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

RESEARCH

Ethnic stereotypes and the underachievement of UK medical students from ethnic minorities: qualitative study

Katherine Woolf,[1] Judith Cave,[1] Trisha Greenhalgh,[2] Jane Dacre[1]

EDITORIAL by Carr and Woodson

[1]Academic Centre for Medical Education, University College London Division of Medical Education, London N19 5LW
[2]Open Learning Unit, Department of Primary Care and Population Sciences, University College London
Correspondence to: K Woolf
k.woolf@medsch.ucl.ac.uk

Cite this as: *BMJ* 2008;337:a1220
doi:10.1136/bmj.a1220

ABSTRACT

Objective To explore ethnic stereotypes of UK medical students in the context of academic underachievement of medical students from ethnic minorities.

Design Qualitative study using semistructured one to one interviews and focus groups.

Setting A London medical school.

Participants 27 year 3 medical students and 25 clinical teachers, purposively sampled for ethnicity and sex.

Methods Data were analysed using the theory of stereotype threat (a psychological phenomenon thought to negatively affect the performance of people from ethnic minorities in educational contexts) and the constant comparative method.

Results Participants believed the student-teacher relationship was vital for clinical learning. Teachers had strong perceptions about "good" clinical students (interactive, keen, respectful), and some described being aggressive towards students whom they perceived as quiet, unmotivated, and unwilling. Students had equally strong perceptions about "good" clinical teachers (encouraging, interested, interactive, non-aggressive). Students and teachers had concordant and well developed perceptions of the "typical" Asian clinical medical student who was considered over-reliant on books, poor at communicating with patients, too quiet during clinical teaching sessions, and unmotivated owing to being pushed into studying medicine by ambitious parents. Stereotypes of the "typical" white student were less well developed: autonomous, confident, and outgoing team player.

Conclusions Asian clinical medical students may be more likely than white students to be perceived stereotypically and negatively, which may reduce their learning by jeopardising their relationships with teachers. The existence of a negative stereotype about their group also raises the possibility that underperformance of medical students from ethnic minorities may be partly due to stereotype threat. It is recommended that clinical teachers be given opportunities and training to encourage them to get to know their students as individuals and thus foster positive educational relationships with them.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a1220

INTRODUCTION

Medical students from ethnic minorities make up about 30% of the medical student population in the United Kingdom[1]; however, UK medical students and doctors from ethnic minorities significantly underperform in assessments compared with their white counterparts.[2-8] Some evidence suggests that the gap might be greatest in assessments of clinical knowledge and skills, but it is also present in machine marked tests of basic medical knowledge.[9 10] Students from ethnic minorities enter medical school with slightly lower examination grades than white students, but this only partly explains the gap seen later.[1]

In the United States, academic underperformance of people from ethnic minorities, particularly African-Americans, has been explained by the theory of "stereotype threat," whereby members of negatively stereotyped groups can feel sufficient anxiety at the prospect of being negatively stereotyped that they underperform in tests.[11 12]

We explored stereotype threat and other factors that might affect students from ethnic minorities learning in clinical environments, in a sample of first year clinical medical students and a sample of their teachers. We aimed to explore any stereotypes about medical students from ethnic minorities and to generate hypotheses to explain underachievement in the students.

METHODS

We gathered data from year 3 medical students and their clinical teachers using one to one, face to face interviews initially. It became clear that the students did not feel comfortable discussing ethnicity and so we collected the rest of the data using focus groups of single ethnicity.

Clinical teachers were purposively selected from a sampling frame including senior surgeons, doctors, general practitioners, and tutors in clinical skills. Data from one London hospital's website showed that in 2003 69% of consultants were men, 88% were white, 1% were black, and 9% were Asian. The sampling frame was designed to reflect these demographics. Forty clinical teachers were invited to be interviewed. During the sampling phase participants were assigned a sex and ethnic group on the basis of their names, and they subsequently self reported ethnicity using the 2001 UK census categories.

Students were purposively sampled using the demographics of the year 3 student population as a sampling frame (information on self reported sex and ethnicity was obtained from student records). Forty nine students were invited by email to be interviewed about their experiences of clinical teaching.

Of the 360 medical students of known self reported ethnicity (94.5% of total year group), students from the three largest ethnic groups—white, Indian, and Pakistani—were invited to participate. We organised potential participants by ethnicity and clinical firm grades into six groups: Indian high achieving, Indian low achieving, Pakistani/Bangladeshi high achieving, Pakistani/Bangladeshi low achieving, white high achieving, and white low achieving.

KW (British white female faculty member) carried out the interviews. During the focus groups KW moderated while SB (British black Caribbean female student) took notes. Storytelling and expanding on comments were encouraged.

We used open questions, concerning experiences of teaching and learning in the clinical environment. The groups also included an open question about what it meant to be from an ethnic group (see web extra for questions). Participants were asked what they thought of findings that medical students from ethnic minorities under-perform academically compared with white students.

KW transcribed the data verbatim. We analysed all data using stereotype threat as a theoretical framework. KW searched the data for the ways in which participants portrayed white students and those from ethnic minorities in clinical teaching contexts. KW and JC discussed the stereotypes that emerged from the data and coded them for how they related to different aspects of teaching and learning in the clinical environment. Opposite examples were sought and used to refine the analysis.

RESULTS

Overall, 26 of 40 teachers agreed to be interviewed (65%; see bmj.com). Twenty one of 49 students invited agreed to be interviewed (43%). Six white students and six students from ethnic minorities were interviewed. High achievers were more likely to attend the focus groups than low achievers (19.3% *v* 9.0% of those invited).

Importance of student-teacher relationship to learning

The student-teacher relationship was described as one of the most important factors in determining the quality of learning. Teachers believed that to foster these relationships it was important to find out about students' educational and pastoral needs and to tailor the teaching to those needs. They enjoyed and put most effort into teaching students who interacted with them, asked questions, and seemed enthusiastic whereas quiet students were perceived as unresponsive, unenthusiastic, and unappreciative. Some teachers (mainly white men) described how they behaved antagonistically towards students they thought were not making sufficient effort to learn:

> "We're busy, we're fairly bullish […] If you give me five keen students, they get a fantastic deal. If you give me five quiet reticent students they get a crap deal" (teacher 25: male surgeon, other ethnic group)
>
> "A little bit of fear ain't a bad thing from where I come from. I may push someone over the edge and they'll probably commit suicide and I'll be terribly sorry but that's a risk I will take […] If they're prepared to work together, I will work with them, literally the whole time on the firm, if they're not, don't bother me about it, go and get a life because you're not going to enjoy it" (teacher 18: male surgeon, white)

Students described how they appreciated the teachers' efforts and learnt most when teachers interacted with them. They did not like being taught by teachers who seemed unenthusiastic, or who ignored or humiliated them. They explained that they would not feel able to interact or might not attend those sessions:

> "I think most of when I learn when there's lots of perhaps student-teacher interaction" (student 2: male, Asian)
>
> "[Bullying tactics] will tend to make [students] quieter, they don't tend to ask questions and they tend to leave sessions confused and I think that leads to them thinking 'there's no point in me learning'" (student 1: female, Asian)

Perceptions about "typical" Asian medical students

As well as having perceptions about "good" clinical students, clinical teachers had perceptions about Asian students. Students, including Asians, had similar perceptions.

In terms of their relationship with books, the "typical" Asian student was viewed as conscientious, hard working, and bright. Non-Asian participants were more likely to qualify this picture with a flip side: that being over-reliant on books made the "typical" Asian student inflexible and less able to adapt to new ways of behaving. Thus it was perceived that although Asian students had excellent school examination results, learning clinical medicine required flexible learning and that such students might struggle in these circumstances:

> "Students that are of South Asian or Indian origin tend to be, or come across as being far more academically knowledgeable and they can justify what they're doing and they're very very bright, but actually putting that into practice and both with communication and practical skills doesn't seem to gel that well" (teacher 11: female clinical skills tutor, white)

In terms of their relationships with patients, non-Asian participants perceived the "typical" Asian student as a poor communicator, either because of varying degrees of linguistic problems, which

(allegedly) made them feel under-confident, or because they were culturally more formal than white students:

> "It's much more common to come across an Asian student, even if they're English-born, that has formal relationships with patients than it is to find a, a white British-born person having formal relationships with patients" (teacher 10: female general practitioner, white)
>
> "I've had people [students] who are, for example, from the Far East who are extremely polite you know very polite and so on, but may come across um, in not quite the same, just because they're you know of their culture being extremely polite, may not come across [to patients] as well" (teacher 3: male doctor, Black African)

In terms of their relationships with teachers, the "typical" Asian student was perceived by non-Asian teachers and one non-Asian student as shy, quiet, reserved, and under-confident. Clinical teachers thought this was because Asians were overly respectful of authority:

> "Some of these sweet little Asian girlies are very hard to get through to. I'm quite a physically biggish sort of chap, maybe that's another factor. I'm older, obviously that's a factor. I'm male. I'm … they don't communicate terribly well" (teacher 2: male doctor, white)

The "typical" Asian student was seen by Asian and non-Asian participants as more likely to be studying medicine to conform to their parents' wishes; and non-Asians were more likely to equate this with having led a sheltered life and being less mature and autonomous. Differences in motivation were seen by participants from different ethnic groups as indicating that Asian students would be less likely to be deep learners, caring communicators, active participants in their learning, and ultimately good doctors:

> "You sometimes find that students who are incredibly disillusioned say 'I went in to medicine because of this that and the other, because my parents wanted me to' […] My parents certainly wanted me to become a doctor but I wanted to. Um, it worked that way […] I'm speaking from an Asian background not for anyone else. Um, it's quite well known" (teacher 22: female general practitioner, Asian)
>
> "There's a stigma of sort of ethnic families wanting their children to do best and then there's the whole doctor, lawyer, you know, get the upper, upper rank jobs or whatever they're called and so I suppose if they're thinking 'oh bollocks, I've got to choose between three jobs, I'll choose the doctor then'" (student 12: male, white)
>
> "You do get a lot of Asian families who push their children to be doctors […] Even if we're born and bred here and we've lived here for about 30 years, it's still the kind of thing: if your son's a doctor that's fine" (student 2: male, Asian)

Perceptions about "typical" white students

The idea of the "typical" white student was less well developed; however, white students were perceived by teachers and students as being autonomous learners who were dedicated and self motivated; were tough but sociable team players, as evidenced by their love of rugby; and were confident, outspoken, and good communicators (especially the women). Sometimes white students were perceived as pushy or arrogant, although these characteristics were deemed likely to help students succeed:

> "White female students seem to have, for me, the best communication skills with patients. And be most patient-oriented in their approach" (teacher 11: female clinical skills tutor, white)
>
> "The white people who get into medical school, they're just across the board they're more motivated and are doing it for the right reasons and they always have that in mind" (student B: female, Indian high achieving group).

DISCUSSION

Teachers of clinical medical students, and the students themselves, have strong perceptions about "typical" Asian students, and there is a systematic mismatch between these perceptions and the perception of what makes a "good" clinical student.

The strong theoretical underpinnings of the data analysis were useful in organising the data meaningfully and in generating hypotheses for future testing on the ways in which stereotyping, teacher-student interactions, and performance are related. Both students and teachers were interviewed, which provided triangulation of the results, and the data were analysed by two researchers with different backgrounds, which improved validity and reliability.

Our study has some limitations. The study design was based on what students and teachers said they thought and did, not on direct observation of what they actually did. The triangulation of results suggested these descriptions were valid.

The sex, ethnicity, and age of the interviewer (British female, white) may have affected participants' discussion of certain topics. That students in the one to one interviews felt uncomfortable discussing ethnicity means that potentially important topics may not have been covered. This taboo was addressed by running the focus groups: students in the interviews and focus groups gave similar answers to non-controversial questions, but additional themes arose in the focus groups.

The low participation rate in the focus groups and the relatively low numbers of non-white clinical teachers may have introduced a systematic bias. This is

important as low achieving students were less likely to attend the focus group than high achieving students, and teachers from ethnic minorities may have had different ideas from white teachers. This was an exploratory study designed to open up areas for further research.

Generalisability is a problem inherent in all qualitative studies. For example, the terms "ethnic minority," "non-white," and "Asian" are not considered interchangeable in many contexts; however, in the context of a medical school where most of the students from ethnic minorities are of South Asian origin (Indian, Pakistani, Sri Lankan), when participants spoke of "ethnic minorities" or "non-whites" these terms were interpreted as meaning "Asian"—a term used by many of the Indian and Sri Lankan participants to refer to themselves. The aim of this study, however, was to provide a preliminary exploration of the topic with a view to prompting reflection by teachers, students, and policy makers, and informing future research.

Much of the previous research into ethnicity in medical education has related to measuring differences in performance between ethnic groups.[5 6 9] Although these differences are important, these studies provide little insight into the reasons why this gap exists. Previous studies have found that students from ethnic minorities can experience marginalisation and segregation.[13 14] We looked at the ways in which students from different ethnic groups perceive each other but not the ways in which they thought they behaved towards each other.

One study[12] argued that for stereotype threat to have an effect students have to be concerned that people will make negative assumptions about them on the basis of the stereotype and that by behaving in a particular way they are conforming to that negative stereotype. According to that study this concern can make students sufficiently anxious to negatively affect their performance, or can prompt them to try and behave in a way that counters the stereotype. A male Asian student described how he thought he needed to work extra hard to combat any possibility that he would be negatively stereotyped as an underachiever. A female Indian Muslim student said that she believed that her seniors assumed she was "substandard" on the basis that she wore a headscarf.

Research shows that relying on stereotypes can stop people from searching out information that conflicts with the stereotype (confirmation bias[15]). One study[12] recommends that to combat the effects of stereotyping, teachers should get to know their students individually. Even in the absence of stereotype threat, however, Asian students who are stereotyped and therefore not seen as individuals are less likely to have effective educational relationships with their teachers.[16 17] Students who perceive that they are being stereotyped may find it more difficult to learn, as negative emotions interfere with learning.[16] Clinical teachers with stereotypical views of Asian students may feel less positive about teaching them.

We recommend that clinical teachers should get to know their students as individuals, and that employers should provide the training and infrastructure to help them achieve this task. This will benefit most students but has the additional advantage for students from ethnic minorities of countering the effects of stereotyping, including stereotype threat.

Several studies, mainly with US college students, have measured stereotype threat and investigated the ways in which it can be manipulated experimentally.[11 18] Such studies could usefully be replicated with UK medical student populations.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Medical students from ethnic minorities under-perform academically compared with their white colleagues, for reasons that are unclear

US research has linked underperformance in college students from ethnic minorities to a psychological phenomenon known as stereotype threat—systematic differences in expectations and treatment owing to pervasive negative stereotypes

WHAT THIS STUDY ADDS

In one London medical school, Asian clinical medical students were negatively stereotyped by clinical teachers and students as overdependent on book learning and quiet in class

Participants believed these attributes explained underperformance in the Asian students, raising the possibility that stereotype threat may be occurring

Clinical teachers disliked teaching students they perceived as having these attributes, and students reported being unable to learn from unenthusiastic teachers, suggesting that negative stereotyping can adversely affect learning

We thank Sara Belle for comoderating the focus groups, the participants for taking time to answer our questions, and those who made useful comments on earlier drafts of the paper and during oral presentations of the data. KW's salary is supported by the General Medical Council. JC's salary is supported by Cancer Research UK. The other authors are funded by the Higher Education Funding Council.

Contributors: See bmj.com.

Competing interests: None declared.

Ethical approval: Graduate School Ethics Committee and the NHS Research Ethics Committee.

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

1 McManus IC, Woolf K, Dacre J. The educational background and qualifications of UK medical students from ethnic minorities. *BMC Med Educ* 2008;8:21.

2 Dillner L. Manchester tackles failure rate of Asian students. *BMJ* 1995;310:209.

3 Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ* 2002;324:952-7.

4 Wass V, Roberts C, Hoogenboom R, Jones R, Van der Vleuten C. Effect of ethnicity on performance in a final objective structured clinical

examination: qualitative and quantitative study. *BMJ* 2003;326:800-3.
5 Lumb AB, Vail A. Comparison of academic, application form and social factors in predicting early performance on the medical course. *Med Educ* 2004;38:1002-5.
6 Haq I, Higham J, Morris R, Dacre J. Effect of ethnicity and gender on performance in undergraduate medical examinations. *Med Educ* 2005;39:1126-8.
7 Bessant R, Bessant D, Chesser A, Coakley G. Analysis of predictors of success in the MRCP (UK) PACES examination in candidates attending a revision course. *Postgrad Med J* 2006;82:145-9.
8 Dewhurst N, McManus IC, Mollon J, Dacre J, Vale JA. Performance in the MRCP (UK) examination 2003-4: analysis of pass rates of UK graduates in relation to self-declared ethnicity and gender. *BMC Med Educ* 2007;5:8.
9 McManus IC, Richards P, Winder BC, Sproston KA. Final examination performance of medical students from ethnic minorities. *Med Educ* 1996;30:195-200.
10 Woolf K, Haq I, Higham I, McManus IC, Dacre J. Exploring the underperformance of male and minority ethnic medical students in first year clinical examinations. *Adv Health Sci Educ* 2007 May 9 [ePub ahead of print].
11 Steele CM, Aronson J. Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol* 1995;69:797-811.
12 Steele CM. A threat in the air. How stereotypes shape intellectual identity and performance. *Am Psychol* 1997;52:613-29.
13 Roberts JH, Sanders T, Wass V. Students' perceptions of race, ethnicity and culture at two UK medical schools: a qualitative study. *Med Educ* 2008;42:45-52.
14 Beagan BL. 'Is this worth getting into a big fuss over?' Everyday racism in medical school. *Med Educ* 2003;37:852-60.
15 Eysenck MW, Keane MT. *Cognitive psychology: a student's handbook*. Hove: Psychology Press, 2000.
16 Haidet P, Stein HF. The role of the student-teacher relationship in the formation of physicians. The hidden curriculum as process. *J Gen Intern Med* 2006;21(suppl 1):S16-20.
17 Howe A. Patient-centred medicine through student-centred teaching: a student perspective on the key impacts of community-based learning in undergraduate medical education. *Med Educ* 2001;35:666-72.
18 Cohen GL, Garcia J, Apfel N, Master A. Reducing the racial achievement gap: a social-psychological intervention. *Science* 2006;313:1307-10.

Accepted: 27 June 2008

Commentary: An "ethnic minority" medical student

Hugh Ip

EDITORIAL by Carr and Woodson

Imperial College London, London SW7 2AZ
hughip@gmail.com

Cite this as: *BMJ* 2008;337:a1240
doi:10.1136/bmj.a1240

I began my medical studies in London in 2003, naive, but brimming with enthusiasm. Five years on[1] I have more insight into the challenges medical students face, as well as the particular implications of belonging to an ethnic minority. In their study, Woolf et al describe stereotypical views of South Asian medical students. I think that students and teachers hold similar views of Chinese medical students: conscientious, hardworking, and bright in terms of book learning, but weaker at communicating with patients and teachers.

The authors suggest that academic performance is adversely affected by negative stereotypes (stereotype threat). I feel that my mindset—that of an international student—shielded me from some of these adverse effects. I left behind family and friends in Hong Kong with my eyes wide open. When I landed at Heathrow airport a week before the start of term I knew that I would have to adapt to a new culture. I expected to be perceived differently from the "white British" student and I was prepared to adapt my learning style with a positive attitude.

Maybe my awareness of ethnic stereotypes was heightened by my secondary education in an international school. I was taught by British teachers who deftly facilitated cross cultural learning and friendship between students of almost 30 nationalities. Woolf et al "recommend that clinical teachers should make efforts to get to know their students as individuals." Pressured clinicians, with less time and training, will find it more difficult than my school teachers did.

Hong Kong is a cosmopolitan city, London even more so. During my clinical studies I have probably encountered as many doctors and patients from ethnic minorities as white doctors and patients. I wonder how medical students from ethnic minorities fare in less diverse parts of the United Kingdom. How prevalent are the adverse effects of negative stereotyping and do they vary?

Perhaps the diversity of the medical student population matters too. In facing the pressures of medical studies I have found the support of fellow students invaluable.[2] It is not surprising that ethnically Chinese friends (from various countries) find it easier to empathise with my own struggles—for example, academic and familial—given our shared cultural background. Some white students, however, not only befriend students from ethnic minorities superficially but also go beyond stereotypes to understand the particular challenges they face, for which I am grateful.

Thirty per cent of the UK medical student population come from ethnic minorities. This percentage looks set to rise, in light of the government's widening participation initiative.[3] The study by Woolf et al, of the adverse effects of negative stereotyping on learning, is highly relevant to students from ethnic minorities as well as the patients they will one day serve. The authors offer suggestions for clinical teachers to counter stereotype threat, but medical students have responsibilities too. Although we cannot change the colour of our skin, we can persevere[4] by grasping any learning opportunities that present and making ourselves teachable as best we can.

HI is a former student editor of the *Student BMJ*.
Competing interests: None declared.
Provenance and peer review: Commissioned; not peer reviewed.

1 Ip H. The time of your life. *Student BMJ* 2008;16:178.
2 Mentzer A, Sturgeon D. Dealing with stress. *Student BMJ* 2008;16:272.
3 Ip H, McManus IC. Increasing diversity among clinicians. *BMJ* 2008;336:1082-3.
4 Dawes M. To be a medical student today. *Student BMJ* 2007;15:298.

Exposure to antipsychotics and risk of stroke: self controlled case series study

Ian J Douglas, Liam Smeeth

EDITORIAL by Whitaker

Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT

Correspondence to: I Douglas
ian.douglas@lshtm.ac.uk

Cite this as: *BMJ* 2008;337:a1227
doi:10.1136/bmj.a1227

ABSTRACT

Objectives To investigate the association between use of typical and atypical antipsychotic drugs and incidence of stroke in patients with and without dementia.

Design Self controlled case series.

Setting UK based electronic primary care records in the general practice research database (GPRD).

Participants All patients registered in the database with a recorded incident stroke and at least one prescription for any antipsychotic drug before the end of 2002: 6790 eligible participants were identified and included in the final analysis.

Main outcome measures Rate ratio for stroke in periods of time exposed to antipsychotics compared with unexposed periods.

Results Use of any antipsychotic drug was associated with a rate ratio for stroke of 1.73 (95% confidence interval 1.60 to 1.87): 1.69 (1.55 to 1.84) for typical antipsychotics and 2.32 (1.73 to 3.10) for atypical antipsychotics. In patients receiving any antipsychotic drug, the rate ratios were 3.50 (2.97 to 4.12) for those with dementia and 1.41 (1.29 to 1.55) for those without dementia.

Conclusions All antipsychotics are associated with an increased risk of stroke, and the risk might be higher in patients receiving atypical antipsychotics than those receiving typical antipsychotics. People with dementia seem to be at a higher risk of an associated stroke than people without dementia and use of antipsychotics should, when possible, be avoided in these patients.

INTRODUCTION

Concerns have been expressed about an increased risk of stroke associated with atypical antipsychotic drugs.[1] In 2004 the United Kingdom's Committee on Safety of Medicines (CSM) recommended avoiding the use of atypical antipsychotic drugs among people with dementia.[2]

We used a within person case series design to assess the risk of stroke associated with antipsychotic drugs. We examined whether some or all of the previously observed increased risk of stroke associated with antipsychotic drug use could be attributable to confounding, whether the risk of stroke associated with typical and atypical antipsychotic drug use differs, and whether the risk of stroke associated with antipsychotic drug use is higher among people with dementia.

METHODS

Selection of participants

Patients were selected from those registered with the general practice research database before 2003. Eligible participants all had a first ever incident diagnosis of stroke on or before 31 December 2002 and had been prescribed at least one antipsychotic drug before this date.

Self controlled case series analysis

The self controlled case series method relies on intraperson comparisons in a population of individuals who have both the outcome and exposure of interest. Rate ratios compare the rate of events during exposed periods of time with the rate during all other observed time periods.[3] This method removes the potential confounding effect of characteristics that vary between individuals, such as frailty and risk factors for vascular disease.

For each participant, we identified and classified all prescriptions for antipsychotic drugs before the end of 2002. When possible, we calculated the expected length of exposure after each prescription. Each individual's observation time was then divided into fully exposed periods covered by the expected length of exposure, followed by a sequence of five 35 day periods after treatment up to a maximum of 175 days after the expected end of a treatment period. We also searched each patient's medical record for diagnoses indicating dementia before stroke to allow subgroup analysis in these patients.

We estimated the relative rate ratios with conditional Poisson regression, adjusting for age at stroke in five year bands. We assessed the impact of exposure to any antipsychotic medication, looked at the effect of any antipsychotic drug among patients with and without dementia, and measured the differential effects of typical and atypical antipsychotics among all patients and stratified by dementia status.

RESULTS

We identified 6790 eligible patients (4353 women) in the database with at least one prescription for an antipsychotic drug and a recorded incident stroke between January 1988 and the end of 2002. Of these, 905 patients were prescribed at least one atypical antipsychotic drug and 6334 were prescribed at least one typical antipsychotic drug during the study period. The median age at first exposure to any antipsychotic drug was 80, while median age at the time of first recorded stroke was 81; 1423 patients had a recorded diagnosis of dementia before the incident stroke and these patients were slightly older than those without dementia at the time the first antipsychotic drug was prescribed. In total, 5885 patients received a

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a1227

Case series analysis for antipsychotic drugs: association between exposure and stroke. Figures are rate ratios (95% confidence intervals)

	Any antipsychotic	Typical only†	Atypical only†
All patients (n=6790)			
No on group	6790	5885	456
Exposed *v* unexposed periods	1.73 (1.60 to 1.87)	1.69 (1.55 to 1.84)	2.32 (1.73 to 3.10)
Patients with recorded dementia (n=1423)			
No in group	1423	1208	85
Exposed *v* unexposed periods	3.50 (2.97 to 4.12)	3.26 (2.73 to 3.89)	5.86 (3.01 to 11.38)
Patients without recorded dementia (n=5367)			
No in group	5367	4673	371
Exposed *v* unexposed periods	1.41 (1.29 to 1.55)	1.40 (1.26 to 1.54)	1.90 (1.36 to 2.65)

NA=not applicable (no events).
†Patients who received both typical and atypical antipsychotics not included in this analysis.

prescription for a typical but not an atypical antipsychotic drug during the study period, and 456 patients received prescriptions only for atypical antipsychotics. The age at first recorded exposure to antipsychotics was similar for patients exposed only to typical or atypical antipsychotics. The 449 remaining patients received prescriptions for both typical and atypical antipsychotics. Among patients with dementia, 1212 received only typical antipsychotics and 85 received only atypical antipsychotics. The median duration of total observation period included in the analysis was at least four years for each subgroup.

The rate ratio for stroke among all patients prescribed any antipsychotic drug was 1.73 (95% confidence interval 1.60 to 1.87), comparing exposed with unexposed baseline periods (table). During the periods after treatment the rate ratio fell towards unity (see bmj.com). During exposed periods, the rate ratio was 1.69 (1.55 to 1.84) for patients receiving only typical antipsychotics and 2.32 (1.73 to 3.10) for patients receiving only atypical antipsychotics. During periods of treatment with any antipsychotic drug the rate ratio was 3.50 (2.97 to 4.12) in patients with recorded dementia before stroke and 1.41 (1.29 to 1.55) in patients with no record of dementia before stroke. In patients with dementia and only typical antipsychotic drug prescriptions, the rate ratio for stroke was 3.26 (2.73 to 3.89). This compares with a figure of 5.86 (3.01 to 11.38) in patients with dementia and only treated with atypical antipsychotics. Patients without dementia before stroke and receiving only typical antipsychotics had a rate ratio of 1.40 (1.26 to 1.54) compared with a figure of 1.90 (1.36 to 2.65) in patients without dementia and receiving only atypical antipsychotics. In all analysis subgroups, the rate ratio for stroke subsequently fell towards unity during the phase after treatment (see bmj.com).

DISCUSSION

The previously observed increased risk of stroke associated with use of antipsychotic drugs is not attributable to differences in baseline cardiovascular risk between people prescribed and not prescribed these drugs. The risk of stroke is slightly higher with use of atypical rather than typical antipsychotic drugs. The magnitude of the increased risk of stroke associated with antipsychotic drug use is more than twice as great among people with dementia compared with those without.

Strengths and weaknesses

A key advantage of this study is the use of the self controlled case series design. We censored follow-up for all patients at the end of 2002 as this was when concerns about the effects of antipsychotic drugs in patients with dementia first emerged. This should avoid possible biases arising from altered prescribing habits in the light of these findings.

Our study was large and statistically powerful and used routine clinical data from the UK general practice research database, which is largely representative of the population of the UK and so the results are likely to be highly generalisable.[4] A potential weakness might relate to the quality of the clinical data. Drug prescriptions in the database are generated by practice computers to ensure the accuracy of the electronic prescribing records. Prescription data were highly detailed and recorded before people developed stroke so there was no potential for recall bias. Some patients were probably not taking their prescribed antipsychotics during periods we classified as exposed, though this would result only in a reduced effect estimate for antipsychotic drug exposure. There might have been some inaccuracy in our estimation of exposed and unexposed periods.

We chose not to include less serious outcomes related to stroke, such as transient ischaemic attack, as the ascertainment rate and accuracy of dating for these events is probably less reliable. Over 85% of the strokes identified in this study were recorded with codes that did not specify the subtype of stroke. Nevertheless, we carried out a subgroup analysis in the 233 patients with designated haemorrhagic stroke. The rate ratio for this subgroup, considering exposure to any type of antipsychotic, was 1.14 (0.71 to 1.84), indicating that the effect might be limited to non-haemorrhagic stroke.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Atypical antipsychotics might increase the risk of stroke in elderly patients, but this association could be due to unmeasured or uncontrolled confounding

The extent to which typical antipsychotics are associated with an increased risk of stroke and whether the effect is the same in patients with dementia as those without is not known

WHAT THIS STUDY ADDS

Both typical and atypical antipsychotic drugs are associated with an increased risk of stroke and this association is unlikely to be caused by confounding

The risk of stroke in patients receiving antipsychotics seems to be greater in those with dementia than those without

Conclusion

We have established that all types of antipsychotics carry an increased risk, although the risk might be somewhat higher with the atypical drugs. As the background risk of stroke in elderly patients is relatively high,[5] we reaffirm that the risks associated with antipsychotic drug use in patients with dementia generally outweigh the potential benefits, and use of antipsychotic drugs in these patients should be avoided whenever possible.

Contributors: See bmj.com.
Funding: LS is supported by a Wellcome Trust senior research fellowship in clinical science. Data from the general practice research database were made available through the access for UK academics via Medical Research Council agreement.
Competing interests: None declared.
Ethical approval: Independent scientific advisory group of the general practice research database and the London School of Hygiene and Tropical Medicine ethics committee.
Provenance and peer review: Not commissioned; externally peer reviewed.

1 Gill SS, Rochon PA, Herrmann N, Lee PE, Sykora K, Gunraj N, et al. Atypical antipsychotic drugs and risk of ischaemic stroke: population based retrospective cohort study. *BMJ* 2005;330:445.
2 Committee on Safety of Medicines. *Atypical antipsychotic drugs and stroke: message from Professor Gordon Duff, chairman, (CEM/CMO/2004/1)*. London: Committee on Safety of Medicines, 2006.
3 Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med* 2006;25:1768-97.
4 Office for National Statistics. *Key health statistics from general practice 1998*. London: Office for National Statistics, 2000:149. Series MB6 No 2.
5 Liebetrau M, Steen B, Skoog I. Stroke in 85-year-olds: prevalence, incidence, risk factors, and relation to mortality and dementia. *Stroke* 2003;34:2617-22.

Accepted: 30 June 2008

Effect on birth outcomes of a formalised approach to care in hospital labour assessment units: international, randomised controlled trial

Ellen D Hodnett,[1] Robyn Stremler,[1,2] Andrew R Willan,[2,3] Julie A Weston,[1] Nancy K Lowe,[4] Kathleen R Simpson,[5] William D Fraser,[6] Amiram Gafni,[7] the SELAN (Structured Early Labour Assessment and Care by Nurses) Trial Group

EDITORIAL by Spiby and Renfrew

[1]Lawrence S Bloomberg Faculty of Nursing, University of Toronto, 155 College St, Toronto, ON, Canada, M5T 1P8
[2]Program in Child Health Evaluative Sciences, and Research Institute at SickKids, SickKids Hospital, Toronto, Canada
[3]Dalla Lana School of Public Health, University of Toronto, 155 College St, Toronto, ON, Canada, M5T 3M7
[4]University of Colorado Health Sciences Center, School of Nursing, Denver, CO, USA
[5]Labor and Delivery, St John's Mercy Medical Center, St Louis, MO, USA
[6]Université de Montréal, Département d'obstétrique-gynécologie, Pavillon Roger-Gaudry, Montréal, QC, Canada
[7]Department of Clinical Epidemiology and Biostatistics, and Centre for Health Economics and Policy Evaluation, McMaster University, Hamilton, ON, Canada
Correspondence to: E D Hodnett
ellen.hodnett@utoronto.ca

Cite this as: *BMJ* 2008;337:a1021
doi:10.1136/bmj.a1021

ABSTRACT

Objective To determine if a complex nursing and midwifery intervention in hospital labour assessment units would increase the likelihood of spontaneous vaginal birth and improve other maternal and neonatal outcomes.
Design Multicentre, randomised controlled trial with prognostic stratification by hospital.
Setting 20 North American and UK hospitals.
Participants 5002 nulliparous women experiencing contractions but not in active labour; 2501 were allocated to structured care and 2501 to usual care.
Interventions Usual nursing or midwifery care for a minimum of one hour of care by a nurse or midwife trained in structured care, consisting of a formalised approach to assessment of and interventions for maternal emotional state, pain, and fetal position.
Main outcome measures Primary outcome was spontaneous vaginal birth. Other outcomes included intrapartum interventions, women's views of their care, and indicators of maternal and fetal health during hospital stay and 6-8 weeks after discharge.
Results Outcome data were obtained for 4996 women. The rate of spontaneous vaginal delivery was 64.0% (n=1597) in the structured care group and 61.3% (n=1533) in the usual care group (odds ratio 1.12, 95% confidence interval 0.96 to 1.27). Fewer women allocated to structured care (n=403, 19.5%) rated staff helpfulness as less than very helpful than those allocated to usual care (n=544, 26.4%); odds ratio 0.67, 98.75% confidence interval 0.50 to 0.85. Fewer women allocated to structured care (n=233, 11.3%) were disappointed with the amount of attention received from staff than those allocated to usual care (n=407, 19.7%); odds ratio 0.51, 98.75% confidence interval 0.32 to 0.70. None of the other results met prespecified levels of statistical significance.
Conclusion A structured approach to care in hospital labour assessment units increased satisfaction with care and was suggestive of a modest increase in the likelihood of spontaneous vaginal birth. Further study to strengthen the intervention is warranted.
Trial registration Current Controlled Trials ISRCTN16315180.

INTRODUCTION

Labour assessment units are a routine feature in North American hospitals but remain uncommon elsewhere. They offer an opportunity for primary and secondary prevention of intrapartum complications.

Several studies have found associations between anxiety before active labour and intrapartum complications.[1-4] Women in latent labour who expressed negative feelings about their ability to cope or had high pain ratings were more likely to develop intrapartum complications.[5] Malposition of the fetal head has been associated with a prolonged latent phase, increased pain, higher maternal anxiety, complications in active labour, and higher rates of operative delivery.[6-9] Medical interventions during active labour,

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a1021

such as epidural analgesia, can be effective treatments for problems resulting from malposition, but they do not increase the likelihood that the fetal head will rotate to the occiput anterior position.[10] Simple positioning techniques may encourage rotation and descent.[6]

Although caesarean delivery rates continue to increase,[11] spontaneous vaginal birth is widely regarded as the safest method of birth for healthy mothers and babies at low risk.[12] We determined the effects on the likelihood of spontaneous vaginal birth of a formalised approach to care in labour assessment units, which focused on assessment of and anticipatory guidance on cognitive-emotional state, pain, and fetal position.

Comparisons of maternal outcomes from randomisation until postnatal hospital discharge. Values are numbers (percentages) of women unless stated otherwise

Event	Structured care (n=2497)	Usual care (n=2499)	Odds ratio (CI)
Labour onset:			
Spontaneous	2232 (89.4)	2209 (88.4)	
Induced	255 (10.2)	283 (11.3)	
No labour	8 (0.3)	6 (0.2)	
Oxytocin started after active labour	1553 (62.2)	1587 (63.5)	0.95 (0.77 to 1.12)*
Analgesia or anaesthesia†:			
Regional‡	2112 (84.6)	2159 (86.4)	0.85 (0.62 to 1.08)*
Intramuscular or intravenous opioid	1126 (45.1)	1078 (43.2)	
Nitrous oxide	167 (6.7)	146 (5.8)	
Pudendal, paracervical, saddle block	8 (0.3)	4 (0.2)	
General	16 (0.6)	22 (0.9)	
Other§	6 (0.2)	2 (0.1)	
None	112 (4.5)	112 (4.5)	1.06 (0.54 to 1.58)*
Continuous electronic fetal heart rate monitoring	2117 (84.8)	2160 (86.4)	0.84 (0.59 to 1.08)*
Method of delivery:			
Spontaneous vaginal delivery	1597 (64.0)	1533 (61.3)	1.12 (0.96 to 1.27)*
Instrumental vaginal delivery	341 (13.7)	362 (14.5)	
Vacuum	231	240	
Forceps (low or mid)	110	122	
Caesarean delivery	559 (22.4)	604 (24.2)	0.90 (0.71 to 1.10)*
Perineal trauma requiring suturing:	1336 (53.5)	1350 (54.0)	0.98 (0.82 to 1.13)*
Episiotomy	569	573	
Second degree laceration	764	790	
Third or fourth degree laceration	131	111	
Other	3	0	
Maternal death	1¶	0	
Health problems during postnatal stay:			
Postnatal fever	24 (1.0)	23 (0.9)	
Haemorrhage >1000 cc	51 (2.0)	49 (2.0)	
Transfusion given	13 (0.5)	7 (0.2)	
Other**	10 (0.4)	4 (0.2)	
Length of postnatal hospital stay, median (interquartile range), hours	50.1 (41.4, 63.5)	50.3 (41.2, 64.1), P=0.75‡‡	

*Spontaneous vaginal delivery was primary outcome; prespecified confidence interval 95%. Oxytocin, regional analgesia, electronic fetal heart rate monitoring, and caesarean delivery were "other" outcomes; prespecified confidence interval 99.5%. Perineal trauma requiring suturing was a secondary outcome; prespecified confidence interval 98.75%.
†Some women had more than one form of analgesia or anaesthesia.
‡Epidural analgesia, combined spinal anaesthesia and epidural, or spinal anaesthesia.
§Sterile water injections (n=7) and intrathecal opioid (n=1).
¶Due to undetected haemorrhage from uterine artery after caesarean delivery. Data safety and monitoring committee concluded that death was unrelated to the trial.
**Such as hospital acquired pneumonia; severe pregnancy induced hypertension; septic pelvic thrombophlebitis; severe endometriosis; major delivery complications (tear of small bowel, cystostomy, bladder tear, severe bleeding requiring laparotomy, hysterectomy).
‡‡Prespecified "other" outcome.

METHODS

The study was a multicentre, randomised controlled trial with prognostic stratification by hospital. A group of nurses or midwives at each hospital were trained in the structured approach (see box on bmj.com).

Eligible hospitals had to have a pre-existing, separate labour assessment unit and a spontaneous vaginal delivery rate of 75% or less. In North America the units were staffed by nurses and in the UK by midwives, but the approach was the same—to determine whether a woman should be admitted to the labour ward or sent home to await active labour. Women were eligible for the study if they were nulliparous, had a live singleton fetus in the cephalic position, had no contraindications to labour, were able to give informed consent, and were experiencing contractions but did not meet labour ward criteria for admission.

Treatment protocol

Baseline data were obtained after consent was obtained and before randomisation; the nurse or midwife accessed the trial website to obtain the participant's study group allocation. Women assigned to the experimental group received one to one care by a nurse or midwife trained in structured care. The structured care provider determined fetal position and asked the woman to describe her thoughts during the last contraction and to rate pain on a visual analogue scale. The provider used positioning techniques, comfort measures, and simple cognitive restructuring techniques such as positive visual imagery and reframing negative thoughts, and offered anticipatory guidance about coping with active labour (see box on bmj.com).

Women assigned to the control group received care by a nurse or midwife without training in structured care. One nurse or midwife often provided care to more than one woman.

The length of time women received structured care or usual care was designed to reflect the usual time spent by women in labour assessment units (1-4 hours). In both groups the decision on whether to admit women to the labour ward or to send them home was made as per usual hospital policy. No woman was invited to participate while a trial participant was in the labour assessment unit unless privacy could be assured.

Compliance

Centres were instructed not to randomise women unless providers of usual care and structured care were available. Structured care providers could care for non-study women. Usual care providers could not care for women in the structured care group. Compliance was assessed in two ways. We expected that over 90% in each group would receive their assigned

method of care immediately after randomisation. In addition, providers' reports of their care for women in the structured care group provided evidence of adherence to the main components of the intervention.

Outcomes

The primary outcome was spontaneous vaginal birth. Secondary outcomes were the number of women who had no intrapartum analgesia or anaesthesia, had perineal trauma requiring suturing, and reported negative views of their care. Other study outcomes included the number of women with more than two visits for assessment of labour; use of intrapartum oxytocics, regional analgesia, and electronic fetal heart rate monitoring; length of hospital stay; and indicators of short term and longer term maternal and neonatal morbidity, including postnatal emotional distress, readmission to hospital of mother or baby for delivery related complications during 6-8 weeks after birth, neonatal transfer to a special care nursery, and fetal death or neonatal death.

Trained research nurses or midwives at each hospital abstracted data from the medical records and entered them into forms on the trial website.

Participants were asked to complete a questionnaire 6-8 weeks after the birth, focusing on their health, their baby's health, and their satisfaction with care. The questionnaire included the Edinburgh postnatal depression scale (score >12 indicates postnatal depressive symptomatology).[13] A systematic review identified key factors influencing satisfaction with childbirth,[14] and the questionnaire items were adapted from one of the most reliable and well validated population based surveys of satisfaction with childbirth.[15]

Statistical analysis

We analysed the results according to intention to treat. For the primary outcome we used a significance level of 0.05 (two tailed). We set the significance level for secondary outcomes at 0.0125 and for other study outcomes at 0.005. Because we expected variation owing to the effects of unknown characteristics of the hospitals, the analytical approach allowed the proportion of women experiencing spontaneous vaginal birth and treatment effects to vary between hospitals. For binary outcome variables we compared the groups using a logistic regression model with a random hospital effect for the intercept and slope. We present the odds ratios and accompanying confidence intervals (corresponding to the preset P values for primary, secondary, and "other" outcomes). We used a similar logistic regression model to explore the interaction effects between baseline variables and treatment group on the primary outcome. For length of hospital stay we analysed data using a linear regression model with a random hospital effect for the intercept and slope, using the log of length of stay as the dependent variable. Statistical procedures were done using SAS version 9.1. For ratings of women's views of their care we followed the standard practice of comparing the frequencies with which less than very positive views were reported.[14 15]

RESULTS

Twenty hospitals participated in the trial, eight in Canada, 10 in the United States, and two in the UK. Training in structured care was provided to 505 nurses and midwives; the remaining 1351 were available to provide usual care. The labour assessment units varied in design, size, and staffing.

We enrolled 5002 women between 1 May 2003 and 6 March 2007 (see bmj.com). Immediately after randomisation allocated care was provided to 2412 of 2501 women (96.6%) in the structured care group and to 2497 of 2501 women (99.8%) in the usual care group (see bmj.com). Structured care providers completed forms describing their activities for 2406 of the 2497 women in the structured care group (96.4%). One or more structured care interventions were provided to all but seven women (99.8%).

Outcomes

Primary and secondary outcomes

The rate of spontaneous vaginal delivery was 64.0% (n=1597) in the structured care group and 61.3% (n=1533) in the usual care group (odds ratio 1.12, 95% confidence interval 0.96 to 1.27; table). The groups were comparable for women who had no intrapartum analgesia or anaesthesia and for those requiring suturing for perineal trauma (table). Fewer women allocated to structured care (n=403, 19.5%) rated staff helpfulness as less than very helpful than those allocated to usual care (n=544, 26.4%); odds ratio 0.67, 98.75% confidence interval 0.50 to 0.85. Fewer women allocated to structured care (n=233, 11.3%) were disappointed with the amount of attention received from staff than those allocated to usual care (n=407, 19.7%); odds ratio 0.51, 98.75% confidence interval 0.32 to 0.70.

Other immediate maternal outcomes

Comparable numbers of women in both groups were sent home from the labour assessment unit on more than two occasions. In the structured care group 84.6% (n=2112) of women had regional analgesia, compared with 86.4% (n=2159) in the usual care group. The rate of caesarean delivery in the structured care group was 22.4% (n=559), compared with 24.2% (n=604) in the usual care group. One mother died due to haemorrhage from a uterine artery after caesarean delivery. Other immediate maternal outcomes were comparable between groups.

Mothers' and babies' health 6-8 weeks after discharge

In total, 76.0% (n=1570) of mothers in the structured care group rated their general health as excellent or very good compared with 74.7% (n=1542) in the usual care group. Postnatal depressive symptomatology was present in

WHAT IS ALREADY KNOWN ON THIS TOPIC

Prolonged latent phase labour is associated with increased risk of operative delivery and neonatal morbidity

Hospital labour assessment units are prevalent in North America but uncommon elsewhere

WHAT THIS STUDY ADDS

A formalised approach to care in hospital labour assessment units improves women's views of their care and may increase the likelihood of spontaneous vaginal birth

Labour assessment units may want to consider standardising the care provided to include assessment and interventions for maternal psychological state, pain, and positioning

134 (6.5%) mothers in the structured care group, compared with 149 (7.2%) in the usual care group (odds ratio 0.84, 99.5% confidence interval 0.36 to 1.32). Most mothers rated their baby's health as excellent or very good (95.7% in structured care group, 94.9% in usual care group). Forty four women in the structured care group and 37 women in the usual care group were readmitted for delivery related complications (odds ratio 1.19, 99.9% confidence interval 0.34 to 2.04). Sixty six babies in the structured care group and 83 in the usual care group were readmitted (0.78, 0.37 to 1.20).

DISCUSSION

We evaluated a structured approach to nursing or midwifery care in hospital labour assessment units, which included assessment of and interventions for maternal emotional state, fetal position, and pain, during a minimum of one hour. With the important exception of women's views of their care, results did not reach conventional levels of statistical significance. The trend towards increased likelihood of spontaneous vaginal birth indicates that further refinement of the intervention is warranted.

Compliance was excellent, and reports from the providers of structured care indicated that the intervention was applied appropriately and consistently across and within sites. In this large multicentre trial it would have been prohibitively expensive to directly observe the providers' actions. We took several measures to prevent contamination. Throughout the trial we emphasised the importance of maintaining distinct study groups and the uncertainty of the value of the experimental approach. Staff providing structured care were volunteers who were favourably disposed towards the type of care. Staffing was such that usual care rarely allowed for one to one attention for 1-4 hours, as required in structured care. Complex interventions such as structured care have the advantage of mirroring the real world of practice, in which assessments and interventions are tailored to individual needs. Furthermore, synergistic effects among components of an intervention would be lost if each were evaluated individually.[16] However, complex interventions have the disadvantage of leaving some uncertainty about the importance of each component of the intervention.[17] Our approach reflected current best practice guidelines, by addressing context, collecting the best evidence, developing a conceptual model to explain the links between intervention components and outcome, and standardising the intervention.[16 18] None the less, the labour assessment units did not seem to delay admission to the labour ward, as nearly 60% of participants (n=2897) were not in active labour when admitted (see bmj.com).

Given the low intensity of the intervention, the absence of evidence of risk, the potential population effects if it were adopted, the continuing rise in caesarean delivery rates, and the beneficial effect on satisfaction with care, hospitals with labour assessment units may want to consider incorporating structured care into routine practice.

A combination of structured care plus strict adherence to a policy of delayed admission to the labour ward until clinically indicated may yield greater benefits. Questions remain about the optimum setting for women in latent phase labour and about the characteristics of hospitals that influence the effectiveness of forms of intrapartum nursing or midwifery care.

We thank the participants, staff at the participating hospitals, and the structured care trainers (Melrose East, Adele Hood, Debbie Kaye, Debbie Patrick, Anne Simmonds, Ann Sprague, and Marie-Josée Trépanier); members of the Data Safety Monitoring Committee (see bmj.com; members of the Structured Early Labour Assessment and Care by Nurses Trial Group who enrolled participants and collected data (see bmj.com).
Contributors: See bmj.com.
Funding: Canadian Institutes of Health Research (grant No MCT59614).
Competing interests: None declared.
Ethical approval: Research ethics committees at the University of Toronto and by all participating hospitals.
Provenance and peer review: Not commissioned; externally peer reviewed.

1 Lederman R, Lederman E, Work B, McCann D. Relationship of psychological factors in pregnancy to progress in labor. *Nurs Res* 1979;28:94-7.
2 Beck N, Siegel L, Davidson N, Kormeier S, Breitenstein A, Hall D. The prediction of pregnancy outcome: maternal preparation, anxiety, and attitudinal sets. *J Psychom Res* 1980;24:343-51.
3 Crandon A. Maternal anxiety and obstetric complications. *J Psychosom Res* 1979;23:109-11.
4 Erickson M. The relationship between psychological variables and specific complications of pregnancy, labor, and delivery. *J Psychosom Res* 1976;20:207-10.
5 Wuitchik M, Bakal D, Lipshitz J. The clinical significance of pain and cognitive activity in latent labor. *Obstet Gynecol* 1989;73:35-42.
6 Stremler R, Hodnett E, Petryshen P, Stevens B, Weston J, Willan A. Randomized trial of hands-and-knees position for occipitoposterior position in labor. *Birth* 2005;32:243-51.
7 Pearl ML, Roberts JM, Laros RK, Hurd WW. Vaginal delivery from the persistent occiput posterior position: influence on maternal and neonatal morbidity. *J Reprod Med* 1993;38:955-61.
8 Gardberg M, Laakkonen E, Salevaara M. Intrapartum sonography and persistent occiput posterior position: a study of 408 deliveries. *Obstet Gynecol* 1998 May;91(5 Pt 1):746-9.
9 Rutherford AM. The management of the occipito-posterior position: a prospective study of 145 cases in 1979. *NZ Med J* 1981;94:419-21.
10 Saunders NJ, Spiby H, Gilbert L, Fraser RB, Hall JM, Mutton PM, et al. Oxytocin infusion during second stage of labour in primiparous women using epidural analgesia: a randomised double blind placebo controlled trial. *BMJ* 1989;299:1423-6.

11 Menacker F. Trends in cesarean rates for first births and repeat cesarean rates for low-risk women: United States, 1990–2003. *Natl Vital Stat Rep* 2005;54:1-8.
12 US Department of Health and Human Services. *Healthy people 2010*. Rockville, MD: USDHHS, 2000.
13 Cox J, Holden J. *Perinatal psychiatry: use and misuse of the Edinburgh postnatal depression scale*. London: Gaskell, 1994.
14 Hodnett E. Pain and women's satisfaction with the experience of childbirth: a systematic review. *Am J Obstet Gynecol* 2002;186:S160-72.
15 Brown S, Lumley J. Satisfaction with care in labour and birth: a survey of 790 Australian women. *Birth* 1994;21:4-13.
16 Stephenson J, Imrie J. Why do we need randomised controlled trials to assess behavioural interventions? *BMJ* 1998;316:611-3.
17 Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. *BMJ* 2007;334:455-9.
18 Hawe P, Shiell A, Riley T. Complex interventions: how "out of control" can a randomised controlled trial be? *BMJ* 2004;328: 1561-3.

Accepted: 27 June 2008

Implementation of computerised physician order entry (CPOE) and picture archiving and communication systems (PACS) in the NHS: quantitative before and after study

Simon Collin,[1] Barnaby C Reeves,[2] Jane Hendy,[3] Naomi Fulop,[4] Andrew Hutchings,[5] Eugenia Priedane[5]

[1]Department of Social Medicine, University of Bristol
[2]Department of Clinical Sciences at South Bristol, University of Bristol, Bristol
[3]Health and Care Infrastructure Research and Innovation Centre, Imperial College, London
[4]Health and Health Policy NIHR King's Patient Safety and Service Quality Research Centre, King's College, London
[5]Department of Public Health and Policy, London School of Hygiene and Tropical Medicine

Correspondence to: B C Reeves, Clinical Trials and Evaluation Unit, University of Bristol, Level 7 Bristol Royal Infirmary, Bristol BS2 8HW
barney.reeves@bristol.ac.uk

Cite this as: *BMJ* 2008;337:a939
doi:10.1136/bmj.a939

ABSTRACT

Objective To assess the impact of components of the national programme for information technology (NPfIT) on measures of clinical and operational efficiency.

Design Quasi-experimental controlled before and after study using routinely collected patient level data.

Setting Four NHS acute hospital trusts in England.

Data sources Inpatient admissions and outpatient appointments, 2000-5.

Interventions A system for ordering pathology tests and browsing results (computerised physician order entry, CPOE) and a system for requesting radiological examinations and displaying images (picture archiving and communications system, PACS).

Main outcome measures Requests per inpatient, outpatient, or day case patient for full blood count, urine culture, and urea and electrolytes tests, and plain x ray film, computed tomography, and ultrasonography examinations.

Results CPOE was associated with a reduction in the proportion of outpatient appointments at which full blood count (odds ratio 0.25, 95% confidence interval 0.16 to 0.40), urea and electrolytes (0.55, 0.39 to 0.77), and urine culture (0.30, 0.17 to 0.51) tests were ordered, and at which full blood count tests were repeated (0.73, 0.53 to 0.99). Conversely, the same system was associated with an almost fourfold increase in the use of urea and electrolytes tests among day case patients (3.63, 1.66 to 7.94). PACS was associated with a reduction in repeat plain x ray films at outpatient appointments (0.62, 0.44 to 0.88) and a reduction in inpatient computed tomography (0.83, 0.70 to 0.98). Conversely, it was associated with increases in computed tomography requested at outpatient appointments (1.89, 1.26 to 2.84) and computed tomography repeated within 48 hours during an inpatient stay (2.18, 1.52 to 3.14).

Conclusions CPOE and PACS were associated with both increases and reductions in tests and examinations. The magnitude of the changes is potentially important with respect to the efficiency of provision of health care. Better information about the impact of modern IT is required to enable healthcare organisations to manage implementation optimally.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a939

INTRODUCTION

The rate at which information technology (IT) systems are being ordered and deployed by healthcare providers around the world has far outpaced the growth of the evidence base of clinical and operational benefits associated with such systems.[1 2] In the United Kingdom, an estimated £20bn (€25bn; $37bn) over 10 years is being invested in the National Health Service (NHS) national programme for information technology (NPfIT).[3]

We previously reported the findings of a qualitative study that assessed challenges and progress in implementing NPfIT in four NHS acute hospital trusts in England.[4 5] Here we report a quantitative assessment of the implementation of a system for ordering pathology tests and browsing results (referred to here as computerised physician order entry or CPOE) and a system for requesting radiological examinations and storing and displaying images (referred to as a picture archiving and communications system or PACS) in the same four trusts.

METHODS

Study design—We selected four trusts representing a range of characteristics of NHS hospital trusts (size, financial situation, and state of information technology development). We used a quasi-experimental "controlled before and after cohort" design,[6] with each trust as a unit of the experiment, to quantify the effects of IT systems implemented in 2000-5.

Outcomes—Our outcomes were proxy measures of clinical and operational efficiency derived from a larger set of indicators that had been defined a priori, based partly on consideration of the NHS efficiency map.[7] We classified outcomes as primary or secondary depending on whether a direct causal pathway between

implementation of an IT system and the outcome was plausible or not (see table A on bmj.com). We refer to changes in outcomes as "efficiency gains" where we consider the change to reflect an improvement in clinical or operational efficiency—for example, a reduction in the number of pathology test orders—and as "detrimental" if the opposite.

Data analysis—For a description of data sources see bmj.com. We estimated effects by multiple regression modelling, calculating robust standard errors to take into account clustering of individual records by the common specialties (seven inpatient, 18 outpatient) within the trusts, resulting in 28 clusters for inpatient data and 72 clusters for outpatient data. We assessed effects on length of stay and time to death by Cox regression, after checking the proportional hazards assumption. For details see bmj.com.

RESULTS

Participating trusts and systems implementation—Trust 1 was the only trust to implement a CPOE system. This system provided test ordering (with automated form filling, order sets, warnings of possible test duplication, and user defined rules) and access to previous test results. Trust 4 was the only trust to implement PACS. This system provided web based access to requested and archived images and was implemented together with a new (but separate) system for requesting examinations.

CPOE primary outcomes—Table 1 summarises the results of the comparisons for implementation of CPOE. Evidence for possible efficiency gains was most apparent in the reduction in outpatient tests. This effect was seen for full blood count, urea and electrolytes, and urine culture tests. There was also an effect of CPOE in reducing "repeat" full blood count tests at outpatient appointments. Conversely, it was associated with an almost fourfold increase in the use of urea and electrolytes tests among day case patients.

PACS primary outcomes—Table 2 summarises the effects of implementing PACS. Evidence for possible efficiency gains was apparent in the reduction in repeat plain x ray film exams at outpatient appointments and in the reduction in inpatient computed tomography. Conversely, implementation of PACS was associated with increases in computed tomography requested at outpatient appointments and computed tomography repeated within 48 hours during an inpatient stay. Ultrasonography was not a component of the PACS in trust 4, and there was no evidence of changes in outcomes.

Secondary outcomes—Comparisons of the impact of the two systems on secondary outcomes showed evidence of detrimental effects of CPOE and PACS in reducing the proportion of outpatients discharged, a detrimental effect of CPOE in reducing outpatient attendance, and a beneficial impact of CPOE in reducing inpatient deaths. See bmj.com for more results.

DISCUSSION

Two IT systems showed both benefit and detriment on various efficiency outcomes. We found evidence for an effect of CPOE on five out of 18 primary outcomes and on three out of seven secondary outcomes; and for PACS, on four of 17 primary outcomes and one of eight secondary outcomes. Of the five effects on primary outcomes attributable to CPOE, four were indicative of efficiency gains; for PACS, two out of four. For a discussion of the impact of CPOE and PACS on specific outcomes see bmj.com.

Study in context

CPOE in trust 1 and PACS in trust 4 were considered by managers and end users to have been successful

Table 1 | Implementation of CPOE in trust 1 compared with trusts 2, 3, and 4. Figures are odds ratios, or regression coefficients where specified (95% confidence intervals) for interaction between intervention (in trust 1) and period after intervention (2003-5) and mean change for intervention trust *v* control trusts

Primary outcomes*	Full blood count	Urea and electrolytes†	Urine culture
Inpatient			
Tests per inpatient: non-zero *v* zero response	0.74 (0.48 to 1.16)	0.66 (0.43 to 1.02)	1.14 (0.80 to 1.63)
Change	1.9% *v* 1.1%	7.8% *v* 5.3%	−4.3% *v* 3.7%
Tests per inpatient day: continuous non-zero response	1.00‡ (0.90 to 1.10)	1.03‡ (0.89 to 1.18)	0.93‡ (0.82 to 1.06)
Change	0.05 *v* 0.03	0.08 *v* 0.05	−0.01 *v* 0.05
Tests per day case: non-zero *v* zero response	1.76 (0.78 to 3.99)	3.63 (1.66 to 7.94)	1.29 (0.54 to 3.13)
Change	6.5% *v* 2.2%	8.0% *v* 5.9%	1.9% *v* 1.2%
Test within 48 hours of previous test of same type (inpatients)	0.93 (0.79 to 1.10)	1.07 (0.89 to 1.29)	0.89 (0.70 to 1.12)
Change	1.6% *v* 3.8%	−0.2% *v* 0.5%	−1.4% *v* −0.1%
Outpatient			
Test(s) at outpatient appointment	0.25 (0.16 to 0.40)	0.55 (0.39 to 0.77)	0.30 (0.17 to 0.51)
Change	−1.9% *v* 4.6%	−0.6% *v* 3.6%	−0.5% *v* 1.5%
Test of same type at next outpatient appointment	0.73 (0.53 to 0.99)	0.84 (0.64 to 1.11)	0.73 (0.52 to 1.02)
Change	0.6% *v* 4.3%	4.3% *v* 6.0%	0.4% *v* 2.3%

*See table B on bmj.com for full data for each trust.
†No data contributed by trust 2.
‡Exponent of regression coefficient.

Table 2 | Implementation of PACS in trust 4 compared with trusts 1, 2, and 3. Figures are odds ratios, or regression coefficients where specified (95% confidence intervals) for interaction between intervention (in trust 4) and period after intervention (2003-5) and mean change for intervention trust *v* control trusts

Primary outcomes*	Plain x ray film	Computed tomography	Ultrasonography†
Inpatient			
Exams per inpatient: non-zero *v* zero response	0.90 (0.71 to 1.14)	0.83 (0.70 to 0.98)	0.89 (0.69 to 1.14)
Change	1.0% *v* 4.1%	2.1% *v* 3.0%	−1.3% *v* 0.5%
Exams per inpatient day: continuous non-zero response	0.97‡ (0.90 to 1.05)	1.02‡ (0.91 to 1.14)	0.96‡ (0.85 to 1.09)
Change	0.02 *v* 0.02	0.02 *v* 0.05	−0.01 *v* 0.00
Exams per day case: non-zero *v* zero response	1.01 (0.55 to 1.86)	0.73 (0.31 to 1.73)	1.55 (0.83 to 2.89)
Change	7.0% *v* 5.2%	0.7% *v* 0.7%	0.0% *v* −0.2%
Exam within 48 hours of previous exam of same type (inpatients)	1.02 (0.91 to 1.14)	2.18 (1.52 to 3.14)	1.08 (0.81 to 1.44)
Change	−3.2% *v* −4.3%	1.2% *v* −0.1%	0.2% *v* 0.2%
Outpatient			
Exam(s) at outpatient appointment	0.90 (0.76 to 1.07)	1.89 (1.26 to 2.84)	1.48 (0.60 to 3.66)
Change	1.0% *v* 0.0%	0.2% *v* 0.1%	1.9% *v* −0.1%
Exam of same type at next outpatient appointment	0.62 (0.44 to 0.88)	NA	0.58 (0.19 to 1.82)
Change	−1.2% *v* 4.6%	NA	−10.4% *v* −2.2%

NA=not analysed because of insufficient numbers.
*See table C on bmj.com for full data for each trust.
†Ultrasonography not included in PACS in intervention trust (trust 4).
‡Exponent of regression coefficient.

implementations of these types of healthcare IT system.[8] The NHS is leading the way in terms of the scale and homogeneity of its healthcare IT programme and, although running behind schedule and over budget, the programme continues to receive the support of managers and clinicians alike.[4,5] CPOE and PACS, when fully integrated with the other information technology systems that comprise NPfIT (national electronic health records, patient administration systems, electronic referral, etc), might contribute to more dramatic quantitative changes.

Our study has shown that it is possible to use routinely collected patient level data from disparate sources within large healthcare institutions as a basis for assessing the impact of technological changes on indicators of clinical activity and operational efficiency. Implications for future research are discussed in a full paper at bmj.com.

Limitations

Although our study benefited from a large number of observations, adjustment for clustering by site and specialty gave rise to large standard errors. Hence, although there seemed to be evidence of potentially important effects for many outcomes, few could be measured with sufficient precision in our final analysis. We restricted our analyses to specialties common to all of the participating trusts, but our results remain susceptible to residual confounding because of differences in case mix between trusts. Confounding was a particular concern in the few instances where the indicator data showed substantial differences between trusts. Inclusion of specialty as a covariate in our regression models to control for differences in case mix, however, did not tend to change our point estimates.

We could not verify data quality, although outpatient data from the commissioning datasets have been assessed as reliable.[9] Data on pathology tests and radiology examinations were unlikely to contain important omissions as these were obtained directly from pathology and radiology information systems used routinely to manage all requests. Some omissions might have arisen by using local patient identifiers to join these data with data from the commissioning datasets. We had no means to verify the reliability of this process, but the reasonable consistency of our outcome measures between trusts was reassuring.

Our "repeat investigation" measure was a proxy for redundant tests and was dictated by the data available from routine sources. We did not have the level of detail necessary to determine whether tests repeated within this interval were redundant—for example, redundant tests have typically been identified by chart review, or whether they reflected good clinical practice. Our method might not be equally applicable across specialties, but we found no evidence to the contrary by comparing the distributions of times to retest within specialties. Our choice of interval (48 hours) was

WHAT IS ALREADY KNOWN ON THIS TOPIC

The NHS national programme for IT is expected to contribute considerable gains in efficiency

Evidence of quantifiable efficiency gains attributable to healthcare IT systems is limited

WHAT THIS STUDY ADDS

CPOE and PACS were associated with possible efficiency gains in some areas, particularly in ordering of outpatient pathology tests and requests for repeat plain x ray film examination

CPOE and PACS were also associated with possible efficiency reductions in other areas

Changes in efficiency from healthcare IT systems based on routinely derived indicators are difficult to quantify because they are difficult to interpret and measure

Assumptions of substantial efficiency gains from healthcare IT systems might be unrealistic

consistent with that used in other studies.[10-12] Misclassification of tests as redundant when clinically required, and vice versa, would probably have been non-differential and hence would have caused underestimation of underlying effects.

In addition to improved efficiency of delivery of care, modernisation of IT could improve patients' health outcomes, most obviously through better patient safety. Because our study was designed to take advantage of routinely collected data, we were unable to investigate the impact of CPOE and PACS on health outcomes other than death and overall length of stay.

Conclusions

Efficiency gains from healthcare IT systems are difficult to quantify. Changes in routinely derived indicators are difficult to interpret and measure. We observed both beneficial and detrimental, or at least unexpected, changes so assumptions of substantial efficiency gains from healthcare IT systems might be unrealistic. Given the large overall benefit that would accrue from small efficiency gains occurring in all trusts across the NHS, further research is justified. Although our underlying methods are promising, quantitative research must be closely allied with qualitative research to provide context and to explain observed changes.

Members of the steering group were Aileen Clarke, David King, David Lawrence, Charles Normand, Michael Soljak, and Ken Walton. We thank pathology, radiology, and information technology departments in the trusts and trust managers who provided the data or facilitated this process, and Cristina Masseria, who worked on the project for a year.
Contributors: See bmj.com.
Funding: NHS service and delivery organisation research and delivery programme (SDO/44/2003).
Competing interests: None declared.
Ethical approval: NHS Trent multicentre research ethics committee (MREC/03/4/017) and NHS trust local research ethics committees.
Provenance and peer review: Not commissioned; externally peer reviewed.

1 Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006;144:742-52.
2 Georgiou A, Williamson M, Westbrook JI, Ray S. The impact of computerised physician order entry systems on pathology services: a systematic review. *Int J Med Inform* 2007;76:514-29.
3 Department of Health. *Connecting for health*. London: DH, 2008. www.connectingforhealth.nhs.uk/
4 Hendy J, Reeves BC, Fulop N, Hutchings A, Masseria C. Challenges to implementing the national programme for information technology (NPfIT): a qualitative study. *BMJ* 2005;331:331-6.
5 Hendy J, Fulop N, Reeves BC, Hutchings A, Collin S. Implementing the NHS information technology programme: qualitative study of progress in acute trusts. *BMJ* 2007;334:1360.
6 Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin, 2002.
7 Department of Health. *The efficiency map*. London: DH, 2005. www.dh.gov.uk/assetRoot/04/11/75/50/04117550.pdf
8 Reeves BC, Fulop N, Hendy J, Hutchings A, Collin S, Priedane E, et al. *Evaluation of IT modernisation in the NHS. Service*. London: National Institute for Health Research, 2008. www.sdo.nihr.ac.uk/sdo442003.html
9 *Outpatient data quality report 2003-03 and 2004-05*. London: The Information Centre, 2006. www.ic.nhs.uk/webfiles/publications/outpatientdq/OutpatientDataQualityReport151206_PDF.pdf
10 Bates DW, Kuperman GJ, Rittenberg E, Teich JM, Fiskio J, Ma'luf N, et al. A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests. *Am J Med* 1999;106:144-50.
11 Bates DW, Boyle DL, Rittenberg E, Kuperman GJ, Ma'luf N, Menkin V, et al. What proportion of common diagnostic tests appear redundant? *Am J Med* 1998;104:361-8.
12 Weydert JA, Nobbs ND, Feld R, Kemp JD. A simple, focused, computerized query to detect overutilization of laboratory tests. *Arch Pathol Lab Med* 2005;129:1141-3.

Accepted: 1 June 2008

A tale of two families

Since about the age of 12 years I have been a self appointed family archivist. I really wanted to know how everyone I met fitted into the scheme of things. In the early 1960s I was taken by an elderly cousin of my father to an old churchyard near Closeburn in Dumfriesshire. She showed me the monument for the family of her grandmother's half-brother, a miller, William Rogerson. Born in 1817, he and his wife had eight children between 1843 and 1861. All were interred with him.

William himself died aged 43; according to the certifying doctor the cause was phthisis. Earlier two daughters had died in infancy. His eldest daughter, Ann, died of phthisis aged 31 in 1874. Another daughter, Henrietta, died of "disease of the lungs" at age 21 in 1875. A further daughter, Margaret, died of phthisis in 1877 aged 27, and the only son, William, died of phthisis in 1889 aged 28. Ann was the only one to marry. I have been unable to discover what became of one of her sons, but the other died of phthisis in 1888 aged 20. As far as I can ascertain there were no other descendants.

William's younger half-sister, Mary Ann Rogerson, married a farm worker and had 13 children and 53 grandchildren. At the latest count there were 349 living descendants scattered over the globe.

The contrast between William's family and that of Mary Ann emphasises the terrible scourge that tuberculosis was in Britain in the 19th century. Flynn commented on this period: "As killers … both cholera and typhus were dwarfed by tuberculosis; and tuberculosis scarcely stirred the imagination of any social group in this period. It was so much a part of life, so inevitable, so little understood, that it was accepted mutely.… In the early nineteenth century it may have accounted for one third of all deaths."[1]

Despite the limitations of the diagnostic methods at the time, there is little doubt that William, several children, and at least one grandson died from tuberculosis. It is likely that concern about tuberculosis blighted the lives of the younger children who never married.

In the 20th century the incidence of tuberculosis was greatly reduced in the United Kingdom by efficient case finding and prompt treatment. The disease lost its prominence on the medical radar in the UK but remained a major cause of death in many countries. Even in the developed world the number of affected people is now rising again. The emergence of resistant strains has compounded the difficulties. Particularly alarming are the new "extreme drug-resistant strains" found in 45 countries and recently identified in the UK and the US. These forms of tuberculosis are virtually untreatable, and with them "we have returned to a situation analogous to the pre-antibiotic era."[2] It does not do to underestimate this most competent of pathogens.

Colin R Paterson retired reader in medicine, University of Dundee, Dundee
c.s.paterson@btinternet.com

1 Flynn MW. *Introduction to Edwin Chadwick, report on the sanitary condition of the labouring population of Great Britain (1842)*. Edinburgh: Edinburgh University Press, 1965.
2 Selgelid M. A killer we left to roam. *Guardian* 2008 Mar 23. http://commentisfree.guardian.co.uk/