [14.2 (13.0 to 15.5)]	2.9 (1.8 to 4.3) [4.7 (3.1 to 6.9)]
Total placebo	32 [44]	14.0 (12.5 to 15.7) [13.9 (12.5 to 15.4)]	5.4 (3.9 to 7.3) [6.4 (4.9 to 8.4)]
Total 1 session counselling	33 [45]	14.0 (12.8 to 15.4) [14.2 (12.9 to 15.6)]	4.8 (3.4 to 6.7) [6.0 (4.4 to 8.1)]
Total 6 sessions counselling	28 [42]	13.3 (11.8 to 14.9) [13.9 (12.5 to 15.4)]	3.2 (2.1 to 4.9) [5.0 (3.4 to 7.2)]

Scores of 8-17 indicate mild depression.

Key messages

- Fluoxetine, an anxiolytic antidepressant, is an effective treatment for postnatal depression
- A course of six sessions of a simple form of counselling derived from cognitive-behavioural therapy is more effective than a single session
- The drug and counselling treatments do not interact significantly, and there seems to be no advantage in receiving both
- In primary care, the simplest treatment after a single session of cognitive-behavioural counselling may be fluoxetine as it removes the need for additional counselling
- Many women with postnatal depression are reluctant to take medication, and for them a course of cognitive-behavioural counselling is as effective as an antidepressant drug

Because our recruitment began with community screening we were able to assess how closely our subject sample represented depressed postnatal women on a number of demographic and obstetric variables. There was no evidence of any clinically important bias in the sample or in those who completed treatment, although those entering the trial were presumably more accepting of treatment. Our sample therefore represents those women who would be seen by any service that routinely screened for postnatal depression and offered treatment.

These results justify such screening by showing the effectiveness of two forms of treatment. In women who will accept an antidepressant after a single counselling session, further counselling offers no additional benefit. Many will prefer not to take a drug treatment, however,[18] and for them a course of counselling is equivalently effective.

1 Pitt B. "Atypical" depression following childbirth. *Br J Psychiatry* 1968;114:1325-35.
2 Kumar R, Robson KM. A prospective study of emotional disorders in childbearing women. *Br J Psychiatry* 1984;144:35-47.
3 Watson JP, Elliott SA, Rugg AJ, Brough DI. Psychiatric disorder in pregnancy and the first postnatal year. *Br J Psychiatry* 1984;144:453-62.
4 Cooper PJ, Campbell EA, Day A, Gath D. Non-psychotic psychiatric disorder after childbirth. *Br J Psychiatry* 1988;152:799-806.
5 Cox JL, Murray D, Chapman G. A controlled study of the onset, duration and prevalence of postnatal depression. *Br J Psychiatry* 1993;163:27-31.
6 Cogill S, Caplan H, Alexandra H, Robson K, Kumar R. Impact of postnatal depression on the developing child. *BMJ* 1986;292:1165-7.
7 Murray L. The impact of postnatal depression on infant development. *J Child Psychol Psychiatry* 1992;33:543-61.
8 Holden JM, Sagovsky R, Cox JL. Counselling in a general practice setting: a controlled study of health visitor intervention in the treatment of postnatal depression. *BMJ* 1989;298:223-6.
9 Gregoire AFP, Kumar R, Everitt B, Henderson AF, Studd JWW. Transdermal oestrogen for the treatment of severe postnatal depression. *Lancet* 1996;347:930-3.
10 Paykel ES, Emms EM, Fletcher J, Rassaby ES. Life events and social support in puerperal depression. *Br J Psychiatry*, 1980;136:339-46.
11 Warner R, Appleby L, Whitton A, Faragher B. Demographic and obstetric risk factors for postnatal psychiatric morbidity. *Br J Psychiatry* 1996;168:607-11.
12 Tennant C, Bebbington P, Hurry J. The short-term outcome of neurotic disorders in the community: the relation of remission to clinical factors and to "neutralising" life events. *Br J Psychiatry* 198;139:213-20.
13 Cox JL, Holden J, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh postnatal depression scale. *Br J Psychiatry* 1987;150:782-6.
14 Murray L, Carothers A. The validation of the Edinburgh postnatal depression scale on a community sample. *Br J Psychiatry* 1990;157:288-90.
15 Lewis G, Pelosi AJ, Araya R, Dunn G. Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. *Psychol Med* 1992;22:465-86.
16 Spitzer RL, Endicott J, Robins E. Research diagnostic criteria: rationale and reliability. *Arch Gen Psychiatry* 1978;36:773-82.
17 Hamilton M. Hamilton depression scale. In: Guy W, ed. *ECDEU assessment manual for psychopharmacology*. Rev ed. Rockville, MD: National Institute of Mental Health, 1976:179-92.
18 Whitton A, Warner R, Appleby L. The pathway to care in postnatal depression: Women's attitudes to postnatal depression and its treatment. *Br J Gen Pract* 1996;46:427-8.

(Accepted 16 January 1997)

Cross sectional study of contribution of clinical assessment and simple cardiac investigations to diagnosis of left ventricular systolic dysfunction in patients admitted with acute dyspnoea

Neil D Gillespie, Graeme McNeill, Terence Pringle, Simon Ogston, Allan D Struthers, Stuart D Pringle

Correspondence to: Dr N D Gillespie, Department of Medicine (Section of Ageing and Health), Ninewells Hospital and Medical School, Dundee DD1 9SY.

BMJ 1997;314:936–40

continued over

Abstract

Objective: To assess the comparative contribution of clinical assessment, electrocardiography, and chest radiography to the diagnosis of left ventricular systolic dysfunction in patients admitted to a general medical ward with acute dyspnoea.

Design: Prospective cross sectional study.

Setting: Acute medical admissions ward of a teaching hospital.

Subjects: 71 randomly selected patients admitted with acute dyspnoea.

Main outcome measures: Sensitivity and specificity of each investigation and logistic regression analysis of each variable in identifying left ventricular systolic dysfunction.

Results: Clinical assessment in this cohort of patients with severe dyspnoea was generally sensitive (sensitivity 81%). Patients were divided into three groups on the basis of clinical assessment. In the first group (37 patients) the diagnosis of systolic dysfunction was clear, in the second (22) it was in doubt, and in the third (12) it was unlikely. The sensitivity of clinical assessment in identifying left ventricular systolic dysfunction was 81% and the specificity was 47%. The specificity of diagnosis was improved by electrocardiography (69%) and chest radiography (92%). Logistic regression analysis

showed that isolated pulmonary crepitations were a comparatively poor predictor of left ventricular systolic dysfunction ($\chi^2 = 10.215$, $P = 0.0014$) but that a full clinical examination had reasonable predictive value ($\chi^2 = 24.82$, $P < 0.00001$). The combination of clinical assessment and chest radiography improved the accuracy of diagnosis ($\chi^2 = 28.08$, $P < 0.00001$), as did the combination of clinical assessment and electrocardiography ($\chi^2 = 32.41$, $P < 0.00001$).
Conclusion: Clinical assessment in patients admitted with acute dyspnoea is comparatively accurate. Patients with abnormal results on chest radiography, electrocardiography, and clinical examination have a high likelihood of having left ventricular systolic dysfunction. Echocardiography contributes little more to the diagnosis in these patients and may be more efficiently directed towards patients in whom the diagnosis is still in doubt after clinical assessment, chest radiography, and electrocardiography.

Departments of Cardiology, Clinical Pharmacology, and Epidemiology, Ninewells Hospital, Dundee DD1 9SY
Neil D Gillespie, *research fellow, department of cardiology*
Graeme McNeill, *consultant cardiologist*
Terence Pringle, *consultant cardiologist*
Simon Ogston, *statistician, department of epidemiology*
Allan D Struthers, *professor of clinical pharmacology*
Stuart D Pringle, *consultant cardiologist*

Introduction

Heart failure is increasingly common in an aging population.[1] Emergency admissions are increasing,[2] and patients with heart failure represent about 5% of all hospital admissions.[3] Recent studies have emphasised the difficulty of diagnosing left ventricular systolic dysfunction without echocardiography in outpatients with suspected mild heart failure.[4 5] However, whether echocardiography is as important in patients with severe disease—that is, those admitted as emergencies with acute dyspnoea—is unclear. An accurate diagnosis of left ventricular systolic dysfunction is important to start effective treatments such as angiotensin converting enzyme inhibitors.[6]

Most patients with suspected heart failure are investigated after clinical examination by electrocardiography, chest radiography, and echocardiography (if available). Although nearly all patients have electrocardiography and chest radiography, less than half of those admitted to hospital with suspected severe heart failure have access to echocardiography.[7 8] A critical evaluation of the information supplied by clinical assessment, electrocardiography, and chest radiography may help streamline the selection of patients for echocardiography and use limited resources more efficiently. We assessed the relative contribution of these three investigations to the diagnosis of left ventricular systolic dysfunction in patients admitted to a general medical unit with acute dyspnoea.

Patients and methods

Seventy one randomly selected patients admitted with dyspnoea to the acute medical receiving ward at Ninewells Hospital in Dundee were recruited to the study. Patients were included if they presented with acute dyspnoea or had dyspnoea as a major component of their overall symptoms. Patients with obvious isolated pneumothorax, pneumonia, pulmonary emboli, cor pulmonale, or renal failure were excluded from the study.

Clinical assessment and preliminary investigations

On admission each patient underwent assessment by the receiving medical registrar with chest radiography, electrocardiography, and routine blood testing. Each patient was reviewed by a consultant physician on the following morning and a presumptive diagnosis was given, pending the results of further investigations. The clinical diagnosis of left ventricular systolic dysfunction was based on the clinical decision of both the receiving medical registrar and the consultant physician—that is, it was based on standard routine clinical assessment by the admitting medical team. This diagnosis was given independently of the research fellow (NDG) performing the echocardiography. The following details were recorded from the case notes for subsequent analysis: the presence of pulmonary crepitations, pulse, blood pressure, presence of murmur, jugular venous pulse, presence of oedema (ankle or lung), drug treatment, and findings on electrocardiography.

Echocardiography

Each patient had echocardiography performed on the morning after admission by an experienced echocardiographer whose assessments had been previously validated by comparison with radionuclide ventriculography. The echocardiographer was unaware of the full clinical diagnosis while performing echocardiography. Assessments were performed on a Challenge Sim 7000 device, and each patient was studied on the ward on the morning after their admission. Three measurements were taken while patients were lying in the left lateral decubitus position: M mode assessment of left ventricular dimensions at the tip of the mitral valve leaflets,[9] calculation of fractional shortening, and assessment of regional wall motion index.[10] In addition, Doppler studies of the mitral and aortic valves were performed.

Left ventricular systolic dysfunction was diagnosed before discharge from the information obtained at clinical assessment, echocardiography, and electrocardiography and from radiological findings.

The contribution of each investigation to the overall final diagnosis of left ventricular systolic dysfunction was assessed. For the purposes of analysis the patients were categorised into three groups on the basis of the original clinical assessment. Those in the first group

Table 1 Comparison of results of clinical assessment with those of cardiac investigations. Values are numbers of patients

	Clinical diagnosis		
	Heart failure (n=37; group 1)	Possible heart failure (n=22; group 2)	Non-cardiac cause of dyspnoea (n=12; group 3)
Chest radiography			
Normal results	2	17	9
Lung oedema	27	1	0
Cardiomegaly	6	4	0
Other*	2	0	3
Electrocardiography			
Normal results†	0	10	10
Abnormal results	37	12	2
Echocardiography			
Normal results	0	8	10
Systolic dysfunction	28	11	1
Other‡	9	3	1

*Includes isolated consolidation, reticulonodular shadowing, and old rib fractures.
†Includes bradycardia, prolonged PR interval, intraventricular conduction defects, right axis deviation, tachycardia, broadening of the QRS complex, non-specific ST-T wave changes.
‡Includes valvar disease and isolated left ventricular hypertrophy.

Table 2 Sensitivity, specificity, and predictive accuracy of combinations of clinical examination, electrocardiography, and chest radiography in detecting left ventricular systolic dysfunction. Values are percentages (95% confidence intervals)

Investigation	Sensitivity	Specificity	Predictive accuracy	
			Positive	Negative
Clinical examination	81 (62 to 88)	47 (31 to 69)	0.81 (0.67 to 0.90)	0.47 (0.24 to 0.71)
Electrocardiography	98 (88 to 99)	69 (48 to 85)	0.85 (0.71 to 0.93)	0.95 (0.74 to 0.99)
Chest radiography	71 (55 to 83)	92 (74 to 99)	0.94 (0.80 to 0.99)	0.70 (0.47 to 0.79)
Clinical examination and chest radiography	92 (81 to 98)	91 (71 to 99)	0.96 (0.86 to 0.99)	0.83 (0.63 to 0.95)
Clinical examination and electrocardiography	98 (89 to 99.9)	76 (53 to 92)	0.91 (0.79 to 0.97)	0.94 (0.71 to 0.99)
Clinical examination, chest radiography, and electrocardiography	100 (93 to 100)	95 (73 to 99.9)	0.98 (0.90 to 0.99)	1 (0.80 to 1)

had a clear diagnosis of heart failure on the basis of clinical history, raised jugular venous pressure, crepitations, and the presence of a third heart sound. In the second group the findings of the clinical assessment were less clear, and in the third group the probability of heart failure was considered to be low and the main cause of the dyspnoea was non-cardiac.

For the purposes of interpretation electrocardiograms were either normal or abnormal, with minor abnormalities being included in the normal group. Findings on chest radiography were categorised as being within normal limits, as isolated cardiomegaly, as pulmonary oedema or cardiomegaly, or both, or as other findings. Most findings on chest radiography were easily interpreted by the receiving medical team, but cases of doubt were assessed by a radiologist.

Left ventricular systolic dysfunction was diagnosed echocardiographically on the basis of a fractional shortening of less than 20% or the presence of one or more regional wall motion abnormalities, a dilated left ventricle on M mode, or a subjectively determined reduction in left ventricular systolic function assessed by an experienced echocardiographer.

The contribution of each investigation to the final diagnosis of left ventricular systolic dysfunction was assessed by performing logistic regression analysis on the clinical assessment and on the findings of electrocardiography and chest radiography. The sensitivity, specificity, positive predictive accuracy, and negative predictive accuracy were calculated for each investigation in determining left ventricular systolic dysfunction.

Results

The mean age of the patients was 73 (range 33-95). Thirty patients were men. Twenty three patients had a history of myocardial infarction and 18 a history of hypertension. Most patients admitted to this medical receiving ward had predominantly cardiac causes for their dyspnoea as patients with lung disease are admitted to a separate chest unit.

Table 3 Logistic regression analysis of variables in determining echocardiographic diagnosis of left ventricular systolic dysfunction

	χ^2 Value	df	P value
Lung crepitations	10.215	1	0.0014
Crepitations/third heart sound/raised jugular venous pressure/ clinical failure	24.824	1	<0.00001
Clinical failure and electrocardiographic abnormalities	32.419	2	<0.00001
Clinical failure and radiographic abnormalities	28.08	2	<0.00001
Clinical failure and radiographic and electrocardiographic abnormalities	37.890	3	<0.00001

A total of 45 patients had impaired left ventricular systolic function. Three patients had isolated left ventricular diastolic dysfunction.

Table 1 shows the three groups of patients according to the findings at initial clinical assessment. Twenty two of the 23 patients with a previous myocardial infarction were in groups 1 or 2. Twelve of the patients had left ventricular hypertrophy at echocardiography and 11 of these 12 patients were in groups 1 or 2. Those with severe left ventricular dysfunction identified readily at clinical examination were the largest proportion (37/71 (52%)), but despite the comparative severity of symptoms of the whole group, there were many patients in which the diagnosis was not clear after initial clinical assessment. Thirty three of the 37 patients (89%) in group 1 had radiological evidence of heart failure, while 17 of the 22 patients (77%) in group 2 had normal radiological findings.

Table 2 shows the sensitivities, specificities, and predictive accuracy of combinations of the investigations after initial clinical assessment in diagnosing left ventricular systolic dysfunction.

Logistic regression analysis showed the relative contributions of both electrocardiography and chest radiography to the diagnosis of left ventricular systolic dysfunction (table 3). In the model for clinical failure and electrocardiography the odds ratio for clinical failure was 4.39 and 24.02 for an abnormal electrocardiogram. In the model for clinical failure and chest radiography the odds ratio was 1.72 for clinical failure and 4.34 for chest radiography.

Discussion

There has been a steady increase in patients admitted to acute medical wards over the past 10 years.[2] Many patients have left ventricular systolic dysfunction, and treatment with angiotensin converting enzyme inhibitors is beneficial for them.[6] Such treatment is not without its hazards, however, and a clear diagnosis of left ventricular systolic dysfunction is necessary before treatment begins. Ideally, all patients should have echocardiography performed before treatment begins to confirm the diagnosis and to exclude significant valvar disease, but up to 50% of patients undergoing assessment for suspected left ventricular dysfunction are not investigated by echocardiography.[7 8]

There was a good correlation between clinical assessment and objective assessments of left ventricular function in this cohort of patients. This is likely to be because these patients had disease severe enough to necessitate hospital admission. It may therefore be possible to bypass echocardiography in patients with

little clinical doubt about the diagnosis so long as they do not have an associated heart murmur. We still believe that all patients with a heart murmur should have echocardiography. Otherwise, our proposed approach could possibly result in cases of aortic valve disease being missed, although a previous study in our department showed no cases of unsuspected aortic valve disease in an audit of 400 requests for echocardiography.[11] However, our findings have to be put in perspective as our unit is in a large teaching hospital and the accuracy of clinical assessment may vary from centre to centre and in primary care.

Most patients with obvious clinical heart failure had abnormal chest radiographs. Consequently, little additional information was provided by echocardiography in patients who did not have a heart murmur.

Key messages

- The availability of echocardiography is limited for patients admitted to hospital with acute dyspnoea
- The presence of isolated lung crepitations is a poor predictor of left ventricular systolic dysfunction
- Full clinical assesssment is sensitive in detecting left ventricular systolic dysfunction, with specificity being added by either chest radiography or electrocardiography
- Echocardiography should be reserved for cases with the most diagnostic doubt

Heart failure

In acute medical admissions, heart failure is usually the result of systolic dysfunction due to coronary artery disease or hypertension. Although, echocardiography would be able to identify patients who have additional diastolic dysfunction, there are still no established treatments for isolated diastolic dysfunction.[12] As a result, the limited resources for echocardiography may be better directed to those with an uncertain clinical diagnosis and in whom the radiological findings are inconclusive.

In patients with possible heart failure chest radiography provided diagnostic information in only five of them (22%), while echocardiography confirmed systolic dysfunction in 11 (50%), with a further three having other abnormalities identified. In these patients a clear diagnosis is essential as appropriate treatment of systolic dysfunction and hypertension may prevent progression to overt heart failure, with possible reductions in subsequent hospital admissions.[2]

In the third group—patients with non-cardiac causes for their dyspnoea—echocardiography may be more appropriate than chest radiography in those with abnormal electrocardiograms as many of these patients often undergo both investigations with little additional information provided by chest radiography.

Electrocardiography gave abnormal results in most patients with left ventricular dysfunction. This agrees with the findings of other studies[13] and suggests that the diagnosis of left ventricular systolic dysfunction should be reconsidered in patients with a normal electrocardiogram.[14]

On the basis of these findings, we propose that the diagnosis and treatment of left ventricular systolic function in this setting could be based on the following approaches.

(1) If the diagnosis is clinically obvious act on the clinical diagnosis and obtain an echocardiogram only if there is a heart murmur.

(2) When the diagnosis is not so obvious add chest radiography and electrocardiography, with echocardiography later.

(3) When the diagnosis is not clear use echocardiography as the initial investigation.

Echocardiography

The mode of delivery of echocardiography in this study—that is, performed at the bedside with a portable device—may permit a more rapid assessment of patients to enable a quicker diagnosis and a shorter duration of admission. The increased demand for and availability of echocardiography has clear financial implications, but the cost of smaller simpler devices continues to fall. The British Society of Echocardiography has now issued training guidelines which will help non-cardiologists gain expertise in simple echocardiography.[15] The more widespread use of these devices could result in more patients with heart failure being assessed for treatment with angiotensin converting enzyme inhibitors. Currently, only 20-30% of patients with heart failure receive such treatment.[16]

Conclusions

In conclusion, this study illustrates the contribution of clinical assessment and simple non-invasive cardiac investigations to the diagnosis of left ventricular systolic dysfunction in a cohort of patients admitted with acute dyspnoea. Ideally, all patients admitted to hospital with acute dyspnoea should have access to echocardiography, but current resources prevent this. We suggest that echocardiography is not essential in patients with acute dyspnoea and clear clinical evidence of cardiac failure but no heart murmur. Clinical assessment is sensitive and specificity can be added by either chest radiography or electrocardiography. Echocardiography should be reserved for those cases in which there is the most diagnostic doubt. The observations made in this study may help streamline the selection of patients for echocardiography services, enabling more efficient use of currently available resources.

We thank the medical staff of ward 15, particularly Dr R Duffy, in Ninewells Hospital, Dundee, and Charge Nurse Connolly and the other nursing staff.

Funding: Merck Sharpe and Dohme provided an educational grant to fund the study.

Conflict of interest: None.

1 Dargie HJ, McMurray JJV. Diagnosis and management of heart failure. *BMJ* 1994;308:321-8.
2 Hobbs R. Rising admissions. *BMJ* 1995;310:207.
3 Parameshwar J, Poole-Wilson PA, Sutton GC. Heart failure in a district general hospital. *J R Coll Physicians Lond* 1992;26:139-42.
4 Remes J, Miettinen H, Reunanen A, Pyorala K. Validity of clinical diagnosis of heart failure in primary health care. *Eur Heart J* 1991;12:315-21.
5 Wheeldon NM, MacDonald TM, Flucker CJ, McKendrick AD, McDevitt DG, Stuthers AD. Echocardiography in chronic heart failure in the community. *Q J Med* 1993;86:17-23.
6 SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med* 1991;325:293-302.
7 Zarifis J, Beevers DG, Lip GYH. Hospital patients need open access echocardiography. *BMJ* 1995;311:326.

8 Slovick DI. Open access echocardiograhy should be reserved for equivocal cases. *BMJ* 1995;311:327.
9 Feigenbaum H. Echocardiographic examination of the left ventricle. *Circulation* 1975;51:118-25.
10 Heger JJ, Weyman AE, Wann LS, Dillon JC, Fiegenbaum H. Cross sectional echocardiography in acute myocardial infarction: detection and localisation of regional left ventricular asynergy. *Circulation* 1979;60:531-8.
11 Gillespie ND, Struthers AD, Pringle SD. Changing echocardiography request patterns between 1988 and 1993. *Health Bulletin* 1996;54:395-401.
12 Struthers AD. Health care delivery for chronic congestive heart failure in the United Kingdom. *Current Opinion in Cardiology* 1994;9(suppl 1):S12-5.
13 Davie AP, Francis CM, Love MP, Caruana L, Starkey IR, Shaw TRD, *et al.* Value of the electrocardiogram in identifying heart failure due to left ventricular systolic dysfunction. *BMJ* 1996;312:222.
14 Task Force on Heart Failure of the European Society of Cardiology. Guidelines for the diagnosis of heart failure. *Eur Heart J* 1995;16:741-51.
15 Education and Training Subcommittee of the British Society of Echocardiography. Training in echocardiography. *Br Heart J* 1994;71(suppl 1):2-5.
16 Clark KW, Gray D, Hampton JR. Evidence of inadequate investigation and treatment of patients with heart failure. *Br Heart J* 1994;71:584-7.

(Accepted 16 January 1997)

Examination of attendance patterns before and after introduction of South Africa's policy of free health care for children aged under 6 years and pregnant women

David Wilkinson, Marlene E Sach, Salim S Abdool Karim

Centre for Epidemiological Research in Southern Africa, Medical Research Council, PO Box 658, Hlabisa 3937, South Africa
David Wilkinson, *specialist scientist*
Salim S Abdool Karim, *director*
Hlabisa Hospital, P/Bag X5001, Hlabisa
Marlene E Sach, *nursing sister*

Correspondence to: Dr Wilkinson.

BMJ 1997;314:940–1

President Nelson Mandela's first major policy announcement after election was that all health care for children under 6 years and for pregnant and lactating women would be free in the South African government's health service. Financing mechanisms that balance equity and efficiency in health care are under vigorous debate.[1] To contribute to this debate we evaluated the impact of this new, free care policy on rural mobile clinic services in a health district with a relatively well developed public health service.

Subjects, methods, and results

A mobile clinic team in the Hlabisa health district of KwaZulu/Natal was chosen because the communities served are representative of the whole district and record keeping is excellent. The team visits most clinic points monthly and sees 100-150 patients each day. Most patients are pregnant women receiving antenatal care, children for immunisation and growth monitoring ("under 6 clinics"), and those requiring treatment for illnesses. Around 80% of patients seeking treatment are children. Under 6 clinics have always been free and are included here as a control group. The new policy on free care was implemented in July 1994; until then pregnant women paid 5 rand (£1; $1.60) per antenatal visit and treatment services cost 3 rand.

Quarterly data for clinic attendance were extracted from registers. To determine the impact of free care we compared the mean number of new clients registering for each service (antenatal care, under 6 clinic, treatment services) and the total number of attendances during the 30 month period before the new policy on free care (January 1992 to June 1994) with the corresponding numbers during the 18 month period after free care (July 1994 to December 1995). We also compared proportions of children referred to hospital during these two periods and compared 100 consecutive women registering for antenatal care in early 1994 with the same number in early 1995 for age, gravidity, and gestation at registration.

We found no important changes in attendance patterns before the implementation of the new policy on free care. The number of both new registrations and total visits to the under 6 clinics did not change significantly after the new policy was implemented (table 1). Although the average number of women registering for antenatal care each quarter was similar before and after care became free, the total number of visits each quarter increased. At the same time mean gestation at registering decreased from 28 weeks in 1994 to 26 weeks in 1995 ($P = 0.018$), while age and gravidity did not change. Most significantly, the number of new patients registering for treatment services and the total number of such visits increased substantially after the new policy on free care was implemented. The proportion of children referred to hospital decreased.

Comment

Free care substantially increased the use of treatment services by children but not the use of preventive services (under 6 clinics), which have always been free. The number of pregnant women registering for antenatal care did not increase, probably because most already received it.[2] A reduction in the gestational age at

Table 1 Changes in attendance patterns after implementation of new policy on free health care. Values are mean (SD) numbers of visits per quarter unless stated otherwise

	Antenatal service		Under 6 clinics		Treatment services		
	New	All	New	All	New	All	Referrals (%)
Before new policy	268.9 (34.2)	801.9 (121.7)	570.0 (55.2)	4460.8 (306.9)	440.1 (49.7)	1048.8 (126.8)	197.4 (18)
After new policy	279.3 (30.8)	920.8 (77.0)	556.7 (78.1)	4337.0 (432.5)	637.0 (60.5)	1859.0 (249.2)	189.3 (10.2)
% Change	3.9	14.8*	–2.3	–2.8	44.7*	77.3*	4.3

*P<0.05.

registration is to be welcomed. The new policy on free care was widely popular with communities and achieved some equity by improving access to the treatment services. The reduced referral rate to hospital suggests that the increased workload was due to presentation of either earlier or milder forms of illness. If the decrease is due to earlier presentation of serious illness it is to be welcomed; if it is due to presentation of self limiting illness—previously discouraged by a fee—then the policy is potentially counterproductive. Staff in clinics and hospitals were overworked and stressed by the increased workload—a workload widely perceived by them as largely unnecessary. The demand for treatment services is probably almost limitless and may, if not controlled, gradually steer important healthcare resources away from preventive services and health promotion.

Funding: South African Medical Research Council.
Conflict of interest: None.

1 Gilson L. Health care reform in developing countries. *Lancet* 1993;342:800.
2 Wilkinson D, Cutts F, Ntuli N, Abdool Karim SS. Maternal and child health indicators in a rural South African health district. *S Afr Med J* (in press).

(Accepted 29 November 1996)

Commentary: Should mother and child health services in developing countries be free?

Anthony Costello

Institute of Child Health, London WC1N 1EH
Anthony Costello, *reader in international child health*

Charging users for primary health care services in developing countries has been an intensely controversial issue over the past decade. The World Bank and donor governments have promoted "community financing" through user charges as part of health sector reform.[1] The arguments seem compelling. Potentially, user charges give a perceived value to services and deter frivolous demand; provide incentives for staff; remove the hidden, unofficial charges levied by unscrupulous health workers or parallel markets; and ultimately increase use by improving service quality. In the poorest countries, where government spending on health may be below £2 per head, there would seem to be no alternative.

But many health professionals working in the field doubt these claims. Demand for services by the poor and the vulnerable sections of the population—such as mothers and children—is not "price inelastic" (that is, the demand is sensitive to small increases in price). Appreciable declines in the use of services for antenatal care, maternity care, child health, and sexually transmitted disease have been reported after the introduction of user charges.[2] Declines in demand are greatest among the most socioeconomically deprived sectors.[3] Concerns have been expressed about the extent to which revenue from user charges can finance quality improvements alone (studies show that 2-32% of operating costs can be recovered[4]), the burden of administrative costs, and the potential for perverse incentives—for example, encouraging overprescribing, over-investigation, or corruption. Exemption schemes or differential charges to ensure that the poor sections of the population are not denied access to basic health services are difficult to implement.[5]

Interestingly, South Africa has bucked the international trend. With democracy under Nelson Mandela, the policy change has been in a different direction—that is, from user charges to free services. Wilkinson and colleagues have documented the impact of a new policy on free care for mothers and children in a rural district and shown a substantial increase in the use of treatment services by children and of antenatal care for mothers. Users have generally welcomed the policy change. More detailed study is needed to evaluate the impact on health outcomes such as maternal and child mortality or serious morbidity rates. These indicators might have improved if the increased attendance reflects earlier presentation of potentially serious illness. The picture is not, however, straightforward. Staff have been under stress from the greater workload and feel that many attendances for self limiting diseases are frivolous. The authors conclude pessimistically that "the demand for treatment services is probably almost limitless." This need not necessarily be so if primary maternal and child health services only are considered. In Britain universal free care for mothers and children was introduced in 1948 when health and wealth indicators were not dissimilar from those in modern South Africa. In Britain the demand at the primary level was contained as the health of the nation improved and family size declined. Few would argue for the abolition of this free care policy today.

Wilkinson *et al* have certainly raised important questions for South African managers and professionals. The immediate options might be to improve triage to educate users about the proper use of primary care and to develop a debate within communities about realistic expectations from services.

But the current economic situation in Britain (gross national product per head £12 227) and indeed South Africa (£2026) is a far cry from the 59 low income countries (mean gross national product per head $<$£500) such as Rwanda (£53), Mozambique (£60), and Ethiopia (£67). For these nations user charge schemes, however inequitable, will remain the only option to ensure additional investment in primary health care, unless recurrent costs are subsidised in the medium term by international aid or concessionary loans.

1 World Bank. *Investing in health. World development report*. Oxford: Oxford University Press, 1993.
2 Costello AM de L, Watson F, Woodward D. *Human face or human facade? Adjustment and the health of mothers and children*. Parts 1 and 2. London: Institute of Child Health, 1994. (Institute of Child Health working paper.)
3 McPake B. User charges for health services in developing countries: a review of the economic literature. *Soc Sci Med* 1993;36:1397-405.
4 Carr-Hill RA. Efficiency and equity implications of the health care reforms. *Soc Sci Med* 1994;39:1189-201.
5 Abel-Smith B. Financing health services in developing countries: the options. *NU News on Health Care in Developing Countries* 1993;7:8-15.