Open Access: free full text available online
For the full versions of these articles see bmj.com

RESEARCH

Effectiveness of continuous glucose monitoring in pregnant women with diabetes: randomised clinical trial

Helen R Murphy,[1] Gerry Rayman,[1] Karen Lewis,[1] Susan Kelly,[2] Balroop Johal,[3] Katherine Duffield,[4] Duncan Fowler,[1] Peter J Campbell,[5] Rosemary C Temple[4]

EDITORIAL by Festin

[1]Department of Diabetes and Endocrinology, Ipswich Hospital NHS Trust, Ipswich IP4 5PD
[2]Departments of Diabetes and Obstetrics, Ipswich Hospital NHS Trust
[3]Department of Obstetrics, Ipswich Hospital NHS Trust
[4]Elsie Bertram Diabetes Centre, Norfolk and Norwich University Hospital NHS Trust, Norwich
[5]Wellcome Trust Sanger Institute, Cambridge
Correspondence to: H R Murphy Helen.Murphy@ipswichhospital.nhs.uk

Cite this as: *BMJ* 2008;337:a1680
doi:10.1136/bmj.a1680

ABSTRACT

Objective To evaluate the effectiveness of continuous glucose monitoring during pregnancy on maternal glycaemic control, infant birth weight, and risk of macrosomia in women with type 1 and type 2 diabetes.

Design Prospective, open label randomised controlled trial.

Setting Two secondary care multidisciplinary obstetric clinics for diabetes in the United Kingdom.

Participants 71 women with type 1 diabetes (n=46) or type 2 diabetes (n=25) allocated to antenatal care plus continuous glucose monitoring (n=38) or to standard antenatal care (n=33).

Intervention Continuous glucose monitoring was used as an educational tool to inform shared decision making and future therapeutic changes at intervals of 4-6 weeks during pregnancy. All other aspects of antenatal care were equal between the groups.

Main outcome measures The primary outcome was maternal glycaemic control during the second and third trimesters from measurements of HbA_{1c} levels every four weeks. Secondary outcomes were birth weight and risk of macrosomia using birthweight standard deviation scores and customised birthweight centiles. Statistical analyses were done on an intention to treat basis.

Results Women randomised to continuous glucose monitoring had lower mean HbA_{1c} levels from 32 to 36 weeks' gestation compared with women randomised to standard antenatal care: 5.8% (SD 0.6) *v* 6.4% (SD 0.7). Compared with infants of mothers in the control arm those of mothers in the intervention arm had decreased mean birthweight standard deviation scores (0.9 *v* 1.6; effect size 0.7 SD, 95% confidence interval 0.0 to 1.3), decreased median customised birthweight centiles (69% *v* 93%), and a reduced risk of macrosomia (odds ratio 0.36, 95% confidence interval 0.13 to 0.98).

Conclusion Continuous glucose monitoring during pregnancy is associated with improved glycaemic control in the third trimester, lower birth weight, and reduced risk of macrosomia.

Trial registration Current Controlled Trials ISRCTN84461581.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a1680

INTRODUCTION

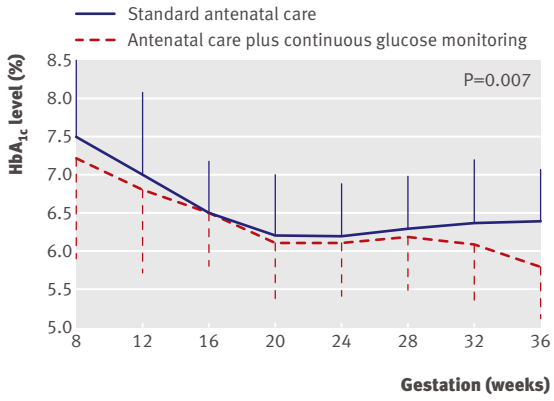
Macrosomia (infant birth weight ≥90th centile for sex and gestational age) remains the commonest complication of pregnancy in women with diabetes, leading to increased risk of perineal lacerations, complications in labour, and delivery by caesarean section.[1] For the infants the risks of immediate complications are increased, as well as the longer term risks of insulin resistance, obesity, and type 2 diabetes.[2 3]

Data from the Netherlands suggest that near optimal glycaemic control during early pregnancy failed to reduce the risk of macrosomia, present in 48.8% infants.[4] This prompted us to consider new strategies, focusing on reducing postprandial hyperglycaemic spikes during the second and third trimesters.

Novel methods of continuous glucose monitoring provide up to 288 measurements a day.[5 6] Detailed data on the magnitude and duration of glucose fluctuations give insights into daily glycaemic control, especially valuable during the physiological changes of pregnancy.[6] Continuous monitoring also provides patients with visual feedback on the consequences on glycaemia of factors such as diet, which can be harnessed as a powerful educational tool. We evaluated the effectiveness of antenatal continuous glucose monitoring on maternal glycaemic control and birth weight and risk of macrosomia in the offspring of mothers with diabetes.

METHODS

From September 2003-6 we carried out a prospective, open label randomised controlled trial at two multidisciplinary diabetic antenatal clinics in the United Kingdom. We compared intermittent self monitoring of blood glucose with continuous glucose monitoring in pregnant women aged 16-45 with type 1 and type 2 diabetes. Overall, 71 of 93 (76%) consecutively approached women, representative of women attending the clinics (see bmj.com), agreed to participate. Forty six (65%) women had type 1 diabetes and 25 (35%) had type 2 diabetes. We offered continuous glucose monitoring for up to seven days at intervals of 4-6 weeks between 8 and 32 weeks' gestation.[6] The sensor was inserted at the hospital and data were downloaded after one week. The women were asked to identify patterns in hypoglycaemia and hyperglycaemia. They were advised to consider these and to



Mean HbA_{1c} levels every four weeks in women receiving standard antenatal care (n=33) or antenatal care plus continuous glucose monitoring (n=38). Vertical lines are standard deviation at each time point

suggest possible solutions such as changes to insulin dose, diet, and activity. The suggested changes were then discussed with the diabetes team.

We advised all women to measure blood glucose levels at least seven times a day, aiming for the targets of 3.5-5.5 mmol/l before meals, <7.8 mmol/l one hour after meals, and <6.7 mmol/l two hours after meals. The women were seen every 2-4 weeks for up to 28 weeks, fortnightly until 32 weeks, and weekly thereafter, with assessments of fetal growth at 28, 32, and 36 weeks. Women with type 2 diabetes were treated with insulin before pregnancy or when pregnancy was confirmed.

HbA_{1c} levels were measured once every four weeks. We calculated the birthweight standard deviation scores for only liveborn singletons.[7] Customised birthweight centiles[8] were calculated using maternal height, weight, ethnicity, and parity; neonatal sex; and gestational age, with twin standards used for twin pregnancies.[9]

Statistical analysis

We compared the baseline characteristics of the women using Fisher exact tests for discrete variables and *t* tests for continuous variables. For glycaemic control, we allocated the HbA_{1c} measurements to windows of four weeks, and compared the two arms using *t* tests. We also used a *t* test to compare the standard deviation scores for birth weight. As a result of the heavily skewed distribution of birthweight centiles, we compared these between the two arms using the non-parametric Wilcoxon rank sum test, and we used Fisher exact tests to compare the presence or absence of macrosomia. We used binomial tests for testing whether the proportion of small for gestational age infants was different from 10%. All analyses were done on an intention to treat basis, with statistical significance set at 5%.

We excluded three infants in each group from the analyses of mean birth weight, birthweight centile, and birthweight standard deviation score owing to miscarriage, neonatal death, and major malformation (see bmj.com). For birthweight standard deviation scores we included only healthy singletons (n=62), as no comparative standard exists for twins. We did the analyses for birthweight centile with twins (n=67), using the appropriate centile reference range for twins,[9] and without twins (n=62).

RESULTS

Overall, 38 women were randomised to continuous glucose monitoring and 33 to standard antenatal care. The mean age of participants was 31.3 (SD 6.1) years, mean body mass index 28.1 (SD 7.4), and mean duration of diabetes 12.8 (SD 0.3) years (see bmj.com). At booking the mean HbA_{1c} level was 7.3% (SD 1.2%) and mean gestational age 9.2 (SD 2.7) weeks. Except for duration of diabetes, the groups were similar.

The continuous glucose monitor was generally well tolerated. The mean number of periods for continuous monitoring in 36 women (two pregnancies ended) was 4.2 (range 0-8), with 29 of the women (80%) wearing the monitor at least once per trimester. Two women withdrew.

Glycaemic control

Although the HbA_{1c} level was consistently lower in the intervention arm no statistical difference was found in mean levels between the two groups at booking or throughout the first two trimesters (figure). Between 28 and 32 weeks' gestation the mean HbA_{1c} level in the intervention arm was 6.1% (SD 0.6) compared with 6.4% (SD 0.8) in the control arm, with a trend towards but not reaching statistical significance (P=0.1). At 32-36 weeks' gestation a further reduction in HbA_{1c} levels occurred in the intervention arm but not the control arm.

Pregnancy outcomes

From 71 pregnancies there were 69 living infants (table). One woman in each arm miscarried in the first trimester and one woman had a termination. Two early neonatal deaths occurred (one in each arm); a twin pregnancy with the death of one twin at 34 weeks' gestation (anencephaly) and another at 28 weeks' gestation (Edward's syndrome). Two sets of twins plus the twin of the anencephalic infant resulted in five healthy twins (offspring of women in intervention arm). Two live infants had malformations, one cardiovascular (control arm) and one chromosomal (intervention arm).

Birth weight

Compared with healthy singletons of women in the control group (n=30), those of women in the intervention group (n=32) had decreased mean birthweight standard deviation scores (0.9 *v* 1.6; effect size 0.7 SD, 95% confidence interval 0.0 to 1.3; P=0.05). The median birthweight centiles for 67 of 69 healthy live infants were 69% in the intervention group and 93% in the control group (P=0.02; table), with twins included.

The difference remained significant after exclusion of the twins (P=0.04).

In total, 31 of 67 (46%) infants were macrosomic, of whom 23 were born to mothers with type 1 diabetes (74%) and eight to mothers with type 2 diabetes (26%). Thirteen of 37 (35%) infants in the intervention arm had macrosomia compared with 18 of 30 (60%) in the control arm (odds ratio 0.36, 95% confidence interval 0.13 to 0.98; P=0.05). Fewer extremely large for gestational age infants were born to mothers in the intervention arm than the control arm: 5/37 (13.5%) *v* 9/30 (30%). This difference did not reach statistical significance (P=0.13). The two largest infants born to mothers in the intervention arm (see bmj.com) were born to mothers who withdrew from the intervention (included in intention to treat analysis).

One infant of a mother in the intervention arm had a birthweight standard deviation score below −2 (see bmj.com). She was delivered at gestational age 34 weeks to a primiparous mother with severe pre-eclampsia. On the basis of the birthweight centile definition of small for gestational age, four of 37 (11%) infants of women in the intervention arm were small for gestational age compared with 0 of 30 infants of women in the control arm (P=0.1). Formal testing showed that the number of small for gestational age infants in the intervention arm was no different from that expected (11% *v* 10%; P=1.0), whereas in the control arm fewer babies than expected were below the 10th centile (0%; P=0.07).

DISCUSSION

Use of supplementary continuous glucose monitoring as an educational tool during pregnancy is associated with improved maternal glycaemic control. The infants of mothers in the intervention arm had a lower birth weight and a reduced risk of macrosomia than those of mothers in the control arm (standard antenatal care).

The observed differences in maternal HbA_{1c} levels, reflecting mean blood glucose levels over the preceding 4-6 weeks, emerged at 28 weeks' gestation but did not reach statistical significance until after 32 weeks. We consistently found improvements in glycaemic profiles that were achieved gradually rather than after one consultation (data not shown), emphasising the importance of using the continuous glucose monitoring profile as a tool to facilitate patient education and shared problem solving. That HbA_{1c} levels do not improve until the third trimester may reflect the delay between improved glycaemic control and clearance of glycated haemoglobin from the circulation.

Several possible explanations may contribute to the observed improvements in maternal and neonatal outcomes. Generally, women measure their blood glucose levels before and after meals, with only the most motivated achieving the suggested 10 daily tests.[10] The continuous glucose monitoring profiles yielded maximal information about the frequency, magnitude, and duration of glucose excursions and may have contributed to increased motivation and compliance with self management, in particular glucose testing after meals among women in the intervention group. The visual impact of continuous monitoring focused input towards reducing hyperglycaemic spikes, which are less apparent from a diary or glucose meter download. The data also provided feedback on the adequacy of prandial and basal insulin levels, day to day variability in diet, physical activity, and management of hypoglycaemia, with positive feedback to reenforce optimal self management.

The relative contribution of hyperglycaemia compared with obesity in the development of macrosomia is controversial.[11] We therefore adjusted the birthweight centiles for maternal height and weight and baseline maternal body mass indices were similar in both groups. The observed differences in birth weight are most likely attributable to the improvements obtained in glycaemic control.

Although we cannot exclude the possibility that tight glycaemic control contributed to fetal growth restriction, only one infant of a mother in the intervention arm had a birthweight standard deviation score below −2. It is of note that no infants of mothers in the control arm were born small for gestational age.

Strengths and limitations of the study

This trial combined recent advances in glucose monitoring with a pragmatic educational approach,

Pregnancy outcome in women with pregestational diabetes allocated to standard antenatal care plus continuous glucose monitoring or to standard antenatal care only (control group). Values are percentages (numbers) unless stated otherwise

Variable	Continuous glucose monitoring (n=38)	Standard antenatal care (n=33)	P value
Mean (SD) birthweight standard deviation score*	0.9 (1.0)	1.6 (1.4)	0.05
Mean (SD) birth weight (g)†	3340 (760)	3630 (500)	0.07
Median birthweight centile†	69	93	0.02
Macrosomia (≥90th centile)	35 (13)	60 (18)	0.05
Extremely large for gestational age (≥97.7th centile)	14 (5)	30 (9)	0.1
Small for gestational age (≤10th centile)	11 (4)	0	0.1

*Scores calculated only for 62 healthy living singletons (32 in intervention arm, 30 in control arm).
†Calculations were done for 67 of 69 healthy living infants (37 in intervention arm, 30 in control arm), after excluding one infant from each group as a result of congenital or chromosomal malformation. Twins are included, with centiles calculated using specific twin standards. Difference in birthweight centiles remained significant (P=0.04) when twins were excluded.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Continuous glucose monitoring is a potentially beneficial educational tool, with benefits on glycaemic control

NICE have called for research during pregnancy, to focus on pregnancy complications and infant outcomes

WHAT THIS STUDY ADDS

Continuous glucose monitoring is associated with improved maternal glycaemic control and reduced infant birth weight and risk of macrosomia

The beneficial effects on pregnancy complications potentially offer longer term health benefits for the infants of mothers with diabetes

suitable for implementation in real life clinical settings. It fills a gap identified by the National Institute for Health and Clinical Excellence for more research into continuous glucose monitoring and diabetes related morbidity, including pregnancy complications and neonatal outcomes. Furthermore, recruitment and retention rates were high and 76% of women participated. Despite the inconvenience of wearing a sensor, 80% of women used continuous glucose monitoring at least once per trimester.

Some limitations should also be considered. Firstly, the women were predominantly of white European ethnicity, which may limit applicability to women from other ethnic backgrounds. Secondly, health professionals were not blinded and therefore we cannot fully exclude the possibility of bias in clinical management. We attempted to standardise antenatal contacts between groups, with no difference in the frequency or duration of clinical appointments, dietary advice, obstetric input, or fetal surveillance, and no clinical input during attachment of the sensor. Thirdly, differences in maternal characteristics, with longer duration of diabetes in the intervention group, perhaps contributed to some of the effect on infant outcomes, although would not explain the improvements in glycaemic control. Fourthly, the sample size was small.

Putting the study in context and implications

Our study has filled an important gap in the evidence base,[12] suggesting that new technologies can be used to reduce the risk of diabetes related complications in pregnancy. Continuous glucose monitoring can provide added benefits to the pregnancy outcomes for women with diabetes and their infants. Additional benefits on intention to treat analysis were a reduction in HbA_{1c} levels in the third trimester and reduced birth weight and risk of macrosomia. The observed reductions in birth weight are particularly important in view of the widespread prevalence of macrosomia, affecting over half of all infants born to mothers with diabetes in the United Kingdom. These large for gestational age infants have an increased longer term risk of insulin resistance, obesity, and type 2 diabetes.

Although rates of macrosomia were reduced in women using continuous glucose monitoring, they remain 3.5 times higher than the general maternity population. This suggests that although continuous monitoring assists women to improve average glycaemic control in late pregnancy, it is still inadequate for achieving optimal day to day glucose control and birth weights comparable with the background population. We have shown that supplementary continuous glucose monitoring provided in a routine clinical setting leads to better glycaemic control and reduced risk of macrosomia. If confirmed by other studies these data have important implications for the antenatal management of women with diabetes as well as the immediate and longer term health of their infants.

We thank the women who participated, Medtronic UK for donating study equipment, and coinvestigators Katharine Stanley, Sue Land, Nicoletta Dozio (Norwich UK), and Damian Morris (Ipswich UK) for comments on the manuscript.

Contributors: See bmj.com.

Funding: This was an investigator initiated study funded by the Ipswich Diabetes Centre Charity Research Fund. HRM also received salary support from Diabetes UK. The study equipment (6 x CGMS Gold monitors and 300 sensors) was donated free of charge by Medtronic UK. The research was sponsored by Ipswich Hospital NHS Trust and was independent of all the study funders.

Competing interests: HRM and GR have received honorariums for speaking at research symposiums sponsored by Medtronic in 2004 and 2005.

Ethical approval: Suffolk and Norfolk local research ethics committees.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Langer O, Berkus MD, Huff RW, Samueloff A. Shoulder dystocia: should the fetus weighing greater than or equal to 4000 grams be delivered by cesarean section? *Am J Obstet Gynecol* 1991;165(4 Pt 1):831-7.
2 Murtaugh MA, Jacobs DR Jr, Moran A, Steinberger J, Sinaiko AR. Relation of birth weight to fasting insulin, insulin resistance, and body size in adolescence. *Diabetes Care* 2003;26:187-92.
3 Van Assche FA, Holemans K, Aerts L. Fetal growth and consequences for later life. *J Perinat Med* 1998;26:337-46.
4 Evers IM, de Valk HW, Mol BW, ter Braak EW, Visser GH. Macrosomia despite good glycaemic control in type I diabetic pregnancy; results of a nationwide study in the Netherlands. *Diabetologia* 2002;45:1484-9.
5 Klonoff DC. Continuous glucose monitoring: roadmap for 21st century diabetes therapy. *Diabetes Care* 2005;28:1231-9.
6 Murphy HR, Rayman G, Duffield K, Lewis KS, Kelly S, Johal B, et al. Changes in the glycemic profiles of women with type 1 and type 2 diabetes during pregnancy. *Diabetes Care* 2007;30:2785-91.
7 Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index, and head circumference fitted by maximum penalized likelihood. *Stat Med* 1998;17:407-29.
8 Gestation Network: West Midlands Perinatal Institute, NHS Perinatal Institute for Maternal and Child Health, Birmingham UK. www.gestation.net/birthweight_centiles/centile_object.htm.
9 Buckler JM, Green M. Birth weight and head circumference standards for English twins. *Arch Dis Child* 1994;71:516-21.
10 Kerssen A, de Valk HW, Visser GH. Do HbA(1)c levels and the self-monitoring of blood glucose levels adequately reflect glycaemic control during pregnancy in women with type 1 diabetes mellitus? *Diabetologia* 2006;49:25-8.
11 Persson B, Hanson U. Fetal size at birth in relation to quality of blood glucose control in pregnancies complicated by pregestational diabetes mellitus. *Br J Obstet Gynecol* 1996;103:427-33.
12 NICE guideline 63: Diabetes in pregnancy: management of diabetes and its complications from the pre-conception to the postnatal period, 2008. www.nice.org.

Accepted: 5 August 2008

Commentary: The technology of continuous glucose monitoring

Mary H Charlton

EDITORIAL by Festin

Diabetes Centre, Birmingham Heartlands Hospital, Birmingham B9 5SS

mary.charlton@heartofengland.nhs.uk

Cite this as: *BMJ* 2008;337:a1733
doi:10.1136/bmj.a1733

The continuous glucose monitor used by Murphy and colleagues can be thought of as diabetology's answer to cardiology's 24 hour electrocardiography monitor. It is used as an investigative tool, is attached to the patient (usually for 72 hours), and requires the user to return to hospital for its removal. The data are downloaded and examined for abnormal glucose levels and trends.

The sensor is inserted through the abdominal wall, so that its glucose oxidase tip lies within subcutaneous tissue, where it measures the glucose level of interstitial fluid, not of blood. The sensor is wired to a palm sized monitor, which can be clipped to clothing. The sensor cannot be disconnected, even temporarily—for example, when showering. The monitor captures and stores the average glucose measurement every five minutes, yielding 288 readings in 24 hours.[1]

The patient is taught to key in the times of food, exercise, insulin, and symptoms of hypoglycaemia. The device also requires at least four self monitored blood glucose readings to be entered each day for calibration, obtained by finger prick.

The device used by Murphy and colleagues does not give "real time" readings; both patient and professional are unaware of the measurements until the end of the 3-7 day period of recording, determined by the life of the sensor. The data are then downloaded. Software ensures adequate correlation between the sensor's interstitial fluid readings and the more reliable[2] self monitored glucose readings.

Continuous glucose monitoring "fleshes out" the information obtained from conventional self monitored glucose readings, helping inform changes to management and providing a graphic educational tool. Such monitoring is often used when patients and professionals cannot reconcile self monitored readings with those of measured HbA_{1c} levels. Its usefulness in this situation seems limited.[3] More specific indications are suspected asymptomatic nocturnal hypoglycaemia especially with "rebound" hyperglycaemia; suspected "dawn phenomenon"—when endogenous corticosteroids drive blood glucose levels up to a misleadingly normal or high level on waking; unawareness of hypoglycaemia; pregnancy, when postprandial hyperglycaemia needs tight control; and proved or suspected gastroparesis, when postprandial hyperglycaemia is unpredictable.

Continuous glucose monitors cost around £2000 (€2509; $3566) and each sensor around £50.[4] The accuracy of the monitor is limited[2] especially at lower glucose levels, and readings still require validation with those of the self monitored blood glucose. A lag exists between changes in blood glucose levels and those of interstitial fluid, and readings from interstitial fluid are slightly lower than those from blood. This is of particular concern at times of rapidly changing blood glucose levels and in hypoglycaemia. The National Institute for Health and Clinical Excellence has yet to appraise the technology.

Real time continuous glucose monitoring displays the most recent glucose reading and alerts the user when a preset glucose range is breached. The advantages of rapid response, particularly to hypoglycaemia, are evident but in practice there are problems with false alarms; the temptation to give extra insulin so often that it accumulates, causing hypoglycaemia; the temptation not to validate with self monitored blood glucose readings; and the psychological morbidity of being attached to a device for a defined period.

Competing interests: None declared.
Provenance and peer review: Commissioned; not peer reviewed.

1 NHS Purchasing and Supply Agency. *Blood glucose systems*. CEP 08008. Reading: NHS PASA, May 2008.
2 Klonoff DC. Continuous glucose monitoring: roadmap for 21st century diabetes therapy. *Diabetes Care* 2005;28:1231-9.
3 Ng JM, Patel J, Gibson H, Jennings PE. Continuous blood glucose monitoring: does it really affect diabetic control? *Pract Diab Int* 2008;25:239-40.
4 Medtronic MiniMed UK. *Insulin pump therapy and continuous glucose monitoring. Price list*. Watford, Herts: Medtronic, Jun 2006.

Does a functional activity programme improve function, quality of life, and falls for residents in long term care? Cluster randomised controlled trial

Ngaire Kerse,[1] Kathy Peri,[2] Elizabeth Robinson,[3] Tim Wilkinson,[4] Martin von Randow,[5] Liz Kiata,[1] John Parsons,[2] Nancy Latham,[6] Matthew Parsons,[2] Jane Willingale,[7] Paul Brown,[7] Bruce Arroll[1]

[1]Department of General Practice and Primary Health Care, School of Population Health, University of Auckland, Private Bag 92019, Auckland, 1001, New Zealand
[2]School of Nursing, University of Auckland
[3]Department of Epidemiology and Biostatistics, School of Population Health, University of Auckland
[4]Health Care of the Elderly, University of Otago, Christchurch, New Zealand
[5]Department of Sociology, University of Auckland
[6]Health and Disability Research Unit, Boston University School of Public Health, Boston, MA, USA
[7]Health Systems, School of Population Health, University of Auckland

Correspondence to: N Kerse
n.kerse@auckland.ac.nz

Cite this as: *BMJ* 2008;337:a1445
doi:10.1136/bmj.a1445

ABSTRACT

Objective To assess the effectiveness of an activity programme in improving function, quality of life, and falls in older people in residential care.

Design Cluster randomised controlled trial with one year follow-up.

Setting 41 low level dependency residential care homes in New Zealand.

Participants 682 people aged 65 years or over.

Interventions 330 residents were offered a goal setting and individualised activities of daily living activity programme by a gerontology nurse, reinforced by usual healthcare assistants; 352 residents received social visits.

Main outcome measures Function (late life function and disability instruments, elderly mobility scale, FICSIT-4 balance test, timed up and go test), quality of life (life satisfaction index, EuroQol), and falls (time to fall over 12 months). Secondary outcomes were depressive symptoms and hospital admissions.

Results 473 (70%) participants completed the trial. The programme had no impact overall. However, in contrast to residents with impaired cognition (no differences between intervention and control group), those with normal cognition in the intervention group may have maintained overall function (late life function and disability instrument total function, P=0.024) and lower limb function (late life function and disability instrument basic lower extremity, P=0.015). In residents with cognitive impairment, the likelihood of depression increased in the intervention group. No other outcomes differed between groups.

Conclusion A programme of functional rehabilitation had minimal impact for elderly people in residential care with normal cognition but was not beneficial for those with poor cognition.

Trial registration Australian Clinical Trials Register ACTRN12605000667617.

INTRODUCTION

Elderly people with high levels of habitual physical activity live longer and have better general health and higher levels of wellbeing than do those who are inactive.[1] However, for very frail elderly people in residential care, the potential for such activity is limited. Sustainable ways to slow functional decline in residential care have not yet been identified. We report the results of a pragmatic cluster randomised controlled trial that follows a successful efficacy trial, which suggested that an intervention based on activities that are meaningful to the individual person can be effective in improving quality of life for elderly people in residential care.[2]

METHODS

Participants and recruitment

Residents in low level dependency residential care in two cities of New Zealand were eligible for this study.[3] We invited homes to participate in random order by using computer generated random numbers.

Eligible residents were aged 65 years and over, able to engage in a conversation about a goal, remember the goal, and participate in a programme to achieve the goal. We excluded residents who were unable to communicate to complete the study measures, had anxiety as their main diagnosis, were acutely unwell, or were in a terminal state. We recruited residents between February and November 2004 and followed each resident for 12 months.

Measures

Trained independent research nurses recruited residents and collected demographic data, health information, and drug use data from the medical and nursing record and did standardised face to face interviews. The primary outcomes were function, self reported (late life function and disability instrument[4]) and observed (timed up and go,[5] elderly mobility scale[6]); quality of life; and falls over 12 months of follow-up. Secondary outcome measures were depressive symptoms,[7] fear of falling,[8] and hospital admissions over the 12 months of the trial (see bmj.com).

Interventions

Two trained gerontology nurses, one at each site, delivered the promoting independence in residential care (PIRC) intervention on the basis of a successful efficacy trial.[2]

Goal setting—The resident, assisted by the gerontology nurse, set a mutually agreed goal that had to meet two criteria: it had to be relevant and meaningful to the resident, and it had to promote progressive increases in physical activity.

Functional assessment and activity programme design—The gerontology nurse then completed a functional assessment and designed an individualised programme of physical activities based on repetitions of activities of

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a1445

Table 1 | Effect of an activity intervention on late life function and disability instrument disability and limitation scales. Values are mean (SE)

Measure and group	Baseline	Change 0-12 months	12 months	P value
Disability overall:				
Activity group	36.3 (0.6)	−1.3 (0.6)	35.4 (0.6)	0.25*
Social group	35.9 (0.5)	−1.6 (0.4)	34.5 (0.6)	
Limitation overall:				
Activity group	58.8 (1.4)	−8.3 (1.9)	49.8 (1.5)	0.27
Social group	57.8 (1.3)	−7.3 (2.0)	50.1 (1.4)	

*For interaction between time and intervention group from repeated measures analysis comparing groups over time; no interaction between cognition and intervention.

daily living, such as rising from a chair, additional walking, or repeated transfers. Exercise activities were planned to be done daily or several times a day in short doses. A prescriptive plan, the promoting independence plan, was constructed and placed on the wall in the resident's room and in the resident's file.

Staff implementation—The gerontology nurse trained the healthcare assistants in implementing the promoting independence plan on a one to one basis for each resident. The healthcare assistants then implemented the plan under the supervision of the usual nursing staff of the facility. For the more independent residents, the healthcare assistants provided minimal supervision. The more dependent residents needed up to 15 minutes of close supervision of the plan twice daily (see bmj.com for a full description of the intervention).

Control group—Residents in control group homes received usual care and were offered two social visits by a social science researcher to control for the attention received by the resident from the gerontology nurse visits.

Statistical analysis

We used generalised linear mixed models, with rest home as a random factor, city as a fixed factor, and time as a repeated measure, to investigate whether the intervention was associated with a change in function or quality of life over time and in particular whether this change differed between the social and intervention groups. We used a generalised linear mixed model with a logit link for the secondary outcome of depression. We investigated time to fall, allowing for multiple falls per resident, by using a Cox proportional hazards model with rest home included as a random effect. We included mental status, depression, age, sex, previous falls, total number of drugs (as a proxy for health status), time in the low level dependency home, and socioeconomic status in the models for the primary outcomes.

RESULTS

Forty one of 46 randomly selected homes participated; 83% of eligible residents participated (330 activity

Table 2 | Effect of an activity intervention on function, mobility and balance, and quality of life. Values are mean (SE)

Measure and group	Baseline	Change 0-6 months	6 months	Change 6-12 months	12 months	P value
Function overall (LLFDI)						
Cognition normal:						
Activity group	44.5 (1.0)	−0.8 (0.5)	43.9 (1.0)	−0.2 (0.3)	43.6 (1.0)	
Social group	44.9 (1.0)	−2.2 (0.6)	42.7 (1.0)	0.3 (0.7)	43.0 (1.0)	0.024*
Cognition impaired:						
Activity group	46.7 (1.4)	−2.1 (1.1)	44.4 (1.5)	−2.8 (1.1)	42.5 (1.7)	
Social control	47.1 (1.2)	−0.3 (1.4)	47.1 (1.3)	0.2 (1.0)	46.8 (1.4)	
Mobility and balance						
TUG (seconds):						
Activity group	35.4 (2.1)	3.0 (1.5)	38.9 (2.2)	0.3 (1.4)	38.5 (2.3)	0.67
Social group	34.6 (2.0)	2.4 (1.0)	36.9 (2.1)	1.2 (0.5)	37.0 (2.1)	
EMS (% score ≤16):		(% increased:% decreased)		(% increased:% decreased)		
Activity group	47.0 (4.01)	9:11	45.8 (4.2)	5:8	43.5 (4.4)	0.48
Social group	46.0 (3.9)	9:8	49.2 (4.0)	5:9	43.9 (4.1)	
Balance (% ≥10 seconds tandem stance):		(% increased:% decreased)		(% increased:% decreased)		
Activity group	42.9 (4.0)	14:13	42.4 (4.4)	10:13	38.9 (4.7)	0.16
Social group	38.2 (3.8)	17:11	47.0 (4.1)	7:17	34.8 (4.3)	
Quality of life						
EuroQol (score/12):						
Activity group	9.7 (0.2)	0.2 (0.1)	9.9 (0.2)	0.2 (0.2)	10.0 (0.2)	0.49
Social group	9.6 (0.1)	0.3 (0.2)	10.0 (0.2)	0.2 (0.2)	10.1 (0.2)	
LSI-Z (score/20):						
Activity group	15.3 (0.4)	−0.4 (0.2)	14.7 (0.4)	−0.3 (0.2)	14.5 (0.4)	0.35
Social group	14.9 (0.4)	−0.3 (0.3)	14.6 (0.4)	0.3 (0.3)	14.9 (0.4)	

*For interaction between time, cognition, and intervention group from repeated measures analysis.
EMS=elderly mobility scale (higher scores=better function); LLFDI=late life function and disability instruments; LSI-Z=life satisfaction index; TUG=timed up and go test.

Table 3 | Effect of activity intervention on depression and aches and pains (secondary outcome measures) in older people in residential care

Secondary outcomes and adverse events	Baseline	6 months	12 months	P value
Geriatric depression scale score >4 (% (SD))				
Cognitively impaired:				0.004*
Activity group	41.9 (7.1)	48.5 (7.7)	63.8 (8.0)	
Social group	38.4 (6.1)	35.2 (6.2)	35.9 (6.6)	
Cognitively normal:				
Activity group	37.1 (3.7)	37.5 (3.8)	37.9 (4.1)	
Social group	44.5 (3.8)	46.8 (4.0)	49.0 (4.2)	
Aches and pains (% (95% CI))				
Activity group	48.8 (40.4 to 54.9)	46.7 (39.3 to 54.9)	42.4 (34.7 to 50.4)	0.75†
Social group	49.5 (42.6 to 56.5)	51.1 (43.8 to 58.4)	48.8 (41.2 to 56.6)	

*P value for interaction between time, cognition, and intervention group from repeated measures analysis.
†P value for interaction between time and intervention group from repeated measures.

group, 352 social group). The mean age of residents was 84 (SD 7) years, and 504 (74%) residents were women. Characteristics of residents were mainly evenly balanced between the groups.

We found no evidence of interactions between depression, group status, and time for any of the primary outcomes. Therefore, all analyses include residents with and without depression. No significant interaction existed for the disability overall and limitation overall scales of the late life function and disability instrument (table 1).

A significant interaction existed between cognition and group status for the function overall scale of the late life function and disability instrument (P=0.024), and we report the results separately for this variable. For residents with normal cognition (abbreviated mental test score ≥7) the activity group deteriorated less in overall function (late life function and disability instrument, total function component score) in the first six months of follow-up (table 2). A similar significant interaction (P=0.015) in the lower limb subscale score of the late life function and disability instrument showed a maintenance in score in the activity group for those with normal cognition (intervention group score 48.8 at baseline, 48.1 at six months, and 47.7 at 12 months; control group 49.5, 45.9, and 46.5) but no differences in the cognitively impaired subgroup. For those with impaired cognition (abbreviated mental test score <7), the activity intervention had no significant effect on function and the mean scores for activity group participants tended to deteriorate faster than those of the social group. Other subscale scores, including advanced lower limb, upper limb, overall disability, and social and personal role subscales on the late life function and disability instrument; mobility and balance measures; quality of life; and time to fall, were not affected by the activity intervention (see table 1 and bmj.com).

More people with impaired cognition in the activity group had pronounced depressive symptoms after the activity intervention than did those with cognitive impairment in the social group (table 3). We did not analyse fear of falling, as at least 40% of the sample had incomplete data for this measure. No difference existed in the proportion of residents who were either admitted into higher level care or died during the 12 months of the trial between the activity group (55/330, 17%) and the social group (47/352, 13%) (P=0.61). We found no evidence of a difference in the rate of hospital admissions between the social group and the activity group (P=0.55), nor any difference in the level of adverse outcomes (P=0.75).

When we considered uptake of the intervention, analyses found that those who achieved their goal (187/330, 57%) were no more likely to have an improvement in function than those who did not (P=0.84). We were unable to accurately ascertain to what degree the promoting independence plan was adhered to. Intervention nurses reported that 145 (44%) of the intervention group residents did few or no activity sessions. We found no greater change in outcomes for those with greater compliance.

DISCUSSION

An activity programme had no impact on overall function for elderly people in residential care. The difference between the activity group and social group in residents with normal cognition was of questionable clinical significance, and no changes occurred in observed function, quality of life, or falls. Neither achievement of goals nor compliance made any difference to improvement in function. Residents with impaired cognition showed no maintenance of function and may have become more depressed.

Other trials showing greater levels of success in changing quality of life or performance based assessments of mobility have tested more vigorous and resource intensive interventions.[9-12] The training approach in this study focused on practising overall functional tasks embedded within daily activities, facilitated by existing staff. Unlike other studies, it did not use an exercise focused approach to intensively work

WHAT IS ALREADY KNOWN ON THIS TOPIC

Improving activity levels for elderly people in residential care can improve quality of life but may increase falls

Sustainable ways to improve activity in residential care have not been identified

Few trials have tested functionally based activity programmes in residential care

WHAT THIS STUDY ADDS

A goal oriented programme based on activities of daily living made no real impression on function or falls in elderly people in long term residential care

For those with impaired cognition this intervention was not helpful and may have increased depressive symptoms

on underlying impairments such as muscle weakness or balance problems. This intervention also did not consider any environmental barriers or psychological factors that might need to be tackled to facilitate changes at the level of disability or quality of life.

Conclusions

An activity programme based on usual activities of daily living, targeted to a personal goal that is meaningful to an individual resident, did not help to preserve physical function in frail elderly people with normal cognition in residential care and may have adversely affected those with poor cognition. Low compliance with activity recommendations was likely. To be successful, such interventions may need a higher intensity of activity and more effective reinforcement by care workers. Interventions should be carefully targeted to those people likely to engage and respond within the residential care context.

We thank the participants who contributed so eagerly and diligently to this research, and we are grateful to the staff of the participating residential care facilities. We also thank our research team Linda Venables, Gaye Purton, and Helen Roud and the many others who contributed.
Contributors: See bmj.com.
Funding: New Zealand Health Research Council and Accident Compensation Corporation. The funders had no involvement in the research design or conduct or in interpretation of results.
Competing interests: None declared.
Ethical approval: The Auckland Ethics Committee, now named the Multiregional Ethics Committee, approved the study in 2004.
Provenance and peer review: Not commissioned; externally peer reviewed.

1 Simons L, McCallum J, Freidlander Y, Simons J. Predictors of mortality in the prospective Dubbo study of Australian elderly. *Aust N Z J Med* 1996;26:40-8.
2 Peri K, Kerse N, Robinson E, Parsons M, Parsons J, Latham N. Does functionally based activity make a difference to health status and mobility? A randomised control trial in residential care facilities (the promoting independent living study; PILS). *Age Ageing* 2008;37:57-63.
3 Peri K, Kerse N, Kiata L, Wilkinson T, Robinson E, Parsons J, et al. Promoting independence in residential care: successful recruitment for a randomised controlled trial. *J Am Med Dir Assoc* 2008;9:251-6.
4 Sayers SP, Jette AM, Haley SM, Heeren TC, Guralnik JM, Fielding RA. Validation of the late-life function and disability instrument. *J Am Geriatr Soc* 2004;52:1554-9.
5 Podsialdo D, Richardson S. The timed "up & go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991;39:142-8.
6 Smith R. Validation and reliability of the elderly mobility scale. *Physiotherapy* 1994;80:745-7.
7 Yesavage J, Brink T, Rose T, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1983;17:37-49.
8 Hill K, Schwarz J, Kalogeropoulos A, Gibson S. Fear of falling revisited. *Arch Phys Med Rehab* 1996;77:1025-9.
9 Gill T, Baker D, Gottschalk M, Peduzzi P, Allore H, Byers A. A program to prevent functional decline in physically frail, elderly persons who live at home. *N Engl J Med* 2002;347:1068-74.
10 Becker C, Kron M, Lindemann U, Sturm E, Eichner B, Walter-Jung B, et al. Effectiveness of a multifaceted intervention on falls in nursing home residents. *J Am Geriatr Soc* 2003;51:306-13.
11 Fiatarone MA, O'Neill EF, Ryan ND, Clements KM, Solares GR, Nelson ME, et al. Exercise training and nutritional supplementation for physical activity in very elderly people. *N Engl J Med* 1994;330:1769-75.
12 Morris J, Fiatarone M, Kiely DK, Belleville-Taylor P, Murphy K, Littlehale S, et al. Nursing rehabilitation and exercise strategies in the nursing home. *J Gerontol* 1999;54:M494-500.

Accepted: 15 July 2008

Retrospective determination of whether famine existed in Niger, 2005: two stage cluster survey

Avid Reza,[1] Basia Tomczyk,[1] Victor M Aguayo,[2] Noel M Zagré,[3] Kadadé Goumbi,[4] Curtis Blanton,[1] Leisel Talley[1]

EDITORIAL by Gundersen

[1]International Emergency and Refugee Health Branch, Centers for Disease Control and Prevention, 4770 Buford Highway NE (MS F-60), Atlanta GA 30341, USA
[2]UNICEF, New Delhi, India
[3]UNICEF, Niamey, Niger
[4]Tuberculosis Programme, United Nations Development Programme, Niger
Correspondence to: A Reza afr6@cdc.gov

Cite this as: *BMJ* 2008;337:a1622
doi:10.1136/bmj.a1622

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a1622

ABSTRACT

Objective To apply the famine scale by Howe and Devereux to the situation in Niger, west Africa, in 2005 to retrospectively determine whether famine existed.
Design Two stage cluster survey.
Setting Survey of households in each of Niger's eight regions.
Participants 4003 households.
Main outcome measures Crude mortality, mortality in children under 5, and the proportion of caregivers adopting coping strategies to deal with insufficient food needs, nationally and regionally.
Results The estimated national crude mortality rate was 0.4 (0.4 to 0.5) deaths per 10 000 per day and under 5 mortality rate was 1.7 (1.4 to 1.9) deaths per 10 000 per day. Nationally, 22.3% (95% confidence interval 19.9% to 24.8%) of caregivers of under 5s did not resort to any coping strategies to deal with insufficient food needs. Reversible coping strategies were, however, used by 5.8% (4.7% to 7.0%) of caregivers, whereas 49.4% (46.9% to 51.8%) relied on irreversible coping strategies and 22.6% (20.0% to 25.4%) on survival strategies.
Conclusion On the basis of the famine scale proposed by Howe and Devereux, most regions in Niger experienced food crisis conditions and some areas approached famine proportions.

INTRODUCTION

The absence of universal benchmarks or criteria to identify famine conditions creates uncertainty about the magnitude of a crisis, resulting in delays and inappropriate responses.[1 2] The scale proposed by Howe and Devereux offers an approach to defining famine, based on intensity and magnitude.[1 2] The intensity scale involves a combination of anthropometric and mortality indicators and descriptors of food

security (coping strategies, stability of the market and food prices) and ascribes a level from 0 (food secure conditions) to 5 (extreme famine conditions) in population areas within a country. The magnitude scale is determined retrospectively by measuring excess mortality caused by the entire crisis.

Media reports left little doubt of a humanitarian crisis in Niger, west Africa, in 2005 but the scale and severity of the crisis and whether it reached famine proportions was disputed. We applied the scale to the crisis to determine if famine had occurred.

METHODS

We carried out surveys in Niger's eight regions: Agadez, Diffa, Dosso, Maradi, Tahoua, Tillaberi, Zinder, and the capital Niamey. We used a stratified two stage cluster survey design: we first selected 26 village level clusters in each region then used systematic random sampling to select 20 households per cluster.

To estimate mortality we included all selected households, regardless of whether they included a child aged 6 to 59 months. The head of the household completed a questionnaire, including a census of the household. The remainder of the questionnaire was administered to caregivers of children under 5 in the household, which asked about the strategies used to cope with insufficient food needs since the beginning of the last "hungry season" in 2005. We also collected anthropometric measurements on children aged 6 to 59 months.

The census was taken to determine the number of people and under 5s present in the household since 21 January 2005—the day of Tabaski, a major religious event. We recorded deaths in the household between 21 January 2005 and the date of the interview, and the month of death. A series of structured questions were used to assign cause of death among under 5s into one of eight categories on the basis of a verbal autopsy.[3] We also asked questions on the child's appearance at the time of death to determine if malnutrition may have contributed.

We stratified the coping strategies for dealing with insufficient food needs into four categories: no coping strategies, reversible adaptive strategies, irreversible coping strategies, and survival strategies (see bmj.com). On the basis of the caregivers' responses we assigned each caregiver to one of the categories. We estimated the proportion of caregivers both regionally and nationally adopting each strategy.

To account for weighting of the variables and the complex sample design we used SAS callable SUDAAN version 9.01. We calculated sampling weights by multiplying the inverse probability for selection of a person within a household multiplied by a post-stratification adjustment on the basis of the census population of the region. The procedure Proc Ratio was used to compute 95% confidence intervals for mortality. We based the number of recall days for each person on the date of the interview minus the start of the recall period (21 January 2005). For live births and deaths during the recall period we estimated the number of recall days as half of the recall period. For proportions we computed confidence intervals using Proc Crosstab. To test the difference in coping strategies between urban and rural regions we ran a *t* test using Proc Descript.

RESULTS

From 17 September to 14 October 2005 data were collected by survey from 4003 of 4160 households. Information on health was gathered on 5309 children under 5. The mean size of households was 6.9, and under 5s constituted 21.0% of household members.

The estimated crude mortality rate nationally was 0.4 deaths per 10 000 per day (95% confidence interval 0.4 to 0.5) whereas regionally it ranged from 0.2 deaths per 10 000 per day (0.1 to 0.3) in Agadez to 0.7 deaths per 10 000 per day (0.5 to 0.9) in Zinder (see bmj.com). The under 5 mortality rate nationally was 1.7 (1.4 to 1.9) deaths per 10 000 per day whereas regionally it ranged from 0.3 (0.1 to 0.7) deaths per 10 000 per day in Agadez to 2.2 (1.4 to 3.0) deaths per 10 000 per day in Zinder (see bmj.com).

Most deaths from known causes in under 5s nationally were reported as diarrhoea (30.2%, 95% confidence interval 22.7% to 38.8%), acute respiratory illness (20.1%, 13.6% to 28.6%), and meningitis (14.1%, 9.7% to 20.1%). Insufficient sample size prevented the determination of regional variations. Of the children who died during the recall period, 51.6% (42.2% to 61.0%) were perceived by their caregiver as malnourished.

Coping strategies

Since the beginning of the last hungry season coping strategies were not used by 22.3% (19.9% to 24.8%) of caregivers nationally (see bmj.com). Also nationally 5.8% (4.7% to 7.0%) of caregivers of under 5s relied on reversible coping strategies, whereas 49.4% (46.9% to 51.8%) relied on irreversible coping strategies and 22.6% (20.0% to 25.4%) on survival strategies. A similar pattern was observed in the seven predominantly rural regions, with most caregivers relying on irreversible coping strategies: the proportion of caregivers relying on such strategies varied from 35.4% (28.3% to 43.1%) in Agadez to 64.7% (58.7% to 70.3%) in Zinder, and the proportion relying on survival strategies ranged from 14.2% (9.7% to 20.4%) in Diffa to 33.0% (24.4% to 43.0%) in Maradi. In contrast with the predominantly rural regions, 56.2% (49.8% to 62.4%) of caregivers in Niamey did not use any coping strategies. Compared with each of the other seven regions, the proportion of caregivers in Niamey not using any coping strategies differed statistically ($P<0.001$ for all seven *t* tests).

WHAT IS ALREADY KNOWN ON THIS TOPIC

A lack of universal criteria to identify famine creates uncertainty about the size of the crisis, resulting in delays and inappropriate responses

The scale and severity of the food crisis in Niger in 2005 and whether it reached famine proportions was disputed

WHAT THIS STUDY ADDS

On the basis of the famine scale by Howe and Devereux, most regions in Niger in 2005 experienced food crisis conditions, and some areas approached famine

The scale provides more objective criteria than do previous approaches and along with early warning systems may help guide the level of response

DISCUSSION

Using the famine scale by Howe and Devereux we retrospectively determined that a food crisis was under way at the time of the household survey in Niger, 2005. This affected children in each region.[4] In four of the eight regions the prevalence of global acute malnutrition among children aged 6 to 59 months exceeded the critical threshold of 15% established by WHO.[5]

Despite the critical levels of acute malnutrition reported in Diffa, Maradi, Tahoua, and Zinder,[4] the prevalence of global acute malnutrition in these regions during the survey was below the 20% threshold (famine conditions).[1] The levels of global acute malnutrition in these four regions were consistent with food crisis conditions. Furthermore, crude mortality rates nationally and regionally remained below the benchmark of 1 death per 10 000 per day, as required for famine conditions.[1] The crude mortality rates in Tahoua, Tillaberi, and Zinder was between 0.5 and 1 death per 10 000 per day—consistent with food crisis conditions, whereas the remaining regions experienced crude mortality rates consistent with conditions for food insecurity. Although crude mortality rates did not indicate famine, the survey results suggest that malnutrition did play an important part in child mortality, given that 51.6% (95% confidence interval 42.2% to 61.0%) of under 5s who died during the recall period were reported by their caregivers as malnourished. In comparison, 15.6% of the global deaths among under 5s in 1994 were attributed to malnutrition.[6] The anthropometry and mortality indicators suggest that Diffa, Maradi, Tahoua, Tillaberi, and Zinder experienced food crisis conditions, whereas the remaining three regions experienced food insecurity conditions.

At the onset of the crisis, food insecurity was undoubtedly a problem; disagreement remained, however, about the magnitude of the crisis.[7 8] The food security assessment carried out in April 2005 by the government of Niger, the Food and Agricultural Organisation, the World Food Program, and the Famine Early Warning Systems Network indicated that about 67% of the people living in agropastoral areas were highly vulnerable to food insecurity.[8] Coverage by the media and reports from non-governmental organisations claimed famine conditions, but the joint food security assessment concluded that the crisis did not reach famine proportions.[8]

The survey results show that during the crisis more than 50% of caregivers of under 5s in all regions except Niamey relied on irreversible or survival coping strategies. In all regions except Niamey irreversible coping strategies were the most common means to deal with the crisis. According to the famine scale, the adoption of such strategies is consistent with a food crisis.[1] In Agadez, Maradi, Tahoua, Tillaberi, and Zinder, however, more than 20% of caregivers of under 5s resorted to survival strategies. Although the famine intensity scale does not specify what percentage of the affected population has to rely on survival strategies to constitute famine conditions, the large proportion of caregivers using survival strategies seems to indicate that in some regions famine may have been imminent.

During the crisis many raised concerns that Unicef, non-governmental organisations, and the media were exaggerating the magnitude of the situation.[8] Our results indicate that the crisis was not just localised but widespread. Most regions experienced food crisis conditions, and the adoption of survival strategies by caregivers further suggests that many regions were potentially on the threshold of a famine.

Limitations

Our study has several limitations. Firstly, an estimate of retrospective mortality is subject to recall bias. We attempted to minimise this bias by limiting the recall period to eight months and by defining the beginning of the period using Tabaski, a well known religious holiday. Secondly, we relied on oral reports of household members to determine cause of death in under 5s. Although we used a series of structured questions based on the WHO criteria for verbal autopsy to improve the accuracy of the data and to help classify cause of death, some misclassification may have occurred. Thirdly, questions on household coping strategies were administered to caregivers of under 5s, therefore these results are not representative of all households nationally. Finally, in the absence of an agreed and standardised analysis tool for assessing coping strategies, we consolidated these strategies into mutually exclusive categories and assigned each respondent to one of them. This approach does not, however, capture the complex range of strategies to which each household resorted.

Conclusion

On the basis of the famine scale by Howe and Devereux, most regions in Niger in 2005 experienced food crisis conditions and some areas approached famine proportions. The absence of a universally accepted definition of famine contributed to disagreements about the magnitude and severity of the problem. Although the utility of the scale needs rigorous assessment, it does provide more objective

criteria than previous approaches. It also moves from a binary concept of "famine" or "no famine" to a graduated approach that captures a range of food insecurity situations.[1] One potential use of the scale is in conjunction with early warning systems, to help guide the level of humanitarian response and to provide a greater opportunity to recognise a potential famine.[2] A determination of the scale and severity of a crisis and whether it reaches famine proportions is important for ensuring a proportionate and timely response.

We thank the field team members who participated in carrying out this national survey; Alichina Kourgueni for his expertise and logistical support; Sybille Gumucio and Robert Ndamobissi for coordinating the survey operations; Jonathan Seikmann for data management; Abba Kakk, Issoufou Lamissi, Bintou Tidjane, Théophile Bansimba, Baro Mohammed, and Hamadou Boureima for supervising and coordinating the field teams; Ibrahim Combasse and Alio Maidaji for translation; Yacouba Mamdou for providing oversight for data entry; Aminatou Alzouma, Foureratou Ekade, Makinta Lalla, Rachidou Katchi Amadou, Ramatou Alou Mamane, and Naroua Mavaki for data entry; and Mark Anderson for his critical input of the manuscript. KG was affiliated with the Ministry of Health, Niger during the survey. The opinions presented in this paper are those of the authors and do not necessarily reflect those of the institutions with which they are affiliated.

Contributors: See bmj.com.

Funding: Office of US Foreign Disaster Assistance and Unicef funded the survey.

Competing interests: None declared.

Ethical approval: Not required as the primary intent of the survey was determined to be non-research.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Howe P, Devereux S. Famine intensity and magnitude scales: a proposal for an instrumental definition of famine. *Disasters* 2004;28:353-72.
2 Sheckler AC, Shortley T, Swindale A, Lautze S. *Synthesis report on the famine forum*. Washington, DC. United States Agency for International Development, 2004.
3 Anker M, Black RE, Coldham C, Kalter HD, Quigley MA, Ross D, et al. A standard verbal autopsy method for investigating causes of death in infants and children. World Health Organization. www.who.int/csr/resources/publications/surveillance/WHO_CDS_CSR_ISR_99_4/en/.
4 Centers for Disease Control and Prevention. Nutritional and health status of children during a food crisis—Niger, 17 Sep to 14 Oct, 2005. *MMWR* 2006;55:1172-6.
5 World Health Organization. *The management of nutrition in major emergencies*. Geneva: WHO, 2000.
6 Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, et al. Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet* 2008;371:243-60.
7 Overseas Development Institute. Humanitarian issues in Niger. ReliefWeb, 2005. www.reliefweb.int/rw/rwb.nsf/db900sid/VBOL-6EWGXX?OpenDocument&query=humanitarian%20issues%20in%20niger%20and%20ODI.
8 Famine Early Warning System Network. Niger: an evidence base for understanding the current crisis. ReliefWeb, 2005. www.reliefweb.int/rw/rwb.nsf/db900sid/egua-6epqul?opendocument&rc=1&cc=ner.

Accepted: 15 July 2008

Overuse of tympanostomy tubes in New York metropolitan area: evidence from five hospital cohort

Salomeh Keyhani,[1] Lawrence C Kleinman,[1] Michael Rothschild,[2] Joseph M Bernstein,[3] Rebecca Anderson,[1] Mark Chassin[4]

EDITORIAL by Haggard

[1]Department of Health Policy, Mount Sinai School of Medicine, One Gustave L Levy Place, Box 1077, New York, NY 10029, USA
[2]Departments of Otolaryngology and Pediatrics, Mount Sinai School of Medicine
[3]Departments of Otolaryngology and Pediatrics, New York University School of Medicine, New York
[4]The Joint Commission, Oakbrook Terrace, IL, USA
Correspondence to: S Keyhani
Salomeh.Keyhani@mountsinai.org

Cite this as: *BMJ* 2008;337:a1607
doi:10.1136/bmj.a1607

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a1607

ABSTRACT

Objectives To compare tympanostomy tube insertion for children with otitis media in 2002 with the recommendations of two sets of expert guidelines.

Design Retrospective cohort study.

Setting New York metropolitan area practices associated with five diverse hospitals.

Participants 682 of 1046 children who received tympanostomy tubes in the five hospitals for whom charts from the hospital, primary care physician, and otolaryngologist could be accessed.

Results The mean age was 3.8 years. On average, children with acute otitis media had fewer than four infections in the year before surgery. Children with otitis media with effusion had less than 30 consecutive days of effusion at the time of surgery. Concordance with recommendations was very low: 30.3% (n=207) of all tympanostomies were concordant with the explicit criteria developed for this study and 7.5% (n=13) with the 1994 guideline from the American Academy of Pediatrics, American Academy of Family Medicine, and American Academy of Otolaryngology—Head and Neck Surgery. Children who had previously had tympanostomy tube surgery, who were having a concomitant procedure, or who had "at risk conditions" were more likely to be discordant.

Conclusions A significant majority of tympanostomy tube insertions in the largest and most populous metropolitan area in the United States were inappropriate according to the explicit criteria and not recommended according to both guidelines. Regardless of whether current practice represents a substantial overuse of surgery or the guidelines are overly restrictive, the persistent discrepancy between guidelines and practice cannot be good for children or for people interested in improving their health care.

INTRODUCTION

Otitis media is the most common illness for which children present to the doctor, and tympanostomy tubes are the most common reason for general anaesthesia in children.[1-3] Only one study has examined the appropriateness of insertion of tympanostomy tubes in children.[4 5] This study, published in 1994, reported that less than half of surgeries among children in the United States were appropriate.

We compared the clinical characteristics of the children in our study with the recommended indications for surgery as codified by the prevailing guideline (1994) at that time and a set of explicit criteria that we developed specifically for this study in 2000. To our knowledge, this

Table 1 | Clinical characteristics of cohort (year before tympanostomy)

Characteristics	Mean (SE)	Median (interquartile range)
Episodes of infection*		
6 months before tympanostomy	3.1 (0.1)	3 (2-4)
1 year before tympanostomy	4.6 (0.1)	4 (3-6)
Length of effusion†		
Consecutive days of effusion‡:		
Bilateral	30.7 (1.7)	16 (0-49)
Unilateral (left)	39.1 (2.0)	26 (2-63.5)
Unilateral (right)	39.8 (2.0)	28 (3-63.5)
Cumulative days of effusion§:		
Bilateral	77 (2.9)	66 (30-109.5)
Unilateral (left)	91.7 (3.2)	78.5 (40.5-131.5)
Unilateral (right)	96.8 (3.1)	88 (46.5-138)

*Children with acute otitis media (n=230).
†Children with otitis media with effusion (n=452).
‡Refers to effusion directly preceding surgery.
§Cumulative effusion over one year

is the first study to examine the appropriateness of insertion of tympanostomy tubes with data collected by independent audits of the records.

METHODS

Study population and data

We did a retrospective cohort study and collected detailed data for a one year period before insertion of a tympanostomy tube for each child.[6] We developed an electronic tool to assist data collection for chart audits of all 1046 children less than 18 years old who were identified by hospital administrative databases as having received tympanostomy in any of five New York metropolitan area hospitals in 2002. This manuscript reports on those 682 children for whom we were able to audit all three medical records: from the primary care physician's office, the otolaryngologist's office, and the hospital.

Development of explicit criteria

The RAND appropriateness method uses a two round modified Delphi process to integrate literature with expert opinion into explicit criteria.[7] We convened an expert panel of four otolaryngologists, four paediatricians, and one family physician and provided them with a detailed literature review. The panel identified relevant clinical factors that we organised into an exhaustive and mutually exclusive list of potential clinical scenarios to represent the range of circumstances that might present to a clinician. The panellists then rated the appropriateness of each scenario (see bmj.com).

The 1994 academy guideline was developed independently of this study by three clinical societies and was published as a clinical practice guideline by the US Agency for Health Care Policy and Research.[8] The 1994 guideline was limited to "healthy" children from their first birthday until before they turn 4 and suggests that insertion of a tympanostomy tube is optional after three months of persistent effusion with bilateral hearing loss and is recommended after four to six months of bilateral effusion.

Analysis

We mapped each child to the detailed clinical scenario rated by the panel that was consistent with the details of the clinical history. We also compared practice with the 1994 academy guideline. As no national guidelines on the surgical management of acute otitis media exist, we limited our analysis to children with otitis media with effusion.

We also examined how alternatives to our expert panel's judgment would affect the appropriateness ratings for the treatment of children with acute otitis media. Finally, we looked at the relation to appropriateness of the presence of conditions that would place a child at risk for poor outcomes, a history of tympanostomy tube surgery, and other procedures done at the time of tympanostomy tube surgery.[5]

RESULTS

Children's mean age was 3.8 years, and 61% were white. More than a quarter (26.5%) of children had previously received tubes, and 21.7% had another surgery done concurrently with the insertion of a tympanostomy tube; 17.3% of children had a clinical condition that could be considered to place them "at risk" of poor developmental outcomes (see bmj.com). Children with acute otitis media on average had about three infections in the six months before tympanostomy. Children with otitis media with effusion had less than 30 consecutive days of bilateral effusion before tympanostomy (table 1). Only 25% of children with otitis media with effusion had bilateral effusions of more than 49 days' duration at surgery. More than a

Table 2 | Appropriateness ratings based on explicit criteria and academy guidelines. Values are numbers (percentages)

	Concordant		Not concordant
	Appropriate	Uncertain	Inappropriate
Ratings according to explicit criteria			
All children (n=682)	48 (7.0)	159 (23.3)	475 (69.7)
Acute otitis media (n=230)	9 (3.9)	110 (47.8)	111 (48.3)
Otitis media with effusion (n=452)	39 (8.6)	49 (10.8)	364 (80.6)
All children: criteria modified to eliminate preference for antibiotic prophylaxis* (n=682)	150 (21.9)	57 (8.3)	475 (69.7)
Excluding cases with potentially extenuating circumstances† (n=341)	31 (9.1)	104 (30.5)	206 (60.4)
Academy guidelines‡			
1994 Academy guidelines (n=172)	13 (7.5)		159 (92.5)
1994 Academy guidelines (age>1 year) (n=533)	30 (5.6)		503 (94.4)

*The expert panel considered indication for surgery for recurrent acute otitis media to be of uncertain appropriateness if no antibiotic prophylaxis was used to suppress reoccurrence; this analysis elevates the ratings for surgeries that failed to meet this aspect of the criteria from uncertain to appropriate (a potentially alternative view); no official guidelines on the surgical treatment of acute otitis media exist.
†Such as history of previous tubes, other surgery/procedure at time of tube insertion and "at risk conditions"; the expert panel considered only the nature and magnitude of ear disease in the decision to insert tubes and not consider these extenuating circumstances.
‡The 1994 Academy guideline covers the management of otitis media with effusion, not recurrent acute otitis media.

Table 3 | Variability of concordance with explicit criteria

Subpopulations of children	Concordant (%)	P value
All children (n=682)	207 (30.3)	
At risk*		
Yes (n=118)	18 (15.2)	<0.0001
No (n=564)	189 (33.5)	
Concomitant procedure:		
Yes (n=148)	34 (22.9)	<0.05
No (n=534)	173 (32.4)	
History of tympanostomy tubes:		
Yes (n=181)	37 (20.4)	<0.001
No (n= 501)	170 (33.9)	

*Includes children with hearing loss independent of otitis media with effusion; language or speech disorder; autism and other developmental symptoms; Down's syndrome or other craniofacial syndromes that include cognitive, speech, or language delay, visual impairment, cleft palate, and developmental delay.

quarter of children who had surgery had normal audiograms.[6]

Of 2268 potential clinical scenarios, we saw 220 (9.7%) in clinical practice (see bmj.com). The explicit criteria classified 7.0% (48 cases) of actual surgeries as appropriate, 23.3% (159 cases) as of uncertain appropriateness, and 69.7% (475 cases) as inappropriate (table 2). Cases that were classified as appropriate, uncertain, and inappropriate had on average 80, 38, and 18 days of effusion. Counting surgeries classified as appropriate or uncertain as concordant, 30.3% were concordant with the explicit criteria. We found no statistically significant differences in appropriateness between hospitals.

Among children with effusion as the reason for surgery, 76% of the inappropriate cases were inappropriate primarily owing to the short duration of the effusion immediately preceding surgery. Low frequency of infection was the most common reason why the cases with acute otitis media were not concordant with the explicit criteria. In the analysis based on academy guidelines, the main reason for discordance with the guideline was again short duration of effusion.

In an additional analysis, we excluded all children with potentially extenuating circumstances. We found that the explicit criteria would consider 9.1% of the surgeries to be appropriate, 30.5% to be of uncertain appropriateness, and 60.4% to be inappropriate. Thus even in a liberal review, more than 60% of cases were not concordant with the panel's findings (table 2). Not surprisingly, concordance with the criteria also was lower for children in each of these three subpopulations than for the population as a whole (table 3).

DISCUSSION

We found that tympanostomy tubes in the United States are often used in a manner inconsistent with expert recommendations.

Implications

The finding that 69% of cases deviated from the practices specifically developed for this study—or that more than 92% of surgeries would have been "not recommended" according to the guideline in force at the time of the surgery—suggests considerable overuse of this procedure. Historically, the major benefit of tympanostomy tubes discussed in the literature pertains to speech and language development. Recent research provides strong evidence that delay in the insertion of tympanostomy tubes is not associated with worse behavioural or developmental outcomes.[9 10] However, these findings do not imply that tubes should be avoided or that there are not health systems for which tubes may be underused.

Limited evidence shows that tubes improve disease specific quality of life.[11-13] If tubes do make children feel better or otherwise improve the quality of their lives in the short term, then the emphasis on long term outcomes and development that has predominated in the guideline may not be sufficient.

Our data show that otolaryngologists treat children differently if they have one or more of three specific circumstances—a history of previously having had tympanostomy tubes, the scheduling of a concomitant surgery, and the presence of one or more of the conditions we identified as putting a child "at risk." About half of our sample had one or more of these circumstances present. The extent to which these circumstances should be extenuating is questionable.

Limitations

In this study, we focused only on children who had received tubes; we did not consider the possible underuse of tympanostomy in some populations. Our timing during implementation of a US federal privacy rule restricted our access to charts for about one third of children. However, even considering the most generous assumption that insertion of a tympanostomy tube was appropriate for every child with data missing, more than 40% of cases would still be considered inappropriate.

To make assessments about the course of otitis media we needed to translate the intermittent clinical assessments available from the charts into the continuous variables (days of effusion) that we used in our analysis.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Tympanostomy tubes are used commonly in the United States to treat otitis media, whereas in the UK this practice has been reduced

A 1994 study which suggested that inappropriate use was common in the US was controversial because of its data sources and the criteria used

WHAT THIS STUDY ADDS

Most insertions of tympanostomy tubes in the New York metropolitan area were for inappropriate reasons according to two different standards

These findings suggest a serious discrepancy between the clinical care of children and the recommendations that experts suggest should be the standard of care in the US

The need to impute findings is an unavoidable limitation; however, we made generous clinical assumptions that would favour a longer duration of effusion and concordance with the explicit criteria and the academy guideline (see bmj.com).

Conclusions

Regardless of whether current practice represents a substantial overuse of surgery or the guidelines are overly restrictive, the persistent discrepancy between guidelines and clinical practice cannot be good either for children or for those interested in improving their health. Substantial overuse would expose children to risk and consume resources that could be better applied to otherwise improving the health of children. Erroneous guidelines could lead clinicians, policy makers, and researchers to ill advised interventions and undermine the value of guidelines in general. Given the ubiquity of this disease and its surgical treatment, resolution of these issues should represent an urgent priority. The UK experience may prove a useful resource for policy makers in the United States as they take on these challenges.

Expert panel: Ellen M Friedman, chief, Pediatric Otolaryngology, Texas Children's Hospital, Houston, TX; G Scott Giebink, professor of pediatrics and otolaryngology, University of Minnesota Medical School, Minneapolis, MN; Gregory Hayden, professor of pediatrics, University of Virginia Health System, Charlottesville, VA; John Hickner, professor and vice chair, Department of Family Medicine, University of Chicago Pritzker School of Medicine; Margaret A Kenna, professor of otology and laryngology, Harvard Medical School, Boston, MA; Jack L Paradise, professor of pediatrics, University of Pittsburgh, Pittsburgh, PA; Seth H Pransky, director, Pediatric Otolaryngology, Children's Hospital, San Diego, CA; Oliver Roddey, Eastover Pediatrics, Charlotte, NC. One panel member from the Seattle, WA, area preferred to remain anonymous.

Contributors: See bmj.com.

Funding: Agency for Health Care Research and Quality (R01 HS 10302). The funding agency had no role in the design or implementation of the study; all views presented are the authors' own and may not reflect the views of AHRQ.

Competing interests: MC was the principal investigator on this study before he left his position as chair of the Department of Health Policy to become the president of the Joint Commission.

Ethical approval: Institutional review boards of all five hospitals. This is a retrospective medical record review and no contact with patients was permitted.

This work was presented at the June 2007 and 2008 annual AcademyHealth meetings.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 American Academy of Pediatrics Subcommittee on Management of Acute Otitis Media. Diagnosis and management of acute otitis media. *Pediatrics* 2004;113:1451-65.
2 American Academy of Family Physicians, American Academy of Otolaryngology-Head and Neck Surgery, American Academy of Pediatrics Subcommittee on Otitis Media with Effusion. Otitis media with effusion. *Pediatrics* 2004;113:1412-29.
3 Boston M, McCook J, Burke B, Derkay C. Incidence of and risk factors for additional tympanostomy tube insertion in children. *Arch Otolaryngol Head Neck Surg* 2003;129:293-6.
4 Kleinman LC, Kosecoff J, Dubois RW, Brook RH. The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States. *JAMA* 1994;271:1250-5.
5 Kleinman LC, Boyd EA, Heritage JC. Adherence to prescribed explicit criteria during utilization review: an analysis of communications between attending and reviewing physicians. *JAMA* 1997;278:497-501.
6 Keyhani S, Kleinman LC, Rothschild M, Bernstein JM, Anderson R, Simon M. Clinical characteristics of New York City children who received tympanostomy tubes in 2002. *Pediatrics* 2008;121:e24-33.
7 Brook RH, Chassin MR, Fink A, Kosecoff J. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986;2:53-63.
8 Stool SE, Berg AO, Berman S, Carney CJ, Cooley JR, Culpepper L. *Otitis media with effusion in young children: clinical practice guideline, number 12*. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services, 1994.(AHCPR Publication No 94-0622.)
9 Paradise JL, Feldman HM, Campbell TF, Dollaghan CA, Colborn DK, Bernard BS. Early versus delayed insertion of tympanostomy tubes for persistent otitis media: developmental outcomes at the age of three years in relation to prerandomization illness patterns and hearing levels. *Pediatr Infect Dis J* 2003;22:309-14.
10 Paradise JL, Feldman HM, Campbell TF, Dollaghan CA, Rockette HE, Pitcairn DL. Tympanostomy tubes and developmental outcomes at 9 to 11 years of age. *N Engl J Med* 2007;356:248-61.
11 Rosenfeld RM, Bhaya MH, Bower CM, Brookhouser PE, Casselbrant ML, Chan KH. Impact of tympanostomy tubes on child quality of life. *Arch Otolaryngol Head Neck Surg* 2000;126:585-92.
12 Chow Y, Wabnitz DA, Ling J. Quality of life outcomes after ventilating tube insertion for otitis media in an Australian population. *Int J Pediatr Otorhinolaryngol* 2007;71:1543-7.
13 Richards M, Giannoni C. Quality-of-life outcomes after surgical intervention for otitis media. *Arch Otolaryngol Head Neck Surg* 2002;128:776-82.

Accepted: 4 August 2008

Dangerously ill list patient

After completing my medical studies 27 years ago, I was posted to a government hospital as a medical officer. On my first day on duty as the emergency medical officer I was told by a senior matron that I had to conduct rounds of all the wards and write progress notes for all the patients on the dangerously ill list.

I started the ward rounds in earnest. In a remote corner of the sprawling, colonial era hospital, I entered a high ceilinged ward that was deserted except for a steel bed in the centre of the room. Placed in the centre of the bed was a table, on top of which was a stool. Standing on the stool was a young man, and perched on his shoulders was another young man, who was cleaning the ceiling fan as part of the preparations for a forthcoming hospital inspection. It was pointed out to me by the nursing officer in charge of the ward that the young man cleaning the fan was on the dangerously ill list.

When I asked the patient about his diagnosis and progress, he confirmed being on the dangerously ill list for anaemia—his haemoglobin concentration was 90 g/l. The nursing officer informed me that, according to the existing medical services rules (framed in the era of British rule), all patients with haemoglobin levels below 100 g/l had to be placed on the dangerously ill list.

Ajit Singh Kashyap consultant, Command Hospital (Central Command), Lucknow, India
Kuldip Parkash Anand deputy director of medical services, Uttar Bharat Area, Bareilly, India
kashyapajits@gmail.com
Cite this as: *BMJ* 2008;337:a229