The *BMJ* is an Open Access journal.
We set no word limits on *BMJ* research articles, but they are abridged for print. The full text of each *BMJ* research article is freely available on bmj.com

RESEARCH

Effect of using an interactive booklet about childhood respiratory tract infections in primary care consultations on reconsulting and antibiotic prescribing: a cluster randomised controlled trial

Nick A Francis,[1 2] Christopher C Butler,[1] Kerenza Hood,[1 2] Sharon Simpson,[1 2] Fiona Wood,[1] Jacqueline Nuttall[1 2]

[1]Department of Primary Care and Public Health, School of Medicine, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4XN
[2]South East Wales Trials Unit, School of Medicine, Cardiff University
Correspondence to: N Francis
francisna@cf.ac.uk

Cite this as: *BMJ* 2009;339:b2885
doi: 10.1136/bmj.b2885

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b2885

ABSTRACT

Objectives To determine whether an interactive booklet on respiratory tract infections in children reduces reconsultation for the same illness episode, antibiotic use, and future consulting intentions, while maintaining parental satisfaction with care.

Design Pragmatic cluster randomised controlled trial with randomisation at the level of the general practice.

Setting 61 general practices in Wales and England.

Participants 558 children (6 months to 14 years) consulting in primary care with an acute respiratory tract infection (7 days or less). Children with suspected pneumonia, asthma or a serious concomitant illness, or needing immediate hospital admission were excluded. Three withdrew and 27 were lost to follow-up, leaving 528 (94.6%) with main outcome data.

Interventions Clinicians in the intervention arm were trained in the use of an 'interactive' booklet on respiratory tract infections and asked to use the booklet during consultations with recruited patients (and provide it as a take-home resource). Clinicians in the control practices conducted their consultations as usual.

Main outcome measures The proportion of children who attended a face –to face consultation about the same illness during the two week follow-up period. Secondary outcomes included antibiotic prescribing, antibiotic consumption, future consulting intentions, and parental satisfaction, reassurance and enablement.

Results Reconsultation occurred in 12.9% and 16.2% of children in the intervention and control arms respectively (absolute risk reduction=3.3% 95% CI −2.7% to 9.3%, P=0.29). Using multilevel modelling (at the practice and individual level) to account for clustering, no significant difference in reconsulting was found (odds ratio 0.75; 95% CI 0.41 to 1.38). Antibiotics were prescribed at the index consultation to 19.5% and 40.8% of children in intervention and control practices respectively (absolute risk reduction=21.3% 95% CI 13.7 to 28.9, P<0.001). A significant difference was still present after adjusting for clustering (odds ratio 0.29; 95% CI 0.14 to 0.60). There was also a significant difference in the proportion of parents who said they would consult in the future should their child develop a similar illness (odds ratio 0.34; 95% confidence interval 0.20 to 0.57). Satisfaction, reassurance, and parental enablement scores were not significantly different between the two groups.

Conclusions Use of a booklet on respiratory tract infections in children within primary care consultations led to important reductions in antibiotic prescribing and reduced intention to consult without reducing satisfaction with care.

Trial registration Current Controlled Trials ISRCTN46104365

WHAT IS ALREADY KNOWN ON THIS TOPIC

Respiratory tract infections in children are largely self-limiting and benefit very little from antibiotic treatment

However, consultation rates continue to be high and antibiotics are still frequently prescribed

WHAT THIS STUDY ADDS

Providing primary care clinicians with a carefully developed booklet on respiratory tract infections in children, and training in its use within the consultation, reduces antibiotic prescribing by around two thirds

Satisfaction among parents receiving this intervention was high, and no significant difference was found between those receiving the intervention and those receiving usual care

Use of this intervention appears to have little impact on re-consulting for the same illness episode, but does reduce future consulting intentions

Clinicians should consider using this intervention in routine consultations with children with respiratory tract infections

INTRODUCTION

Complications of respiratory tract infections are rare, and there is little or no benefit from treatment with antibiotics.[1-5] Nevertheless, antibiotics continue to be overprescribed,[6 7] with children receiving more antibiotics than any other age group.[8] Prescribing for non-specific upper respiratory tract infections, which declined in the late nineties, is once again increasing.[9] Parental beliefs, fears, and expectations play an important role in both consulting behaviour and determining whether an antibiotic is prescribed. Communication within the consultation is central to addressing these.

We set out to determine whether training clinicians in the use of an "interactive" booklet, designed to enhance communication within the consultation, and act as a take-home resource for parents, would have an effect on re-consultation rates and antibiotic prescribing. A cluster design was needed as the intervention is in part directed at the clinicians in the practice.

METHODS

The methods for this cluster randomised controlled trial have already been described.[10] In Wales 49 practices were randomised and, of these, 36 recruited study participants. In England 34 practices were randomised, and 25 of these recruited participants. Practices were randomised using block randomisation with random block sizes and stratification by practice list size, antibiotic prescribing rate for 2005, and country.

Participating clinicians were asked to recruit sequential eligible children (6 months to 14 years) consulting with a respiratory tract infection (cough, cold, sore throat, earache for seven days or less) and their parents. From the power calculation we calculated we would need 524 participants recruited from 60 clusters (practices). See bmj.com.

The intervention

The intervention consisted of an eight page booklet on respiratory tract infections in children, designed to be used within the consultation and then provided to parents as a take-home resource, and online training for clinicians in its use. The booklet (www.equipstudy.com) acted as an evidence based information resource for parents, an aide-memoir for clinicians, a tool to help set realistic expectations, and a prompt to enhance communication within the consultation.

Clinicians in practices randomised to the control arm of the study were asked to conduct the consultation in their usual manner.

Measures

Baseline data, including age, prior duration of illness, and symptoms, were collected by clinicians at the time of recruitment. We asked clinicians to collect non-identifiable data on all "potentially eligible patients" (including those not approached, those who were approached but found to be ineligible, and those who declined participation) in order to assess for possible selection bias. Follow up was via a telephone administered questionnaire with the child's parent or guardian, fourteen days after recruitment. Telephone interviewers were blinded to treatment arm and were asked to record becoming unblinded (by a parent talking about receiving a booklet for example).

Reconsultation during the two weeks following the index consultation was the primary outcome. Antibiotic prescribing, antibiotic consumption, future consulting intentions, parental satisfaction, perception of the usefulness of information received, reassurance and enablement, were secondary outcomes.

Analysis

Following missing data and range checks we obtained summary statistics and performed univariate analyses. The primary analysis was intention to treat, conducted by fitting two-level (practice and patient) random intercept logistic regression models. Similar models were fitted for the secondary outcomes.

Sensitivity analyses were conducted by adding the stratifying variables, age, prior duration of illness, and any symptoms found to be significantly associated in univariate analyses at the 10% level into each model as covariates. Exploratory analyses were conducted by including factors likely to influence reconsulting and antibiotic prescribing into the models and examining the interaction factors to look for subgroup effects.

RESULTS

Eighty-three practices were randomised, and 61 of these recruited a total of 558 eligible patients between October 2006 and April 2008.

Intervention and control practices, and randomised practices that did and did not recruit participants, were similar in terms of list size, antibiotic prescribing history, and location. Patients recruited by intervention and control practices were similar in terms of age, gender, duration of illness, and symptoms. Patients were recruited by intervention and control practices at a similar rate. We achieved a follow-up rate of 94.6% (93.4% intervention, 95.8% control) for the primary outcome data. Telephone interviewers reported becoming aware of the participant's treatment arm in 34 of 509 interviews (6.7%).

There was no significant difference between the intervention and control groups in the odds of reconsulting in primary care during the two weeks following registration (table). Children in the intervention arm were significantly less likely to receive a prescription for antibiotics at the index consultation, less likely to take antibiotics during the first two weeks, and parents were less likely to report that they would consult in the future with a similar illness in their child. There were no significant differences in terms of satisfaction, level of reassurance, parental enablement, or the parent's rating of the "usefulness of any information received in the consultation." Similar results were found at the univariate level with a non-significant difference in reconsulting (absolute risk reduction=3.3%, 95% CI –2.7% to 9.3%, P=0.29), and significant differences in antibiotic prescribing (absolute risk reduction=21.3%, 95% CI 13.7 to 28.9, NNT=4.7, P<0.001), antibiotic consumption (absolute risk reduction=20.6%, 95% CI 12.7% to 28.5%, NNT=4.9, P<0.001) and future consulting intentions (absolute risk reduction=21.1%, 95% CI 13.1% to 29.2%, NNT=4.7, P<0.001).

Sensitivity analyses did not result in meaningful changes to the results.

In the antibiotic prescribing model, we found that the intervention was more effective in above average prescribing practices. There were no other significant interaction effects. See bmj.com.

Adverse events

Seven patients (three intervention, four control) were admitted to hospital or observed in a paediatric assessment unit. One patient (control) had a longstanding diagnosis of asthma, and was removed from all analyses. The longest hospital admission (two nights) was in an intervention patient who had febrile convulsions. The remaining admissions were one night or less. See bmj.com for comparison of recruited and non-recruited patients.

Effect of the intervention on patient outcomes

	Number (%) experiencing the outcome		Odds ratio from multilevel modelling (95% CI)
	Intervention	Control	
Outcomes with data collected from telephone administered and postal questionnaires	n=256	n=272	
Primary outcome: primary care reconsultation within first two weeks* (intracluster correlation coefficient=0.06)	33 (12.9)	44 (16.2)	0.75 (0.41 to 1.38)
Antibiotic prescribed at index consultation (intracluster correlation coefficient=0.24)	50 (19.5)	111 (40.8)	0.29 (0.14 to 0.60)
Outcomes with data collected from telephone administered questionnaire only	n=246	n=263	
Antibiotics taken within first two weeks (including antibiotics prescribed after index consultation)	55 (22.4)	111 (43.0)	0.35 (0.18 to 0.66)
Parent intends to consult if their child has similar illness in future	136 (55.3)	201 (76.4)	0.34 (0.20 to 0.57)
Parental enablement score (≥5)	99 (40.2)	94† (35.9)	1.20 (0.84 to 1.73)
Satisfaction‡	222 (90.2)	246 (93.5)	0.64 (0.33 to 1.22)
Reassurance§	177 (72.0)	198 (75.3)	0.84 (0.57 to 1.25)
Usefulness of information received¶	210 (85.4)	224 (85.2)	1.01 (0.60 to 1.68)

*Parental report that child attended a face to face consultation with a primary care clinician in their general practice, or with an out-of-hours provider, in the two weeks after registration.
†No=262 for this group as one parent was unable to complete enablement questions because of language problems.
‡Proportion of parents who reported being very satisfied or satisfied with the consultation.
§ Proportion of parents who reported feeling very reassured after their consultation.
¶Proportion of parents who reported that information they received in the consultation was very useful or useful.

DISCUSSION

Clinicians' use of an interactive booklet on respiratory tract infections in children within primary care consultations resulted in a significant reduction in antibiotic prescribing and consumption and high levels of parental satisfaction. Use of the intervention did not result in a significant reduction in the proportion of children who reconsulted in the two weeks following the index consultation (although there was a trend towards less reconsulting in the intervention group). Fewer parents in the intervention arm said they would consult in the future should their child develop a similar illness. No significant differences were found in terms of parental satisfaction, reassurance, enablement or perception of the "usefulness" of any information received about their child's illness.

Study strengths and weaknesses

Practices were broadly representative of UK general practice, so likely to be highly generalisable. We achieved the target sample for both clusters and patients, with a high follow-up rate.

Cluster randomised designs can increase risk of selection bias, which can occur at the level of the cluster (through differential dropout) or the individual. We found little evidence of selection bias. See bmj.com.

The non-significant difference in scores of parental enablement and usefulness of information received are surprising. The patient enablement instrument was designed for first person use in routine general practice and may not have been sensitive enough.

Clinicians in the control practices may have altered their behaviour (towards providing more information than usual). This may have attenuated any effect that changes in the behaviour of doctors in intervention practices might have had on parental satisfaction, enablement and usefulness of information received.

Clinicians did not have any involvement in measuring outcomes and telephone interviewers were kept blind to trial arm in 93% of all interviews. Therefore, there is unlikely to be a significant degree of ascertainment bias. Children (and their parents) were not blinded to treatment group, but were not informed of the arm to which they had been assigned prior to obtaining consent.

We did not measure treatment fidelity as we wanted the assessors to remain blinded to study arm where possible. However, suboptimal fidelity of intervention delivery is likely to dilute the treatment effect and therefore may have led to a type II error regarding reconsultations, but is unlikely to have led to a type I error regarding the positive findings.

Comparison with other published work

Our findings are consistent with a study which found use of a leaflet on lower respiratory tract infection in adults resulted in a reduction in antibiotic use of nearly 25%.[11] These researchers also demonstrated a reduction in reconsultations from use of a leaflet,[12] while we did not demonstrate a statistically significant reduction. This may be because the underlying reconsultation rate in our study was lower, and lower than that used in our sample size calculation. Other studies that have evaluated the use of booklets on minor illnesses have generally found little benefit. This may be because they were provided outside the context of the consultation,[13-17] did not encourage "interactive" use of the material within the consultation,[18 19] or provided negative messages (such as "don't use antibiotics") rather than positive ones (how best to manage the illness).[20]

Interpretation of the results

We demonstrated statistically and clinically significant reductions in antibiotic prescribing and consumption, which have important implications for policy makers, practitioners, and ultimately patients. It is not yet clear how the reduction in prescribing was mediated, but it is likely that it was through a combined effect on clinician and parental behaviour.

We are grateful to all the patients, parents, and clinicians who participated in the development of the intervention and the trial. We gratefully acknowledge the support received from the Primary Care Research Network and from all participating research networks. We thank the administrative staff in the Department of Primary Care and Public Health and the South-East Wales Trials Unit who worked hard to the ensure the success of the study. We also gratefully acknowledge the contribution made by members of the independent Trial Steering Committee.

Contributors: See bmj.com.

Funding: We gratefully acknowledge funding from the Medical Research Council and the Welsh Assembly Government in the form of a joint Health Services Fellowship for NF. Funding for the development of the training website was from an educational grant from Pfizer UK. The South-East Wales Trials Unit is funded by the Welsh Office for Research and Development. All authors declare that this work was conducted independently from the study funders.

Sponsorship: The study was sponsored by Cardiff University.

Competing interests: None declared.

Ethical approval: The study was approved by the South East Wales Local Research Ethics Committee (reference number 04/WSE04/109).

1 Arroll B, Kenealy T. Antibiotics for the common cold and acute purulent rhinitis. *Cochrane Database Syst Rev* 2005;3:CD000247.
2 Del Mar C, Glasziou Paul P, Spinks A. Antibiotics for sore throat. *Cochrane Database Syst Rev* 2006;4:CD000023.
3 Smith SM, Fahey T, Smucny J, Becker Lorne A. Antibiotics for acute bronchitis. *Cochrane Database Syst Rev* 2004;4:CD000245.
4 Spurling GKP, Fonseka K, Doust J, Del Mar C. Antibiotics for bronchiolitis in children. *Cochrane Database Syst Rev* 2007;1:CD005189.
5 Glasziou PP, Del Mar C, Sanders S, Hayem M. Antibiotics for acute otitis media in children. *Cochrane Database Syst Rev* 2004;1:CD000219.
6 Ashworth M, Charlton J, Cox K, Gulliford M, Latinovic R, Rowlands G. Why has antibiotic prescribing for respiratory illness declined in primary care? A longitudinal study using the General Practice Research Database. *J Public Health* 2004;26:268-74.
7 Shapiro E. Injudicious antibiotic use: an unforeseen consequence of the emphasis on patient satisfaction? *Clin Ther* 2002;24:197-204.
8 Akkerman AE, van der Wouden JC, Kuyvenhoven MM, Dieleman JP, Verheij TJM. Antibiotic prescribing for respiratory tract infections in Dutch primary care in relation to patient age clinical entities. *J Antimicrob Chemother* 2004;54:1116-21.
9 Thompson PL, Spyridis N, Sharland M, Gilbert RE, Saxena S, Long PF, et al. Changes in clinical indications for community antibiotic prescribing for children in the UK from 1996-2006: will the new NICE prescribing guidance on upper respiratory tract infections be ignored? *Arch Dis Child* 2008;doi:10.1136/adc.2008.147579.
10 Francis N, Hood K, Simpson S, Wood F, Nuttall J, Butler C. The effect of using an interactive booklet on childhood respiratory tract infections in consultations: study protocol for a cluster randomised controlled trial in primary care. *BMC Family Practice* 2008;9:23.
11 Macfarlane J, Holmes W, Gard P, Thornhill D, Macfarlane R, Hubbard R. Reducing antibiotic use for acute bronchitis in primary care: blinded, randomised controlled trial of patient information leaflet. *BMJ* 2002;324:91-4.
12 Macfarlane JT, Holmes WF, Macfarlane RM. Reducing reconsultations for acute lower respiratory tract illness with an information leaflet: A randomized controlled study of patients in primary care. *Br J Gen Pract* 1997;47:719-22.
13 Morrell DC, Avery AJ, Watkins CJ. Management of minor illness. *BMJ* 1980;280:769-71.
14 Hansen BW. A randomized controlled trial on the effect of an information booklet for young families in Denmark. *Patient Educ Couns* 1990;16:147-50.
15 Usherwood TP. Development and randomized controlled trial of a booklet of advice for parents. *Br J Gen Pract* 1991;41:58-62.
16 Heaney D, Wyke S, Wilson P, Elton R, Rutledge P, Sommerville A, et al. Assessment of impact of information booklets on use of healthcare services: randomised controlled trial. *BMJ* 2001;322:1-5.
17 Little P, Somerville J, Williamson I, Warner G, Moore M, Wiles R, et al. Randomised controlled trial of self management leaflets and booklets for minor illness provided by post. *BMJ* 2001;322:1214-7.
18 Finkelstein JA, Huang SS, Kleinman K, Rifas-Shiman SL, Stille CJ, Daniel J, et al. Impact of a 16-community trial to promote judicious antibiotic use in Massachusetts. *Pediatrics* 2008;121:e15-23.
19 Little P, Rumsby K, Kelly J, Watson L, Moore M, Warner G, et al. Information leaflet and antibiotic prescribing strategies for acute lower respiratory tract infection: a randomized controlled trial. *JAMA* 2005;293:3029-35.
20 Taylor JA, Kwan-Gett TSC, McMahon EM, Jr. Effectiveness of a parental educational intervention in reducing antibiotic use in children: a randomized controlled trial. *Pediatr Infect Dis J* 2005;24:489-93.

Accepted: 4 March 2009

What factors predict differences in infant and perinatal mortality in primary care trusts in England? A prognostic model

Nick Freemantle,[1] J Wood,[2] C Griffin,[3] P Gill,[1] M J Calvert,[1] A Shankar,[4] J Chambers,[5] C MacArthur[1]

ABSTRACT

Objective To identify predictors of perinatal and infant mortality variations between primary care trusts (PCTs) and identify outlier trusts where outcomes were worse than expected.

Design Prognostic multivariable mixed models attempting to explain observed variability between PCTs in perinatal and infant mortality. We used these predictive models to identify PCTs with higher than expected rates of either outcome.

Setting All primary care trusts in England.

Population For each PCT, data on the number of infant and perinatal deaths, ethnicity, deprivation, maternal age, PCT spending on maternal services, and "Spearhead" status.

Main outcome measures Rates of perinatal and infant mortality across PCTs.

Results The final models for infant mortality and perinatal mortality included measures of deprivation, ethnicity, and maternal age. The final model for infant mortality explained 70% of the observed heterogeneity in outcome between PCTs. The final model for perinatal mortality explained 80.5% of the between-PCT heterogeneity. PCT spending on maternal services did not explain differences in observed events. Two PCTs had higher than expected rates of perinatal mortality.

Conclusions Social deprivation, ethnicity, and maternal age are important predictors of infant and perinatal mortality. Spearhead PCTs are performing in line with expectations given their levels of deprivation, ethnicity, and maternal age. Higher spending on maternity services using the current configuration of services may not reduce rates of infant and perinatal mortality.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b2892

WHAT IS ALREADY KNOWN ON THIS TOPIC

There is substantial heterogeneity in infant and perinatal mortality between primary care trusts (PCTs) in England

Around 30% of PCTs with the worst health and deprivation indicators have been given Spearhead status, requiring special attention

No study has attempted to account for between-PCT variability in infant and perinatal mortality on the basis of known population risk factors and PCT spending

WHAT THIS STUDY ADDS

Between 70% and 80% of between-PCT variability in infant and perinatal mortality can be explained by a combination of deprivation, ethnicity, and maternal age

Differences in PCT spending, either between-PCT or over time, do not reliably explain differences in rates of infant and perinatal mortality

Although having higher rates of infant and perinatal mortality, Spearhead PCTs do not have results out of line with the risks in their populations. Neither of the two PCTs identified as having higher than expected rates of perinatal mortality had Spearhead status

INTRODUCTION

There is increasing interest in the levels of performance of primary care trusts (PCTs). Those with particularly poor performance data have been assigned "Spearhead" status by the Department of Health.[1] It is unclear whether worse outcomes arise from poor service provision and lack of expenditure or from patient demographics such as deprivation or ethnicity. It is also unclear whether variation in NHS service provision contributes to variation in outcome.

[1]School of Health & Population Sciences, University of Birmingham, Birmingham B15 2TT
[2]Department of Chemical Sciences and Pharmacy, University of East Anglia, Norwich NR4 7TJ
[3]Sandwell Primary Care Trust, Kingston House, West Bromwich B70 9LD
[4]Health and Social Surveys Research Group, Department of Epidemiology and Public Health, UCL-Gower Street Campus, London WC1E 6BT
[5]Heart of Birmingham PCT, Edgbaston, Birmingham B16 9NX
Correspondence to: N Freemantle
N.Freemantle@bham.ac.uk

Cite this as: *BMJ* 2009;339:b2892
doi: 10.1136/bmj.b2892

PCT characteristics included as candidate variables in model building process for both infant mortality and perinatal mortality

- Infant mortality events (the number of deaths of infants (1 year of age or younger)) and rates per 1000 live births[2]
- Perinatal mortality events (deaths occurring during late pregnancy (at ≥24 completed weeks' gestation), during childbirth, and up to seven completed days of life) and rates per 1000 births[2]
- English Indices of Deprivation score for each PCT[3]
- Spearhead status of PCT[1]
- Ethnicity for the PCT population[2]—mixed, black, Indian, Pakistani, Bangladeshi/other Asian, Chinese, white (comparator)
- Mother's age at birth[2]—<16 years, <18 years, >35 years
- PCT expenditure (estimated) on maternity and reproductive health (except fertility) per birth in England[4]

The aim of this study was to develop multivariable prognostic models to identify potential causes of variability in the rates of infant and perinatal mortality between the 303 PCTs in England and to identify PCTs with worse than expected outcomes. Potential causes of variability between PCTs included population characteristics, such as ethnicity and deprivation, and health service funding for maternity services.

METHODS

Data sources

Included variables and their sources are described in the box. We only had partial information on parity and smoking so did not include these because of the potential for confounding by inclusion.

Statistical analysis

Following the general approach described by Harrell et al,[5] we developed prognostic models identifying predictors of infant or perinatal mortality in order to examine the extent to which these variables may account for observed variation between PCTs. We developed Poisson mixed models, in which the observed number of events was the response variable. See bmj.com.

The prespecified characteristics of the PCT were examined in the resulting statistical models. Backward model selection was used to derive separate models for infant mortality and perinatal mortality. The α level required to remain in the model was 5%.

Data from the three years were combined, apart from spending on maternity services which was available for each year included, so we examined the year on year effect of changes in spending on infant and perinatal mortality rates.

We used the models to derive predicted event rates for each PCT for infant mortality and perinatal mortality, and compared predicted and observed rates. Outliers worthy of further attention for both outcomes were defined as those PCTs for which the observed rate differed from the expected rate by more than three studentised residual errors.

For each final model we calculated the extent to which heterogeneity (extra-Poisson variability) between PCTs was "explained" by the included parameter terms. Model validation assessing potential optimism due to over-fitting was conducted using the bootstrap algorithm.[5]

RESULTS

Rates of infant and perinatal mortality and PCT level demographics

We identified data for all 303 PCTs in England. Spearhead status was designated in 88 (29%). Because of changing boundaries, data on deprivation status was not available for two PCTs. All other data were complete. There was substantial and striking variability across the PCTs (table 1).

We found no statistically significant predictive effect between yearly spending on outcome for either perinatal mortality or infant mortality, so all subsequent analyses were conducted using the three year period combined. See bmj.com.

PCT level predictors of infant mortality and perinatal mortality

Deprivation, Pakistani population, and maternal age <18 years were significant predictors of increased levels of infant mortality. The final fitted model explained 70.0% of the between-PCT heterogeneity in infant mortality.

Black ethnicity and deprivation were strong determinants of increased perinatal mortality, maternal age >35 years at birth was associated with decreased rates of perinatal mortality, and we observed a weak detrimental effect for the PCT level birth rates associated with Pakistani ethnicity. The final fitted model explained 80.5% of the between-PCT heterogeneity in perinatal mortality. See bmj.com.

Table 1 | Infant and perinatal mortality per 1000 live births, and candidate explanatory variables by primary care trust in England

Variable	Median (IQR)	Minimum	Maximum
Infant mortality per 1000 live births	4.81 (3.80-5.89)	1.40	10.83
Perinatal mortality per 1000 live births	7.81 (6.69-9.07)	3.93	16.66
Deprivation index	19.45 (13.34-28.20)	5.09	58.67
Maternity spend per birth (£)	4569 (3893-5217)	1897	8904
Birth rate by maternal age (%):			
<18 years	1.95 (1.32-2.86)	0.41	6.05
<16 years	0.16 (0.10-0.27)	0.00	0.75
>35 years	19.14 (15.01-23.29)	9.08	39.48
Birth rate by ethnicity (%):			
Mixed race	0.86 (0.56-1.42)	0.23	4.83
Black	0.38 (0.18-1.21)	0.02	25.90
Pakistani	0.18 (0.06-1.11)	0.01	40.76
Indian	0.51 (0.21-1.66)	0.04	38.02
White	96.83 (91.47-98.42)	29.13	99.45
Bangladeshi or other Asian	0.28 (0.13-0.75)	0.03	34.33
Chinese	0.28 (0.17-0.47)	0.07	2.25

IQR=interquartile range.

Ethnic categories: white = white British + white Irish + white other; mixed race = mixed white-Caribbean + mixed white-African + mixed white-Asian + mixed other; black = black Caribbean + black African + black Other.

Observed versus predicted infant and perinatal mortality by PCT

For infant mortality, no PCTs had observed rates that differed by more than three studentised residual errors from their predicted rate from the multivariable analysis (fig 1). For perinatal mortality, in two PCTs the observed rate was substantially higher than the expected rate (fig 2, table 2). Neither of the trusts with extreme results was categorised as Spearhead status.

Model validation

The models for infant and perinatal mortality were validated using the approach described by Harrell et al.[5] For infant mortality the estimate of model optimism derived from the bootstrap process was 4.6%, and for perinatal mortality it was 3.9%.

DISCUSSION

We developed prognostic models that used available data to predict differences in infant and perinatal mortality at the level of a PCT. The results from both models show clearly the importance of deprivation, ethnicity, and maternal age as risk factors. These results were achieved through the application of a parsimonious model fitting strategy designed to avoid optimism due to overfitting. The validation process we conducted confirmed that the level of optimism was low in each fitted model, indicating that the results are sound and may be generalisable.

Our analyses aimed to develop models to account for systematic variation (heterogeneity) in infant and perinatal mortality between PCTs. Caution is required in attempting to interpret the parameter estimates from the statistical models since relationships observed at the PCT level may not apply directly to individual subjects because of ecological confounding. Further, care is required in interpreting the relative risks provided, which are often subject to log transformation to aid model fit, and which relate to a 1 unit change in the candidate predictive variable. Since several of the variables are associated with very low birth rates in the population, the relative risks imply a very large difference in risk but are applied to a very low incident risk of events.

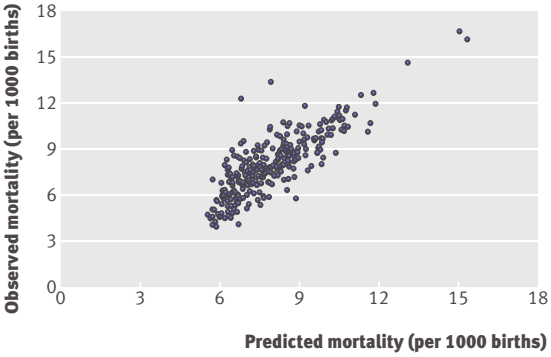


Fig 1 | Observed versus predicted infant mortality by primary care trust in England

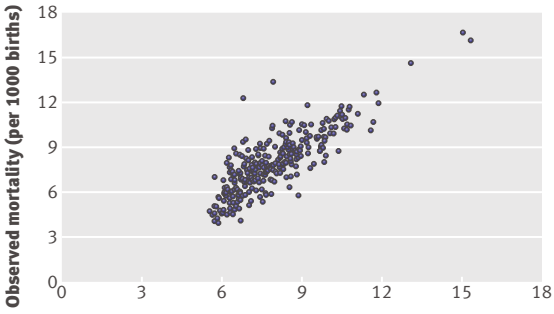


Fig 2 | Observed versus predicted perinatal mortality by primary care trust in England

The Index of Multiple Deprivation 2004[3] performed well in our study, on its own explaining 54% of the heterogeneity between PCTs in the infant mortality model and 57% of the heterogeneity for the perinatal mortality model.

We found no evidence of an effect of spending on maternity services by PCT for infant or perinatal mortality in repeated measures models across the three years of data or aggregated over all three years. Resources need to be directed to interventions which are effective; it is well established that determinants of infant mortality outside health services have a more profound effect than the provision of health care per se.[6]

The raised risk of infant and perinatal mortality among Pakistanis could be linked to consanguineous marriages,[7] but this remains controversial.[8 9] In common with other predictors included in the models, we cannot be certain whether the observed relationship is causal or the result of Pakistani ethnicity being related to another factor or factors not otherwise captured in the models. Other work based on individual subject data has noted an increased risk of infant and perinatal mortality among mothers of Pakistani origin.[10]

Except for two PCTs, both models show high predictive values and suggest that all trusts, including those with Spearhead status, had perinatal and infant outcomes consistent with the demographic composition of the communities they serve. Further local scrutiny is required in the case of these two apparently poorly performing trusts.

For the first time, estimated budgetary data on the amount PCTs spend on maternity and reproductive services were available and incorporated in the analysis. Programme budgeting is a developing tool for commissioning public health programmes and health services. It allows PCTs to compare expenditure and health outcomes in a systematic way.[11 12] See bmj.com.

The implications of this study are that national monitoring of Spearhead PCTs' performance against key health outcomes such as infant mortality that are not adjusted for key demographic factors (such as deprivation, ethnicity, and maternal age) are unlikely to be useful or fair. The concept of "added value,"

Table 2 | Details of primary care trusts in England where observed perinatal mortality differed by more than three studentised residual errors from predicted mortality

Trust name	Studentised residual deviance	Spearhead status	Total perinatal mortality		Perinatal mortality rate/ 1000 births	
			Observed	Predicted	Observed	Predicted
South Hams and West Devon PCT	3.98	No	34	18.83	12.29	6.81
Wyre Forest PCT	3.96	No	41	24.26	13.41	7.93

commonly used in the education sector to assess year on year improvements, may present a better approach.

Our study is original. Previous authors have described studies addressing related but discrete topics.[10 13-15]

Limitations of study

We undertook an ecological analysis of factors that predict infant and perinatal mortality, two related measures that include many of the same deaths. Caution is warranted about drawing any conclusions on causation because an association at the PCT level does not guarantee that the association will hold at the individual level. Furthermore, it may not be possible to assess the strength of the exposure-outcome relationship using ecological data. Our prior understanding of the likely risk factors for infant and perinatal mortality and the very strong and consistent effects of deprivation in the models makes it highly plausible that deprivation has a direct negative effect. The effect of maternal age, albeit strong statistically, is not in line with our prior understanding and thus may be considered likely to be confounded by other unmeasured factors aliased to that factor. Limitations with available data meant that we were not able to include mothers' smoking behaviour or parity in the statistical models. Estimated data on PCT spending on maternity services may be subject to some inaccuracy. Nevertheless, the results are valuable, enabling us to draw inferences about the experiences of whole communities and in doing so provide information on the level of avoidable deaths experienced within communities.

Contributors: See bmj.com.

Funding: Funded by the Birmingham Health and Wellbeing partnership.

Competing interests: None declared.

Ethical approval: Not needed.

1 Department of Health. *Health inequalities: revised list of Spearhead group primary care trusts*. www.dh.gov.uk/en/Publicationsandstatistics/Lettersandcirculars/Dearcolleagueletters/DH_4138963 (accessed 12 March 2008).
2 National Centre for Health Outcomes Development (NCHOD). www.nchod.nhs.uk/ (accessed 12 June 2007).
3 Noble M, Wright G, Dibben C, Smith GAN, McLennan D, Anttila C, et al. *Indices of deprivation 2004. Report to the Office of the Deputy Prime Minister*. London: Neighbourhood Renewal Unit, 2004.
4 Department of Health. *Programme budgeting category 18—maternity and reproductive health*. www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/Programmebudgeting/index.htm (accessed 12 March 2008).
5 Harrell FE Jr, Lee KL, Mark DB. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and measuring and reducing errors. *Stat Med* 1996;15:361-87.
6 Marmot M, Friel S, Bell R, Houweling TAJ, Taylor S. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet* 2008;372:1661-9.
7 Bundey S, Alam H, Kaur A, Mir S, Lancahire RJ. Why do UK-born Pakistani babies have high perinatal and neonatal mortality rates? *Paediat Perinat Ep* 1991;5:101-14.
8 Ahmad WIU. Reflections on the consanguinity and birth outcome debate. *J Public Health Med* 1994;16:423-8.
9 Alwan A, Modell B, Bittles, Czeizel A, Hamamy H. *Community control of genetic and congenital disorders*. Alexandria: World Health Organization, 1997.
10 Bakeo AC. Investigating variations in infant mortality in England and Wales by mother's country of birth, 1983-2001. *Paediatric and Perinatal Epidemiology* 2006;20:127-39.
11 Reddy UM, Ko CW, Willinger M. Maternal age and risk of stillbirth throughout pregnancy in the United States. *Am J Obstet Gynecol* 2006;195:764-70.
12 Brambley P, Jackson A, Muir Gray JA. Better allocation for better health and heathcare: The first annual population review. NHS National Knowledge Service. February 2007.
13 Adam T, Lim SS, Mehta S, Bhutta ZA, Fogstad H, Mathai M, et al. Cost effectiveness analysis of strategies for maternal and neonatal health in developing countries. *BMJ* 2005;331:1107.
14 Glinianaia SV, Rankin J, Bell R, Pearce MS, Parker L. Temporal changes in the distribution of population risk factors attenuate the reduction in perinatal mortality. *J Clin Epidemiol* 2005;58:1299-307.
15 Nixon J, Ulmann P. The relationship between health care expenditure and health outcomes: evidence and caveats for a causal link. *Eur J Health Econom* 2006;7:7-18.

Accepted: 11 March 2009

Tell your mum about it

I was a third year medical student experiencing hands-on clinical medicine for the first time. I started my placement in a district general hospital with a teaching session by a consultant surgeon. He had a fantastic collection of clinical slides, and the ultimate challenge was to describe the signs on the images.

Feeling daunted and not knowing where or how to start, I hesitated. The consultant said, "Describe it as you would be telling your mother about it on the phone."

I described exactly what I was seeing in simple terms, as if I was talking to my mother. With that, I started with descriptions of the basic lumps and bumps and thought to myself that it was not as bad as I had feared.

Years passed, with several clinical exams and finals, and I religiously thought of describing lumps and bumps with the site, size, shape, etc. The day of the clinical part of my MRCS came, and the terminology for describing lumps and bumps escaped me with the adrenaline rush, but out of the sky fell the words, "Tell your mum about it," and with that I was able to rattle off an accurate description and diagnosis of the lump I was asked to examine.

Today, I use the "Tell your mum" technique to teach medical students and to explain things to patients sometimes, to remind me to avoid medical jargon and explain things simply.

Maryam Alfa-Wali clinical research fellow, Chelsea and Westminster Hospital, London **fawalim@yahoo.com**

Cite this as: *BMJ* 2009;339:b2367

Quality of care in for-profit and not-for-profit nursing homes: systematic review and meta-analysis

Vikram R Comondore,[1] P J Devereaux,[2] Qi Zhou,[2] Samuel B Stone,[3] Jason W Busse,[2 4] Nikila C Ravindran,[5] Karen E Burns,[6 7] Ted Haines,[2] Bernadette Stringer,[2] Deborah J Cook,[2] Stephen D Walter,[2] Terrence Sullivan,[8] Otavio Berwanger,[9] Mohit Bhandari,[2] Sarfaraz Banglawala,[3] John N Lavis,[2] Brad Petrisor,[3] Holger Schünemann,[2 10] Katie Walsh,[2] Neera Bhatnagar,[11] Gordon H Guyatt[2]

EDITORIAL by Konetzka

[1]Department of Medicine, University of British Columbia, Vancouver, BC, Canada V5Z 1M9
[2]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada L8N 3Z5
[3]Department of Surgery, McMaster University
[4]The Institute for Work and Health, Toronto, ON, Canada M5G 2E9
[5]Department of Medicine, Division of Gastroenterology, University of Toronto, Toronto, M5T 2S8
[6]St Michael's Hospital, Toronto, M5B 1W8
[7]Keenan Research Centre and Li Ka Shing Knowledge Institute, Toronto, M5B 1W8
[8]Cancer Care Ontario, Toronto, M5G 2L7
[9]Department of Clinical Epidemiology, Federal University of Rio Grande do Sul, Porto Alegre-RS, Brazil
[10]Italian National Cancer Institute Regina Elena, Rome, 00144, Italy
[11]Health Sciences Library, McMaster University
Correspondence to: P J Devereaux philipj@mcmaster.ca

Cite this as: *BMJ* 2009;339:b2732
doi: 10.1136/bmj.b2732

ABSTRACT

Objective To compare quality of care in for-profit and not-for-profit nursing homes.

Design Systematic review and meta-analysis of observational studies and randomised controlled trials investigating quality of care in for-profit versus not-for-profit nursing homes.

Results A comprehensive search yielded 8827 citations, of which 956 were judged appropriate for full text review. Study characteristics and results of 82 articles that met inclusion criteria were summarised, and results for the four most frequently reported quality measures were pooled. Included studies reported results dating from 1965 to 2003. In 40 studies, all statistically significant comparisons ($P<0.05$) favoured not-for-profit facilities; in three studies, all statistically significant comparisons favoured for-profit facilities, and the remaining studies had less consistent findings. Meta-analyses suggested that not-for-profit facilities delivered higher quality care than did for-profit facilities for two of the four most frequently reported quality measures: more or higher quality staffing (ratio of effect 1.11, 95% confidence interval 1.07 to 1.14, $P<0.001$) and lower pressure ulcer prevalence (odds ratio 0.91, 95% confidence interval 0.83 to 0.98, $P=0.02$). Non-significant results favouring not-for-profit homes were found for the two other most frequently used measures: physical restraint use (odds ratio 0.93, 0.82 to 1.05, $P=0.25$) and fewer deficiencies in governmental regulatory assessments (ratio of effect 0.90, 0.78 to 1.04, $P=0.17$).

Conclusions This systematic review and meta-analysis of the evidence suggests that, on average, not-for-profit nursing homes deliver higher quality care than do for-profit nursing homes. Many factors may, however, influence this relation in the case of individual institutions.

WHAT IS ALREADY KNOWN ON THIS TOPIC

The quality and appropriateness of care delivered in nursing homes is a major concern for the public, policy makers, and media

Controversy exists about whether for-profit compared with not-for-profit ownership affects quality of care

WHAT THIS STUDY ADDS

Most studies suggest a trend towards higher quality care in not-for-profit facilities than in for-profit homes, but a large proportion of studies show no significant trend

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b2732

INTRODUCTION

Concerns about quality of care in nursing homes are widespread among academic investigators,[1-4] the lay press,[5-10] and policy makers.[11 12] Whether a facility is owned by a for-profit or a not-for-profit organisation may affect structure, process, and outcome determinants of quality of care. In the United States, two thirds of nursing homes are for-profit institutions; in the United Kingdom, more than half of healthcare beds belong to independent nursing homes for older people, most of which are operated by for-profit institutions.[13] In Europe, nursing home ownership patterns are evolving as nations with previously dominant public healthcare systems now seek privatisation.[14]

Several investigators have assessed the relation between for-profit/not-for-profit status and quality of care.[15] If quality or appropriateness of care varies significantly by ownership, this should influence government policies related to regulatory assessments and the use of public funds for nursing homes. The objective of this systematic review and meta-analysis was to examine the quality of care in for-profit and not-for-profit (privately and publicly owned) nursing homes.

METHODS

Search strategy

We searched 18 bibliographical databases, personal files, PubMed, and SciSearch; reviewed references; and consulted with experts. We searched databases from inception to April 2006. Search terms included nursing home specific terms combined with ownership terms.

Study selection

Our inclusion criteria were: patients—those residing in nursing homes in any jurisdiction; intervention—for-profit status of the institutions; comparator—not-for-profit status; outcomes—measures of quality of care in for-profit and not-for-profit nursing homes.

Many quality of care instruments have been proposed, although none has been universally accepted.[16] We accepted any quality of care measure defined as representing "quality of care" or "appropriateness of care" and evaluated the following commonly used measures separately. (1) Number of staff per resident or level of training of staff—

Results of testing of a priori hypotheses to explain heterogeneity

Outcome	Summary study characteristics	Interaction P value: FP-NFP *v* FP-private NFP	Above median *v* below median appropriate adjustment score	Presence *v* absence of inappropriate adjustment, among studies with adjusted analysis	Data collection before or during 1987 *v* after 1987
More extensively trained staff or more staff	13 studies had poolable data, from 1971-2002; 3 removed for data overlap; 10 meta-analysed—4 collected data after 1987, 1 used primary data, 1 had data from Canada (remainder from United States)	0.64 for FP-private NFP; ratio of effect sizes 1.09 (95% CI 1.07 to 1.12, P<0.001, I^2=0%)	0.15	0.99	0.66
Lower pressure ulcer prevalence	16 studies had poolable data, from 1987-2003; 5 removed for data overlap; 11 meta-analysed—2 used primary data, 1 had data from Canada (remainder from United States)	0.76 for FP-private NFP comparison; ratio of effect sizes 0.89 (0.82 to 0.98, P=0.02, I^2=39.3%)	0.42	0.54	All meta-analysed data collected after 1987
Lower physical restraint prevalence	13 studies had poolable data from 1987-2003; 5 removed for data overlap; 8 meta-analysed—1 used primary data	0.84 for FP-private NFP comparison; ratio of effect sizes 0.94 (0.78 to 1.14, P=0.53, I^2=84.9%)	0.86	0.13	All meta-analysed data collected after 1987
Fewer deficiencies on government surveys	13 studies had poolable data from 1976-2003; 6 removed for data overlap; 7 meta-analysed—2 collected data before 1987	0.56 for FP-private NFP comparison; ratio of effect sizes 0.92 (0.79 to 1.06, P=0.25, I^2=63.1%)	0.80	0.54	0.11; for data collected after 1987, pooled effect size 0.73 (95% CI 0.54 to 0.97, P=0.03, I^2=67.9%) favouring NFP homes; for data collected before or during 1987, pooled effect size 1.09 (0.94 to 1.25, P=0.25, I^2=0%)

FP=for-profit; NFP=not-for-profit.

studies have consistently shown a positive association between staffing and measures of quality.[17-19] (2) Physical restraints—although use of physical restraints can prevent patients from injuring themselves, restraints diminish a patient's self esteem and dignity. (3) Pressure ulcers—these are preventable and are associated with pain and risk of infection. (4) Regulatory (government survey) deficiencies—deficiency citations by a regulatory body cover many aspects of nursing home care; their strength lies in providing an overall measure of quality. We defined a nursing home as a home for elderly people in which most residents need daily nursing care.

We screened the titles and abstracts of all citations, and retrieved eligible studies for full text review. With study results masked we evaluated each article to determine eligibility.

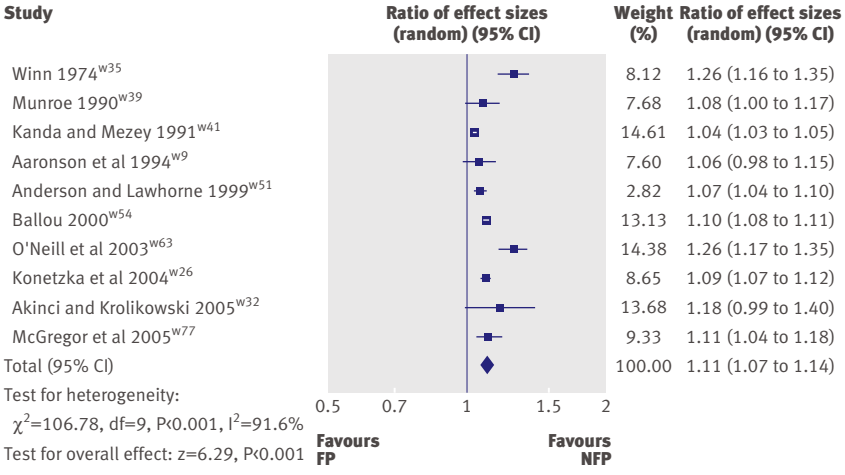


Fig 1 | Ratio of effect sizes for staffing quality in for-profit (FP) and not-for-profit (NFP) nursing homes. Ratios listed represent effect size in NFP homes compared with that in FP homes. Ratio>1 indicates that NFP homes had more, or higher quality, staffing (that is, favours NFP)

Data extraction and study quality evaluation

We collected data on geographical area, year, data source, unit of measurement, and quality of care measure. We developed and applied a 0-5 scale for evaluating appropriate adjustments and a yes/no scale for inappropriate adjustments. We explored whether appropriate and inappropriate adjustment explained heterogeneity.

Statistical analysis

We classified studies into three categories. (1) "All statistically significant differences favour one ownership type"—studies fulfilled two requirements: at least one outcome with P<0.05 favours either for-profit or not-for-profit and all outcomes with P<0.05 favour the same funding structure. (2) "Most but not all significant differences favour one ownership type"—studies fulfilled two requirements: at least four quality measures have P<0.05 and three times as many outcomes with P<0.05 favour one ownership as favour the other. (3) "Mixed results"—all other results.

We pooled outcomes separately for the most frequently used quality of care measures: number of staff or level of training of staff, pressure ulcers, physical restraints, and regulatory deficiencies. We converted effect measures to odds ratios. We avoided repetition of data on the same resident from different studies by preferentially using data from the larger dataset when necessary. We examined funnel plots for evidence of publication bias. We applied a meta-regression to each pooled outcome to evaluate potential sources of heterogeneity.

Hypotheses to explain heterogeneity

Our a priori hypotheses for sources of potential heterogeneity included analysis of privately owned and publicly owned nursing facilities in the same

category, appropriate and inappropriate adjustments, the year of data collection, geography and political environment, and primary compared with secondary data collection. We did meta-regression for each potential cause of heterogeneity.

RESULTS

Of the 8827 articles screened, we selected 956 for blinded full text review. We found 82 studies, spanning 1965 to 2003, comparing for-profit and not-for-profit nursing homes.[w1-w82] We found 40 studies in which all statistically significant analyses (P<0.05) favoured not-for-profit homes and three in which all statistically significant analyses favoured for-profit homes. Similarly, 34 studies compared for-profit and privately owned not-for-profit nursing homes. In 16 of these, all statistically significant comparisons favoured higher quality in privately owned not-for-profit homes; none had all statistically significant analyses favouring higher quality in for-profit homes (see bmj.com).

We meta-analysed data for the four most commonly used quality measures. The table presents a summary of the characteristics of studies meta-analysed, along with the results of sensitivity analyses to explain heterogeneity among studies in each meta-analysis. Two meta-analyses showed statistically significant results favouring higher quality care in not-for-profit nursing homes.

We found more or higher quality staffing in not-for-profit homes (ratio of effect 1.11, 95% confidence interval 1.07 to 1.14, P<0.001, I^2=91.6%) (fig 1). We found a similar result favouring not-for-profit homes when assessing staffing hours alone, with a ratio of effect of 1.11 (1.08 to 1.14, P<0.001, I^2=70.3%), an absolute increase in hours of 0.42 (0.31 to 0.53) hours/resident/bed/day, and a relative increase in hours of 11% (8% to 14%). We found a lower prevalence of pressure ulcers in not-for-profit homes (odds ratio 0.91, 95% confidence interval 0.83 to 0.98, P=0.02, I^2=52.1%), with an absolute risk reduction of 0.59% (0.13% to 1.12%) and a relative risk reduction of 8.4% (1.9% to 16%) (fig 2).

The remaining two meta-analyses showed non-statistically significant differences. We found less use of physical restraints in not-for-profit homes (odds ratio 0.93, 0.82 to 1.05, P=0.25, I^2=74.6%) and fewer deficiencies in governmental regulatory assessments in not-for-profit homes (ratio of effect 0.90, 0.78 to 1.04, P=0.17, I^2=59.8%) (see bmj.com).

Funnel plots for the four meta-analyses did not suggest publication bias. A priori hypotheses did not explain the observed heterogeneity (table).

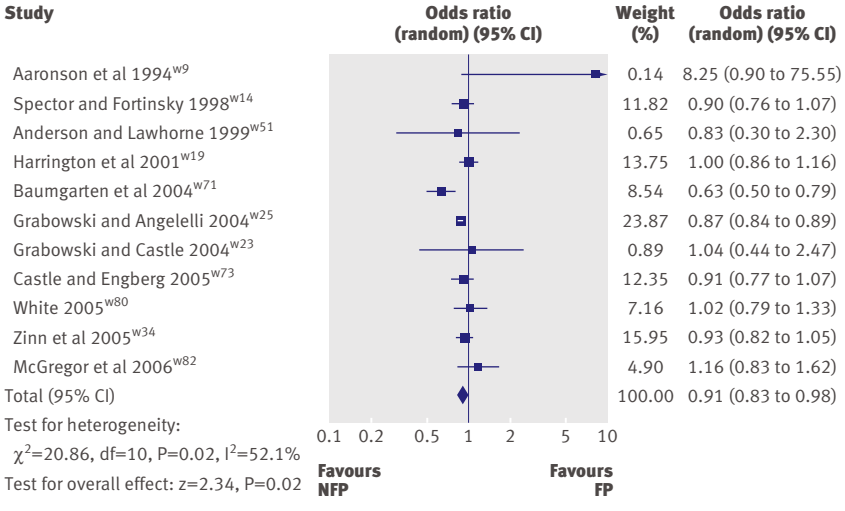


Fig 2 | Odds ratios (OR) comparing pressure ulcer prevalence in for-profit (FP) and not-for-profit (NFP) nursing homes. OR<1 indicates lower risk of pressure ulcers in NFP facilities than in FP facilities, suggesting that NFP facilities deliver higher quality care

DISCUSSION

Our systematic review identified 82 studies comparing quality of care in for-profit and not-for-profit nursing homes. More studies had all statistically significant analyses showing higher quality in not-for-profit nursing homes than in for-profit nursing homes. Many studies, however, showed no significant differences in quality by ownership, and a small number showed statistically significant differences in favour of for-profit homes. This pattern held true when we compared for-profit homes with both privately owned and publicly owned not-for-profit facilities. Pooled analyses of the four most commonly used quality measures showed statistically significant results favouring higher quality care in not-for-profit homes for staffing and prevalence of pressure ulcers and non-statistically significant differences favouring not-for-profit homes in physical restraint use and regulatory agency deficiencies. The large observed heterogeneity was not explained by our a priori hypotheses.

Previous systematic reviews

Two previous systematic reviews have compared quality of care in for-profit and not-for-profit nursing homes. In 1991 Davis and colleagues found that many studies showed that higher quality of care was provided in not-for-profit nursing homes; however, methodological weaknesses in the included studies limited the conclusions that could be drawn.[20] In 2002 Hillmer and colleagues also concluded that not-for-profit facilities provided better quality care than for-profit facilities.[21]

Strengths and weaknesses of this review

We did a comprehensive search, which identified 60 studies not included in previous reviews. We assessed studies spanning four decades. We compared quality of care in both for-profit versus not-for-profit nursing homes and for-profit versus privately owned not-for-profit nursing homes, did pooled analyses of quality of care measures, and found largely consistent results.

Our review has limitations resulting from the characteristics of the studies included. No randomised trials have compared quality of care across nursing home ownership. Most studies are from the United

States, which raises questions of generalisability. Studies are also limited in that no standard definition of quality of care exists. Even when the same measures were used, standardised approaches were lacking. Several eligible studies used administrative databases, which further limits the comprehensiveness and quality of the data.

Heterogeneity

The gradient between studies in which all significant measures favoured not-for-profit (40 studies) and those in which all measures favoured for-profit (3) is large. All four meta-analyses favoured not-for-profit institutions, and two reached statistical significance. However, 37 studies had mixed results and considerable heterogeneity was present in the results of the meta-analyses. This suggests that although the average effect is clear, that effect probably varies substantially across situations. The variability is probably explained, in part, by factors that vary within categories of for-profit and not-for-profit homes, including management styles, motivations, and organisational behaviour. For example, for-profit facilities owned and operated by investor owned corporations versus small private businesses or single proprietors. We have partially mitigated this problem with our a priori hypotheses. None of these hypotheses, however, explained the substantial heterogeneity of our results.

Significance of this study

In the long term care market, in which funding is often provided by the government at fixed rates, both for-profit and not-for-profit facilities face an economic challenge. For-profit nursing homes have a strong incentive to minimise expenditures.[22] Minimising expenditures may lead to lower quality staffing and higher rates of adverse events, which may be reflected in citations for deficiency.

Our results are based on observational studies, which cannot demonstrate causality. Furthermore, given their variability, the results do not imply a blanket judgment of all institutions.

Further research and conclusions

Additional work is needed to compare the costs between for-profit and not-for-profit facilities and to evaluate the consistency of these findings outside of the United States and Canada. The available studies did not allow comparison of subcategory of for-profit ownership. Nursing home management companies further complicate the relation between ownership and quality of care. These all warrant further research.

We acknowledge the outstanding work of Deborah Maddock, Denise Healey, Shelley Anderson, Michelle Murray, Monica Owen, and Laurel Grainger who coordinated this study. We thank our foreign article reviewers Janek Brozek, Matthias Briel, Toshi Furukawa, Marjuka Makela, Ben de Mol, Paola Muti, Patricia Smith, Kristian Thorlund, and David Wei. We appreciate the work of Dana Keilty, Navneet Binepal, Tony Soeyonggo, and Minji Kim, who blinded articles for us. We thank Christina Lacchetti, Michael Levy, and Rajesh Hiralal, who reviewed articles for us, and Diane Heels-Ansdell for her statistical help. We also thank the authors of included studies who did additional analyses for our systematic review: Chappin White, Robert Weech-Maldonado, and Ann L Gruber-Baldini.

Contributors: See bmj.com.

Funding: Atkinson Foundation Grant; the study sponsor did not contribute to the study design. JWB is funded by a Canadian Institutes of Health research fellowship award. DJC, MB and JNL are supported, in part, by their respective Canada Research chairs. PJD is supported by a Canadian Institutes of Health Research new investigator award. HS is funded by a European Commission: The Human Factor, Mobility and Marie Curie Actions scientist reintegration grant (IGR 42192).

Competing interests: None declared.

Ethical approval: Not needed.

1 Georgiades S. Quality nursing home care in Cyprus: are elder residents content with their treatment? *J Gerontol Soc Work* 2008;50(3-4):3-24.
2 Konetzka RT, Stearns SC, Park J. The staffing-outcomes relationship in nursing homes. *Health Serv Res* 2008;43:1025-42.
3 Mukamel DB, Glance LG, Li Y, Weimer DL, Spector WD, Zinn JS, et al. Does risk adjustment of the CMS quality measures for nursing homes matter? *Med Care* 2008;46:532-41.
4 Castle NG, Liu D, Engberg J. The association of Nursing Home Compare quality measures with competition and occupancy rates. *J Healthc Qual* 2008;30(2):4-14.
5 BBC News. Nursing home residents 'in pain'. 2007. http://news.bbc.co.uk/2/hi/health/6471485.stm.
6 Welsh M, Benzie R. No minimum standard for nursing care. 2008. www.thestar.com/News/Ontario/article/445152.
7 The Toronto Star. Troubling report on nursing homes. 2008. www.thestar.com/article/453994.
8 Covello L. Nursing home quality: more hype than hazard? 2008. www.foxbusiness.com/portal/site/fb/menuitem.5b2f8f9bb693bd972f08aa8738d48a0c/?vgnextoid=ac387b770dbba110VgnVCM10000086c1a8c0RCRD&redirected=true.
9 Duhigg C. At many homes, more profit and less nursing. 2007. www.nytimes.com/2007/09/23/business/23nursing.html?_r=1&pagewanted=all&oref=slogin#.
10 CBC News. Another seniors home in B.C. not up to standard, NDP says. 2007. www.cbc.ca/canada/british-columbia/story/2007/10/04/bc-senior.html.
11 European Commission. Long-term care in the European Union. 2008. http://ec.europa.eu/employment_social/spsi/docs/social_protection/2008/hc_ltc_en.pdf.
12 US Department of Health and Human Services. Nursing home compare. www.medicare.gov/NHCompare.
13 Kerrison SH, Pollock AM. Care for older people in the private sector in England. *BMJ* 2001;323:566-9.
14 Tyszko P, Wierzba WM, Kanecki K, Ziolkowska A. Transformation of the ownership structure in Polish Healthcare and its effects. *Cent Eur J Med* 2007;2:528-38.
15 Shapiro E, Tate RB. Monitoring the outcomes of quality of care in nursing homes using administrative data. *Can J Aging* 1995;14:755-68.
16 Mukamel DB. Risk-adjusted outcome measures and quality of care in nursing homes. *Medical Care* 1997;35(4):367-85.
17 Abt Associates. *Appropriateness of minimum nurse staffing ratios in nursing homes: report to Congress*. (Phase II final, volume 1.) Baltimore, MD: Centres for Medicare and Medicaid Services, 2001 (available from: www.allhealth.org/BriefingMaterials/Abt-NurseStaffingRatios(12-01)-999.pdf).
18 Castle NG, Engberg J. The influence of staffing characteristics on quality of care in nursing homes. *Health Serv Res* 2007;42:1822-47.
19 Hutt E, Radcliffe TA, Liebrecht D, Fish R, McNulty M, Kramer AM. Associations among nurse and certified nursing assistant hours per resident per day and adherence to guidelines for treating nursing home acquired-pneumonia. *J Gerontol A Biol Sci Med Sci* 2008;63:1105-11.
20 Davis MA. On nursing home quality: a review and analysis. *Med Care Rev* 1991;48:129-68.
21 Hillmer MP, Wodchis WP, Gill SS, Anderson GM, Rochon PA. Nursing home profit status and quality of care: is there any evidence of an association? *Med Care Res Rev* 2005;62:139-66.
22 Friedman BS, Hattis PA, Bogue RJ. Tax exemption and community benefits of not for-profit hospitals. *Adv Health Econ Health Serv Res* 1990;11:131-58.

Accepted: 21 April 2009

Comparison of direct and indirect methods of estimating health state utilities for resource allocation: review and empirical analysis

David Arnold,[1] Alan Girling,[2] Andrew Stevens,[2] Richard Lilford[2]

See also ANALYSIS, p 371

[1]Cardiff University Medical School, Cardiff
[2]School of Health and Population Sciences, University of Birmingham, Birmingham B15 2TT
Correspondence to: R Lilford
r.j.lilford@bham.ac.uk

Cite this as: *BMJ* 2009;339:b2688
doi: 10.1136/bmj.b2688

ABSTRACT

Background and objective Utilities (values representing preferences) for healthcare priority setting are typically obtained indirectly by asking patients to fill in a quality of life questionnaire and then converting the results to a utility using population values. We compared such utilities with those obtained directly from patients or the public.

Design Review of studies providing both a direct and indirect utility estimate.

Selection criteria Papers reporting comparisons of utilities obtained directly (standard gamble or time trade off) or indirectly (European quality of life 5D [EQ-5D], short form 6D [SF-6D], or health utilities index [HUI]) from the same patient.

Data sources PubMed and Tufts database of utilities.

Statistical methods Sign test for paired comparisons between direct and indirect utilities; least squares regression to describe average relations between the different methods.

Main outcome measures Mean utility scores (or median if means unavailable) for each method, and differences in mean (median) scores between direct and indirect methods.

Results We found 32 studies yielding 83 instances where direct and indirect methods could be compared for health states experienced by adults. The direct methods used were standard gamble in 57 cases and time trade off in 60 (34 used both); the indirect methods were EQ-5D (67 cases), SF-6D (13), HUI-2 (5), and HUI-3 (37). Mean utility values were 0.81 (standard gamble) and 0.77 (time trade off) for the direct methods; for the indirect methods: 0.59 (EQ-5D), 0.63 (SF-6D), 0.75 (HUI-2) and 0.68 (HUI-3).

Discussion Direct methods of estimating utilities tend to result in higher health ratings than the more widely used indirect methods, and the difference can be substantial. Use of indirect methods could have important implications for decisions about resource allocation: for example, non-lifesaving treatments are relatively more favoured in comparison with lifesaving interventions than when using direct methods.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Health state utilities play a crucial role in the allocation of health care resources

Utilities may be obtained directly (usually from patients) or, more often, indirectly, by using a quality of life questionnaire, the results of which are converted to utilities using "weights" (tariffs) obtained from the general public

Different direct and indirect methods yield different utility values

WHAT THIS STUDY ADDS

Indirect methods as a group produce consistently lower utilities (worse recorded health) than the direct group of methods

This difference may be larger than the differences between methods within each group

Reliance on indirect methods will result in less resources being allocated to life saving treatments than if direct methods were used

Conversion of indirect utilities to direct utilities is only partly successful

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b2688

INTRODUCTION

For resources to be allocated fairly, according to benefit gained per unit cost, health economists and policy makers use a common currency of benefit. This standardisation is done by attaching different utilities to different health states.[1] Utilities are captured on a scale where 1 represents perfect health, 0 represents death, and states worse than death have negative values.

Measurement of utilities is a controversial area. Two groups of methods exist (see bmj.com). The first is based on mapping preferences directly onto the utility scale. This can be done by means of a trade off (standard gamble or time trade off) or visual analogue scale.[2 3] We refer to these as direct measures of utility.

The second is based on mapping preferences onto the utility scale indirectly via a generic health related quality of life questionnaire. Questionnaire responses are converted to utilities by means of "tariffs" or "weights" derived from previous exercises in which possible health states have been calibrated through a trade off method from a sample of the general population. We refer to this group of methods as indirect methods of utility measurement.[4-7]

Different methods of utility estimation yield systematically different values.[8-14] Standard gamble, time trade off, and visual analogue scale have all been compared across studies in a systematic way.[15-21] However, the relation between direct and indirect utility measures has not been systematically documented. There is a widely held impression among health economists that direct methods tend to yield higher utilities (reflecting better reported health) for given health states than do indirect methods. The aims of the current study were to assess whether direct methods yield higher utility values than indirect methods; to quantify the magnitude of any such differences; and to describe the relation between direct and indirect measures.

METHODS

Literature search

We searched for papers in PubMed (covering all diseases) that mentioned a direct and an indirect method in the title or abstract. We identified those in which

Table 1 | Distribution of preference values across all disease states assessed by included studies

	Current patients*				Hypothetical patients				
	Studies	States	Range	Mean (SD)	Studies	States	Range	Mean (SD)	Overall mean (SD)
Time trade off	19	50	0.52, 0.99	0.83 (0.10)	2	10	−0.17, 0.77	0.45 (0.27)	0.77 (0.20)
Standard gamble	16	46	0.45, 1.00	0.83 (0.12)	2	11	0.47, 0.90	0.70 (0.14)	0.81 (0.14)
EQ-5D	19	52	−0.01, 0.93	0.65 (0.15)	3	15	−0.52, 0.81	0.35 (0.34)	0.59 (0.24)
SF-6D	6	13	0.55, 0.69	0.63 (0.04)	—	—	—	—	0.63 (0.04)
HUI-2	2	5	0.55, 0.95	0.75 (0.17)	—	—	—	—	0.75 (0.17)
HUI-3	10	37	0.34, 0.89	0.68 (0.11)	—	—	—	—	0.68 (0.11)

Each disease state is associated with just one study. The statistics in each row (mean, SD, range) refer to the sample of disease states.
*The majority of studies reported mean utility values across samples of patients. Two studies of current patients reported only median utilities. These have been included in the table nevertheless.

the same group of patients had contributed both direct and indirect utilities. We reviewed the Tufts database of utilities[22] for any studies that compared the utilities within the same group of respondents.

Data extraction

We extracted information about the disease topic, methods of elicitation of direct utilities (standard gamble, time trade off, or both); generic questionnaires used (EQ-5D, SF-6D, HUI-2, HUI-3); and mean or median utility values for each method. Each study was classified according to whether the respondents were patients ("current patients") or were asked to imagine the experience of the condition ("hypothetical patients").

Statistical methods

The sign test was used for paired comparisons of direct and indirect methods. Average relations between direct and indirect utilities were fitted using least squares regression lines. In each case the scores from the direct method were regressed on the scores from the indirect method, to test the hypothesis that direct utilities could be predicted by applying a linear correction to the indirect utilities. We assessed the predictive value of the fitted lines, allowing for between-subject variation in both utilities.

RESULTS

The search returned 32 studies. Four of the 32 studies were about health states experienced by children and were excluded. This left 28 studies covering a wide range of diseases and encompassing data from 4688 respondents. Altogether there were 83 instances in which direct and indirect methods could be compared. Sixty-eight of the 83 comparisons (from 25 studies) were based on current patients, and 15 (from three studies) were based on hypothetical patients.

For direct utilities, standard gamble alone was used in 23 of the 83 comparisons, time trade off alone in 26, and both in 34. The most popular indirect methods were EQ-5D (n=67) and HUI-3 (n=37), one or both of which figured in all but four of the 83 comparisons.

The utility values reported by the individual studies are averages over samples of respondents, with sample sizes ranging from three to 1011 (median 62). Table 1 summarises the distribution of the reported mean or median utility values.

More detailed analyses were undertaken for current and hypothetical comparisons that reported EQ-5D, HUI-3, or both. The direct methods (time trade off, standard gamble) produced significantly higher values than the indirect methods (EQ-5D, HUI-3) in every case where a statistical comparison was feasible (table 2). By contrast, the differences in utility values between each of the direct methods and between each of the indirect methods were not statistically significant.

The discrepancy between individual direct and indirect measures is reflected in the figure. If direct and indirect methods gave the same results, then the points would be distributed equally above and below the 45° line in each panel. The great majority of points in all panels, however, fell above this line. In each panel, the broken line represents the predicted direct utility score from a regression on the indirect score, as computed from the "current patient" comparisons. Table 2 shows the slopes for these lines and those based on hypothetical comparisons. The lines represent average relations only, with statistically significant ($P<0.05$) departures from the line in all but one instance, which was based on very low sample numbers. This finding suggests that the variation between participants within studies was not sufficient to account for the discrepancies between the plotted points and the fitted lines.

Table 2 | Pairwise comparisons between and within direct and indirect methods

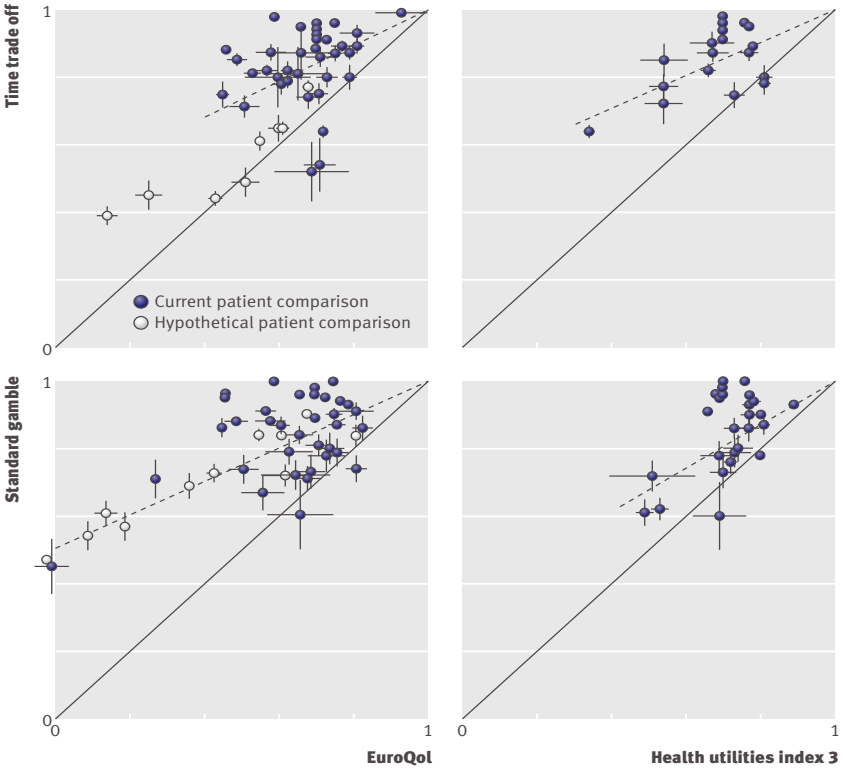
Comparison	Number of states	Mean difference (SD)	Number of independent groups	Sign test	Regression slope (SE)	Predictive P value
Current patients						
Time trade off – EQ-5D	40	0.17 (0.13)	26	P<0.001	0.54 (0.07)	P<0.001
Standard gamble – EQ-5D	39	0.20 (0.15)	27	P<0.001	0.50 (0.05)	P<0.001
Time trade off – HUI-3	24	0.19 (0.09)	13	P=0.023	0.48 (0.06)	P<0.003
Standard gamble – HUI-3	29	0.15 (0.11)	16	P=0.004	0.65 (0.08)	P<0.002
Standard gamble – Time trade off	28	0.02 (0.06)	16	P=0.804	—	—
HUI-3 – EQ-5D	25	0.03 (0.13)	14	P=0.424	—	—
Hypothetical patients						
Time trade off – EQ-5D	10	0.13 (0.12)	2	NA	0.81 (0.01)	P<0.994
Standard gamble – EQ-5D	11	0.15 (0.11)	2	NA	0.50 (0.06)	P<0.042
Standard gamble – Time trade off	6	0.22 (0.05)	1	NA	—	—

NA=not applicable. Second column shows numbers of states contributing to mean and SD of the difference.
Sign tests and regression analyses are based on aggregate health states obtained from averaging data within independent groups of participants. Regression slopes refer to least squares lines for predicting direct utility from indirect utility, constrained to pass through the point (1,1). Predictive P-values assess the goodness-of-fit of the lines, taking variation within studies into account.

DISCUSSION

Principal findings

The versatile and convenient indirect methods of utility measurement yield different results from those obtained by the direct methods. The indirect methods yielded



Direct utilities against indirect utilities. Plotted points are means (if available) or medians from health-states within 28 studies. Vertical and horizontal lines represent standard errors cited (or deduced) within the studies. Broken lines are regressions of direct or indirect utilities from current patient comparisons. In top left panel, one (hypothetical) point lies off the scale, with EQ-5D=−0.52, time trade off=−0.17

systematically lower utility values than direct methods for a wide range of diseases. The differences in utility values between the direct and indirect groups of methods were sometimes substantial, and tended to be greater than the differences between the individual methods making up the group.

Limitations

As with all systematic reviews, the results are constrained by the published literature. The studies covered many different diseases and several different methods of utility elicitation allowing a general trend to be identified. However, the heterogeneity of methods and disease states precludes any generalisable summary of effect size that would apply to a different spectrum of patients or methods.

Explaining the findings

Our first potential explanation is that the generic questionnaires used to obtain the indirect utilities impose constraints: respondents are forced to encapsulate their potentially complex condition within five to eight categories. The questionnaires do not allow respondents to report, for example, potentially positive aspects of their situations that would boost utility values.

Our second hypothesis is that the respondents who contribute trade off values for indirect utility elicitation are different to the patients that participate in direct methods. The general population used to obtain tariffs for indirect utility estimation is spread across a wider age range than patient populations typically used in direct estimation. Young people in good health and people with diseases may have very different perspectives on the relative merits of remaining in a given health state and making a trade off involving death.

Those participating in elicitation of indirect tariffs are always asked to make a hypothetical choice whereas direct utilities are usually based on the experience of people who actually have the condition. However, the difference between direct and indirect utilities seems to remain even when direct values are elicited from a sample of hypothetical patients.

Finally, the results may respond to suggested methodological improvements in deriving indirect utilities from questionnaire responses, which would tend to increase the indirect utility attached to mild health states.[23]

Direct or indirect?

There is no universally accepted theoretical basis for choosing direct or indirect methods.[1] Some people think that utilities should be derived from patients who really know what the condition is like. Others think that the citizen's perspective is more relevant, as a locus for a decision about use of society's resources, and because citizens can be asked to factor in societal objectives, under conditions of uncertainty that more closely conform to "the axioms of utility theory". However, a preference for direct or indirect utilities does not necessarily result from these considerations: direct utilities derived hypothetically from citizens and indirect utilities calibrated through a survey of the general population can both provide the advantages of the citizen's perspective.

Implications for resource allocation

It seems that indirect methods give consistently lower levels of utility than direct methods. This means that there is more headroom for utility gain with indirect methods. The utility of death, however, is fixed at zero. Thus, in comparison with direct methods, indirect methods will favour the allocation of healthcare resources away from interventions that prevent or delay death in favour of those that alleviate non-fatal conditions. It could be argued that the popularity of indirect methods for informing rationing decisions simply expresses a legitimate societal attitude in favour of non-lethal over lethal conditions. However, it could equally be argued that the public would rather give more weight to delaying death. In that case indirect methods might risk undervaluing both personal and societal preferences.

Implications for decision makers

Those who prefer direct methods, but who wish to exploit the convenience of indirect methods, might propose using a correction to map indirect utilities onto the putatively more valid direct utilities. Our results show that linear adjustments of indirect utilities can achieve only a partial conversion to direct utility scores.

This paper adds weight to the recommendation to be cautious when using utilities of any type. In the construction of health economic models it may be prudent

to extend the range of uncertainty beyond the confines of statistical confidence limits and conduct a sensitivity analysis.

Because direct and indirect methods can lead to noticeable differences in elicited utilities, priority setting institutions should avoid using a mixture of methods for different decisions, otherwise a motivated choice of method might be used to distort the outcome in a preferred direction.

We thank Stirling Bryan and Tracey Roberts for helpful discussion. We also thank the referees Steven McPhail and Cindy Lam for their detailed and perceptive comments.

Contributors: RL conceived the idea for the paper, DA performed the literature reviews and data extraction, and AG repeated the data extraction and carried out statistical analyses. All authors prepared the manuscript and RL is guarantor.

Funding: This work was supported through the MATCH programme (EPSRC Grant GR/S29874/01) although the views expressed are entirely the authors' own.

Competing interests: None declared.

Ethical approval: Not required.

1 Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. Oxford: Oxford University Press, 1996.
2 Stiggelbout AM, Eijkemans MJ, Kiebert GM, Kievit J, Leer JW, De Haes HJ. The 'utility' of the visual analog scale in medical decision making and technology assessment. Is it an alternative to the time trade-off? *Int J Technol Assess Health Care* 1996;12:291-8.
3 Torrance GW, Feeny D, Furlong W. Visual analog scales: do they have a role in the measurement of preferences for health states? *Med Decis Making* 2001;21:329-34.
4 Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ* 1986;5:1-30.
5 Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;21:271-92.
6 Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes* 2003;1:54.
7 EuroQol group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
8 Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;14:1523-32.
9 Boonen A, van der HD, Landewe R, van TA, Mielants H, Dougados M, et al. How do the EQ-5D, SF-6D and the well-being rating scale compare in patients with ankylosing spondylitis? *Ann Rheum Dis* 2007;66:771-7.
10 Macran S, Weatherly H, Kind P. Measuring population health: a comparison of three generic health status measures. *Med Care* 2003;41:218-31.
11 Grutters JP, Joore MA, van der HF, Verschuure H, Dreschler WA, Anteunis LJ. Choosing between measures: comparison of EQ-5D, HUI2 and HUI3 in persons with hearing complaints. *Qual Life Res* 2007;16:1439-49.
12 van Stel HF, Buskens E. Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease. *Health Qual Life Outcomes* 2006;4:20.
13 Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004;13:873-84.
14 Bryan S, Longworth L. Measuring health-related utility: why the disparity between EQ-5D and SF-6D? *Eur J Health Econ* 2005;6:253-60.
15 Robinson A, Dolan P, Williams A. Valuing health status using VAS and TTO: what lies behind the numbers? *Soc Sci Med* 1997;45:1289-97.
16 Lamers LM, Stalmeier PF, Krabbe PF, Busschbach JJ. Inconsistencies in TTO and VAS values for EQ-5D health states. *Med Decis Making* 2006;26:173-81.
17 Badia X, Monserrat S, Roset M, Herdman M. Feasibility, validity and test-retest reliability of scaling methods for health states: the visual analogue scale and the time trade-off. *Qual Life Res* 1999;8:303-10.
18 Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *J Health Econ* 1997;16:155-75.
19 Puhan MA, Schunemann HJ, Wong E, Griffith L, Guyatt GH. The standard gamble showed better construct validity than the time trade-off. *J Clin Epidemiol* 2007;60:1029-33.
20 Martin AJ, Glasziou PP, Simes RJ, Lumley T. A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *Int J Technol Assess Health Care* 2000;16:137-47.
21 Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ* 2002;11:447-56.
22 Tufts Medical Center. Center for the evaluation of value and risk in health.
2008. https://research.tufts-nemc.org/cear/search/search.aspx.
23 Craig B, Busschbach J. The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. *Population Health Metrics* 2009;7:3.

Accepted: 18 February 2009

Not good enough

We had just arrived in France for a week's skiing. I checked my email in the hotel lobby, and there it was—the outcome of my specialty trainee level 3 (ST3) interview. "You were well below the standard we expect of registrars in this deanery. You will not proceed to ST3 but instead will go onto an ST2-3 transition post." Suddenly, my guaranteed run-through training had suffered a year long hiccup.

Although I'm not one of those people who always knew they had to be an orthopaedic surgeon, I did always want to be a doctor. I got into my first choice of medical school. I achieved the degree I aimed for. I secured my first choice junior medical and surgical jobs. I had even, by luck, revised the right topics for the Membership of the Royal College of Surgeons examination. My medical career was progressing smoothly, and arrogantly I thought the next step was mine for the taking. As I trudged off to the slopes, my pride was hurt, my ego dented, and I was angry and upset.

A week of exercise in the fresh air is a great medicine, and when I returned to work I felt entirely different. I was going to grasp the opportunity of this extra training and seek out every chance to develop my experience. I would make certain that no one could say I wasn't ready for ST3 in August 2009. Eight months on, and it's turning out to be a fantastic year.

Although I may be biased, I think I have matured as a surgeon and, more importantly, as a doctor. I've gained a huge breadth of operative experience. I've managed a rota and written papers. I'm definitely better at saying "I don't know" and asking for help and feedback from my peers. I have been inspired, supported, and encouraged by the consultants I've worked with. I went on Remedy UK's London march (a group set up by junior doctors to lead protest at the government's reforms of medical training) and am the first to say that the changes to training have tragically led to the loss of many talented potential surgeons.

However, I cannot deny that I am extremely grateful for the opportunity I have been given this year. I've realised that, although it is difficult to hear, sometimes it is simply best to be told you're just not good enough.

Christopher Brown ST2 doctor in trauma and orthopaedics, Royal Berkshire Hospital, Reading cnbrown@doctors.org.uk

Cite this as: *BMJ* 2009;339:b2947

pico

Performance evaluation of a new rapid urine test for chlamydia in men: prospective cohort study

Elpidio-Cesar Nadala,[1] Beng T Goh,[2] Jose-Paolo Magbanua,[3] Penelope Barber,[4] Alison Swain,[4] Sarah Alexander,[5] Vivian Laitila,[1] Claude-Edouard Michel,[3] Lourdes Mahilum-Tapay,[1] Ines Ushiro-Lumb,[2] Catherine Ison,[5] Helen H Lee[3]

[1]Diagnostics for the Real World (Europe), Cambridge Science Park, Cambridge CB4 0WG
[2]Departments of Genitourinary Medicine and Virology, Barts and The London NHS Trust, London E1 1BB
[3]University of Cambridge, Department of Haematology, Diagnostics Development Unit, NHS Blood and Transplant Site, Cambridge CB2 2PT
[4]Brook in Birmingham, Birmingham B1 1BL
[5]Sexually Transmitted Bacteria Reference Laboratory, Health Protection Agency Centre for Infections, London NW9 5HT

Correspondence to: H H Lee
hl207@cam.ac.uk

Cite this as: *BMJ* 2009;339:b2655
doi: 10.1136/bmj.b2655

STUDY QUESTION To evaluate the performance of a new rapid urine test for chlamydia as a potential diagnostic and screening tool in men.

SUMMARY ANSWER With a novel signal amplification system in combination with FirstBurst, a collection device for first void urine, the *Chlamydia* Rapid Test achieved a high level of sensitivity (82.6%) and specificity (98.5%) compared with a nucleic acid amplified assay, the polymerase chain reaction.

Participants and setting

The evaluation sites comprised of a young people's sexual health centre (site 1) and a genitourinary medicine clinic (site 2). Eligible participants were aged at least 16, had not taken antibiotics in the previous month, and gave informed consent.

Design, size, and duration

The participants included 1211 men with a mean age of 18.2 and 29.8 at site 1 and site 2, respectively. The sample size was based on a predicted prevalence of 10% and a sensitivity rate of 85%. Participants were recruited from March to November 2007.

Main results and the role of chance

Compared with polymerase chain reaction, the sensitivity, specificity, positive predictive value, and negative predictive value of the new rapid test was 82.6%, 98.5%, 84.1%, and 98.3%, respectively. Samples that yielded discordant results were tested with the APTIMA CT kit (Gen-Probe). In addition, 100 randomly selected specimens negative by polymerase chain reaction and 20 concordant positive samples were tested blinded to minimise potential bias of testing discordant samples only. The organism load ranged from 7.3 x10^2 to 6.9 x10^6 plasmids/ml and correlated with the signal strength of the *Chlamydia* Rapid Test.

Bias, confounding, and other reasons for caution

The urine samples for the rapid test were obtained with FirstBurst immediately after informed consent was given. For testing with polymerase chain reaction, the patients waited for at least two hours before providing a second urine sample with the standard cup in accordance with manufacturer's instructions. This order of sample collection was necessary to evaluate the rapid test without requiring the patients to wait for the one to two hours imposed by traditional testing algorithm, but was not thought to bias the results from previous findings.

Generalisability to other populations

The new *Chlamydia* Rapid Test for men showed high sensitivity and specificity and produced results within an hour of sample collection. This test is suitable as a primary diagnostic tool, especially when patients need to be tested and treated on site. It could also be used as a screening tool in areas with high prevalence of infection, in settings where access to nucleic acid amplification testing is limited, or where return rate is low. It is therefore a useful tool to enable more men to be screened and treated in various settings. The test recognises the genus specific chlamydial lipopolysaccharide and thus can detect certain variants that a number of nucleic acid based assays do not detect.

Study funding/potential competing interests

The study was funded by a Wellcome Trust grant to the University of Cambridge, the National Institutes of Health, and additional support from the NIHR Cambridge Biomedical Research Centre. J-PM, C-EM, and HHL are equity holders in Diagnostics for the Real World, which markets the rapid test technologies developed at the University of Cambridge.

This is a summary of a paper that was published on bmj.com *BMJ* 2009;339:b2655

PERFORMANCE OF CHLAMYDIA RAPID TEST VERSUS POLYMERASE CHAIN REACTION

	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Site 1 (n=454)	90.0 (18/20)	98.2 (426/434)	69.2 (18/26)	99.5 (426/428)
Site 2 (n=757)	80.9 (72/89)	98.7 (659/668)	88.9 (72/81)	97.5 (659/676)
Total (n=1211)	82.6 (90/109)	98.5 (1085/1102)	84.1 (90/107)	98.3 (1085/1104)

BMJ pico: advice to authors

For Research articles, we routinely post the full, version only on bmj.com, and prepare an abridged version for the print journal.

To increase readership of research articles in the print *BMJ* and to give authors more control over the abridging, we are piloting a new way of abridging research articles for the print *BMJ*—publishing what is essentially an evidence abstract called BMJ pico. We hope that you will want to take part in this pilot if your research article is accepted. There is no need to prepare a BMJ pico in advance, however—please wait until we have offered to publish your article.

pico

Euthanasia and other end of life decisions and care provided in final three months of life: nationwide retrospective study in Belgium

Lieve Van den Block,[1 2] Reginald Deschepper,[1 6] Johan Bilsen,[1 4] Nathalie Bossuyt,[3] Viviane Van Casteren,[3] Luc Deliens[1 5]

EDITORIAL by Byock

[1]Vrije Universiteit Brussel, End-of-Life Care Research Group, Laarbeeklaan 103, 1090 Brussels, Belgium
[2]Vrije Universiteit Brussel, Department of General Practice, 1090 Brussels, Belgium
[3]Scientific Institute of Public Health, Department of Epidemiology, 1040 Brussels, Belgium
[4]Vrije Universiteit Brussel, Department of Public Health, 1090 Brussels, Belgium
[5]Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Centre, 1007 MB Amsterdam, Netherlands
[6]Vrije Universiteit Brussel, Department of Critical Care, 1090 Brussels, Belgium

Cite this as: *BMJ* 2009;339:b2772
doi: 10.1136/bmj.b2772

STUDY QUESTION What is the relation between the care provided in the final three months of life and the prevalence and types of end of life decisions in Belgium?

SUMMARY ANSWER End of life decisions including euthanasia or physician assisted suicide are not related to a lower use of palliative care in Belgium, but instead often occur within the context of multidisciplinary care.

Participants and setting

We collected data via the Sentinel Network of General Practitioners, an epidemiological surveillance system representative of all Belgian general practitioners. Each week physicians reported all non-sudden deaths of patients (>1 year) in their practice; 1690 non-sudden deaths were registered.

Design

We performed a two year nationwide mortality retrospective study in 2005-6 (sentinel network monitoring end of life care or SENTI-MELC study).

Primary outcomes

Care provided in the final three months of life and end of life decisions as reported by general practitioners. Multivariable regression analysis controlled for age, sex, and cause and place of death.

Main results and the role of chance

Use of specialist multidisciplinary palliative care services was associated with intensified symptom alleviation (odds ratio 2.1, 95% confidence interval 1.6 to 2.6), continuous deep sedation forgoing food/fluid (2.9, 1.7 to 4.9), and the total of decisions explicitly intended to shorten life (1.5, 1.1 to 2.1) but not with euthanasia or physician assisted suicide in particular. More patients in inpatient palliative care units died after euthanasia, physician assisted suicide, or life ending drugs without their explicit request compared with patients in hospitals or care homes ($P<0.05$), but effects were not significant after adjustment for differences in patients' characteristics. To a large extent receiving spiritual care was associated with higher frequencies of euthanasia or physician assisted suicide than receiving little spiritual care (18.5, 2.0 to 172.7).

Bias, confounding, and other reasons for caution

Our study used only observational data, which made it possible to study associations but not cause and effect relations. We also had to rely on general practitioners to report end of life care and decision making. Medical practice in hospitals might have been difficult for general practitioners to judge, especially concerning those aspects that are part of standard practice and generally less often discussed with other professionals, such as intensified alleviation of symptoms.

Generalisability to other populations

Generalisability depends on the legal situation and the customary ethical and moral belief systems within palliative care, medical practice, and the broader society in other countries. Our results do show, however, that making life shortening decisions and a philosophy of palliative care can and commonly do coexist. Sometimes far reaching decisions to reduce patients' end of life suffering are part of provision of palliative care.

Study funding/potential competing interests

The first author is a postdoctoral researcher of the Fund for Scientific Research in Flanders, Belgium. Support for the study came from the Research Council of the Vrije Universiteit Brussel in Belgium and the Institute for the Promotion of Innovation by Science and Technology in Flanders. The Belgian Sentinel Network of GPs is supported by the Flemish and Walloon Ministry of Welfare, Public Health and Family. The sponsors had no role in design and conduct of the study, in the collection, analysis, and interpretation of the data, in the writing of the report or in the decision to submit the article for publication. The researchers are independent from the funders.

CARE PROVIDED IN FINAL THREE MONTHS OF LIFE AS POSSIBLE DETERMINANTS OF END OF LIFE DECISIONS IN 1690 CASES OF NON-SUDDEN DEATH

	Euthanasia or physician assisted suicide	End of life decisions with explicit life shortening intent*	Continuous deep sedation without nutrition or hydration
No (%) who used multidisciplinary palliative care services			
No	8 (0.9)	90 (9.8)	21 (2.3)
Yes	13 (2.0)	90 (14.1)	39 (6.1)
No (%) who received spiritual care			
Not or to (very) small extent	1 (0.2)	64 (10.5)	21 (3.5)
Average	7 (1.9)	58 (15.4)	23 (6.2)
To (very) large extent	4 (3.8)	14 (13.3)	2 (1.9)

*Includes non-treatment decisions with explicit life shortening intention, euthanasia, physician assisted suicide, and life ending drugs without explicit request from patient