1. UN summit 2005: gridlock at the gabfest • EDITORIAL
   Page 1051
   The Lancet

2. Pragmatism over opium production in Afghanistan • EDITORIAL
   Page 1052
   The Lancet

3. Children's health coming of age in Europe • EDITORIAL
   Page 1052
   The Lancet

4. Outbreak from a high-toxin intruder: *Clostridium difficile* • DISCUSSION
   Pages 1053-1054
   Torbjörn Norèn

5. Antimicrobial treatment for cholera • DISCUSSION
   Pages 1054-1055
   Jacqueline L Deen and John D Clemens

6. Reducing cancer drug doses in obese patients: dogma disputed • DISCUSSION
   Pages 1056-1057
   Lars Holmberg and Rulla Tamini

7. Is poverty or wealth at the root of HIV? • DISCUSSION
   Pages 1057-1058
   James D Shelton, Michael M Cassell and Jacob Adetunji

8. The metabolic syndrome—a new worldwide definition • DISCUSSION
   Pages 1059-1062
   K George MM Alberti, Paul Zimmet, Jonathan Shaw and for the IDF Epidemiology Task Force Consensus Group

9. Developing an open relationship with the drug industry • DISCUSSION
   Pages 1062-1064
   Patrick Vallance
<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Section</th>
<th>Pages</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.</td>
<td>As Niger's emergency eases, another crisis looms</td>
<td>NEWS</td>
<td>1065-1066</td>
<td>Clare Kapp</td>
</tr>
<tr>
<td>11.</td>
<td>Why did helping Niger take so long?</td>
<td>NEWS</td>
<td>1067-1068</td>
<td>Samuel Loewenberg</td>
</tr>
<tr>
<td>12.</td>
<td>Prejudice in a portrayal of Huntington's disease</td>
<td>BOOK REVIEW</td>
<td>1069-1070</td>
<td>Nancy S Wexler and Michael D Rawlins</td>
</tr>
<tr>
<td>13.</td>
<td>Fighting back against fat</td>
<td>BOOK REVIEW</td>
<td>1070</td>
<td>Anne Harding</td>
</tr>
<tr>
<td>14.</td>
<td>Empathy</td>
<td>DISCUSSION</td>
<td>1071</td>
<td>Rhodri Hayward</td>
</tr>
<tr>
<td>15.</td>
<td>Leena Peltonen</td>
<td>DISCUSSION</td>
<td>1071</td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>Jan Moor-Jankowski</td>
<td>PERSONAL REPORT</td>
<td>1072</td>
<td>Stephen Pincock</td>
</tr>
<tr>
<td>17.</td>
<td>Inadequate anaesthesia in lethal injection for execution</td>
<td>CORRESPONDENCE</td>
<td>1073</td>
<td>Jonathan I Groner</td>
</tr>
<tr>
<td>18.</td>
<td>Inadequate anaesthesia in lethal injection for execution</td>
<td>CORRESPONDENCE</td>
<td>1073-1074</td>
<td>Mark JS Heath, Donald R Stanski and Derrick J Pounder</td>
</tr>
<tr>
<td>19.</td>
<td>Inadequate anaesthesia in lethal injection for execution</td>
<td>CORRESPONDENCE</td>
<td>1074</td>
<td>Robyn S Weisman, Jeffrey N Bernstein and Richard S Weisman</td>
</tr>
<tr>
<td>20.</td>
<td>Inadequate anaesthesia in lethal injection for execution – Authors' reply</td>
<td>CORRESPONDENCE</td>
<td>1074-1076</td>
<td>Teresa A Zimmers, David A Lubarsky, Jonathan P Sheldon and Leonidas G Koniaris</td>
</tr>
<tr>
<td>21.</td>
<td><strong>Research fraud</strong> • CORRESPONDENCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Page 1076</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HS Bawaskar</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>22.</th>
<th><strong>Integration of HIV/AIDS and family planning</strong> • CORRESPONDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Pages 1076-1077</em></td>
</tr>
<tr>
<td></td>
<td>Jane Galvão and Jessica Halverson</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>23.</th>
<th><strong>Integration of HIV/AIDS and family planning</strong> • CORRESPONDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Page 1077</em></td>
</tr>
<tr>
<td></td>
<td>Jianghong Li</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>24.</th>
<th><strong>Industry-funded bioethics articles</strong> • CORRESPONDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Pages 1077-1078</em></td>
</tr>
<tr>
<td></td>
<td>James C Coyne</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>25.</th>
<th><strong>Industry-funded bioethics articles</strong> • CORRESPONDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Page 1078</em></td>
</tr>
<tr>
<td></td>
<td>Alexander C Tsai</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>26.</th>
<th><strong>Department of Error</strong> • ERRATUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Page 1078</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>27.</th>
<th><strong>Toxin production by an emerging strain of <em>Clostridium difficile</em> associated with outbreaks of severe disease in North America and Europe</strong> • ARTICLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Pages 1079-1084</em></td>
</tr>
<tr>
<td></td>
<td>Michel Warny, Jacques Pepin, Ai Qi Fang, George Killgore, Angela Thompson, Jon Brazier, Eric Frost and L Clifford McDonald</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>28.</th>
<th><strong>Single-dose ciprofloxacin versus 12-dose erythromycin for childhood cholera: a randomised controlled trial</strong> • ARTICLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Pages 1085-1093</em></td>
</tr>
<tr>
<td></td>
<td>Debasis Saha, Wasif A Khan, Mohammad M Karim, Hafizur R Chowdhury, Mohammed A Salam and Michael L Bennish</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>29.</th>
<th><strong>Comparison of the effect of two systems for the promotion of exclusive breastfeeding</strong> • ARTICLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Pages 1094-1100</em></td>
</tr>
<tr>
<td></td>
<td>Sonia Bechara Coutinho, Pedro Israel Cabral de Lira, Marilia de Carvalho Lima and Ann Ashworth</td>
</tr>
</tbody>
</table>
| 30. | **Nitrous oxide and risk of surgical wound infection: a randomised trial** • ARTICLE  
*Pages 1101-1107*  
Edith Fleischmann, Rainer Lenhardt, Andrea Kurz, Friedrich Herbst, Béla Fülesdi, Robert Greif, Daniel I Sessler, Ozan Akça and on behalf of the Outcomes Research Group |
| 31. | **Relation between chemotherapy dose, oestrogen receptor expression, and body-mass index** • ARTICLE  
*Pages 1108-1110*  
Marco Colleoni, Sigui Li, Richard D Gelber, Karen N Price, Alan S Coates, Monica Castiglione-Gertsch, Aron Goldhirsch and for the International Breast Cancer Study Group |
| 32. | **Presbycusis** • REVIEW ARTICLE  
*Pages 1111-1120*  
George A Gates and John H Mills |
| 33. | **Genetic association studies** • REVIEW ARTICLE  
*Pages 1121-1131*  
Heather J Cordell and David G Clayton |
| 34. | **End-of-life: a Catholic view** • DISCUSSION  
*Pages 1132-1135*  
Hazel Markwell |
| 35. | **A 21-year-old girl with recurrent abdominal pain after a robbery** • SHORT COMMUNICATION  
*Page 1136*  
Albert Selva-O’Callaghan, Jose Gómez-Acha, Pere Munne and Miquel Vilaridel-Tarrés |
UN summit 2005: gridlock at the gabfest

The 2005 World Summit, a “high-level plenary meeting of the 60th session of the General Assembly of the United Nations”, was the grand title of a dreary and lacklustre meeting held in New York on Sept 14–16.

The summit has been widely derided (a New York Times columnist called it “the greatest assembly in history—of hypocrites”), mainly for its watered-down outcome document, which was painfully agreed on the eve of the summit. The main results highlighted in the 35-page document included an agreement to create a new human rights council and a new peace-building commission (the membership of which countries began squabbling over immediately), a statement condemning terrorism (undefined), and a commitment to protect citizens “from genocide, war crimes, ethnic cleansing and crimes against humanity”, when their own countries have failed to do so.

With the release of the Volcker report on the Iraq “oil-for-food” scandal, several people calling for the resignation of Secretary-General Kofi Annan, and the publication of a tell-all memoir by a former UN official entitled The UN gang: a memoir of incompetence, corruption, espionage, anti-Semitism, and Islamic extremism at the UN Secretariat, all occurring virtually simultaneously, the meeting was not set to begin on a high note. Before the summit could get underway, reporters were asking whether it could already be labelled a fiasco. The draft document was called weak and held up as proof that the UN cannot be effective in the 21st century.

For those most concerned about the fate of the Millennium Development Goals (MDGs), with their aim to halve world poverty by 2015, there were one or two potentially bright spots. UN delegates committed an additional US$50 billion to the MDGs over the next 5 years, with every developing country agreeing to create a national plan by the end of 2006 for achieving the MDGs. And in what was thought to be the first time he had ever acknowledged the MDGs, US president George Bush affirmed the USA’s commitment to them in his opening address to the Assembly. Furthermore, he said the USA would end all trade barriers, if other nations would do the same, in accordance with the 2002 Doha trade negotiations. Perhaps giving 0.7% of the gross national product to developing countries suddenly did not seem like so much, given that the cost of rebuilding the US Gulf Coast after hurricane Katrina is now predicted to exceed $200 billion—an amount greater than the Millennium Project’s estimation of what is needed from all rich nations put together to meet the MDGs.

Perhaps more promising, and certainly more interesting than the endless drone of speeches at the UN, was the emergence of grassroots efforts on the part of private organisations to step in to solve global problems. These efforts, in which business leaders, philanthropists, academics, and private citizens are undertaking projects in the areas of poverty reduction, climate change, religious tolerance, and accountable governance, were highlighted across town, on the west side of Manhattan, where former US president Bill Clinton was hosting the inaugural meeting of the Clinton Global Initiative (where, to be fair, talk was not in short supply, either). At this meeting, participants (who each paid $15 000 to attend) were all expected to make a commitment, in writing, to some project aimed at alleviating global problems. Since the net financial worth of those assembled in the ballroom for the opening session of the meeting probably dwarfed any amount needed to fund the MDGs, we have high hopes for the former president’s forum.

By carving out innovative ways outside government to tackle big issues, Clinton is following in the footsteps of another former president, Jimmy Carter. By their actions, they are putting into practice what many ordinary people seem to be realising as well: conceding global problem-solving to giant lumbering bureaucracies like the UN increasingly looks like a bad bet. Now that the UN gathering has mercifully ended, the world will see whether or not money and action will take the place of all that talk.
Pragmatism over opium production in Afghanistan

On Sept 18, 2005, Afghanistan held elections to choose a national assembly and local councils for its 34 provinces—an important first step towards creating a democratic forum for debate about the country’s future. Key to this debate is a reconstruction agenda that is littered with broken promises, and threatened by the continuing battle to curb the cultivation and production of opium.

Afghanistan’s economy, culture, and political life are dominated by the trade in opium, which accounts for 60% of its gross domestic product. An estimated 350,000 rural households depend on an income from poppy cultivation, and alongside financing warlords and their militia, the opium produced accounts for up to 80% of the heroin consumed illegally worldwide. Crop eradication and alternative livelihood programmes have repeatedly failed to eliminate drug crops in Afghanistan. However, these tactics have succeeded in destroying livelihoods and health, and fuelling poverty.

The Senlis Council, an international drug-policy think tank, has proposed a possible solution: a strictly supervised licensing system in Afghanistan, to complement longer-term alternative development initiatives, allowing the cultivation of opium for the production of essential medicines such as morphine and codeine. The plan is modelled on programmes in Turkey and India that have helped reduce illegal opium production through US Congress-backed licensing schemes. Such a system could break the vicious circle of the drug economy in Afghanistan by moving the opium trade into a legal system controlled by and benefiting the state, and giving it a legitimate source of income to aid stability and economic development.

The preliminary results of this feasibility study on opium licensing are due to be presented at a conference in Kabul next week. This may be the only chance Afghanistan has to solve its drug problem, while providing a pragmatic and dynamic solution to its future peace, and meeting the vital public health objective of supplying essential medications to the developing world.

Children’s health coming of age in Europe

After The Lancet’s Child Survival series, together with a renewed commitment to tackling child mortality by UNICEF and WHO, it is heartening to see children’s health emphasised in WHO’s European Health Report, 2005. What is disheartening is that, in a region containing some of the world’s wealthiest countries, widening health inequalities remain the principal determinant of mortality.

The report builds on WHO’s 2004 action plan for children’s health and the environment. A major challenge is addressing heterogeneity in a region that extends from Greenland to Kamchatka, and includes 880 million people, almost half of whom live in poverty. Hence recommendations are broad, multifaceted, and mindful of local practicalities, including the importance of political and economic factors. One cause for concern is the proposal that public-health interventions are justified only if local effectiveness can be predicted. Because many member states lack sufficient data to make such judgments, waiting for this information is an invitation to inaction.

To promote action WHO encourages member states to share experiences, as do Sonia Bechara Coutinho and colleagues in their study of breast-feeding support in today’s Lancet. Here is the essence of children’s health: if a community acts collectively to support the care of its children, those children will enjoy better health, and in time may be better able to contribute to their community. Such is the rationale behind the UK’s 2003 green paper Every Child Matters, and an upcoming conference in Brussels on Nov 9–10 titled “Future Europeans”.

Two themes are common to these UK, European, and WHO initiatives. One is a call to involve children in planning and making decisions about issues that concern them. The second is for all agencies working with children to share examples of best practice. By both respecting and protecting children, these efforts emphasise their unique value and contribution to society.

In highlighting children’s health, WHO reminds us of opportunities to improve not only immediate health outcomes for this group, but also to lay the foundation for a healthier future for all generations.
Outbreak from a high-toxin intruder: Clostridium difficile

In today’s *Lancet*, we feel the “cold wind of another threatening epidemic” emerging in the face of our frequent use of broad-spectrum antibiotics within the hospital environment. Michel Warny and colleagues focus on the appearance of a highly toxin-producing *Clostridium difficile* genotype (NAP1/027) as a cause of a recent five-fold increase of nosocomial diarrhoea in Quebec, Canada.1 The great achievement of this study is the novel approach in relating clinical aspects such as morbidity and mortality to the known virulence factor of toxin production, as well as the sporulation capacity to nosocomial spread of disease. Despite the past, rather confusing array of different typing methods;1 the use of the three latest and most reliable ones give confidence to the epidemiological results.

Since 1978, *C difficile* has been clearly connected to antibiotic-associated diarrhoea (CDAD, *C-difficile*-associated diarrhoea). In the tracks of increased use of broad-spectrum antibiotics, the growing incidence of the disease has raised attention all over the western world. Different outbreaks with epidemic strains of *C difficile* have been well documented over the years.1 The progress of typing techniques has revealed dominant genotypes in the UK1 and now in Canada.1

For decades, standard treatment has been vancomycin or metronidazole with equal efficacy. Drawbacks, such as the rise of vancomycin-resistant enterococci, have limited the use of vancomycin. The appearance of metronidazole-resistant clinical *C difficile* isolates1 has discouraged the use of this first-line treatment. Recent reports of accelerated incidence, morbidity, and failure of the standard treatments1 support the clinical impression that we are dealing with a more aggressive disease. Efforts to investigate current strategies1 and alternative treatments, such as binding of toxin with polymers3 and vaccination,1 are underway to meet this therapeutic challenge.

It is a logical assumption, made by Warny and colleagues, that higher levels of toxin will increase bowel damage, frequency of diarrhoea, and morbidity. The increase of diarrhoea would in turn facilitate the spread of spores and promote epidemics. The preceding work by Pepin and colleagues;5 from the same Quebec area in Canada in 1991–2003, presented an increase of CDAD incidence from 35 per 100 000 to 156 per 100 000, and, in patients aged over 65 years, to 102–866 per 100 000. Mortality increased from 4% to 13% within a month of diagnosis and the progression to complicated disease was 79% lower when the infection was treated with vancomycin instead of metronidazole. In the subsequent years (2003–04), Warny and colleagues link this outbreak to a predominant *C difficile* strain (NAP1/027, toxinotype III, North American PFGE type 1 and PCR ribotype 027) in 67% of isolates from health-care facilities and in 37% of community-acquired isolates. Deletion in the gene tcdC, a putative downregulator of toxin A and B genes, is a presumptive explanation why all NAP1/027 isolates had a strikingly elevated production of toxin (16 times for toxin A and 23 times for toxin B) compared with non-dominant isolates. The in-vitro conditions do not necessarily mimic the situation in the bowel. Nevertheless, the efficacy of colonisation and morbidity caused by NAP1/027 could explain the local epidemic and possibly underestimate its role, considering the hyperproduction of toxins tested in “friendly” in-vitro media.

Compared with our stable and fairly high incidence of CDAD in Sweden of 50 per 100 000 from 1995,10 my experience in 2000 for Örebro county in Sweden is similar to that in Quebec. A dominant PCR ribotype SE17 strain (serogroup C) contributed to a mortality of 13%,11 similar to that in Warny and colleagues’ article.

Essentially all antimicrobials can induce CDAD by disrupting the gut ecology, and agents such as clindamycin and the broad-spectrum cephalosporins are the most incriminated villains.12 Warny and colleagues propose that the increased use of quinolones in hospital settings led to the present rise in CDAD incidence in Quebec.1 Previous data from the Quebec area1 showed a 45% increase in quinolone use between 1991 and 2003, but the incidence of CDAD per 1000 patient-days of antibiotic used still associate cephalosporins, clindamycin, and macrolides with the greatest incidence. Other studies
have shown that guidelines restricting the use of cephalosporins will reduce the incidence of CDAD.1,2 Now may be the right time to implement interventions against quinolone use.

The possibility of CDAD epidemics calls for more comparative studies on virulent strains and antibiotic use. By highlighting the clinical aggravation of CDAD together with the spread of a virulent C difficile strain, Wamy and colleagues’ article is of great value, supporting the new treatment strategy of neutralising the pathogenic bowel toxin, either by a toxin-binding polymer or a vaccine.

Torbjörn Norén
Department of Infectious Diseases, Orebro University Hospital and Department of Clinical Medicine, Orebro University, S-701 85 Orebro, Sweden
torbjorn.noren@orebroll.se

I declare that I have no conflict of interest. I am the National Coordinator for Sweden in the coming multicentre trial of a Polymer Alternative for CDAD treatment (Genzyme COOP, USA)


Antimicrobial treatment for cholera

Cholera is caused by strains of Vibrio cholerae belonging to serogroups O1 and O139. Toxin-mediated, massive outpouring of electrolyte-rich isotonic fluid into the bowel can lead to volume depletion and shock. Rehydration therapy, through intravenous or oral routes, markedly decreases case-fatality rates.1,2 Although antimicrobial therapy has not proven useful in prevention,3 it can reduce the total volume of stool passed and shorten both the duration of diarrhoea and the period of faecal excretion of V cholerae.4 Antibiotics are therefore recommended for all cases of suspected cholera with severe dehydration as a useful adjunct to fluid and electrolyte replacement, which remains the principal focus of treatment.5 The standard regimen for children is a 3-day (12-dose) course of either tetracycline or erythromycin, but neither is ideal because tetracycline therapy carries the risk of permanent discoulouration of teeth and erythromycin ingestion is frequently associated with vomiting and abdominal cramps. Additionally, both regimens require 3 days of compliance. In adults, single-dose therapy with doxycycline is well established for cholera.6

Today’s Lancet includes a paper from Bangladesh on an equivalence trial comparing single-dose ciprofloxacin versus the conventional multidose regimen of erythromycin for childhood cholera caused by V cholerae O1 or O139 and associated with severe dehydration.8 The authors are to be congratulated for the conduct of this challenging study. The results showed no difference between the groups in the rate of clinical cure, the primary outcome for the trial. Analyses of several secondary outcomes were discordant, with less stool output and a lower frequency of vomiting, but a lower frequency of bacteriological success (defined as inability to isolate cholera vibrios from stools after the second day following presentation for care) and a longer duration of faecal excretion of cholera vibrios in the ciprofloxacin group. The study also provides some assurance about the absence of arthropathy during the 4–6 weeks of follow-up in the ciprofloxacin-treated group. Because a single-dose regimen would simplify management and because generic ciprofloxacin is now sold at affordable prices in

Published online September 14, 2005
DOI:10.1016/S0140-6736(05)67326-6
See Articles page 1085.
Comment

many developing countries, the authors argue for the use of this regimen.

Although these findings are potentially important, there may be reasons to pause before considering them definitive. The study is weakened because it was not double-blinded. Awareness of the assigned study agent could have affected the ascertainment of outcomes, especially as the primary outcome (resolution of watery stools within 48 h without recurrence in the next 3 days), as well as several secondary outcomes, were subjectively assessed.

The persuasiveness of this study may also be diminished by the comparison of single-dose ciprofloxacin only with multidose erythromycin. Previously, The Lancet published a trial by the same research group in Bangladesh, showing that single-dose azithromycin was equivalent to standard erythromycin therapy in rates of clinical cures and bacteriological successes when used for severely dehydrating childhood cholera, and was also associated with a lower frequency of vomiting. By contrast with the study in today’s Lancet, the previous trial was double-blinded. These previous findings cause substantial ambiguity about whether ciprofloxacin, azithromycin, or both should be recommended as a single-dose alternative to conventional multidose therapy with erythromycin.

The clinical pertinence of the present trial’s findings would have been enhanced if single-dose azithromycin had been included as a compared regimen. A trial comparing single-dose regimens of these two antibiotics would be important to consider for the future.

Also needed are economic analyses of the incremental costs and benefits of adding different antibiotic regimens to rehydration in the treatment of severe childhood cholera. Not all cholera-endemic countries currently recommend the routine use of adjunctive antibiotic therapy, multidose or single-dose, for patients with severe cholera.

Meanwhile, cholera continues to disrupt the health and lives of populations around the world. Although the burden of cholera mortality worldwide cannot be accurately determined from routine public-health statistics, in part due to the disinclination of some countries to report cholera for fear of economic reprisals, most cholera experts estimate a continuing toll of at least 100 000 cholera deaths a year. Cholera outbreaks in west Africa over the past several weeks underscore the disease’s relentless impact on public health. While we continue to find ways to make basic rehydration therapy more available to vulnerable populations at risk, as well as to improve adjunctive treatment with antibiotics, we should also focus on preventive strategies, such as vaccination, to take advantage of the synergy between treatment and prevention approaches in minimising the effects of this dreaded disease.

Jacqueline L Deen, *John D Clemens
International Vaccine Institute, Kwanak, Seoul 151-818, Korea
j Clemens@ivi.int

We declare that we have no conflict of interest.

Oncologists often reduce the dose of chemotherapy in obese patients because of concern about overdosing. In today’s *Lancet*, investigators from the International Breast Cancer Study Group (IBCSG) report that dose reduction of adjuvant systemic chemotherapy in premenopausal women with oestrogen-receptor-negative breast cancer decreases disease-free and overall survival compared with when the full dose is given. As with all interesting scientific investigations, the report raises several stimulating questions. Indeed, the degree to which a study raises new issues can be a measure of how successful the investigation was.2

A first question is whether the result is just another subgroup finding. Subgroup analyses are inherently problematic.3,4 Multiple testing might produce spurious positive findings. Small subgroups yield low statistical power, increasing the likelihood of missing relevant differences. However, we can be somewhat reassured about the IBCSG findings: their a-priori hypothesis was based on a scientifically justified suspicion of heterogeneity in treatment effects between different strata. Thus the findings are a result of hypothesis testing rather than data dredging. The investigators relied on interaction tests rather than only on statistical significance testing in different subgroups. The findings make sense in view of other knowledge about breast cancer. However, none of these circumstances—and especially not the theoretical probability of the hypothesis—is in itself a criterion of truth,7 and so we must consider alternative explanations. Dose reduction was not randomly assigned in this study, making the analysis susceptible to bias and confounding. However, for bias to explain most of these findings, there would have to be a third factor strongly associated with both dose reduction and the development of oestrogen-receptor-negative (but not oestrogen-receptor-positive) tumours, and these relations would have to be strongest in obese women. The restrictions of the study base (randomised clinical trials with specific eligibility criteria and defined chemotherapy treatments) reduce variability in the study population and thus the potential for selection bias and confounding. Nevertheless, at least some confounding could have been assessed with additional models adjusted for potential confounders, including prognostic factors.

The next question is whether the findings should change clinical practice. The IBCSG investigators and others5–9 argue that the empirical data driving dose reduction in obese patients are weak. If this is true, the results of the current study might indeed contribute to correct a historical mistake. For overall survival, the numbers needed to give a full dose to save one life at a 10-year horizon would only be six in obese patients. At the lower bound of the 95% confidence interval for the difference in overall survival at 10 years, the numbers needed to treat with full dose would be 25, still a reasonable number.

A third question is whether the IBCSG findings should make us re-evaluate the evidence implicating obesity as a negative prognostic factor in women with breast cancer;10 is it just a matter of undertreatment? The IBCSG data raise the issue that treatment and dose should definitively be accounted for in future research of obesity and prognosis in breast cancer. However, it is unlikely that reduced dosing alone can explain the whole story. Not all obese women receive chemotherapy; of those that do, only a fraction will receive a reduced dose, and only some 30% of them have an oestrogen-receptor-negative tumour. Other explanations of the prognosis relation between obesity and breast cancer might come from non-clinical factors, such as diet and exercise.11 Recent data from the Women’s Intervention Nutrition Study suggest positive effects of dietary fat reduction on...
breast cancer prognosis, interestingly with a tendency for a larger effect in women with oestrogen-receptor-negative tumours.\(^3\)

A fourth—and perhaps the most interesting—question is to understand the biological mechanism producing these data. Why is there an interaction with hormone receptor status? Does obesity modify the response to chemotherapy within the oestrogen-receptor-negative subset? Obesity has been hypothesised to affect breast cancer prognosis through its effect on hormone levels, mainly oestrogens. The results of the current study should encourage us to consider oestrogen-independent mechanisms, such as insulin resistance, to explain these interactions. Adipocytokines, including leptin and adiponectin, are produced by adipose tissue, are known to modify insulin resistance, and have been reported to be associated with metastasis and angiogenesis.\(^4\)

We suspect that many other study groups have data available to test whether the findings by Colleoni and colleagues\(^1\) are the result of random variation, bias, or confounding, and to study toxicity from chemotherapy in obese patients. Given the effects seen, the issue has some urgency. If the interactions seen in the IBCSG studies are internally valid, this line of research adds one more piece to the puzzle about body-mass index and tumour biology. It would be interesting to see whether the same association reported by Colleoni also occurs in postmenopausal women. The only way to know more is if scientists find new and ingenious ways to address the intriguing hypotheses raised by the IBCSG data.\(^1\)

Is poverty or wealth at the root of HIV?

Poverty and lack of economic opportunity are commonly cited as important contributors to the AIDS epidemic. Indeed, an essay in The Lancet last year asked whether poverty reduction was the only sustainable solution to preventing AIDS.\(^1\) Thus recent findings from the Tanzania 2003–04 HIV/AIDS indicator survey may come as a surprise.\(^2\) The evidence is just the opposite (figure). This nationally representative survey measured wealth in terms of physical characteristics of the household and household possessions. Household wealth is strongly positively related to HIV prevalence. Indeed the difference in prevalence for women between the lowest and highest wealth quintile is four-fold. These findings are similar to those reported for Kenya\(^3\) last year. Notably, HIV prevalence is highest in some of the most economically advanced countries in Africa (eg, South Africa, Botswana). A positive relation between wealth and HIV risk has been noted before,\(^4\) but has been upstaged by the focus on poverty.\(^3\)

The poor, especially women, are vulnerable to sexual exploitation. So why this strong relation in the opposite direction? Part of the reason must be that household wealth relates to urban residence, and HIV is higher in urban areas. Also, HIV prevalence is partly a function of survival, and wealthier people with HIV probably survive somewhat longer. On the other hand, people with HIV...
eventually tend to lose wealth because of loss of employment and increased expenses related to disease, thus blunting a positive relation between wealth and HIV. Perhaps wealth simply enables people and especially men to have more sexual partners. However, in Tanzania neither number of partners nor sex with a prostitute in the past 12 months were related to HIV prevalence in men. Thus none of these explanations appear adequate to explain the observed wealth–HIV relation.

Another explanation seems crucially important—the role of established concurrent sexual partnerships in generalised heterosexual epidemics. According to this idea, serial monogamy and sporadic one-off sexual encounters might not contribute as much to new infections as networks of longer-term concurrent or overlapping partnerships. For example, a person may have a primary relationship, and an additional stable secondary relationship in another town. The primary partner might have no other partner, but the secondary partner might have one or more concurrent partners, and thus link to a larger network. Even though the average number of partners per person may not be especially high, HIV risk is. Once one person in a network characterised by concurrent partnerships has HIV, everyone becomes at high risk, both because more people are more often exposed to the virus and because recently infected individuals have many-fold higher viral loads and are much more infectious.

Wealth is the key for such networks, because wealth is associated with the mobility, time, and resources to maintain concurrent partnerships. Clearly such relationships might often have a strong economic element, but poverty itself may not be a major factor. Similarly, wealth and social interaction are inextricably linked, and wealth might increase the number of opportunities for partnerships to develop.

It is interesting that in both Tanzania and Kenya the positive relation between HIV and wealth is if anything stronger for women. We tend to think of men using their economic means to achieve more partners, and better-off wives may be infected by their better-off husbands. But it appears women to some extent also have concurrent relationships. Indeed a concurrent heterosexual network must include both men and women. Perhaps wealth allows for such behaviour in women as well, in part by increasing mobility and social interaction. Or women might improve their economic situation by having more than one concurrent partner. In any case it appears that paradoxically both wealth and economic disadvantage (or at least desire for economic advancement) play pivotal roles in HIV transmission.

What does this relationship mean for HIV programmes? First, it calls for increased attention to the economic dynamic of sexual risk in a wide variety of HIV interventions including behaviour change, condom promotion, voluntary counselling and testing, treatment, and support. But perhaps even more importantly, as described by Halperin and Epstein, it reinforces the importance of promoting social norms to foster fidelity and specifically supporting a franker discussion and understanding of the dangers of having overlapping sexual partnerships.

"James D Shelton, Michael M Cassell, Jacob Adetunji
Bureau for Global Health, United States Agency for International Development, Washington DC 20523, USA
JShelton@USAID.GOV
Our views do not necessarily reflect those of USAID, and we declare that we have no conflict of interest.

The metabolic syndrome—a new worldwide definition

The metabolic syndrome (visceral obesity, dyslipidaemia, hyperglycaemia, and hypertension), has become one of the major public-health challenges worldwide.1 There has been growing interest in this constellation of closely related cardiovascular risk factors. Although the association of several of these risk factors has been known for more than 80 years,2 the clustering received scant attention until 1988 when Reaven described syndrome X: insulin resistance, hyperglycaemia, hypertension, low HDL-cholesterol, and raised VLDL-triglycerides.3 Surprisingly, he omitted obesity, now seen by many as an essential component, especially visceral obesity.1 Various names were subsequently proposed, the most popular being metabolic syndrome.1

The cause of the syndrome remains obscure. Reaven proposed that insulin resistance played a causative role,3 but this remains uncertain. Lemieux et al suggested visceral obesity and the hypertriglyceridaemic waist phenotype as a central component,4 but this too has been contested. Several different factors are probably involved, many related to changes in lifestyle.1

The ultimate importance of metabolic syndrome is that it helps identify individuals at high risk of both type 2 diabetes and cardiovascular disease (CVD). Several expert groups have therefore attempted to produce diagnostic criteria. The first attempt was by a WHO diabetes group in 1999, which proposed a definition that could be modified as more information became available.5 The criteria had insulin resistance or its surrogates, impaired glucose tolerance or diabetes, as essential components, together with at least two of: raised blood pressure, hypertriglyceridaemia and/or low HDL-cholesterol, obesity (as measured by waist/hip ratio or body-mass index), and microalbuminuria. The European Group for the Study of Insulin Resistance6 then produced a modification of the WHO criteria excluding people with diabetes and requiring hyperinsulinaemia to be present. Waist circumference was the measure of obesity, with different cutoffs for the other variables.

A fresh approach came from the US National Cholesterol Education Program: Adult Treatment Panel III in 2001, with a focus on cardiovascular disease risk.7 The specific remit was to facilitate clinical diagnosis of high-risk individuals. It was less glucocentric than the definition from WHO and the European Group for the Study of Insulin Resistance, requiring the presence of any three of five components: central obesity, raised blood pressure, raised triglycerides, low HDL-cholesterol, and fasting hyperglycaemia.

The different definitions inevitably led to substantial confusion and absence of comparability between studies. One difficulty has been that the conceptual framework used to underpin the metabolic syndrome (and hence drive definitions) has not been agreed on. Opinions have varied as to whether the metabolic syndrome should be defined to mainly indicate insulin resistance, the metabolic consequences of obesity, risk for CVD, or simply a collection of statistically related factors. Prevalence figures for the syndrome have been similar in any given population regardless of which definition is used, but different individuals are identified.8 What matters, of course, is which produces the best prediction of subsequent diabetes and CVD. Thus Adult Treatment Panel III was superior to WHO in the San Antonio Study, but WHO gave better prediction of CVD in Finnish men.9,10

Another problem with the WHO and the Adult Treatment Panel definitions has been their applicability to different ethnic groups, especially as relates to obesity cutoffs.11 For example, the risk of type 2 diabetes is apparent at much lower levels of adiposity in Asian populations than in European populations.12 With current metabolic syndrome definitions, particularly Adult Treatment Panel III, suspiciously low prevalence figures in Asian populations resulted,12 suggesting the need for ethnic-specific cutoffs, at least for obesity.

The International Diabetes Federation (IDF) felt there was a strong need for one practical definition that would be useful in any country for the identification of people at high risk of both type 2 diabetes and cardiovascular disease (CVD). Several expert groups have therefore attempted to produce diagnostic criteria. The first attempt was by a WHO diabetes group in 1999, which proposed a definition that could be modified as more information became available. The criteria had insulin resistance or its surrogates, impaired glucose tolerance or diabetes, as essential components, together with at least two of: raised blood pressure, hypertriglyceridaemia and/or low HDL-cholesterol, obesity (as measured by waist/hip ratio or body-mass index), and microalbuminuria. The European Group for the Study of Insulin Resistance then produced a modification of the WHO criteria excluding people with diabetes and requiring hyperinsulinaemia to be present. Waist circumference was the measure of obesity, with different cutoffs for the other variables.

A fresh approach came from the US National Cholesterol Education Program: Adult Treatment Panel III in 2001, with a focus on cardiovascular disease risk. The specific remit was to facilitate clinical diagnosis of high-risk individuals. It was less glucocentric than the definition from WHO and the European Group for the Study of Insulin Resistance, requiring the presence of any three of five components: central obesity, raised blood pressure, raised triglycerides, low HDL-cholesterol, and fasting hyperglycaemia.

The different definitions inevitably led to substantial confusion and absence of comparability between studies. One difficulty has been that the conceptual framework used to underpin the metabolic syndrome (and hence drive definitions) has not been agreed on. Opinions have varied as to whether the metabolic syndrome should be defined to mainly indicate insulin resistance, the metabolic consequences of obesity, risk for CVD, or simply a collection of statistically related factors. Prevalence figures for the syndrome have been similar in any given population regardless of which definition is used, but different individuals are identified. What matters, of course, is which produces the best prediction of subsequent diabetes and CVD. Thus Adult Treatment Panel III was superior to WHO in the San Antonio Study, but WHO gave better prediction of CVD in Finnish men.9,10

Another problem with the WHO and the Adult Treatment Panel definitions has been their applicability to different ethnic groups, especially as relates to obesity cutoffs. For example, the risk of type 2 diabetes is apparent at much lower levels of adiposity in Asian populations than in European populations. With current metabolic syndrome definitions, particularly Adult Treatment Panel III, suspiciously low prevalence figures in Asian populations resulted, suggesting the need for ethnic-specific cutoffs, at least for obesity.

The International Diabetes Federation (IDF) felt there was a strong need for one practical definition that would be useful in any country for the identification of people at high risk of both type 2 diabetes and cardiovascular disease (CVD). Several expert groups have therefore attempted to produce diagnostic criteria. The first attempt was by a WHO diabetes group in 1999, which proposed a definition that could be modified as more information became available. The criteria had insulin resistance or its surrogates, impaired glucose tolerance or diabetes, as essential components, together with at least two of: raised blood pressure, hypertriglyceridaemia and/or low HDL-cholesterol, obesity (as measured by waist/hip ratio or body-mass index), and microalbuminuria. The European Group for the Study of Insulin Resistance then produced a modification of the WHO criteria excluding people with diabetes and requiring hyperinsulinaemia to be present. Waist circumference was the measure of obesity, with different cutoffs for the other variables.

A fresh approach came from the US National Cholesterol Education Program: Adult Treatment Panel III in 2001, with a focus on cardiovascular disease risk. The specific remit was to facilitate clinical diagnosis of high-risk individuals. It was less glucocentric than the definition from WHO and the European Group for the Study of Insulin Resistance, requiring the presence of any three of five components: central obesity, raised blood pressure, raised triglycerides, low HDL-cholesterol, and fasting hyperglycaemia.

The different definitions inevitably led to substantial confusion and absence of comparability between studies. One difficulty has been that the conceptual framework used to underpin the metabolic syndrome (and hence drive definitions) has not been agreed on. Opinions have varied as to whether the metabolic syndrome should be defined to mainly indicate insulin resistance, the metabolic consequences of obesity, risk for CVD, or simply a collection of statistically related factors. Prevalence figures for the syndrome have been similar in any given population regardless of which definition is used, but different individuals are identified. What matters, of course, is which produces the best prediction of subsequent diabetes and CVD. Thus Adult Treatment Panel III was superior to WHO in the San Antonio Study, but WHO gave better prediction of CVD in Finnish men.9,10

Another problem with the WHO and the Adult Treatment Panel definitions has been their applicability to different ethnic groups, especially as relates to obesity cutoffs. For example, the risk of type 2 diabetes is apparent at much lower levels of adiposity in Asian populations than in European populations. With current metabolic syndrome definitions, particularly Adult Treatment Panel III, suspiciously low prevalence figures in Asian populations resulted, suggesting the need for ethnic-specific cutoffs, at least for obesity.

The International Diabetes Federation (IDF) felt there was a strong need for one practical definition that would be useful in any country for the identification of people at high risk of both type 2 diabetes and cardiovascular disease (CVD). Several expert groups have therefore attempted to produce diagnostic criteria. The first attempt was by a WHO diabetes group in 1999, which proposed a definition that could be modified as more information became available. The criteria had insulin resistance or its surrogates, impaired glucose tolerance or diabetes, as essential components, together with at least two of: raised blood pressure, hypertriglyceridaemia and/or low HDL-cholesterol, obesity (as measured by waist/hip ratio or body-mass index), and microalbuminuria. The European Group for the Study of Insulin Resistance then produced a modification of the WHO criteria excluding people with diabetes and requiring hyperinsulinaemia to be present. Waist circumference was the measure of obesity, with different cutoffs for the other variables.

A fresh approach came from the US National Cholesterol Education Program: Adult Treatment Panel III in 2001, with a focus on cardiovascular disease risk. The specific remit was to facilitate clinical diagnosis of high-risk individuals. It was less glucocentric than the definition from WHO and the European Group for the Study of Insulin Resistance, requiring the presence of any three of five components: central obesity, raised blood pressure, raised triglycerides, low HDL-cholesterol, and fasting hyperglycaemia.

The different definitions inevitably led to substantial confusion and absence of comparability between studies. One difficulty has been that the conceptual framework used to underpin the metabolic syndrome (and hence drive definitions) has not been agreed on. Opinions have varied as to whether the metabolic syndrome should be defined to mainly indicate insulin resistance, the metabolic consequences of obesity, risk for CVD, or simply a collection of statistically related factors. Prevalence figures for the syndrome have been similar in any given population regardless of which definition is used, but different individuals are identified. What matters, of course, is which produces the best prediction of subsequent diabetes and CVD. Thus Adult Treatment Panel III was superior to WHO in the San Antonio Study, but WHO gave better prediction of CVD in Finnish men.9,10

Another problem with the WHO and the Adult Treatment Panel definitions has been their applicability to different ethnic groups, especially as relates to obesity cutoffs. For example, the risk of type 2 diabetes is apparent at much lower levels of adiposity in Asian populations than in European populations. With current metabolic syndrome definitions, particularly Adult Treatment Panel III, suspiciously low prevalence figures in Asian populations resulted, suggesting the need for ethnic-specific cutoffs, at least for obesity.
Central obesity, as assessed by waist circumference, was agreed as essential (panel), because of the strength of the evidence linking waist circumference with cardiovascular disease and the other metabolic syndrome components, and the likelihood that central obesity is an early step in the aetiological cascade leading to full metabolic syndrome. The waist circumference cutoff selected was the same as that used by European Group for the Study of Insulin Resistance, and lower than the main Adult Treatment Panel III recommendations, because most available data suggest an increase in other cardiovascular disease risk factors in Europids (white people of European origin, regardless of where they live in the world) when the waist circumference rises above 94 cm in men and 80 cm in women. Ethnic-specific waist circumference cutoffs have been incorporated into the definition (table), and have been based on available data linking waist circumference to other components of the metabolic syndrome in different populations. The levels of the other variables were as described by Adult Treatment Panel III, except that the most recent diagnostic level from the American Diabetes Association for impaired fasting glucose (5.6 mmol/L [100 mg/dL]) was used. Although this new definition will still miss substantial numbers of people with impaired glucose tolerance (because an oral glucose-tolerance test is not required), it retains the simplicity of the instrument.

The consensus group also recommended additional criteria that should be part of further research into metabolic syndrome, including: tomographic assessment of visceral adiposity and liver fat, biomarkers of adipose tissue (adiponectin, leptin), apolipoprotein B, LDL particle size, formal measurement of insulin resistance and an oral glucose-tolerance test, endothelial dysfunction, urinary albumin, inflammatory markers (C-reactive protein, tumour necrosis factor α, interleukin 6), and thrombotic markers (plasminogen activator inhibitor type 1, fibrinogen). These factors should be combined with assessment of CVD outcome and development of diabetes so better predictors can be developed.

Researchers and clinicians should use the new criteria for the identification of high-risk individuals and for research studies. Preventive measures are obviously needed in the people identified. Mounting evidence suggests that lifestyle modification with weight loss and increased physical activity will be beneficial, although specific studies in metabolic syndrome are needed. There

---

**Panel: International Diabetes Federation: metabolic syndrome definition**

**Central obesity**

Waist circumference*—ethnicity specific (see table 1)

Plus any two:

**Raised triglycerides**

> 150 mg/dL (1.7 mmol/L)

Specific treatment for this lipid abnormality

**Reduced HDL-cholesterol**

< 40 mg/dL (1.03 mmol/L) in men

< 50 mg/dL (1.29 mmol/L) in women

Specific treatment for this lipid abnormality

**Raised blood pressure**

Systolic ≥ 130 mm Hg

Diastolic ≥ 85 mm Hg

Treatment of previously diagnosed hypertension

**Fasting plasma glucose†**

≥ 100 mg/dL (5.6 mmol/L) in women

≥ 126 mg/dL (11.1 mmol/L) in men

Previously diagnosed type 2 diabetes

If above 5.6 mmol/L or 100 mg/dL, oral glucose tolerance test is strongly recommended, but is not necessary to define presence of syndrome

*If body-mass index is over 30 kg/m², central obesity can be assumed and waist circumference does not need to be measured. In clinical practice, impaired glucose tolerance is also acceptable, but all reports of prevalence of metabolic syndrome should use only fasting plasma glucose and presence of previously diagnosed diabetes to define hyperglycaemia. Prevalences also incorporating 2-h glucose results can be added as supplementary findings.

†In clinical practice, impaired glucose tolerance is also acceptable, but all reports of prevalence of metabolic syndrome should use only fasting plasma glucose and presence of previously diagnosed diabetes to define hyperglycaemia. Prevalences also incorporating 2-h glucose results can be added as supplementary findings.

**Data are pragmatic cutoffs and better data are required to link them to risk. Ethnicity should be basis for classification, not country of residence. In USA, Adult Treatment Panel III values (102 cm male, 88 cm female) are likely to continue to be used for clinical purposes. In future epidemiological studies of populations of Europid origin, regardless of where they live in the world, prevalence should be given, with both European and North American cutoffs to allow better comparisons.**

---

**Table: Ethnic-specific values for waist circumference**

<table>
<thead>
<tr>
<th>Ethnic group</th>
<th>Waist circumference (as measure of central obesity)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Europids</strong></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>≥ 94 cm</td>
</tr>
<tr>
<td>Women</td>
<td>≥ 80 cm</td>
</tr>
<tr>
<td><strong>South Asians</strong></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>≥ 90 cm</td>
</tr>
<tr>
<td>Women</td>
<td>≥ 80 cm</td>
</tr>
<tr>
<td><strong>Chinese</strong></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>≥ 90 cm</td>
</tr>
<tr>
<td>Women</td>
<td>≥ 80 cm</td>
</tr>
<tr>
<td><strong>Japanese</strong></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>≥ 85 cm</td>
</tr>
<tr>
<td>Women</td>
<td>≥ 90 cm</td>
</tr>
<tr>
<td><strong>Ethnic south and central Americans</strong></td>
<td>Use south Asian recommendations</td>
</tr>
<tr>
<td><strong>Sub-Saharan Africans</strong></td>
<td>Use European data until more specific data are available</td>
</tr>
<tr>
<td><strong>Eastern Mediterranean and middle east (Arab) populations</strong></td>
<td>Use European data until more specific data are available</td>
</tr>
</tbody>
</table>

---

** definitions. Measurement of insulin resistance was deemed impractical, although it is clear that several metabolic syndrome components, especially waist circumference and triglycerides, are highly correlated with insulin sensitivity.4**

---

Comment
are suggestions from the Finnish Diabetes Prevention Study that individuals with metabolic syndrome show less development of diabetes with lifestyle advice.\(^{12}\) In many people, however, pharmacological intervention will be needed. There is no specific treatment for the metabolic syndrome so individual abnormalities will have to be attended to. Again, long-term studies will help establish whether existing or newer agents, such as agonists for the peroxisome-proliferator-activated \(\alpha/y\) receptors or cannabinoid-1 receptor blockers,\(^{18}\) could be of specific benefit.

Recently, the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD) have published a provocative discussion paper on the syndrome.\(^{19}\) They raise several interesting questions, based on a critique of the earlier WHO and Adult Treatment Panel III criteria: 1) is it indeed a syndrome, particularly as the precise cause is unknown, 2) does it serve a useful purpose, and 3) is it labelling (and medicalising) people unnecessarily? Additionally, it has been suggested in an editorial that recognition of the metabolic syndrome has been largely driven by industry to create new markets.\(^{20}\)

A major part of the ADA/EASD\(^{19}\) stance is based on pure semantics, but the IDF (and the cardiovascular community) feel strongly that this clustering of closely related risk factors for CVD and type 2 diabetes is indeed a very good basis for calling this a syndrome. Many examples exist of conditions being given a name even when the precise underlying cause or causes are unknown (eg, type 2 diabetes). The IDF feels that it serves a useful purpose to focus on people, in both the community and clinical settings, who are at high risk of developing CVD and type 2 diabetes, particularly using the new IDF criteria proposed above.

Indeed, the ADA has just reinvented and redefined the condition of “prediabetes” for people who only have a 50% chance of developing diabetes.\(^{21}\) We also emphasise most strongly in our longer article\(^{13}\) that treatment must be focused on lifestyle change—and on the individual components if the former fails. This is a far cry from a condition claimed to be invented by industry.\(^{21}\) The metabolic syndrome concept has been around for over 80 years.\(^{1}\) The burgeoning epidemic of type 2 diabetes and CVD worldwide, particularly in the developing world seem adequate reasons for identifying and treating people with the syndrome.

We would stress that the new IDF criteria are not the final word, but hopefully will help identify people at increased risk, and through further research will lead to more accurate predictive indices.

**K George M M Alberti, Paul Zimmet, Jonathan Shaw, for the IDF Epidemiology Task Force Consensus Group**

Department of Endocrinology and Metabolism, St Marys Hospital, London W2 1NY, UK (KGMMA); and International Diabetes Institute, Melbourne, Victoria, Australia (PZ, JS)

George.Alberti@newcastle.ac.uk

The International Diabetes Federation Consensus Group was supported by an unrestricted educational grant from AstraZeneca. KGMMMA receives consultancy fees from AstraZeneca, GlaxoSmithKline, Novartis, and Servier. PZ has received consultancy fees from Novartis, GlaxoSmithKline, Bristol Myers Squibb, Bayer AG, Abbott, and Merck, and has received payment for speaking from E Merck, Sanofi-Aventis, AstraZeneca, Kissei, and Fournier. JS has received consultant fees from Merck, Eli Lilly, and Novo Nordisk.

12 Tan CE, Ma S, Wai D, Chew SK, Tai ES. Can we apply the National Cholesterol Education Program Adult Treatment Panel definition of the metabolic syndrome to Asians? Diabetes Care 2004; 27: 1182–86.
Developing an open relationship with the drug industry

Last year I joined the research advisory board of the drug company GlaxoSmithKline and get paid for that work. I was asked to write this article by The Lancet and my fee for writing will be diverted to a charity. You need to know these things before you read on.

A ward round can reveal that many patients are taking ten or more drugs. A scan of a newspaper identified no fewer than six stories suggesting therapeutic breakthroughs. Therapeutic interventions, central to the practice of medicine, have spilled over into daily news—ranging from the adoption of fluoxetine as a drug for well-being in the 1990s to the pursuit of cardiovascular disease prevention resulting in 2003 for calls for a “polypill” for the entire population. Yet despite this enthusiasm for drugs from doctors and patients, paradoxically the reputation of the drug industry is at an all-time low—the industry is often portrayed as aiming for profit above all else. And it is not just the moral highgrounders who are voicing concern. Read this, from the business section of a major newspaper: “most drug failures are a by-product of the way the industry is structured: it develops drugs as fast as possible and employs an army of salesmen to sell like crazy before the patent expires. It ignores the fact that the side-effects of a drug are often not known until it has been taken by hundreds of thousands of patients.” If this picture is correct, is industry alone to blame or are the medical profession and academia complicit in helping industry pursue profit above all else? This question is the theme of a report by Carl Elliott.

Let us get one thing straight: the drug industry works within a system that demands it makes a profit to satisfy shareholders. Indeed it has a fiduciary duty to do so. The best way to make a lot of money is to invent a drug that produces a dramatically beneficial clinical effect, is far more effective than any existing options, and has few unwanted effects. Unfortunately most drugs fall short of this ideal. Does this stop doctors from prescribing them, or patients’ groups from demanding availability for all? Clearly not. Even if we consider novel drugs rather than me-too products, recent examples provide some insights: the interferons for multiple sclerosis, drugs for dementia, and the inhibitors of cyclo-oxygenase 2 (COX-2).

Interferon β was potentially an exciting scientific advance and seemed to produce detectable biological effects in patients with multiple sclerosis. However, you needed an MRI to detect the change and the extent to which structural changes translated into clinical benefit, and improvement in quality of life, was unclear. In 2000, the UK National Institute for Clinical Excellence released an early statement that “on the basis of a very careful consideration of the evidence their [the interferons] modest clinical benefit appears to be outweighed by their very high cost”. The outcry was immediate, loud, and successful. Doctors, nurses, carers, and a patients’ group lobbied Government and the drug was made available within the UK National Health Service (NHS), albeit with

Rights were not granted to include this image in electronic media. Please refer to the printed journal.

References

Comment

some attempts to limit access and collect data. More recently, a similar outcry was seen from the Alzheimer’s Society when it was suggested that drugs for dementia might be made unavailable on the NHS.

The COX-2 class of drug was novel, was based on exciting discoveries, and in many ways looked like a potentially major therapeutic advance. As a class, the coxibs emerged to take 44% of the market share for non-steroidal anti-inflammatory drugs (NSAIDs) in some places, and were warmly welcomed by doctors and specialist societies, such that “retail sales increased disproportionately to the nonselective NSAIDs”. Although the advice has now changed, in late 2004 a visitor to the internet site of the American College of Rheumatology might have been forgiven for assuming that COX-2 inhibitors should be used in preference to standard non-steroidal anti-inflammatories for arthritis, with documents entitled “COX-2 inhibitors have dramatically changed the NSAID landscape” and the COX-2 inhibitors appearing above NSAIDs in the table of recommended treatments.

Even before the recent problems of increased cardiovascular risk with coxibs were identified, some articles were more enthusiastic than those of independent review groups which concluded that the place of these drugs in practice was unclear. An analysis of a Medicare database suggests that coxibs were adopted as the preferred NSAID by 55% of physicians within 180 days after they were marketed and that physician preference rather than patient characteristics seemed to be the major driver for prescription. Yet after the identification of increased cardiovascular risk with rofecoxib, ire was directed one way: towards the industry that developed and marketed the medicines. “We’ve been duped”, the various groups protest. Are they/we innocent parties manipulated by industry, are we willing co-conspirators, or are we just incapable of making a rational judgment about drug treatments on the basis of the evidence provided? The reality is that, at least for rofecoxib, the decreased risk of a gastrointestinal bleed seems to be almost exactly offset by the increased risk of a heart attack—a rather dramatic example of how all medicines have risk and benefits.

Elliot certainly points the finger of blame at doctors, academia, and medical journals. He documents cases of ghost writing (authors putting their name to articles written by someone else on behalf the drugs industry or other third party), bogus educational activities sanctioned for continuing medical education, and generous industry funding of specialist societies; he claims that it was “stunningly inept” of the American Medical Association to launch an industry-funded campaign to educate doctors on the ethics of industry gifts; and he concludes that there has been an enormous betrayal of public trust, and that disclosure of conflict of interest is an “empty ritual” that has been an “utter failure”. He calls for universities to treat ghost-written articles as cases of scientific fraud and for journal editors to refuse to publish editorials, review articles, and ethics essays written by authors funded by the industry whose products they are addressing. He does not hold back: “Pharma needs to make a profit . . . But to surrender our impartiality to that mission is a betrayal of everything that universities are supposed to stand for. The cost . . . is being paid for in human lives.”

There can be no doubt that the drug industry influences prescribing practice, and the enormous sums of money spent on advertising work well. There can also be no doubt, (and this is not discussed by Elliot) that the work of the drug industry has been central to improvements in patients’ care. Indeed, if the effects on length of hospital stay are taken into account (or could be incorporated into a financial model), drugs probably remain the most cost-effective intervention in health care. Who could deny that the industry-sponsored study, ISIS-II, which showed that streptokinase and aspirin produce major benefits after myocardial infarction, changed practice for the better? The prognosis for people infected with HIV has been dramatic in those able to access drugs, and development of new vaccines for infectious diseases holds real promise for prevention in neglected disease areas. As a profession, we are sometimes slow to adopt advances or recognise changes in paradigms: impotence was of little interest to the medical profession until sildenafil was introduced, yet presumably was a concern to patients. Also, it seems unrealistic and wrong to assume that the drug industry can or should develop and test new treatments without interaction with practising doctors and clinical academics. Indeed to do so would significantly decrease the chance of alleviating disease. So what can we learn from Elliot? No doubt industry could be further controlled by regulation, but perhaps we should also look to regulate ourselves as the purchasers, the spenders of public funds, and the group in which
patients place their trust when they are most vulnerable.

The first lesson seems obvious—interactions with industry are important for medical advance, but they need to be open and unambiguous, and there is an institutional responsibility to ensure this is the case. Inferences should be drawn from attempts to hide interactions. These responsibilities are as true for patients’ organisations as they are for professional bodies and universities.

The second lesson is that some activities are just unacceptable and we—academia and industry—need to stamp these out. Ghost authorship (especially with a financial sweetener) is widely considered as wrong and breaches guidelines laid down by the International Committee of Medical Journal Editors.10 Universities and hospitals should issue guidance that makes their position on ghost authorship clear, and it should be a condition of employment that individuals must state clearly in any published work whether they have received remuneration, the nature of the remuneration (personal or institutional), and reason for the payment (eg, to cover consumables, travel, or simply for reading an article and agreeing to be named as an author).

The third lesson is that we need to find a way to interpret the opinions of clinical “opinion leaders” who receive funds from industry. I suggest that the opening sentence of each article should define the relation with industry and the first slide of any lecture or presentation should do the same. It is not good enough to have it as a footnote or lost on the acknowledgments slide; and the declaration should highlight those relations that could be perceived as directly relevant to the article or lecture. Starting with information on funding will help the reader or listener place the work in context and will not diminish the research or opinion of those who present high-quality evidence.

The fourth lesson is that those making official policy judgments and purchasing decisions for public providers of health care should not have current financial ties with industry—it just does not look right.

Finally, it would take the heat out of decision making if health-care providers allocated some of their drug budget for rigorous evaluation of selected new drugs (ie, those for which important therapeutic uncertainties remain) rather than being pressurised into spending the same money on simply introducing the drugs into practice in advance of evidence and with no attempt to collect valuable data on efficacy, safety, or tolerability. 2% of the drug budget would seem a reasonable start.

Patrick Vallance
Division of Medicine, University College London, London WC1E 6JJ, UK
patrick.vallance@ucl.ac.uk
I am on the research advisory board of GlaxoSmithKline. The Division of Medicine at UCL holds grants from pharmaceutical companies as well as from research councils and major charities.

1 Wald NJ, Law MR. A strategy to reduce cardiovascular disease by more than 80%. BMJ 2003; 326: 1419.
5 Meikle J. Minister intervenes in row over drugs to treat Alzheimer’s. Guardian March 14, 2005: 8.
As Niger’s emergency eases, another crisis looms

Aid is now belatedly easing the food shortages in western Africa, but the UN is already warning of another impending disaster: this time in Africa’s south. Frustrated that the world has not learned the lessons of previous tragedies, aid agencies are calling for a new approach. Clare Kapp reports.

The influx of international aid into Niger and the pending harvest has eased the plight of 3 million people at risk of starvation. But as the crisis recedes in the Sahel region, the UN has sounded the alarm about the deadly combination of drought, poverty, and HIV/AIDS in southern Africa.

The UN estimates that up to 10 million people in Lesotho, Malawi, Mozambique, Swaziland, Zimbabwe, and Zambia will need assistance during the next 6 months. Aid groups such as CARE International warn that the scale and complexity of the southern African crisis will dwarf that of the Sahel.

Zimbabwe is particularly at risk because of the accelerating economic and agricultural collapse, compounded by President Robert Mugabe’s recent clampdown on shack dwellers and street traders, which left some 700 000 people without a home or a job. The UN forecasts that up to 4 million people may need aid but has been unable to launch an appeal for funds because the government refuses to acknowledge the emergency.

The prognosis for Malawi is also dire, with more than 4 million people at risk of food shortages by the end of the year thanks to the worst harvest in 13 years. UN Under Secretary-General for Humanitarian Affairs, Jan Egeland, told a press conference in New York in late August that 45% of children under 5 years old were stunted due to malnutrition, diarrhoea, and other diseases and launched an urgent appeal for funds to avert another emergency.

Britain has promised some additional help for Malawi and is urging its European partners to do likewise. But the pledges are only a tiny proportion of the amounts needed.

“The warning signs are already very clear”, says World Food Programme (WFP) regional director Mike Sackett. “Massive international assistance is needed but we simply cannot respond in time unless we get immediate donations . . . We are facing a triple threat comprising food insecurity, weakened governance, and the sky high HIV/AIDS prevalence rate”, Sackett told The Lancet.

Southern Africa accounts for nine of the ten countries with the world’s highest HIV/AIDS prevalence; Swaziland is hardest hit with an estimated rate of some 40%. “Southern Africa is unfortunately quite unique”, Sackett says. “Even though in some parts of the region the HIV prevalence rates are levelling off, because of the time lag, increasing numbers of people are getting sick, increasing numbers of people are dying, there are increasing numbers of orphans in the care of elderly grandparents, and increasing numbers of child-headed households. All these conditions exist regardless of the drought.”

Since the last big crisis in 2002/03, southern African countries have made progress in diversifying crops, creating better safety nets for the region’s estimated 4 million vulnerable orphans, and boosting the number of people on antiretroviral therapy. But the UN says that these advances are fragile and that the region needs a long-term structural plan, embracing the agricultural, food, and health sectors, to escape a downward spiral of poverty, disease, and death. For instance, it cites the reluctance of doctors to start patients on antiretroviral drugs if there is no guarantee they will have enough food.

While aid agencies are frustrated that the world seems not to have learned from the Niger tragedy—or others before it—some comfort is to be found in pledges totalling some US$150 million (out of the US$500 million needed) for a UN humanitarian emergency fund to enable aid workers to move into the field in a matter of days.

“We need this emergency fund so we can save more lives earlier at a lower cost”, Egeland told a news conference at the UN summit in New York.

“It still often takes so much time to get the money from the world to reach Niger in time, to reach the anti-locust teams with funding in time, to reach

Rights were not granted to include this image in electronic media. Please refer to the printed journal.
40% of children in Niger suffer from malnutrition, according to WHO

Darfur before mortality goes up. Too often we are too late because we have to wait for funding”, he says.

Egeland tried in vain in May to attract international attention to the devastation wreaked by drought and a locust plague in Niger. Shamed into action by television images of skeletal children, the international community rushed aid into Niger only in July—9 months after the first appeals from the government.

The UN has raised less than half the $2 billion it needs this year to help feed more than 25 million of the most vulnerable in the northwestern Sahel countries; Horn of Africa nations like Ethiopia and Eritrea; war-ravaged Sudan and Congo; and drought-stricken southern Africa.

But on a continent where one in three of the 900 million population are hungry, there is a thin dividing line between perpetual crisis and acute emergency.

Niger’s President, Mamadou Tandja, caused an outcry when he accused aid groups and political opponents of exaggerating the situation. Tandja insisted in a BBC interview that his people looked “well-fed” and there was no famine in Niger.

To a certain extent, he was right. Food shortages are relatively normal for the nations of the Sahel, an arid region at the southern edge of the Sahara Desert. The drought and worst locust invasion in 15 years were not the root causes of the emergency, but rather exacerbated existing conditions.

Niger ranked last on the recently published UN Development Programme’s poverty index, with more than 60% of the 11·5 million population surviving on less than $1 a day. According to WHO statistics, 40% of Niger’s children suffer from malnutrition, 40% from stunting, and 14% from wasting.

Vaccine-preventable diseases such as measles and meningitis are common, as is cholera. Malaria is a leading cause of morbidity.

The situation is made worse by the fact that more than 50% of the population have no access to health services. There is one health centre per 25 000 people; one pharmacy per 22 500; 1 maternity bed per 577 births; and 1 paediatric bed per 13 540 children under 15, according to WHO. Only 43% of the population have access to clean water, and 18% to sanitation, and the country is classed as being “far behind” in achieving nearly all the Millennium Development Goals.

Ironically, the onset of vital rains in Niger boosted hopes of a reasonable harvest later this year, but exacerbated the current threat of cholera, malaria, and diarrhoea, according to the UN.

The WFP has narrowed its initial assistance to focus on the 1·7 million people deemed most vulnerable and increasingly wants to concentrate on the “poorest of the poor”. But Médecins Sans Frontières (MSF) warned that the situation is worsening in hard-hit areas.

In a Sept 8 press statement, MSF said that a nutritional and retrospective mortality survey in the Zinder region of Niger in August found one in five children suffering from malnutrition. Mortality rates in the area for children under the age of 5 years were 4–1 deaths per 10 000 people per day—or more than double the emergency threshold—since the beginning of the crisis in January, 2005, the survey found. This rose between April and August 2005 to 5·3 deaths per 10 000 people per day.

But amid the gloom, MSF also reported good news, with the advent of a ready-to-eat therapeutic food called Plumpy Nut which needs no cooking and no addition of potentially contaminated water, enabling far more children to be nourished than in the past.

In an interview on the MSF website, nutritional specialist Milton Tectonidis said the organisation gave children two portions of Plumpy Nut per day and saw them on an outpatient basis once a week, reserving hospital care for the most acutely malnourished.

“The experience in Niger may make the combination of outpatient and inpatient care the definitive strategy for MSF. I don’t think we can go back again”, said Tectonidis.

MSF called for a new strategy to cope with the long-term food insecurity in the Sahel, saying that therapeutic food—like Plumpy Nut—for severely malnourished children should be considered an essential medicine and integrated into regular health-care services rather than just used in emergencies.

But it also said that one of the basic problems afflicting Niger was the lack of a free health system, with children with severe malaria, acute respiratory infections, and diarrhoea kept at home until they became severely malnourished.

“If children were treated sooner, we could prevent many cases of severe malnutrition. The lack of access to care is one of the primary causes of this crisis”, said MSF.

Clare Kapp
When the BBC broadcast horrific images of children starving in Niger in July this year it made big news. More than 5 million people in Niger and three other West African countries faced a catastrophic food shortage. An estimated half a million children, all of whom were facing debilitating malnutrition or death, were predicted to be among the worst hit. The world was shocked into action.

Millions of dollars suddenly flowed in from Europe and America. Donations to the UN emergency food agency, the World Food Programme (WFP), increased more than ten-fold after the broadcast. By August, food relief convoys had started rolling, and hundreds of thousands of people began receiving sustenance. It was a testament to the power of the developed world to relieve suffering—when it chooses.

Unfortunately, much of the assistance has been too little and too late. Development experts believe that fundamental problems in the way the international aid organisations operate prevented food aid getting to Niger’s population on time. And, left unresolved, these problems will mean that similar scenarios are played out among vulnerable countries in southern Africa, which, the UN has just warned are facing a deadly mix of droughts, governance problems, and HIV/AIDS.

At issue is the whole mechanism through which aid is gathered and distributed. Rapid responses to impending crises depend on media attention to drum up sufficient political momentum. As Hannah Crabtree of the Johannesburg-based relief group Action Aid, explains: “Until people start dying it is very difficult to raise any money for them.”

The timeline of action in Niger illustrates this point only too well. International aid agencies issued their first alerts of the burgeoning crisis in November, 2004, in the wake of a drought and locust infestation. But the lack of press coverage meant nobody heeded the warnings. Between November and March, the UK press ran only a few small stories on the issue, and even less appeared in US papers. It was not until the images of dying children hit the airwaves this summer that a public outcry arose and governments began to pay attention.

The food crisis reached a critical stage at the moment when the leaders of the world’s most developed economies met in Gleneagles, Scotland. Yet although long-term development and debt relief for Niger and other African countries was much discussed, the encroaching hunger crisis was never mentioned.

“International donors do not give based on need, but on political interest and media attention”, explains Nicolas De Torrente, executive director of Médecins Sans Frontières (MSF) in the USA. “Those are the two drivers and those were absent in Niger until July when the BBC arrived.”

Now, the WFP and private aid agencies are trying to work out what can be done to stop these issues dogging future aid efforts. Some major questions have surfaced: did the WFP recognise the extent of the crisis early enough and explain its urgency to donors? Once they asked for money, was it enough? Why was there such a long wait to distribute free food (which even WFP officials agree was too little and too late)?

Answers to these questions are made all the more pertinent by the fact that most believe the crisis was avoidable. The locust plague and drought that precipitated Niger’s problems accounted for only an 11% drop in food production, which by itself, say development experts, shouldn’t have caused a disaster on this scale.

Furthermore, the food shortage did not occur in the context of an unsolvable morass of war, political corruption, or sudden natural disaster. There was no ongoing fighting in any of the countries facing famine, and with the exception of Mauritania, which recently had a coup, the governments were all regarded as stable.
Neither was it a problem of money. The sums requested by aid agencies were tiny. Early calls by WFP for $16 million were completely ignored, except for a few hundred thousand dollars received from Luxembourg. Now that there is a full blown crisis, the WFP is asking for $57 million.

Who is to blame?
The focus on markets by international donors is partly responsible for the mass starvation that swept through Niger, say development experts and other relief workers.

For the first half of this year, Niger’s government, along with WFP, attempted to deal with the food shortage by providing subsidies, in the belief that the crisis could be contained and managed. But grain traders in the surrounding countries simply hiked up prices and sold the food elsewhere. Jean-Jacques Graisse, the WFP’s Senior Deputy Executive Director, explained the institution’s reasoning: “We were desperately hoping that the markets would work, and the markets did not work properly”, he says.

The world’s richest countries, on which Africa depends for emergency aid, have an ideological commitment to free-market development which often comes at the expense of urgently needed assistance. “The donors are as obtuse as obtuse can be. Markets won’t save the poorest of the poor without additional help, and they don’t understand the distinction”, says Jeffrey Sachs, the influential economist who heads the Earth Institute at Columbia University. “Africa alone in the world lives and dies on food aid because its agriculture is broken and the donors have contributed to that.”

The aid institutions themselves have also come under fire for their performance in West Africa. MSF, in particular, accuses the WFP of worrying too much about disrupting the markets and not enough about delivering food to hungry people. The WFP’s market-based programmes, which include the Food-for-Work programme and the Food-for-Training programme, did not even start in Niger until the last week of July, due to a lack of funding from the donor countries on which WFP depends. Even when WFP finally dug into its own relatively small reserves, it could not find the necessary quantities of food in the region. The Niger government did not allow WFP to begin delivering free food until August, according to WFP spokesperson Carolyn Hurford.

The problem, in hindsight, appears to be that the Niger government, international donors, and the WFP, took far too long to understand the severity of the food shortage and to respond to the acute malnutrition. They neither asked for enough money nor distributed it in the right way.

“People weren’t tracking access issues. They were looking more at availability issues. That’s where the collective ‘we’ got it wrong”, says Titon Mitra, director of Emergency Response for CARE International in Geneva. The private aid agencies, he says, were bound by their dependence on donor countries responding to the problem.

The solution to breaking this cycle of missed opportunities may be breaking with the traditional aid system. The first group to respond to the West African crisis, MSF, was also the least reliant on appealing to governments. MSF made a conscious decision to break from the dependence on government donors two decades ago. Now, unique among aid groups, it raises 80% of its funds from private donors. That gave the group the flexibility to begin relief operations in April, when it instituted a $13 million feeding programme.

Yet between long-term development programmes, which have a notoriously ineffective history, and the frenzied triage that takes place once hunger sets in, finding a middle road—and an effective system of delivering timely aid—remains elusive.

In what he hopes is a groundbreaking compromise, Sachs has reached a tentative agreement with the World Bank to provide emergency donations for seed and fertiliser in time for the approaching growing season. “It’s really important they we not just be fighting emergency after emergency”, he says.

But emergencies are looming, and the relief agencies fear that the West African debacle will be repeated, on an even grander scale. In the coming year a massive food deficit is expected across vast swathes of both western and southern African, affecting 35 million people. Aid workers are fretting about drawing the world’s attention to it before it is too late and the crisis becomes full blown.

“We know we are going to go into a very critical phase in Africa in the next 6 months. How do we get donors interested now?” asks Mitra of CARE.

“Unfortunately”, he laments, “it takes an image of an emaciated child to catalyse the world into action.”

Samuel Loewenberg
Prejudice in a portrayal of Huntington’s disease

Along the northeast shores of Lake Maracaibo, in Venezuela, there are communities with a staggeringly high prevalence of Huntington’s disease. They live as outcasts, in abject poverty, against a backdrop of the skyscrapers of Maracaibo itself. These people are descended from a woman who lived in Venezuela in the early 19th century; and her genetic inheritance has now passed through ten generations.

In Barranquitas, for example, families affected by Huntington’s disease are ostracised from the main town; they live in corrugated iron shacks with earth floors. When you walk through their enclave you soon see the entire spectrum of this terrible disease. It ranges from apparently normal children, to a few affected children, to obviously affected adolescents and young adults, and to bed-ridden men and women in their 30s and 40s who are awaiting death. Yet these families possess great dignity. Their worn clothes are spotlessly clean; the warmth of the invitation into their homes is genuine; and their offer of refreshments is more than a courtesy. This community scrapes a living by fishing in the lake, but fish stocks have dwindled over the years and survival is precarious. They have few health-care facilities. Yet it was samples from these people’s DNA that showed, in 1983, where the genetic abnormality was located (chromosome 4) in Huntington’s disease; and it was these same people’s DNA that was used to identify, in 1993, the precise nature of the genetic defect (a repeat CAG sequence).

The contrast between the worlds of these families with Huntington’s disease around Lake Maracaibo, and that of the hero—or perhaps anti-hero—of Ian McEwan’s novel Saturday, could hardly be greater. The novel follows 24 hours in the life of a London neurosurgeon named Henry Perowne. As well as being wealthy and successful, Perowne is a self-serving, self-satisfied, arrogant, name-dropper. During the course of these 24 hours, Perowne has an altercation with a man—the villain of the book—known only as Baxter. We are introduced to Baxter as he runs from a “gentleman’s club” in London, accompanied by two thugs. He is short, dressed in a black suit, and everything about him is portrayed as sinister: “He’s a fidgety, small faced young man with thick eyebrows and dark brown hair, razored close to the skull. The mouth is set bulbously adding to the effect of a muzzle. The general simian air is compounded by sloping shoulders and the built-up trapezoids suggest time in the gym, compensating for his height perhaps. He gives an impression of fretful impatience, destructive energy, waiting to be released.”

McEwan amusingly reinforces the stigma and stereotypes from which families with Huntington’s disease suffer, and which make them hide both their inheritance and their destiny.

As the book progresses, we learn that Baxter is aware of his symptoms; he knows there is no cure; and he understands his fate. Like his father, he will develop progressively incapacitating chorea and increasingly severe cognitive impairment. The disease will ultimately kill him. Later in the day, Baxter appears at Perowne’s home with a knife and threatens his family.

Throughout the book, McEwan attributes Baxter’s movements, thoughts, and actions to his Huntington’s disease. His volatility, lability, and his emotional freedom to be dangerous and murderous are all features of a man who has nothing to lose. He is dying anyway. His “simian air” is melded into a face constantly beset by movements; the failure of the tracking movements of his eyes are ridiculed—even though Perowne describes these as manifestations of lost neuronal connections. And even when Baxter becomes enraptured with a poem, it is held up as an example of emotional and dangerous instability: “Baxter is hovering behind them, making frenetic little dips of his body. He’s becoming manic, he’s tripping over his words, and shifting weight rapidly from one foot to the other . . . Now Baxter nips forwards and seizes it [the poetry book], waves it in the air, as if he could shake meaning from it. ‘I’m having this’ he cries. ‘You said I could take anything I want. So I’m taking this’. . . It’s the essence of a degenerating mind, periodically to lose all sense of continuous self, and therefore any regard for what others think of your lack of continuity.” And Baxter’s willingness to hope, to allow himself to be duped by Perowene into thinking he could join a clinical trial of RNA interference, is further proof of his cognitive failure.

Baxter is the worst caricature of someone with Huntington’s disease.
His calumny is blamed entirely on the length of the CAG repeat on his fourth chromosome, and his villainy is due only to his disease. McEwan sadly reinforces the stigma and stereotypes from which families with Huntington’s disease suffer, and which make them hide both their inheritance and their destiny.

Statistics show that people with Huntington’s disease are no more prone to violence or crime than anyone else in the population. Their bizarre and uncontrollable movements may be frightening, but they are no more dangerous than anybody else. One of us (Nancy Wexler) is at risk for Huntington’s disease. Her mother taught high school biology; three of her uncles were professional jazz musicians; and her maternal grandfather sold lingerie. All five died from Huntington’s disease—but none were in any way violent.

As well as McEwan’s pitiless stereotyping of Huntington’s disease, the reader is also subjected to the most extraordinary personal and pharmacological name-dropping. Among many other individuals, even UK Prime Minister Tony Blair is made to put in an unflattering (and completely unnecessary) cameo appearance. The pharmacological name-dropping extends to dopamine, GABA, glutamic acid decarboxylase, choline acetyltransferase, and RNA interference that intersperse the text without the slightest indication of what the phrases mean. For most general readers such references will be unintelligible.

In brief

Book  Fighting back against fat
In Fed Up, physician-journalist Susan Okie diagnoses the multitude of factors that are driving the childhood obesity epidemic. Too much food and too little exercise, of course, but it’s a bit more complicated than that.

In the USA, about 15% of children aged older than 6 years are obese, while another 15% are overweight. Given the societal and economic pressures at work to keep children wolfing junk food while planted in front of the television or computer, these children are shaping up to be the first generation in the nation’s history with a life expectancy shorter than that of their parents. And while the USA leads the world in this particular race, the rest of the world is quickly catching up.

Okie lays out what seems like every current piece of research on this topic, from the role of genes and maternal weight during pregnancy in obesity to the effectiveness of school-based exercise programmes and the feasibility of gastric bypass surgery for morbidly obese young people. She interviews leading researchers, clinicians, teachers, and community leaders. Most importantly, she spends time with kids at school, at home, and in the doctor’s office, and relates how young people and their families are trying to develop healthier habits, often with encouraging success.

Given the difficulty grown-ups have keeping slim, it’s daunting to imagine how kids can manage in a world where junk food is marketed so pervasively and, one could argue, insidiously. Vending machines sell chips, sweets, and sodas on the grounds of most US schools, and in ten schools have “pouring rights”, or an exclusivity contract, with beverage makers like Pepsi or Coke. The dual mission of the US Department of Agriculture’s school lunch programme—to subsidise farmers and to nourish children—means offerings are so heavy on meat, potatoes, and dairy that one student tells Okie she longs for salad every day.

There is, though, an irony about this book. Although McEwan doesn’t make it explicit, we only needed to exercise some amateur sleuthing skills to discover that Perowne’s house is set in London’s Fitzroy Square. In the corner of the square there is a statue of Francisco de Miranda. It is poignant that, so close to Perowne’s house, there stands a memorial to the man who tried to deliver independence to the Venezuelan people; for it is thanks to these people’s DNA that we have learned about the genetic basis of Huntington’s disease—and the source of Baxter’s fate.

Nancy S Wexler, Michael D Rawlins
M.D.Rawlins@ncl.ac.uk
wexlern@pi.cpmc.columbia.edu

Anne Harding
anne_harding@yahoo.com
Empathy

Empathy is widely seen as the essential corrective to the modern dehumanisation of the patient. In his Harveian Oration on “Science, Society and the Perplexed Physician” at London’s Royal College of Physicians, UK, in 2000, Lord Turnberg suggested that development of empathy was increasingly important because of the desensitising effects of clinical training. Yet this nostalgia for a lost empathic relationship between doctor and patient is curious, since the word was only invented at the beginning of the 20th century.

Empathy was coined in 1904 by the occult novelist and aesthete Vernon Lee (1856–1935) (aka Violet Paget), who was fascinated by ideas of spiritual possession and identification. In Otillie (1883), a supernatural fiction, she described a brother and sister’s mysterious unspoken bond. In later works, she explored haunting and the ways that the identities of the dead could supposedly shape the minds of the living. Empathy was held up as an aesthetic parallel to these psychic processes. Lee saw it as the projection of our “energies, activities and feelings” into a material work of art and emphasised its equivalence to the process of Einfühlung identified by the German experimental psychologist, Theodor Lipps (1851–1914). Lee and Lipps both saw empathy as a form of transference, like John Ruskin’s “pathetic fallacy”, in which human attributes were ascribed to inanimate objects. This notion is very different to the practitioner’s understanding of empathy today.

The reversal in the meaning of empathy is largely the result of efforts by therapists and psychoanalysts, notably Carl Rogers (1902–87), to develop a more open form of psychotherapy. In Client-Centred Therapy (1951), Rogers suggested that therapeutic success could only be achieved through the cultivation of a genuine interest in the patient. As this idea of empathy entered the psychotherapeutic mainstream, it was transformed into a natural category—a human attribute—whose absence (according to Diagnostic and Statistical Manual of Mental Disorders IV) could be seen as indicative of an underlying personality disorder. Yet despite this process of naturalisation, it is still difficult to distinguish between personal sympathy and egotistic projection. At one level, the ambivalent etymological origins of empathy help to remind us of the unstable agenda behind our own pursuit of empathy.

Rhodri Hayward
Wellcome Trust Centre for the History of Medicine at UCL
rhodri.hayward@ucl.ac.uk

Leena Peltonen

Leena Peltonen is a leading molecular geneticist and a pioneer in the use of genetically isolated populations in the identification of genes linked with disease. She is academy professor at the University of Helsinki and National Public Health Institute, Finland. She is president of The Human Genome Organisation (HUGO) and also leads the Nordic Centre of Excellence. As one of the 22 founding members of the new Scientific Council of the European Research Council (ERC), she will be working to support research funding in Europe.

What has been the greatest achievement of your career?
The identification and molecular characterisation of some 20 human disease genes.

And the greatest embarrassment?
20 years ago, I predicted a limited contribution of genetic information in our understanding of human diseases.

If you had not entered your current profession, what would you have liked to do?
To have been a journalist or a paediatrician (which I almost became before the research lured me to the laboratory).

Who was your most influential teacher, and why?
My first supervisor in research: professor Kari Kivirikko from the small medical school of Oulu, Finland. He embodied the power of enthusiasm and passion towards new knowledge.

What is the best piece of advice you have received, and from whom?
"Reach for the moon, if you fail, you may still land on a star" (from my children).

What complementary therapies have you tried? Did they work?
I always take “Influ Zink” with vitamin C when I sense a forthcoming cold, I am sure the effect is all psychological.

Do you believe there is an afterlife?
As a geneticist, yes, in your children and students.

What is your worst habit?
Making too immediate and frank comments of my surroundings.

What was your first experiment as a child?
To measure how long it took an ant to carry a needle to the ant hill.

What was the most memorable comment you ever received from a referee?
“This is a great manuscript”, a comment I received only once.

What invention has most improved your life?
The cellular phone has transformed the concept of motherhood for working mothers, you are always reachable. I remember guiding my 10-year-old son at 0400 h in Tokyo to find his goalie gloves in Helsinki.
Jan Moor-Jankowski

Primatologist and immunologist who founded New York University’s Laboratory for Experimental Medicine and Surgery in Primates. He was born on Feb 5, 1924, in Warsaw, Poland, and died of a stroke on Aug 27, 2005, in New York, USA, aged 81 years.

Jan Moor-Jankowski, director of New York University’s Laboratory for Experimental Medicine and Surgery (LEMSIP) between 1965 and 1995, believed strongly that animal research should be done openly. “It is not necessary to hide [animal research]”, he told Scientific American in 1996. “I find that open discussion in a democracy is a basis for formulating judgment.” On that principle, recalls Louis Dinetz, who served as assistant director of LEMSIP in the early 1990s, Moor-Jankowski was committed to having a totally open research facility, “with a standing invitation to animal rights organisations to visit LEMSIP and observe how the research animals were treated”. James Mahoney, who worked with Moor-Jankowski for 20 years at LEMSIP, puts it this way: “He was an animal researcher and he used primates for research; but although he was committed to that, he had a great sympathy for those who were appalled by such use, and tried to meet them half way.”

Moor-Jankowski’s professional life was marked by several hard-fought battles. In 1983, as chief editor and founder of the Journal of Medical Primatology, he published a letter from Shirley McGreal, chairwoman of the International Primate Protection League, which criticised an Austrian drug company’s plans to capture wild chimps for hepatitis research. The publication of the letter prompted the firm, Immuno AG, to sue for libel against Moor-Jankowski, the journal publisher, McGreal, and the publisher of New Scientist, which had written a news story about the plan. The suit triggered a 7-year legal battle that was eventually thrown out in a ruling by the New York Court of Appeals. “That law suit was a protracted ordeal”, recalls McGreal. She remembers Moor-Jankowski as a “charming person, an impressive person”, who became increasingly interested in the protection of primates during his life. For much of the duration of the libel case, Moor-Jankowski was the sole defendant after the others settled. In 1996, he explained to The Scientist magazine why he persevered. “As a very young boy I fought the Germans for freedom”, he said. “I didn’t want to stand up for muzzling.”

In fact, Moor-Jankowski had been just 15 years old when the German army invaded his native Poland, and that same year he joined the Polish army. When the country was overrun, he joined the resistance and was repeatedly incarcerated by the Nazis. In 1944, in Berlin, he impersonated a German officer as part of a Polish underground scheme to help Jews and other deportees move between Berlin and Poland to escape persecution. Later that year, an explosive bullet burst in his knee and he was moved from hospital to hospital until, in 1945, he escaped to Switzerland.

He earned his medical degree in Switzerland, writing a thesis on a flexible leg brace he invented for himself and wore throughout his life, but his main interest was in blood types, a subject that would remain one of his prime research areas. In 1959, at the University of Cambridge, Cambridge, UK, he began studying primates as models of human immunity, and in 1965, at the invitation of New York University, he established LEMSIP, which became a centre for the research of blood diseases, hepatitis and, later, AIDS.

But on Aug 9, 1995, 30 years after LEMSIP was founded, Moor-Jankowski’s leadership came to an end when he was fired by the University and barred from the laboratory. The dismissal came soon after the US Department of Agriculture informed the university that Moor-Jankowski, a member of the university’s animal use oversight committee, had reported violations at another of its laboratories. The university at the same time was working to divest itself of LEMSIP, which eventually closed in December, 1997. “I think he should be remembered as a fighter”, said Mahoney, who now works as a consultant to primate sanctuaries worldwide. “Once he believed in something he’d see it through to the bitter end. He didn’t seem to care if it had a bad outcome for his personal position.”

Moor-Jankowski was awarded the William J Brennan Defense of Freedom Award by the Libel Defense Resource Center in 1994. In 1995, he was elected to the French Academy of Medicine, succeeding Linus Pauling as the only American member. He was decorated with the French National Order of Merit and the Polish Order of Merit. He was given a full military burial in Poland and is survived by his wife, Deborah; his children, Bernard, Sarah, and Tadeusz; two grandsons; and two great-granddaughters.

Stephen Pincock
stephenpincock@gmail.com
Inadequate anaesthesia in lethal injection for execution

When asked by editors of The Lancet to assess for publication "Inadequate anaesthesia in lethal injection for execution" by Leonidas Koniaris and colleagues (Apr 16, p 1412),1 I expressed strong support for the article because it contained the largest series of post-mortem lethal injection thiopental concentrations published to date. Furthermore, the finding that many inmates had low serum thiopental concentrations, and were possibly awake during execution, was new and vital information that could profoundly affect the public discourse on lethal injection.

However, after more research, I became concerned that the statement that "21 (43%) [inmates] had [thiopental] concentrations consistent with consciousness" may be erroneous because of a lack of equipoise in the study. In their zeal to "prove" that thiopental concentrations during execution were low, Koniaris and colleagues may have erred in their reporting of the crucial measurement of the elapsed time between the moment of death and the retrieval of blood samples, stating that the samples were collected the "same day or next day". In fact, a graph provided to reviewers, but not included in the paper, suggests that most samples were obtained 12 or more hours after death. This graph clusters nine samples exactly 1 day after death, and 15 or more at about 0·5 days, suggesting that these times were rounded off. Most importantly, only two samples seem to have been obtained within a few hours of execution.

The elapsed time is critical because thiopental—a lipoidal-soluble and ultra-short-acting anaesthetic agent—redistributes into fat and muscle, even after death. In addition, a lethal injection is a unique clinical event, in that death occurs within a few minutes of injection of a large bolus of this drug, therefore a steady-state is not present. Under these circumstances, post-mortem serum concentrations are not reliable if a substantial amount of time has elapsed, because the high concentration of drug in the blood rapidly diffuses across a concentration gradient into the surrounding tissues after death. To state that "thiopental concentrations did not fall with increased time between execution and blood sample collection... consistent with data showing that thiopental is quite stable in stored human plasma" is erroneous since few samples were taken within the first few hours after death. Furthermore, there is a huge difference between the behaviour of thiopental in a corpse (where it diffuses out of the blood and into tissues in the body) and in a test tube of serum (where it has nowhere else to go). Other studies, not cited by Koniaris and colleagues, suggest that post-mortem serum thiopental concentrations in thiopental-caused deaths are lower in blood than in tissue23 and could be unreliable.4

Although Koniaris and colleagues’ conclusion that lethal injection has “led to the unnecessary suffering of at least some of those executed” is probably true, it is not supported by the data presented. Clearly, public review of lethal injection is warranted for several reasons, but so is more careful scrutiny of how and when post-mortem blood samples are obtained.

I declare that I have no conflict of interest.

Jonathan I Groner
gronerj@chis.osu.edu

Children’s Hospital, 700 Children’s Drive, Columbus, OH 43205, USA


Using post-mortem blood samples taken as many as 2 days after death, Leonidas Koniaris and colleagues1 conclude that 43% of inmates undergoing lethal injection with 2 g intravenous thiopental had blood thiopental concentrations consistent with awareness. Koniaris and colleagues do not present scientifically convincing data to justify their conclusion that so large a proportion of inmates have experienced awareness during lethal injection. Indeed, published and unpublished data, and clinical experience, contradict their conclusions.

It is widely accepted that concentrations of a drug in post-mortem blood might not reflect the concentrations present at the time of death because of post-mortem drug redistribution—ie, site-dependent and time-dependent changes in drug concentration that occur after death.7

These problems are particularly significant with thiopental, a highly lipophilic drug. Thiopental can take many minutes to reach equilibrium in highly perfused compartments, and longer in less well perfused tissues.3 When death ensues before equilibrium, as is the case during lethal injection, post-mortem passive diffusion from blood into tissues can cause thiopental concentrations in blood to decline. Results of studies on post-mortem drug diffusion effects suggest that this is a likely explanation for low concentrations of thiopental in blood sampled several hours to days after death. The absence of samples drawn in the first hours after death, the use of samples drawn from different anatomical sites, and the failure to characterise accurately the time between death and blood-drawing probably contributed to Koniaris and colleagues’ flawed conclusions. Notably, Koniaris and colleagues have retracted three critical data points and recognise that they incorrectly estimated the times between autopsy and blood sampling in numerous cases (T Zimmers, written communications), eroding support for their statement that “[t]hiopental concentrations did not fall with increased...
time between execution and blood sample collection.” Our ongoing analysis of 45 lethal injections from Oklahoma, including 18 data points in which blood sampling occurred within 100 min of death, shows a rapid and time-dependent decline in thiopental concentrations, underscoring the pitfalls of neglecting post-mortem redistribution.

There are other flaws in the paper. Clinical studies have shown that 2 g thiopental, if effectively delivered, creates unconsciousness for longer than 10 min. Koniaris and colleagues have not provided an adequate description of how they defined thiopental blood concentrations that allow consciousness using publications by one of us (DRS).

They have, however, correctly reflected on well recognised technical issues that increase the risk of consciousness during lethal injection. These problems include, but are not limited to: unqualified individuals charged with achieving venous access, poor supervision of intravenous drug delivery systems, poor control over timing and sequence of drug delivery, and inclusion of drugs that are unnecessarily dangerous when given by unqualified people in an execution setting. These problems are completely avoidable and warrant, indeed demand, rectification.

Post-mortem thiopental concentrations from blood drawn shortly after death can be quantitatively reliable and, in conjunction with autopsy and witness data, can provide evidence of a prisoner’s potential risk of consciousness. Unfortunately, the data provided by Koniaris and colleagues do not support the conclusion that 43% of inmates undergoing lethal injection are at risk of awareness.

Leonidas Koniaris and colleagues Research Letter¹ raises several important ethical and pharmacokinetic issues. The finding that 43 of 49 executed inmates had post-mortem thiopental concentrations in blood below that considered adequate for surgical anaesthesia raises the issue of awareness and suffering before death. Post-mortem drug concentrations are extremely difficult to interpret and there is substantial variability in results depending on timing, anatomical origin of the specimen, and physical and chemical properties of the drug. Despite these limitations, the consistent finding of low thiopental concentrations in executed prisoners from four different locations requires further assessment.

The distribution of thiopental in a dying prisoner is likely to be very different from its distribution in a ventilated and oxygenated patient. The development of cellular hypoxia and metabolic acidosis will increase the proportion of drug in the non-ionised form and the rate and extent of rapid distribution into muscle and fat. In an experimental dog model, Brodie and colleagues¹ showed that inhalation of carbon dioxide such that arterial pH decreased to 7.09–6.80 produced a 40–75% decrease in thiopental plasma concentrations. When the animals’ blood was allowed to recover to a normal pH, the thiopental concentrations increased.

The pharmacokinetics of thiopental are extremely complex. The end of anaesthetic effect of thiopental is believed to be due to the rapid distribution of drug into muscle and fat. If acidosis causes a more rapid distribution of thiopental away from its receptor sites, the current thiopental protocol might not provide adequate thiopental anaesthesia during the execution of prisoners.

We declare that we have no conflict of interest.

Robyn S Weisman, Jeffrey N Bernstein, Jeffrey N Bernstein, Richard S Weisman
rweisman@med.miami.edu

1 University of Miami, School of Medicine, 1601 NW 12th Avenue, Miami, FL 33136, USA (RSW, RSW); and Florida Poison Information Center—Miami, Miami, FL, USA (JNB)


¹ *Department of Anesthesiology, Columbia University, 620 West 168th St, New York, NY 10032, USA (MJSH); Stanford University School of Medicine, Stanford, CA, USA (DRS); and University of Dundee, Dundee, UK (DJP)

Authors' reply

Jonathan Groner and Mark Heath and colleagues mainly cast doubt on one aspect of our report—i.e., the accuracy of our finding of low post-mortem concentrations of thiopental in blood. Their concerns are unfounded. Both Groner and Heath and colleagues claim that our unpublished analysis of time to sampling versus thiopental was flawed. In general, states perform executions in the evening, and autopsies are done the next morning, accounting for the clustering of data around 12 h. In our initial submission, we did estimate time to collection for certain North Carolina samples and all eight Georgia samples, but have since received precise times from death to autopsy, plotted in the figure (upper panel). With or without these revised data or the time points contended by Heath
and colleagues (circled), no significant relation can be elucidated (Pearson’s correlation $p=0.56$, Spearman’s $p=0.11$), confirming our previous statement that concentrations in blood did not fall with increased time between execution and blood sample collection. We neither retract any data nor alter any of our conclusions.

Heath and colleagues characterise data from Oklahoma executions as showing “a rapid and time-dependent decline in thiopental concentrations”. The Oklahoma Office of the Chief Medical Examiner provided us with investigative reports of 53 of the 56 executions done from 2000 to 2004. The 18 executions in which blood was collected within 100 min of death were done in 2001 and 2002. If thiopental were to undergo “passive diffusion” along a concentration gradient from blood to tissue, a progressive decline in concentrations in blood might be seen during this initial time period. Analysis of all 18 samples taken in the first 100 min shows no significant linear or exponential relation (Pearson $r=-0.29$, $p=0.25$; Spearman $r=-0.04$, $p=0.89$) (figure, lower panel).

A truly rigorous analysis of post-mortem thiopental pharmacodynamics would require multiple, timed samples of blood and tissues taken from individual executions, before and after death. No such data were available and so we made no claims about post-mortem thiopental redistribution in our article, but rather used the individual levels as surrogate measures of levels in life.

Heath and colleagues and Groner misrepresent not only the unpublished data, but also the existing scientific literature. Thiopental does indeed rapidly distribute from blood to tissues in life.1 In terms of post-mortem redistribution, however, Heath and colleagues and Groner have the concept entirely backwards. After death, concentrations of thiopental in blood have been shown to increase (not decrease) in a similar way to virtually all other drugs examined; the only exception described by Pounder is an inhaled solvent, toluene.1 In one of several reports describing thiopental concentrations before and after death, blood thiopental rose from 2.48 mg/L in life to 5.94 mg/L 4 h after death.1 Furthermore, post-mortem cardiac/femoral blood thiopental ratios measured 1–1.9 in one study1 whereas cardiac/aortic blood thiopental ranged from 5.2 mg/L to 19 mg/L and femoral from <2.5 mg/L to 5 mg/L in samples collected from multiple anatomical sites taken on average 14 h after five individual Georgia executions (resulting in cardiac/femoral ratios of 2.17–6.04 [mean 4.11], unpublished data). High cardiac/peripheral venous drug ratios such as these are a hallmark of post-mortem redistribution from tissues into blood, not vice versa.2 Thus thiopental, a lipophilic, organic acid of moderate volume of distribution, is comparable to other barbiturates known to concentrate in solid organs during life, thereby providing a gradient for passive diffusion back into the blood after death.3 These results suggest that post-mortem concentrations of thiopental in blood, including those taken after execution, might actually overestimate concentrations in life.

It is perilous for Heath and colleagues and Groner to extrapolate from clinical experience to lethal injection. No anaesthesiologist practices bolus dosing of thiopental in potentially sedative-hypnotic, resistant, otherwise healthy, unpremedicated, profoundly hyperadrenergic individuals anticipating a peculiarly painful and public death. Modelling thiopental efficacy and pharmacodynamics in executions is further compounded by the profound physiological derangements associated with pancuronium and potassium administration.

Our evidence clearly indicates the potential for awareness during the unnecessarily complicated and error-prone process of lethal injection. It remains morally incumbent on American jurisdictions to demonstrate and document that inmates are not suffering. Until they do so, we stand by our contention that executions by lethal injection should be halted.

JPS practices capital post-conviction defence.

Subsequent to our original publication, DAL provided paid expert testimony in a Missouri case regarding lethal injection. Neither of the other authors has any conflict of interest.
Research fraud

It was disheartening to read the various doubts raised about published data by an Indian author in The Lancet (July 30, p 353 and 354) and the BMJ. India is a developing country and a large amount of raw data is available to an interested physician. However, it is almost impossible to do a randomised trial here, especially in rural areas. Illiteracy is rampant and it is therefore difficult to counsel patients and relatives about entering into a randomised trial. In my experience, not a single patient or relative is prepared to give written consent. Moreover, medical ethics committees are not easily approachable to rural doctors.

R B Singh claimed that he was unable to provide the original data sheets because they had been consumed by termites. However, this excuse would not stand up in a consumer court; by law all patients’ records should be kept for at least 3 years. I really wonder why such important data were not kept in a safety cupboard to prevent the known damage by termites.

If any editor is considering publishing a randomised trial from rural India, he or she should think twice. However, if editors want to give a chance to researchers working in such areas, they should look first to originality. Forcing researchers to comply with strict scientific methods can unfortunately result in fraud and fabrication of data.

I declare that I have no conflict of interest.

H S Bawaskar
himmatbawaskar@rediffmail.com

Bawaskar Hospital and Research Center, Maharashtra 402301, India

Integration of HIV/AIDS and family planning

Ann Duerr and colleagues (July 16, p 261) state the importance of integrating family planning and prevention of mother-to-child HIV transmission. We agree. The fields of family planning and HIV/AIDS have many intersections, but only recently have those involved in research and care begun collaborating. Experts now recognise that incorporating these two arenas is crucial to prevention efforts and to the provision of comprehensive services.

We would push Duerr and colleagues’ argument even further and assert that integration addresses not only one, but all four mother-to-child HIV transmission prevention strategies proposed by the UN Interagency Task Team on Mother-to-Child Transmission of HIV Infection. These are: (1) prevention of HIV infection in all people, especially young women; (2) prevention of unintended pregnancies in HIV-infected women; (3) prevention of HIV transmission from HIV-infected women to their infants through antiretroviral therapy, safe delivery practices, and counselling and support on infant-feeding methods; and (4) provision of care and support to HIV-infected women, their infants, and families.

Duerr and colleagues cite health benefits as rationale for integrating family planning and HIV/AIDS. We stress that integration has advantages beyond health benefits and streamlining of services. Integration allows for continuity of care (which enhances service quality and patient follow-up), provides access to more people, and increases the provision of comprehensive services. By “mainstreaming” HIV/AIDS into family planning, integration also helps to reduce stigma surrounding HIV, and heightens patients’ knowledge of healthy behaviours. Furthermore, integration is a rights-based approach, thereby promoting the principles established at the International Conference on Population and Development (ICPD) in Cairo, 1994.

The International Planned Parenthood Federation (IPPF) continues to make progress in incorporating HIV/AIDS into sexual and reproductive health services through its regional offices and its 149 member associations, which provide more than 30 million service every year.

Many challenges arise in incorporating family planning and HIV, such as start-up costs and the need for staff sensitisation and training. However, there are also many opportunities for scaling-up HIV/AIDS activities, which can over-ride these and other challenges. Some of these challenges, as mentioned by Duerr and colleagues, can be addressed through operations research, while others can be addressed by greater prioritisation of such integrated efforts.

We declare that we have no conflict of interest.

*Jane Galvao, Jessica Halverson
jgalvao@ippfwhr.org

International Planned Parenthood Federation/
Western Hemisphere Region, New York, NY 10005, USA

In their Viewpoint on the integration of family planning into strategies to prevent mother-to-child transmission of HIV in sub-Saharan Africa, Ann Duerr and colleagues do not recognise sex inequality as a challenge to HIV prevention. The high rate of unwanted and repeat pregnancy in sub-Saharan Africa begs an important question: to what extent is it attributed to male dominance in family formation and reproduction?

Research shows that sex inequality is deeply embedded in the patrilineal family system, and it manifests itself in every aspect of reproduction in Ghana. Although both men and women in highly polygonal areas view sexual relations shortly after childbirth as harmful to the infant and being blamed for unwanted pregnancies, women are often left with the dilemma of either refusing the husband’s sexual demand and facing punishment for disobedience or giving in and being blamed for unwanted pregnancies. Men insist that it is men who make the decisions about contraception because they see it as a means to control women’s sexuality. Globally speaking, improvement in sex equality will also contribute significantly to HIV prevention, especially in countries where the transmission pathway is increasingly shifting towards heterosexual contact (eg, China), and where the risk of HIV infection is linked to patriarchal culture. Patriarchal values perpetuate women’s subordination, which in turn leads to female vulnerability to HIV infection, both within and outside marriage, by way of unsafe and coercive sex. Poverty and inferior social position force many young women into commercial sex work in China. An unregulated sex industry with unprotected sex workers has the dangerous potential to form a fertile ground for widespread HIV transmission, as is already occurring in China. Yet sex inequality is seldom mentioned in global responses to the HIV/AIDS pandemic.

I declare that I have no conflict of interest.

Jianghong Li
jianghongli@ichr.uwa.edu.au

Division of Popular Sciences, Telethon Institute for Child Health Research, University of Western Australia, PO Box 955, West Perth, Western Australia 6872, Australia


Industry-funded bioethics articles

In his Viewpoint (July 30, p 422), Carl Elliott suggests that journals should not publish bioethics papers for which there has been industry support. The goal is to halt, as he puts it, “influence peddling”.

Elliott identifies a National Depressive and Manic-Depressive Association (NDMDA) consensus statement on the use of placebos in clinical trials of the treatment of mood disorders as an example. However, he does not provide any documentation of bias or bought opinions, but leaves readers to presume that, because the NDMDA-sponsored conference was underwritten by support from ten pharmaceutical companies, the report must be irredeemably tainted. It is a pity that Elliott did not make a more compelling case by subjecting to critical scrutiny the composition of the consensus panel or, even more to the point, the transparent and well documented arguments presented in the article. Surely, if bias warranted that the article should have been barred from being published, the evidence should be obvious and persuasive to an open-minded reader.

The consensus panel included representatives of the US National Institute of Mental Health, university-based clinical trialists and researchers of psychotropic medication, and a sprinkling of bioethicists, together representing a broad range of opinions. Some of these panellists received reimbursement for their travel and expenses, and if this bought their opinions, they obviously...
come cheap. The key conclusion of the paper is that claims for an antidepressant or mood regulator should be based on demonstration of superiority to placebo, not just equivalence to an established medication. Placebo-controlled trials are appropriate, provided they have adequate scientific merit and do not pose unacceptable risk to patients. Can some flaw in this conclusion be linked to the payments to panelists for travel and expenses?

Another paper cited by Elliott is an industry-underwritten study of the influence of gifts from pharmaceutical companies. The points of view expressed in this paper are very similar to his own. By Elliott’s standards, should we reject the argument of the industry-underwritten article or accept the same argument because Elliott declares no financial support? The article by Katz and colleagues concluded that: “The power of gift-giving, both large and small, must be acknowledged if appropriate regulatory policies are to be created and enforced.” Was this conclusion unsound until regurgitated by Elliott?

Somewhere in all of this a place must remain for paying attention to the quality of evidence and argument, particularly when they are presented with sufficient transparency to make an assessment. Arguably, one should at least exhaust the potential of such material before leaping to inferences based on financial support. Perhaps the consistency with which industry-underwritten articles are characterised by bias, distortion, and outright lies justifies a blanket prohibition against further publishing of such articles. Some documentation of such a strong claim would seem to be warranted.

I declare that I have no conflict of interest.

James C Coyne
jcoyne@mail.med.upenn.edu
Department of Psychiatry, University of Pennsylvania School of Medicine, 3400 Spruce St 11 Gates, Philadelphia, PA 19104, USA


Carl Elliott rightly asks the question: “Is an industry-funded bioethicist a bioethicist that we can trust?” However, en route to his (appropriate) exhortation for medical and bioethics journals to deny publication to industry-funded editorialists, Elliott skims over a few issues that would benefit from more discussion.

I question what he means by the “impartiality of bioethics”. Are any bioethics truly impartial? One’s personal prejudices and interests are certainly bound to influence one’s interpretation of the data that underlie biomedical research papers, much less the arguments and ideas that are the substrate for bioethical rumination. Furthermore, the psychological literature has persuasively shown the pervasiveness of cognitive partialities that lead us to infer a wide variety of biases in others while denying them in ourselves.2

Perhaps Elliott meant that a biotechnical stance that is independent of industry financial support is more impartial than a financially dependent one. That would be true. However, financial conflicts of interest are not the only conflicts that may raise concerns. In Miriam Schuchman’s book The Drug Trial, she describes a morally unhappy situation involving researchers, drug companies, and hospitals in which the primary conflicts are non-financial. I do not raise this issue in order to make the point that if journals cannot ferret out the only conflicts of interest then perhaps they should ignore financial ones, but rather to make the general point that we would all benefit from a more nuanced discussion of conflicts of interest.

In response to Elliott’s Viewpoint, Patrick Vallance (published online July 7) emphasises the need for transparency and disclosure of industry sources of funding. Such sentiments are noble, but ignore the limitations of disclosure policies. Disclosure may perversely worsen the provision of biased information because the closer the may feel licensed to stray ethically once the required disclosure has been made.3

Vallance also suggests “rigorous evaluation of selected new drugs”. I agree. 15 years ago, the Jackson Hole Group suggested that an independent research institute, funded neither by a pharmaceutical industry motivated to peddle pills nor by a health insurance industry motivated to restrict their use, would certainly be useful when doing head-to-head trials of new patent-protected agents versus older generic agents versus placebo, weighing their risks and benefits as well as their costs. A pharmaceutical industry truly interested in showing value might even go so far as to sponsor such studies on its own.

I declare that I have no conflict of interest.

Alexander C Tsai
act2@case.edu
Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA


4 Vallance P. Developing an open relationship with the drug industry. Lancet 2005; published online July 7. DOI:10.1016/S0140-6736(05)66835-3.

Toxin production by an emerging strain of Clostridium difficile associated with outbreaks of severe disease in North America and Europe

Michel Warny, Jacques Pepin, Aiqi Fang, George Killgore, Angela Thompson, Jon Brazier, Eric Frost, L Clifford McDonald

Summary
Background Toxins A and B are the primary virulence factors of Clostridium difficile. Since 2002, an epidemic of C difficile-associated disease with increased morbidity and mortality has been present in Quebec province, Canada. We characterised the dominant strain of this epidemic to determine whether it produces higher amounts of toxins A and B than those produced by non-epidemic strains.

Methods We obtained isolates from 124 patients from Centre Hospitalier Universitaire de Sherbrooke in Quebec. Additional isolates from the USA, Canada, and the UK were included to increase the genetic diversity of the toxinotypes tested. Isolate characterisation included toxotyping, pulsed-field gel electrophoresis (PFGE), PCR ribotyping, detection of a binary toxin gene, and detection of deletions in a putative negative regulator for toxins A and B (tcdC). By use of an enzyme-linked immunosassay, we measured the in-vitro production of toxins A and B by epidemic strain and non-dominant strain isolates.

Findings The epidemic strain was characterised as toxinotype III, North American PFGE type 1, and PCR-ribotype 027 (NAP1/027). This strain carried the binary toxin gene cdtB and an 18-bp deletion in tcdC. We isolated this strain from 72 patients with C difficile-associated disease (58 [67%] of 86 with health-care-associated disease; 14 [37%] of 38 with community-acquired disease). Peak median (IQR) toxin A and toxin B concentrations produced in vitro by NAP1/027 were 16 and 23 times higher, respectively, than those measured in isolates representing 12 different PFGE types, known as toxinotype 0 (toxin A, median 848 μg/L [IQR 504–1022] vs 54 μg/L [23–203]; toxin B, 180 μg/L [137–210] vs 8 μg/L [5–25]; p<0.0001 for both toxins).

Interpretation The severity of C difficile-associated disease caused by NAP1/027 could result from hyperproduction of toxins A and B. Dissemination of this strain in North America and Europe could lead to important changes in the epidemiology of C difficile-associated disease.

Introduction Clostridium difficile infection results in a broad spectrum of disease ranging from mild diarrhoea to severe life-threatening conditions. Colonic injury and inflammation results from the production of two protein toxins: toxin A and toxin B. Isolates that produce neither toxin are non-pathogenic.1 Moreover, some isolates produce toxin B only and can cause pseudomembranous colitis.2 A small group of isolates produce a separate binary toxin in addition to toxins A and B; however, the role of this binary toxin in C difficile-associated disease is not known. Whether variations in disease severity can result from differences in toxin A and B production is unknown.

The genes encoding toxin A and toxin B are part of the pathogenicity locus (PaLoc), which is a short chromosomal segment carried by pathogenic strains of C difficile. Transcription analysis studies in reference strain C difficile VPI 10463 (toxinotype 0) have shown that production of toxin A and toxin B is co-regulated and growth-dependent.3 The logarithmic phase is associated with strong expression of the genes encoding toxin A (tcdA), toxin B (tcdB), a positive regulator (tcdD), and a holin-like protein (tcdE).4 The inverse is seen during the stationary phase, suggesting that tcdC negatively regulates toxin expression.5 Variations in the PaLoc sequence can be detected by toxotyping;6 22 toxotype variants (I–XXII), in addition to the reference toxotype 0, have been characterised. Analysis of two large isolate collections showed that 78–88% isolates were toxotype 0 and 2–3% were toxotype III.6,7 Whether some toxotypes are more virulent than others is not known.

In 2002, hospitals in Montreal and southern Quebec, Canada, began experiencing an epidemic of C difficile-associated disease. Between 2003 and 2004, about 14 000 nosocomial cases of the disease were reported.7 In January, 2005, 30 hospitals in Quebec reported rates of nosocomial disease higher than 15 per 10 000 patient-days, at least five times greater than the historical average.8 At the Centre Hospitalier Universitaire de Sherbrooke in Quebec, the proportion of patients with C difficile-associated disease who died within 30 days after diagnosis rose from 4.7% in 1991–92 to 13.8% in 2003, suggesting increased virulence of C difficile.9 The incidence of C difficile-associated disease per 100 000 individuals aged 65 years or more in Sherbrooke increased from 102 in 1991–92 to 210 in 2002 and 866 in 2003.10
Because toxins A and B are the primary virulence factors of *C difficile*, we postulated that increased virulence could be due to increased toxin production. Thus, we attempted to identify the epidemic strain at Sherbrooke and we compared toxin A and B production in this strain with that in other contemporary isolates.

**Methods**

We regarded *C difficile*-associated disease as health-care-associated in haemodialysis patients, in residents of a long-term care facility, and in hospital inpatients if symptoms developed more than 72 h after admission or within 2 months of discharge from a health-care facility. All other cases were regarded as community-acquired.

We obtained isolates from 124 consecutive patients with *C difficile*-associated disease seen at Sherbrooke between June, 2004, and April, 2005. Faecal samples were diluted ten times in phosphate-buffered saline, centrifuged, and filtered. The supernatant was tested for cytotoxicity by use of MRC-5 cells. *C difficile* was isolated with cycloserine cefoxitin fructose agar (CCFA) plates (Quelabs, Montreal, Canada) after ethanol treatment. Additional isolates from recent outbreaks were obtained from the US Centers for Disease Control and Prevention (n=17), Hôpital Maisonneuve-Rosemont, Montreal (n=7), and the UK Anaerobe Reference Laboratory (n=6). This selection was meant to include isolates geographically distant from Sherbrooke to increase the genetic diversity of the sample.

*C difficile* isolates were characterised by toxinotyping, pulsed-field gel electrophoresis (PFGE), and PCR ribotyping with methods previously described.\(^4\)\(^1\) PFGE and PCR-ribotyping have greater discriminating power than toxinotyping. PFGE dendograms were created with BioNumerics 4.0 (Applied Maths Austin, TX). Additionally, we detected deletions in tcdC by PCR using in-house primers.\(^1\) Finally, the B fragment (*cdtB*) of binary toxin gene was detected by PCR.\(^1\)

**In-vitro production of toxin A and toxin B** was measured by enzyme-linked immunosorbent assays (ELISA) in culture supernatants from a 1 mL subsample of the culture. All isolates were initially frozen at −80°C; aliquots of cell suspensions were then prepared by the anaerobic culture of isolates, undertaken three times for 20–24 h at 37°C in 50 mL Acambis proprietary broth medium (Cambridge, MA, USA). A sample of the third culture was mixed with an equal volume of glycerol and stored at −80°C until testing. Frozen aliquots were thawed and subcultured three times as done previously to ensure equilibration of every isolate with the culture medium and to dilute the storage medium. Finally, a sample of the third culture (5% inoculum) was cultured for 3 days, during which growth and toxin production were measured. Variations in inoculum cell densities had no effect on cell density at the stationary phase. The kinetics of the logarithmic phase and stationary phase were assessed in separate experiments. We measured growth and toxin production at 24 h, 48 h, and 72 h. Growth was measured by absorbance at 600 nm (Genesys 20, Thermo Electron Corporation, Rochester, NY, USA). One optical density unit corresponded to 1×10⁶ colony-forming units (CFU) per mL. Spores were detected by light microscopy.

We measured toxin A and B concentrations by capture ELISA with specific polyclonal antibodies prepared at Acambis as previously described.\(^1\)\(^2\) Briefly, microplates were coated overnight with antitoxin A or antitoxin B goat IgG (2 μg/mL) in carbonate-bicarbonate buffer (pH 9·8). Plates were blocked with 2·5% skim-milk buffer in Dulbecco’s phosphate-buffered saline and 0·05% Tween 20 (blocking buffer) at 37°C for 90 min. We then prepared culture supernatants, standards and controls, using blocking buffer, and these were incubated for 1 h at 37°C (100 μL/well). After washing, cultures were incubated with antitoxin A or antitoxin B mouse-specific IgG for 1 h at 37°C and then detected by goat anti-mouse IgG coupled to alkaline phosphatase (Southern Biotech, Birmingham, AL, USA), with the use of diolamine as substrate. Purified toxins A and B were used as standards (with 95% purity, as measured by high-pressure liquid chromatography). The coefficients of variation (CV) of toxin production by the control strain *C difficile* ATCC 43255 were 8·2% for toxin A production and 26% for toxin B production, based on five independent cultures and ELISA readings. Interassay CV of toxin B ELISA was 13·6%. Isolate testing for toxin production was masked. Culture medium, primary antibodies, and toxin standards were developed at Acambis for industrial production of a toxoid vaccine against *C difficile*.

Differences in cell growth were assessed with the Student’s t test. We compared toxin production using the Mann-Whitney test, since the distribution of toxin production values was skewed.

**Role of the funding source**

The study had no funding source. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

**Results**

To identify the epidemic strain at Sherbrooke, we characterised an initial group of *C difficile* isolates obtained from 13 consecutive patients with *C difficile*-associated disease who presented with the infection during June and July, 2004. 12 (92%) isolates were toxotype III, all showing identical or highly related PFGE patterns. This strain carried an 18-bp *tdcC* deletion and the *cdtB* gene. We sequenced the *tdcC* deletion and found that it was identical to one previously described.\(^1\) The PFGE and PCR-ribotype patterns of these isolates belong to the North American PFGE type 1 (NAP1) and PCR-ribotype 027,
respectively; this genotype (NAP1/027) also encompassed all eight isolates that were toxinotype III, had tcdC deletions, and had the binary toxin gene from other locations (France, UK, and USA [Maine, Georgia, Pennsylvania, New Jersey]). Figure 1 shows the representative PFGE patterns and the resulting dendrogram. Isolates that were toxinotype III, had the tcdC deletion, contained the binary toxin gene, and had PCR-ribotype 027 exhibit two PFGE patterns with 94% similarity (NAP1a and NAP1b; figure 1). By contrast, the six toxinotype 0 isolates from Canada and other countries each present a different PFGE pattern with less than 80% similarity (figure 1).

After the initial 13 isolates were characterised, we toxinotyped and tested 111 additional isolates from consecutive patients for deletions in tcdC. Of 86 isolates that were probably health-care-associated, 58 (67%) were toxinotype III. Of 38 isolates that were probably community-acquired, 14 (37%) were toxinotype III. All but one of the toxinotype III strains showed the 18-bp tcdC deletion.

To compare the production of toxins A and B, the test group of 15 NAP1/027 strains included seven from Sherbrooke, one from Montreal, Canada, two from the UK, four from the USA, and one from Paris (CD196). Historically, toxinotype 0 represents 78–88% of hospital toxigenic isolates; therefore, we investigated whether NAP1/027 produces more toxins A and B than a control group of 25 contemporary isolates of toxinotype 0, randomly selected from the same geographical areas where the toxinotype III isolates were recovered (13 from the USA, nine from Canada, and three from the UK). With 80% similarity by PFGE as a cut-off, all isolates were negative for both the tcdC deletion and stxB.

Growth kinetics were very similar and showed a peak in cell density at 24 h in both groups (figure 2). Cell density was 17% higher in NAP1/027 than in control strains, at 24 h, mean (SD) absorbance at 600 nm was 1.77 (0.41) in controls and 2.07 (0.51) in NAP1/027 (p=0.045, Student’s t test). At 48 h, the difference in density was not significant (p=0.07). Spores were detected at 48 h and 72 h in 18 (72%) toxinotype 0 and 14 (93%) NAP1/027 isolates. In this batch culture method, isolates reached the stationary phase by 24 h (data not shown). Toxin A and toxin B production was much faster and greater in NAP1/027 than in controls. In toxinotype 0, only 14 (56%) isolates produced more than 0.5 μg/L of toxin A at 24 h.

At 24 h, median toxin A concentration was 4 μg/L in toxinotype 0 (IQR 0–64) and 543 μg/L in NAP1/027 (430–873; figure 3). In 11 (44%) toxinotype 0 isolates, toxin A production was not detectable after 24 h (<0.5 μg/L), whereas toxin B ranged from 2 to 3 μg/L. At 24 h, median toxin B concentration was 3 μg/L in toxinotype 0 (2–6 μg/L) and 149 μg/L in NAP1/027 (137–210; figure 3). ELISA values correlated with toxin band intensity by SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis; data not shown). Toxin A and toxin B production did not differ significantly between isolates from Canada and other areas (data not shown). At 48 h, median toxin A was 54 μg/L (IQR 23–203) in toxinotype 0 and 848 μg/L (504–1022) in NAP1/027. Median toxin B at 48 h was 8 μg/L (5–25) in toxinotype 0 and 180 μg/L (137–210) in NAP1/027. Thus, the peak median (IQR) concentration of toxin A was 16 times higher in NAP1/027 than in toxinotype 0 control strains; toxin B concentrations were 23 times higher in the epidemic strain than in the control strain (toxin A, median 848 μg/L [504–1022] vs 54 μg/L [23–203]; toxin B, 180 μg/L [137–210] vs 8 μg/L [5–25]; p<0.0001 for both toxins, Mann-Whitney test).

Figure 1: PFGE analysis of C difficile study isolates from various geographical locations

<table>
<thead>
<tr>
<th>Origin</th>
<th>Toxinotype</th>
<th>PFGE pattern</th>
<th>Similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>III</td>
<td>NAP1b/027</td>
<td>100%</td>
</tr>
<tr>
<td>France</td>
<td>III</td>
<td>NAP1b/027</td>
<td>94%</td>
</tr>
<tr>
<td>UK</td>
<td>III</td>
<td>NAP1b/027</td>
<td>61%</td>
</tr>
<tr>
<td>USA</td>
<td>III</td>
<td>NAP1a/027</td>
<td>57%</td>
</tr>
<tr>
<td>Canada</td>
<td>III</td>
<td>NAP1a/027</td>
<td>67%</td>
</tr>
<tr>
<td>USA</td>
<td>III</td>
<td>NAP1a/027</td>
<td>78%</td>
</tr>
<tr>
<td>Canada</td>
<td>III</td>
<td>NAP1a/027</td>
<td>63%</td>
</tr>
<tr>
<td>Canada</td>
<td>III</td>
<td>NAP1a/027</td>
<td>62%</td>
</tr>
</tbody>
</table>

Figure 2: Growth curves of toxinotype 0 and toxinotype III (NAP1/027)
Mean cell density and SDs are shown.
deletion in this unusual toxin production profile needs further investigation.

In-vitro toxin production in different \textit{C. difficile} isolates was investigated elsewhere\textsuperscript{17} by use of an early phenotyping method.\textsuperscript{17} A correlation between toxin production and type was recorded, as well as a correlation between toxin A and toxin B production. However, toxin A production was reported as either low (\(\leq 40 \, \mu\text{g}/\text{L}\)) or high (\(>40 \, \mu\text{g}/\text{L}\)) and toxin B was semiquantified with a cytotoxicity assay rather than ELISA, which made comparison with our data difficult.

Historically, about 6\% of \textit{C. difficile} isolates have carried the binary toxin genes and these are toxinotype variants (ie, non-toxinotype 0).\textsuperscript{6,13,18,19} By contrast, we found that NAP1/027 uniformly carries the genes for binary toxin. Previous studies have not established conclusively that patients infected with binary-toxin-producing isolates present a more severe disease than those infected with other strains.\textsuperscript{4,8–21} In the USA and Europe, binary-toxin-producing isolates belonged to at least nine toxinotype variants, but the presence of a deletion in \textit{tcdC} was not investigated.\textsuperscript{6,13,18} The role of the binary toxin in the virulence of NAP1/027 remains unclear and we did not measure binary toxin production in vitro. Although strains that are positive for binary toxin but negative for toxins A and B are not pathogenic in hamsters, their culture supernatant causes fluid secretion in rabbit ileum, but not mucosal damage.\textsuperscript{18} The severity of colonic inflammation in patients with \textit{C. difficile}-associated disease in Quebec\textsuperscript{4} does not support the hypothesis that the binary toxin plays a key role in NAP1/027-induced disease since this toxin seems to act mostly on the ileum. Further studies are needed to determine the pathogenic role of the binary toxin in human beings.

The epidemic strain at Sherbrooke and presumably in many hospitals in Quebec belongs to PFGE type NAP1, PCR-ribotype 027, and is toxinotype III. Until recently, toxinotype III strains represented 2–3\% of hospital isolates of \textit{C. difficile}.\textsuperscript{4,6} The ability of this strain to produce binary toxin in vitro was reported in 1988.\textsuperscript{14} With restriction endonuclease analysis, the same genotype as NAP1/027 (also known as type BI) was found in only 14 of more than 6000 US historic isolates obtained before 2001.\textsuperscript{22} To our knowledge, NAP1/027 was not reported to cause either severe disease or outbreaks until recently, when it was identified as the cause of several outbreaks in the USA, some predating the Quebec epidemic.\textsuperscript{23} Moreover, retrospectively the strain has been identified in isolates from sporadic US cases obtained in the early 1980s.\textsuperscript{24} The finding of an association between NAP1/027 (or BI) and high toxin production in the context of an epidemic associated with a high case-fatality ratio confirms the suspicion that the epidemic in Quebec is caused by a more virulent strain. In Sherbrooke, between 2003 and 2004, as many as a sixth of inpatients with health-care-

**Figure 3: In vitro production of toxins A and B by \textit{C. difficile} isolates**

Median concentration and IQRs are shown. \textit{C. difficile} strains included 25 toxinotype 0 and 15 NAP1/027 strains (toxinotype III) from various locations.

**Discussion**

This study reports the emergence of NAP1/027, an epidemic strain of \textit{C. difficile} implicated in outbreaks associated with severe disease. In NAP1/027, toxin concentration peaked early in the stationary phase, indicating that the bulk of toxin production occurred during the logarithmic phase. The finding that NAP1/027 can produce 16 times more toxin A and 23 times more toxin B than control strains is of importance since these proinflammatory and enterotoxic proteins are the primary virulence factors of \textit{C. difficile}. Other researchers have shown that toxin B was ten times more toxic than toxin A in damaging the epithelium in human colonic explants.\textsuperscript{16} If the ratio of toxin A to toxin B in vitro is similar in the colonic lumen, the virulence of NAP1/027 could result mainly from increased toxin B production.

The small differences in cell densities between NAP1/027 and toxinotype 0 strains cannot explain the magnitude of the variation in toxin production. As mentioned earlier, \textit{PaLoc} includes \textit{tcdC}, a gene for a putative negative regulator of toxin A and toxin B gene expression,\textsuperscript{7} and a deletion in \textit{tcdC} might cause increased toxin production by defective repression of toxin gene expression. The accelerated kinetics of toxin A and toxin B production is consistent with our hypothesis. However, the implication of the 18-bp
associated *C. difficile* as a direct or indirect consequence of this infection.23

A dominant strain that was PFGE type NAP1, toxinotype III, and contained a tcdC deletion and ctdB was also discovered in samples from Montreal and from outbreaks associated with increased morbidity, frequent need for colectomy, and mortality in the USA.21,24,25 In the UK, where the number of reported cases of *C. difficile*-associated disease doubled over 3 years,26 NAP1/027 is the cause of ongoing outbreaks in at least three hospitals where a high case-fatality ratio has been noted. Although the detailed distribution of NAP1/027 in the UK remains to be determined, data recently released by the UK Department of Health are worrying; in 2004, nationwide rates of *C. difficile* reports per 1000-bed-days for patients aged 65 years or more were 1·69 for general acute-care trusts and 1·96 for specialist trusts.27 In the Netherlands, NAP1/027 was identified in two severe outbreaks also associated with fatalities.28 The outbreaks in Quebec and the UK have been widely reported in national media and have raised public concern leading to increased emphasis on surveillance and control measures.

The mechanisms behind the emergence and high transmissibility of NAP1/027 are not fully understood, but it seems plausible that the severe diarrhoea induced by this strain could help the dissemination of spores in the hospital environment by incontinent patients. In this study, none of the patients with community-acquired disease had been admitted during the preceding year, suggesting that NAP1/027 could have spread from hospitals into the community. Importantly, antimicrobial susceptibility testing of contemporary and historic isolates of NAP1/027 indicates a substantial increase in resistance to all fluoroquinolones.29 A cohort study of Sherbrooke inpatients recorded that fluoroquinolone use (especially ciprofloxacin) has emerged as the major risk factor for *C. difficile*-associated disease in the context of the ongoing epidemic.30 Fluoroquinolones are now the most widely prescribed antibiotics in many developed countries,30 and the acquisition of fluoroquinolone resistance has been thought to promote the emergence of NAP1.22 A substantial increase in the proportion of patients who fail to respond to metronidazole and a doubling of the frequency of postmetronidazole relapses have been noted,31 which could also promote the dissemination of this strain.

In conclusion, the epidemic *C. difficile* variant strain NAP1/027 produces substantially more toxin A and toxin B than most hospital strains. Evidence indicates that this emerging toxinotype III strain is highly transmissible and more virulent, which could represent a major shift in the epidemiology of *C. difficile*-associated disease. Clinicians need to be vigilant in the prevention, diagnosis, and treatment of the disorder.

Conflict of interest statement
M Warny and A Fang are Acambis employees and are involved in the development of a vaccine against *C. difficile*. M Warny owns Acambis stocks.

Acknowledgments
We thank Lois Wiggs (Centers for Disease Control and Prevention) and Micaela Gal (Anaerobe Reference Laboratory, Cardiff, UK) for their assistance with isolate characterisation; Annie-Claude Labbé (Hôpital Maisonneuve-Rosemont, Montreal, Canada) for providing isolates; Mohammad Hassan Roostaei (University of Sherbrooke) for toxinotyping isolates; and Michael Annunziato, Fuqin Ma, and Jie Zhang (Acambis) for developing the capture ELISAs and for characterising toxin A and toxin B standards.

References


Single-dose ciprofloxacin versus 12-dose erythromycin for childhood cholera: a randomised controlled trial

Debasish Saha, Wasif A Khan, Mohammad M Karim, Hafizur R Chowdhury, Mohammed A Salam, Michael L Bennish

Summary
Background Single-dose ciprofloxacin is effective for the treatment of severe cholera in adults. We assessed whether single-dose ciprofloxacin would be as effective as 3-day, 12-dose erythromycin in achieving clinical cure in children with severe cholera.

Methods We did a randomised, open-label, controlled trial in children age 2–15 years with *V cholerae* O1 or O139 present in stool on dark-field microscopy. Children received either a single 20 mg/kg dose of ciprofloxacin (n=90) or 12.5 mg/kg of erythromycin (n=90) every 6 h for 3 days, and remained in hospital for 5 days. The primary outcome was clinical success of treatment, defined as cessation of watery stools within 48 h of start of drug treatment. Analysis was per protocol. This study is registered as an International Standard Randomised Controlled Trial, number ISRCTN00142272.

Findings Of 180 children randomised 162 completed the study. Treatment was clinically successful in 60% (47/78) of children treated with ciprofloxacin and in 55% (46/84) of those treated with erythromycin (difference 5% [95% CI –10 to 21]). Children receiving ciprofloxacin vomited less often (58% vs 74%; difference 16% [2 to 30]), had fewer stools (15 vs 21; 6 [0 to 9]), and less stool volume (152 vs 196 mL/kg; 43 mL/kg [13 to 87]) than those receiving erythromycin. Bacteriological failure was more common in ciprofloxacin-treated patients (58% vs 30%; 28% [13 to 43]) than erythromycin-treated patients.

Interpretation Single-dose ciprofloxacin achieves clinical outcomes similar to, or better than, those achieved with 12-dose erythromycin treatment in childhood cholera, but is less effective in eradicating *V cholerae* from stool.
The fluoroquinolones have been used sparingly in children because of concerns about arthropathy. The fluoroquinolones have, however, been used in children for treatment of multiple-resistant Salmonella enterica serovar Typhi and Shigella dysenteriae type 1 infection, and studies done at the Diarrhoea Treatment Centres of the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B), and elsewhere, have not reported arthropathy in these children. We postulated that single-dose ciprofloxacin would be as effective as 3-day, 12-dose erythromycin in achieving clinical cure in children with severe cholera.

Methods

Patients

Study patients were enrolled between May 13, 2001, and June 30, 2002, at two diarrhoea treatment centres operated by the ICDDR, B: one in the capital city Dhaka, and the other in the rural Matlab District. Consecutive patients presenting for care at either site from 0600 h to 2200 h were screened for eligibility for study. Children eligible for study were age 2–15 years with a history of watery diarrhoea for 24 h or less. They also had clinical features of severe dehydration as defined by WHO criteria, which requires two of the following four signs in a child with diarrhoea: lethargy or unconsciousness; sunken eyes; inability to drink or drinking poorly; skin pinch goes back very slowly. Inclusion criteria also included identification of V cholerae O1 or O139 on initial dark-field examination of stool and subsequent isolation from a culture of stool, and documented severe diarrhoea as defined by a watery stool output of 20 mL/kg or more bodyweight during a 4 h observation period after completion of initial rehydration.

Children were ineligible for study if they had received an antimicrobial drug (including the study drugs) potentially effective in the treatment of V cholerae within 72 h before screening (as a matter of course patients coming to the ICDDR,B treatment centre usually bring with them any medicines they have been taking); had concomitant infections needing antimicrobial treatment or a concomitant illness that might interfere with the assessment of outcome or safety of the study drugs, including interfering with joint examination; or had cardiac or hepatic impairment (aspartate aminotransferase, alanine aminotransferase, or serum bilirubin >3 times upper limit of normal for age). Children were also excluded if they had enrolled in this study or another clinical study within the previous month or if they had known hypersensitivity to fluoroquinolones or macrolides.

Children eligible on initial screening were admitted to the study wards in the two treatment centres. Their weight (designated as initial weight) and vital signs were recorded; a physical examination, including assessment of dehydration according to WHO guidelines, was undertaken; and a stool specimen was obtained for rapid determination of V cholerae O1 or O139 infection with dark-field microscopy for detection of characteristic motile vibrios and inactivation of motility with antisera to V cholerae O1 and O139. Children were then rehydrated over 2–4 h with an intravenous polyelectrolyte solution containing sodium 133 mmol/L, chloride 98 mmol/L, potassium 13 mmol/L, and bicarbonate 48 mmol/L. Hydration was maintained over a 4 h observation period with rice-based oral rehydration solution containing sodium 90 mmol/L, chloride 80 mmol/L, potassium 20 mmol/L, citrate 10 mmol/L, rice powder 50 g/L, and additional intravenous solution if needed.

The ethics review committee of the ICDDR, B, and the institutional review board of the New England Medical Center approved this study. Parents or guardians provided written informed consent and children provided assent if they were 7 years or older.

Procedures

At the end of the observation period, children fulfilling eligibility criteria were randomly assigned to receive either a single 20 mg per kg bodyweight (maximum dose 750 mg) dose of ciprofloxacin oral suspension or erythromycin ethylsuccinate oral suspension in a dose of 12.5 mg per kg bodyweight (maximum individual dose 500 mg) every 6 h for 3 days. On enrolment children were assigned a consecutive study number that had been randomly pre-allocated to either of the two interventions by a computer-generated list prepared by individuals not otherwise involved in the study. Random numbers and treatment were stratified by the two treatment sites and, for both sites, a block randomisation with a block size of eight was used. Allocation of treatment was concealed by having both study drugs stored in identical sealed boxes that were only opened after a patient had been enrolled in the study and assigned a study number. The first dose of the study drug was given on enrolment, and each study day was defined as the 24 h counted from the start of study treatment.

Participating children remained in the treatment centre for 5 days or until resolution of their watery diarrhoea, whichever was longer. The parents or guardians were requested to bring their children to the respective treatment centres for two follow-up assessments 10–14 days and then 4–6 weeks after study enrolment. Physical examinations and interim histories (patients’ day to day progress was noted in a predesigned case record form) were obtained daily while in hospital, with a focus on signs of dehydration and signs or symptoms of arthropathy or arthritis. Clinical assessments were repeated more often if indicated. Vital signs were obtained every 6 h. Weight-for-age was calculated as a median of the National Centre for Health Statistics median for age. Stool volume (measured to a precision of 20 mL) and stool consistency, defined as watery (stool that could be poured like water), soft (stool that could not be poured like water but took the shape of a container), or...
formed (stools that retained their shapes) were assessed every 6 h, as were volume of urine, vomitus, and oral and intravenous fluid intake. For purposes of this study, diarrhoea was deemed present if watery stool was being passed.

We obtained stool specimens (or a rectal swab sample if a stool specimen was not available) before initial administration of the study drug, on study day 3, and at a follow-up visit for identification of *V cholerae* O1 and O139, *Salmonella*, *Shigella*, and *Campylobacter jejuni*. Rectal swab samples were taken before drug administration and daily while in hospital for identification of only *V cholerae* O1 and O139. Microscopic examination of stool was done before start of study treatment and on study day 4 for semi-quantitative measurement of leucocytes and erythrocytes, and for detection of enteric parasitic infections. A blood sample was obtained after rehydration but before the study drug was given, and on study day 5 for complete blood count and measurement of serum electrolytes and creatinine, total bilirubin, aspartate aminotransferase, alanine aminotransferase, and alkaline phosphatase concentrations. Antimicrobial susceptibility of *V cholerae* O1 and O139 was ascertained for tetracycline, co-trimoxazole, chloramphenicol, erythromycin, furazolidone, and ciprofloxacin by the disk-diffusion method. Minimum inhibitory concentrations (MIC) of the infecting strains of *V cholerae* O1 and O139 for ciprofloxacin and azithromycin were determined with the E test (AB Biodisk, Solna, Sweden). To identify the basic pharmacokinetic properties of ciprofloxacin in these children, blood samples were obtained from the first 40 patients enrolled in the study either before, 0–60 min after, 1–2 h after, or 4–12 h after the first dose of the study drug.

The rate of clinical success was the primary outcome measure of the study. Clinical success was defined as cessation of watery stool within 48 h of the start of the study drug without recurrence in the subsequent 72 h that children remained in hospital. Secondary outcome measures were bacteriological success of treatment, defined as the inability to isolate *V cholerae* from stool or rectal swab samples after study day 2; duration of diarrhoea, defined as the time from administration of the study drug to the last 6-h period in which watery stool was passed; duration of faecal excretion of *V cholerae* O1 or O139; total number of stools; total volume of watery stool; proportion of patients needing intravenous fluids after study drugs were given; frequency and volume of vomiting during the study; rates of adverse events; and pharmacokinetics of ciprofloxacin in children with cholera.

**Statistical analysis**

The primary aim of the study was to show that clinical cure with treatment with single-dose ciprofloxacin oral suspension was equivalent to that with a 12-dose, 3-day treatment regimen with erythromycin. The anticipated proportion of patients in which erythromycin resulted in clinical cure was assumed to be 0.80. With this assumption, and with a power of 90% and a two-sided significance level of 5%, we needed 80 patients in each group to reject the hypothesis that the proportion of patients in which ciprofloxacin resulted in cure was equal to or differed by more than 0.15 from the proportion of patients in the erythromycin group. On the assumption that 10% of enrolled patients would not provide valid data for the analysis of the primary outcome, we enrolled 180 patients. In view of the rate of clinical cure shown in the erythromycin group (55%), the actual power of the study to identify the postulated difference was 83%.

SPSS version 10.0 (SPSS, Chicago, Illinois, USA), EpiInfo 2000 version 1.1 (Centers for Disease Control and Prevention, Atlanta, Georgia, USA), and Confidence Interval Analysis version 2.1.2 (BMJ Books, London, UK) were used for statistical analysis. We used the χ² test with continuity correction to compare differences in proportions between the treatment groups, and Fisher’s exact test was done if the predicted size of any cell was five or less. To assess the

![Figure 1: Trial profile.](www.thelancet.com)
Table 1: Patient characteristics by treatment group

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Ciprofloxacin (n =78)</th>
<th>Erythromycin (n =84)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (months)</td>
<td>88 (41)</td>
<td>91 (42)</td>
</tr>
<tr>
<td></td>
<td>84 (54–120)</td>
<td>84 (60–129)</td>
</tr>
<tr>
<td>Illness characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration (h)</td>
<td>9 (4)</td>
<td>10 (4)</td>
</tr>
<tr>
<td></td>
<td>8 (5–12)</td>
<td>8 (5–12)</td>
</tr>
<tr>
<td>Number of stools since onset</td>
<td>11 (7)</td>
<td>10 (8)</td>
</tr>
<tr>
<td></td>
<td>10 (5–15)</td>
<td>8 (5–15)</td>
</tr>
<tr>
<td>Number of vomiting episodes since onset</td>
<td>8 (6)</td>
<td>11 (13)</td>
</tr>
<tr>
<td></td>
<td>6 (4–12)</td>
<td>8 (5–12)</td>
</tr>
<tr>
<td>Weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>On arrival (kg)</td>
<td>16·6 (7·1)</td>
<td>16·7 (7·1)</td>
</tr>
<tr>
<td></td>
<td>14·2 (12·2–19·8)</td>
<td>15·0 (12·8–20·5)</td>
</tr>
<tr>
<td>At discharge (kg)</td>
<td>18·4 (7·5)</td>
<td>18·7 (7·3)</td>
</tr>
<tr>
<td></td>
<td>16·2 (12·4–22·2)</td>
<td>17·3 (13·4–23·0)</td>
</tr>
<tr>
<td>% increase in weight</td>
<td>11·3 (5·6)</td>
<td>11·4 (6·0)</td>
</tr>
<tr>
<td></td>
<td>10·7 (7·2–12·8)</td>
<td>10·1 (7·4–13·2)</td>
</tr>
<tr>
<td>Weight for age*</td>
<td>73·3 (19·5)</td>
<td>75·1 (21·8)</td>
</tr>
<tr>
<td></td>
<td>73·5 (66·3–78·1)</td>
<td>69·4 (64·5–80·0)</td>
</tr>
<tr>
<td>Laboratory findings after correction of dehydration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Packed-cell volume (%)</td>
<td>36·4 (4)</td>
<td>37 (4)</td>
</tr>
<tr>
<td></td>
<td>36·3 (39–39)</td>
<td>36·3 (33–39)</td>
</tr>
<tr>
<td>Serum creatinine (µmol/L)</td>
<td>81 (25)</td>
<td>88 (34)</td>
</tr>
<tr>
<td></td>
<td>75 (64–95)</td>
<td>81 (66–96)</td>
</tr>
<tr>
<td>Total bilirubin (µmol/L)</td>
<td>5 (4)</td>
<td>5 (3)</td>
</tr>
<tr>
<td></td>
<td>5 (3–6)</td>
<td>5 (3–6)</td>
</tr>
<tr>
<td>Alkaline phosphatase (µU/L)</td>
<td>239 (79)</td>
<td>250 (157)</td>
</tr>
<tr>
<td></td>
<td>218 (173–283)</td>
<td>236 (188–265)</td>
</tr>
<tr>
<td>Alanine aminotransferase (µU/L)</td>
<td>27 (23)</td>
<td>27 (15)</td>
</tr>
<tr>
<td></td>
<td>22 (17–30)</td>
<td>23 (16–32)</td>
</tr>
<tr>
<td>Aspartate aminotransferase (µU/L)</td>
<td>35 (19)</td>
<td>36 (15)</td>
</tr>
<tr>
<td></td>
<td>30 (24–37)</td>
<td>32 (25–43)</td>
</tr>
<tr>
<td>During 4-h observation period after correction of dehydration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of stools</td>
<td>6 (3)</td>
<td>7 (4)</td>
</tr>
<tr>
<td></td>
<td>5 (4–8)</td>
<td>5 (4–8)</td>
</tr>
<tr>
<td>Episodes of vomiting</td>
<td>2 (2)</td>
<td>1 (2)</td>
</tr>
<tr>
<td></td>
<td>1 (1–2)</td>
<td>1 (1–2)</td>
</tr>
<tr>
<td>Fluid balance (ml kg⁻¹ h⁻¹) during the 4 h observation period</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stool</td>
<td>13·6 (7·8)</td>
<td>13·6 (3)</td>
</tr>
<tr>
<td></td>
<td>11·7 (8·2–17·4)</td>
<td>11·8 (8·6–16)</td>
</tr>
<tr>
<td>Vomit</td>
<td>2·7 (3·6)</td>
<td>2·7 (4·8)</td>
</tr>
<tr>
<td></td>
<td>1·3 (0–4)</td>
<td>0·5 (0–3·4)</td>
</tr>
<tr>
<td>Intravenous infusion</td>
<td>2·9 (6·9)</td>
<td>1·9 (5·2)</td>
</tr>
<tr>
<td></td>
<td>0·0 (0–0)</td>
<td>0·0 (0–0)</td>
</tr>
<tr>
<td>Oral rehydration solution</td>
<td>10·6 (4·5)</td>
<td>10·5 (4·6)</td>
</tr>
<tr>
<td></td>
<td>9·2 (7·2–13·3)</td>
<td>9·7 (7–13·3)</td>
</tr>
<tr>
<td>Number with pathogen isolated (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V cholerae O1</td>
<td>69 (88%)</td>
<td>72 (86%)</td>
</tr>
<tr>
<td>V cholerae O139</td>
<td>9 (12%)</td>
<td>12 (14%)</td>
</tr>
<tr>
<td>Number with other enteric pathogens isolated (%)†</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 (8%)</td>
<td>12 (14%)</td>
</tr>
</tbody>
</table>

Data are mean (SD) and median (IQR), unless otherwise stated. * Based on discharge weight percent of National Centre for Health Statistics Median for age. † Ciprofloxacin group: Salmonella spp (n=4); Campylobacter jejuni (n=2); Erythromycin group: Shigella flexneri (n=3); Campylobacter jejuni (n=8); Aeromonas salmonis (n=2).

Table 1: Patient characteristics by treatment group

significance of differences in continuous variables we used the two-tailed Student’s t test for normally distributed data, and the Mann-Whitney U test for non-normal data. Variables were assumed to be non-normally distributed if, for either treatment group, the standard deviation was ≥0.5 of the measured value, and a non-parametric test was applied. All tests of significance were two-tailed. The binomial method was used to calculate differences in medians between study groups and the confidence intervals for those differences; the Newcombe method was used to ascertain confidence intervals for differences in proportions. Patients eligible for the analysis of the primary outcome were all enrolled patients who met eligibility requirements, who did not vomit within 30 min of administration of the first dose of drug, and who stayed in hospital long enough (5 days) for determination of clinical success, or in whom treatment had been deemed to fail before this time point. All patients who received a dose of drug were included in the safety analysis. The initial analysis was done blinded to treatment allocation and was per protocol. However, we also analysed the primary outcome and the vomiting episode in each group on an intention-to-treat basis.

This study is registered as an International Standard Randomised Controlled Trial, number ISRCTN00142272.

Role of the funding source
The investigators were solely responsible for the study design; clinical management of the patients; data collection, analysis, and interpretation; writing of the manuscript; and the decision to submit for publication. A draft version of the manuscript was provided to Bayer AG for comment, but no changes were made to the manuscript in response to comments from Bayer. The investigators had initially designed this as a double-blinded study; Bayer AG was not able to provide drugs in a fashion that would mask the treatment. Bayer also arranged for regular monitoring of the study by a contract research organisation to ensure compliance with good clinical practice.

Results
Of 273 patients screened for enrolment into the study 180 met the eligibility criteria (figure 1). The most common reason for screened patients not being enrolled was absence of *V cholerae* O1 or O139 on dark-field assessment of stool; no patient was excluded because of antimicrobial use before admission. 90 patients were randomly assigned to each treatment group. 12 of those assigned to receive ciprofloxacin were excluded from the primary analysis and six patients in the erythromycin group were excluded from the primary analysis. Patient characteristics by treatment group were well matched on admission (table 1).

The clinical success in nearly two-thirds of the patients who received ciprofloxacin, and around a half of those who received erythromycin (table 2). If an intention-to-treat analysis is used to analyse outcome, and treatment was assumed to have failed in the 12 patients in the ciprofloxacin group and the six patients in the erythromycin group not included
in the above analysis, there was no significant difference between treatment groups (52% vs 51%; difference 1%, 95% CI –13% to 15%).

Children treated with ciprofloxacin had significantly fewer stools during the 5 days of study and a lower total median stool volume than children treated with erythromycin (table 2). Diarrhoea duration did not differ between groups (figure 2). Fewer children receiving ciprofloxacin vomited and they had a lower total volume of vomitus than did children receiving erythromycin (table 2). On an intention-to-treat analysis, the difference in the proportion of patients in each group who vomited remained significant (59% vs 74% in ciprofloxacin and erythromycin groups, respectively; difference 16% [95% CI 2% to 29%]).

Erythromycin resulted in a more rapid clearance of *V cholerae* O1 and O139 than did ciprofloxacin (figure 2). By study day three, *V cholerae* O1 or O139 could be isolated from 45 (58%) of the 78 patients treated with ciprofloxacin, and 25 (30%) of the 84 patients treated with erythromycin (difference 28% [13% to 42%]). All isolates of *V cholerae* O1 and O139 were susceptible to both ciprofloxacin and erythromycin as assessed by the disk-diffusion method, and all were inhibited by 0·25 μg/mL or less of ciprofloxacin (50th and 90th percentiles: 0·023 μg/mL and 0·047 μg/mL, respectively) and 1·00 μg/mL or less of erythromycin (both 50th and 90th percentiles are 0·75 μg/mL). The peak median serum ciprofloxacin concentration was 1·6 μg/mL (IQR 1·1–2·6), observed within 1–4 h of drug administration, and the median trough concentration was 1·2 μg/mL (IQR 0·7–1·6). When serum concentrations of ciprofloxacin in individual patients were compared with the MIC of the *V cholerae* isolate from that patient, peak concentrations were 50 times greater and trough concentrations were 45 times greater than the MIC.

74 of 78 (95%) children treated with ciprofloxacin returned for the first follow-up visit and 71 (91%) for the second follow-up visit. None had a history of watery diarrhoea or increased stool frequency after discharge. *V cholerae* O1 or O139 was not isolated from any of the children at either visit, and there were no reports of arthropathy. At the second follow-up visit, *V cholerae* O1 El Tor serotype Inaba was isolated from two patients who on entry into the study had *V cholerae* O1 El Tor serotype Ogawa. Because *V cholerae* strains can interconvert between serotypes, whether these two isolates represent continued excretion of the original infection or a new infection is not clear.

No child (or their parents on their behalf) reported joint symptoms at the follow-up visits, and no child had evidence of joint abnormalities on physical examination. In the erythromycin group, 75 (89%) children returned for first follow up and 66 (79%) returned for the second follow-up visit.

Treatment clinically failed in 69 of the 162 patients in the study. Children in whom treatment failed according to the predetermined criteria were younger, had a lower

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Ciprofloxacin (n=78)</th>
<th>Erythromycin (n=84)</th>
<th>Difference* (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical success, n (%)</td>
<td>47 (60%)</td>
<td>46 (55%)</td>
<td>5% (-10 to 21)</td>
<td>0·58</td>
</tr>
<tr>
<td><strong>Secondary outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteriological success, n (%)</td>
<td>33 (42%)</td>
<td>59 (70%)</td>
<td>28% (13 to 43)</td>
<td>&lt;0·0001</td>
</tr>
<tr>
<td>Diarrhoea duration (h)</td>
<td>50 (34)</td>
<td>52 (25)</td>
<td>6 (–6 to 12)</td>
<td>0·25</td>
</tr>
<tr>
<td>Number of stools after start of study drug</td>
<td>26 (38)</td>
<td>31 (30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of patients who vomited after start of study drug (%)</td>
<td>15 (8–31)</td>
<td>21 (22–38)</td>
<td>4 (0 to 9)</td>
<td>0·05</td>
</tr>
<tr>
<td>Fluid balance after start of study drug (mL/kg)†</td>
<td>254 (301)</td>
<td>315 (312)</td>
<td>43 (13 to 87)</td>
<td>0·05</td>
</tr>
<tr>
<td>Vomiting</td>
<td>7 (0–40)</td>
<td>19 (0–45)</td>
<td>9 (0 to 11)</td>
<td>0·19</td>
</tr>
<tr>
<td>Intravenous fluids</td>
<td>0 (0–112)</td>
<td>0 (0–181)</td>
<td>0 (0 to 0)</td>
<td>0·31</td>
</tr>
<tr>
<td>Oral rehydration solution</td>
<td>228 (200)</td>
<td>258 (217)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Primary outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical success, n (%)</td>
<td>47 (60%)</td>
<td>46 (55%)</td>
<td>5% (-10 to 21)</td>
<td>0·58</td>
</tr>
<tr>
<td><strong>Secondary outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteriological success, n (%)</td>
<td>33 (42%)</td>
<td>59 (70%)</td>
<td>28% (13 to 43)</td>
<td>&lt;0·0001</td>
</tr>
<tr>
<td>Diarrhoea duration (h)</td>
<td>50 (34)</td>
<td>52 (25)</td>
<td>6 (–6 to 12)</td>
<td>0·25</td>
</tr>
<tr>
<td>Number of stools after start of study drug</td>
<td>26 (38)</td>
<td>31 (30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of patients who vomited after start of study drug (%)</td>
<td>15 (8–31)</td>
<td>21 (22–38)</td>
<td>4 (0 to 9)</td>
<td>0·05</td>
</tr>
<tr>
<td>Fluid balance after start of study drug (mL/kg)†</td>
<td>254 (301)</td>
<td>315 (312)</td>
<td>43 (13 to 87)</td>
<td>0·05</td>
</tr>
<tr>
<td>Vomiting</td>
<td>7 (0–40)</td>
<td>19 (0–45)</td>
<td>9 (0 to 11)</td>
<td>0·19</td>
</tr>
<tr>
<td>Intravenous fluids</td>
<td>0 (0–112)</td>
<td>0 (0–181)</td>
<td>0 (0 to 0)</td>
<td>0·31</td>
</tr>
<tr>
<td>Oral rehydration solution</td>
<td>228 (200)</td>
<td>258 (217)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data are mean (SD) and median (IQR) or number (%). * Differences are calculated using medians as the data are not normally distributed. The differences between medians in the two treatment groups are estimated by the median of all possible differences between patients in the two groups rather than the arithmetic difference in the two population medians. † Based on discharge weight.

---

Figure 2: Kaplan-Meier plots of duration of diarrhoea, by treatment group and study day (A), and of isolation rates of *V cholerae* O1 and O139, by patient group and study day (B)
admission weight-for-age, had a greater stool volume during the observation period, and gained less weight during the study (table 3). When analysed by logistic regression, adjustment for weight-for-age did not affect the probability of success of treatment with the two different regimens (unadjusted odds ratio 0·80; adjusted odds ratio 0·79).

### Discussion

Clinical cure rate did not differ between children with severe cholera given single-dose ciprofloxacin treatment and those given the 3-day, 12-dose course of erythromycin currently recommended by WHO. In most secondary outcome measures—stool frequency, median stool volume, and proportion of patients who had vomiting—ciprofloxacin was significantly more effective than erythromycin. With a predetermined definition of clinical success—resolution of diarrhoea within 48 h of the start of treatment and no recurrence during 5 days stay in hospital—the 60% cure rate achieved with ciprofloxacin was substantially less than the 94% rate of clinical cure achieved with single-dose ciprofloxacin treatment in adults with severe cholera.

The difference in rates of clinical cure is consistent with previous findings in cholera-endemic areas of childhood cholera, being more severe than adult cholera both in terms of total stool volume per unit of bodyweight and in duration of illness. By comparison with previous trials in children, however, single-dose ciprofloxacin in this study did quite well. The mean duration of diarrhoea was shorter than the mean duration of diarrhoea in most of the 19 treatment regimens assessed in the seven previous randomised controlled trials of antimicrobial treatment for childhood cholera that we could identify (table 4).

Single-dose ciprofloxacin was inferior to erythromycin in achieving bacteriological eradication in this study. The results achieved were also substantially inferior to those reported when single-dose ciprofloxacin was used in adult cholera patients, where the rate of bacteriological success was 95%. This difference might be explained in part by diminished susceptibility of *V cholerae* isolates recovered from patients enrolled in this study compared with isolates from patients in the study of single-dose ciprofloxacin in adults. In our study only 50% of all isolates were inhibited by 0·023 µg/mL or less of ciprofloxacin, whereas in the earlier study all isolates were inhibited by 0·003 µg/mL.

Peak serum concentrations of ciprofloxacin in children in this study were similar to those reported in adults in the earlier study, and they were still far in excess of the MIC of the infecting strain of *V cholerae* O1 or O139 (median 50 times that of infecting strain). We did not measure stool concentrations of drug in this study, and it is possible that the high stool drug concentrations observed in adults receiving single-dose ciprofloxacin were not achieved in children in the current study. Still, in view of the similar peak serum concentrations there is no reason to think that stool concentrations would have differed substantially, and the increased MIC could have had a role in the prolonged excretion of *V cholerae* compared with our previous observations in adults. Increases in MIC have been shown to result in diminished efficacy of fluoroquinolones in treating enteric infections, even though strains remain susceptible in vitro when tested by the disk-diffusion method or with standard breakpoints.
for MIC testing.\textsuperscript{25} In this study bacteriological failure and clinical failure were linked: 38% of the patients receiving ciprofloxacin in whom treatment was clinically successful were classified as having bacteriological failure (continued excretion of \textit{V cholerae} after 48 h on drug therapy), compared with 87% of those in whom treatment was not clinically successful.

The public health importance of this prolonged excretion of \textit{V cholerae} is uncertain. Concentration of organisms in stool is probably greatest early in the course of illness,\textsuperscript{26,27} and thus exposure of others in the community is probably highest before treatment has been started. Additionally, most \textit{V cholerae} infections are asymptomatic,\textsuperscript{28} and for every person identified with an infection that can cause a clinical syndrome similar to that caused by \textit{V cholerae} O1 or O139. Because all patients enrolled in this study had culture-confirmed \textit{V cholerae} infection, and their clinical symptoms were consistent with cholera, we think that enterotoxigenic \textit{E coli} infections probably did not play a major part in clinical illness in these patients.

The practical advantages of single-dose ciprofloxacin when compared with 12-dose 3-day erythromycin treatment are self-evident. Despite concerns based on experiments in juvenile animals of joint toxicity when fluoroquinolones are used in children, this adverse effect has not been a problem when the fluoroquinolones are used in India and Bangladesh where cholera is endemic. In Bangladesh, the cost of treating a 10 kg child with the dose of ciprofloxacin used in this study is US$0–11 (if used in the tablet form, since it is not available in suspension formulation); treatment with erythromycin suspension in the dose used in this study was available in suspension formulation); treatment with...
would cost US$0·60, five and a half times higher than that of ciprofloxacin. Development of resistance if the fluoroquinolones are used as first-line drugs has also been a concern. But resistance is much more likely to arise from the widespread misuse of fluoroquinolones and other antimicrobial agents in the largely unregulated private pharmacies that are commonplace in developing countries, rather than from controlled use in medical facilities for specific and well defined indications, such as clinically severe cholera.

WHO recommendations are usually followed by developing countries in establishing national treatment guidelines. They recommend tetracycline or erythromycin in 12 doses over 3 days as the treatment of choice for childhood cholera. Other drugs that have been used have substantial limitations—either a narrow therapeutic ratio (chloramphenicol) or high resistance (co-trimoxazole, ampicillin). Patterns of resistance can rapidly change, however. Strains of *V* cholerae O1 resistant to tetracycline, co-trimoxazole, furazolidone, and erythromycin have been identified in Bangladesh. Reports from India have identified small numbers of *V* cholerae O1 that are resistant to ciprofloxacin or that have increased MIC, such as was shown in this study. These reports reinforce the need for routine surveillance of resistance to guide empiric antimicrobial treatment of cholera.

Patients in whom treatment clinically failed were younger, perhaps suggesting a less robust host response than in those in whom clinical cure was achieved. These patients may not have had previous exposure to *V* cholerae that elicits some degree of pre-existing immunity. Patients in whom treatment failed also had a lower weight-for-age, which could result either from malnutrition or more severe disease with consequent greater loss of body water, rather than loss of lean body mass, than those in whom treatment was successful. That patients in whom treatment failed also had a greater volume of watery stool during the observation period, and had less weight gain during their stay in the treatment centre, suggests that there could have been ongoing increased fluid deficits in these children, rather than malnutrition.

Ciprofloxacin and azithromycin, when used in single doses, have been shown to be as (or more) effective clinically than erythromycin or tetracycline. Both drugs, given their efficacy, their current affordability, and their high therapeutic ratio, should be considered as options for first-line treatment for childhood cholera in areas where *V* cholerae infections are caused by strains susceptible to these drugs.

Conflict of interest statement
Bayer AG supported the travel of W A Khan to present the results of the study at the 43rd Interscience Conference on Antimicrobial Agents and Chemotherapy, Chicago, USA, in 2003.

Acknowledgments
This research was undertaken in collaboration with New England Medical Center grant number GR-00067 and supported by a grant from Bayer AG (IMPACT No. 10110 /AP 176). MLB was supported by grants from the National Institute of Allergy and Infectious Diseases of the United States National Institute of Health (U01 AI45508-01, K24 AI0001671), and the Wellcome Trust (Wellcome 62925). ICDDR, B, acknowledges with gratitude the commitment of New England Medical Center and Bayer AG to the Centre’s research efforts. We thank Sabeena Ahmed for maintaining the drug inventory and dispensing study medication and for providing laboratory support; Humayun Kabir for assistance with data entry; T S Van D, Singeck for advice on statistical methods; and the nurses and staff of the clinical research wards at Dhaka and Matlab hospitals of ICDDR, B, for providing excellent care and support to the participating children and their attendants.

References

Contributors
W A Khan, M A Salam, and M I Bennish conceived and designed the study. M I Bennish obtained funding for the study. W A Khan, M M Karim, H R Chowdhury, and D Saha enrolled patients into the study and supervised their management under the overall supervision of M A Salam. D Saha, W A Khan, and M I Bennish did the data analysis. D Saha and M I Bennish drafted the manuscript, and all authors reviewed the manuscript and approved the final version.


Comparison of the effect of two systems for the promotion of exclusive breastfeeding

Sonia Bechara Coutinho, Pedro Israel Cabral de Lira, Marilia de Carvalho Lima, Ann Ashworth

Summary

Background Promotion of breastfeeding is an important child-survival intervention, yet little is known about which promotional strategies are the most effective. We aimed to compare the effects on rates of breastfeeding of two systems for promotion of breastfeeding in Brazil—a hospital-based system and the same system combined with a programme of home visits.

Methods In February, 2001, maternity staff from two hospitals in Pernambuco, Brazil, were trained according to the Baby-Friendly Hospital Initiative (BFHI). In a randomised trial between March and August, 2001, 350 mothers giving birth at these hospitals were assigned ten postnatal home visits to promote and support breastfeeding (n=175) or no home visits (n=175). Breastfeeding practices were studied on days 1, 10, 30, 60, 90, 120, 150, and 180 by researchers unaware of group allocation. The primary outcome measure was the rate of exclusive breastfeeding from birth to 6 months. Analyses were by intention to treat.

Findings The hospital-training intervention achieved a high rate (70%) of exclusive breastfeeding in the hospitals, but this rate was not sustained at home and at 10 days of age only 30% of infants were exclusively breastfed The patterns of exclusive breastfeeding in the two trial groups for days 10–180 differed significantly (p<0.0001), with a mean aggregated prevalence of 45% among the group assigned home visits compared with 13% for the group assigned none.

Interpretation The BFHI achieves high rates of exclusive breastfeeding in hospital; however, in Brazil at least, the rates fall rapidly thereafter. Reliance on the BFHI as a strategy for breastfeeding promotion should be reassessed. A combination of promotional systems (hospital-based and in the community) is needed.

Introduction

The Lancet’s Child Survival series1–5 drew attention to the unacceptably high rates of child mortality that continue in low-income countries and poor areas of middle-income countries. Most of the 10–8 million child deaths during the year 2000 were from preventable causes, especially neonatal disorders, pneumonia, and diarrhoea.6 If the few interventions for which there is sufficient evidence of effect (level 1)6 or limited evidence of effect (level 2) were fully implemented, 63% of deaths of children younger than 5 years could be prevented. If 90% of infants were exclusively breastfed at 0–5 months and continued to be breastfed from 6 months to 11 months, there would be an estimated 13% reduction in child deaths worldwide.7 This potential reduction in mortality is higher than for any other level 1 intervention. Current rates of exclusive breastfeeding are far below 90% in most countries, and in some, for example in Latin America, even the duration of breastfeeding is short.

The third paper in the Child Survival series highlighted the need to consider systems necessary to put an intervention in place.1 In relation to breastfeeding promotion, there is little information as to which strategies are the most effective in promoting exclusive breastfeeding and achieving high and equitable coverage.8 We report a randomised trial comparing the effect on rates of exclusive breastfeeding of two systems to promote breastfeeding in northeastern Brazil. The interventions were a hospital-based system, in which maternity staff were trained with the course content for the Baby-Friendly Hospital Initiative (BFHI), and a combination of this hospital-based system and a community-based system providing ten postnatal home visits. We also examined whether the effect applied equally among families below and above the poverty line and how it was related to maternal education, since the most disadvantaged infants are more likely to be exposed to health risks than those who are more affluent.8

Methods

Study site and participants

The study was done in the urban areas of Palmares and three neighbouring small towns (Catende, Água Preta, and Joaquim Nabuco) in the interior of the State of Pernambuco, northeastern Brazil. Their combined population is 135 000. The area is hilly and lies 130 km southwest of Recife, the State capital. The climate is hot and humid, and the economy of the region is mostly based on growing and processing sugar cane. Poverty is widespread. The adult female illiteracy rate is around 26%, and the infant mortality rate in 2000 was 76·5 per 1000 livebirths. HIV/AIDS incidence is thought to be very low. Palmares has three public maternity hospitals, although one did not function from August, 2000, when floods destroyed the infrastructure and equipment, until
December, 2001. More than 90% of births occur in hospital, and most women in the four towns give birth in Palmares. Midwives are responsible for routine births, and doctors are called for caesarean or emergency deliveries. The usual stay is 24 h after vaginal births and 48 h after caesarean deliveries.

In the preintervention study and in the randomised trial, all singleton infants were eligible except those with congenital anomalies or serious illness necessitating intensive care and those whose mothers had serious disease or mental illness or were planning to leave the area within 6 months. Infants weighing less than 2500 g at birth were excluded from the preintervention cohort.

**Design and objectives**

Figure 1 outlines the stages of the study. Preintervention data were obtained for a cohort of infants born in the three hospitals between January and August, 1998. The findings showed that maternity practices were poor, rates of exclusive breastfeeding were very low (median duration 0 days), and the duration of any breastfeeding was short (median 116 days). The second stage (February, 2001) provided maternity staff at the two functioning hospitals with training used by the BFHI. The third stage (March to August, 2001) was a randomised trial in which mother-infant pairs in the hospitals where staff had been trained were randomly assigned either ten postnatal home visits or no home visits. For each cohort, breastfeeding data were collected prospectively for 6 months.

The main objective was to compare the hospital-based intervention (BFHI training of maternity staff) with a combined hospital-based and community-based intervention (BFHI training and postnatal home visits). The primary outcome measure was rates of exclusive breastfeeding from birth to 6 months.

For the trial, mother-infant pairs in the maternity wards in March to August, 2001, were randomised in blocks of ten per town by use of a random numbers table (Epi-Info 6.04). The random numbers were generated by two maternity-based research assistants. Concealment was achieved by drawing numbers from envelopes at the time of assignment. The invitation to participate in the research was given at the maternity hospital before assignment.

WHO definitions were used: infants were classified as exclusively breastfed if they received only breastmilk (no water, other liquids, or solids) and as breastfed if they received breastmilk plus other food or liquid (including other milk). Other milk was defined as any non-breastmilk.

20 h of training were provided for health professionals and support staff of the two functioning hospitals in Palmares in early February, 2001, and 90% of the midwives and nursing assistants attended. Doctors were invited but did not participate. The training programme was the 18 h UNICEF/WHO course for training Baby-Friendly Hospitals and 2 h focusing on how to listen, to learn from mothers, to establish good relationships, to build mothers’ confidence, and to offer support, taken from the WHO/UNICEF Breastfeeding Counselling Course. Training was led by one of us (SBC) who is an accredited lactation counsellor and former BFHI assessor. Copies of the UNICEF norms and routines for the encouragement of breastfeeding were offered to the hospital managers, together with posters, educational folders for mothers, and a Ministry of Health videotape on lactation management. Two copies of the book Helping mothers to breastfeed were given to each maternity hospital.

The Ministry of Health has a national programme to deploy community health agents to make home visits, but this programme was not fully established in the study area. Five women were therefore recruited to serve as home visitors for the study. Their educational background (secondary school) was similar to that of community health agents in the national programme and, as for community health agents, personal breastfeeding experience was not a prerequisite. The home visitors received the same 20 h training as the maternity staff plus
5 days in which they studied Helping mothers to breastfeed in depth and practised how to discuss key topics with mothers by use of an illustrated booklet.

For the hospital-based intervention, from March, 2001, maternity staff in Palmares were expected to support, guide, and encourage all mothers to initiate and maintain exclusive breastfeeding throughout their hospital stay and at home for 6 months, and to continue breastfeeding for at least 2 years. Skin-to-skin contact in the delivery room, breastfeeding within the first 30 min, rooming-in, help with positioning, correct breastfeeding technique, and no bottles or pacifiers were expected norms, together with other features of Baby Friendly steps 4–9. Staff were expected to show the video daily, to talk to mothers individually answering their questions and discussing doubts, and to advise them to return to hospital if they experienced any breastfeeding difficulties at home.

For the combined hospital and community intervention, from March, 2001, the home visitors were expected to make home visits to mothers who had given birth in Palmares and who had been randomly assigned home visits. They were expected to visit ten times—four times during the first month (on days 3, 7, 15, and 30), every 2 weeks during the second month, and once a month during the third to sixth months. Each mother was to be given the illustrated booklet. At each visit, the home visitors were expected to encourage exclusive breastfeeding for 6 months and continued breastfeeding for at least 2 years, to answer questions and discuss doubts, and to use the booklet as a basis for discussions of key topics relevant to the infant’s age. Whenever possible, they observed the positioning of the infant at the breast, flow of milk, and the baby’s satisfaction; if there were difficulties that they could not resolve, they were expected to refer the mother for more specialist help at the hospital. If other family members were present, their attitude towards exclusive breastfeeding was assessed and their support was sought, including help with household chores. Each visit had a mean duration of 30 min, with the initial visits taking longer than later ones.

Data were collected in the trial by four researchers who were not aware of group allocation and were unconnected with the delivery of the interventions. Mothers in the trial were not close neighbours, so discussion with other mothers is unlikely, but we did not formally assess whether masking was maintained.

Information on maternity-ward practices was obtained through interviews with mothers in the maternity ward (preintervention cohort) or at home (trial) and included delivery-room practices, rooming-in, assistance given in establishment of breastfeeding (positioning, manual expression), and advice given about feeding other liquids and use of bottles and pacifiers.

Breastfeeding practices up to 6 months were assessed, starting in the maternity ward (day 1). In the preintervention study, households were visited twice a week for 6 months. For the trial, data were obtained at home on days 10, 30, 60, 90, 120, 150, and 180. Information was obtained by means of a structured questionnaire on breastfeeding and use of water, tea, other liquids, other milk, pacifiers, and bottles in the previous 24 h. The time of their first introduction was also recorded.

Information on sociodemographic and environmental characteristics was obtained at delivery by means of precoded, structured questionnaires. The data included information about income, parental education and literacy, family structure, household possessions (television, radio, refrigerator), housing quality, water supply, sanitation, and waste disposal.

Permission was granted before the study by the Ethical Committee of the Federal University of Pernambuco. All mothers gave fully informed written consent.

Statistical analysis
For the preintervention study, 364 mother-infant pairs were recruited. For the trial, we calculated that a sample of 130 mother-infant pairs per group would give 90% power for a 15% difference at 6 months in the prevalence of exclusive breastfeeding to be detected at the 5% significance level (two-sided comparison), with the assumption that the proportion of women exclusively breastfeeding in the non-visited group would be 5%. To allow for possible losses, a target of at least 175 per group was set.

All recording forms were precoded and checked daily for completeness and consistency. Double data entry was verified by use of Epi-Info version 6.04. Statistical analyses were by intention to treat. The patterns over time (days 10–180) of the proportions breastfeeding were compared between groups by use of randomisation tests applied to the sum of the log odds-ratios at each timepoint. The test statistics were calculated with the Gauss Computer Package (version 3.2.38). χ² tests were used for proportions (except where indicated) with the Statistical Package for the Social Sciences (version 8.0).

Role of the funding source
The sponsor of the study had no role in study design; collection, analysis, or interpretation of data; or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit the paper for publication.

Results
In the preintervention study, 364 mother-infant pairs were recruited and 46 (13%) were lost to follow-up at 6 months (figure 1). In the trial, 350 mother-infant pairs were recruited and 20 (6%) were lost to follow-up (one sudden infant death; one congenital malformation diagnosed after recruitment; and 18 moved from the area, 13 in the non-visited group and five in the visited group). In both the preintervention study and the trial, the mother-infant pairs lost did not differ from those who remained for any of the variables studied.
In both the preintervention cohort and the trial, more than half the families had incomes below the poverty line of 0.5 minimum salaries per person per month (equivalent to US$60 preintervention and $40 at the trial). Many were living in environments with no indoor toilet (preintervention 42%; trial 33%) or waste disposal (preintervention 32%; trial 27%). The proportion of adolescent mothers was similar in the two parts of the study (preintervention 36%; trial 33%), and for many women the baby was their first (preintervention 37%; trial 38%). Most mothers had received at least some antenatal care (preintervention 76%; trial 73%), and for many women the birth was their first (preintervention 37%; trial 38%).

Of the four home visits planned for the first month, on average five (82.6%) were completed. 99.6% were completed. Of the six subsequent planned visits, on average five (82%) were completed.

Of the six subsequent planned home visits, on average five (82.6%) were completed. 99.6% were completed. Of the six subsequent planned home visits, on average five (82%) were completed.

Table 1: Characteristics of the study population (with and without home visits)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>Home visits (n=175)</th>
<th>No home visits (n=175)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Group</td>
<td>Home visits (n=175)</td>
<td>No home visits (n=175)</td>
</tr>
<tr>
<td>Family income per head*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.5 $ minimum wage</td>
<td></td>
<td>107 (61%)</td>
<td>102 (58%)</td>
</tr>
<tr>
<td>≥0.5 $ minimum wage</td>
<td></td>
<td>68 (39%)</td>
<td>73 (42%)</td>
</tr>
<tr>
<td>Waste collection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>125 (71%)</td>
<td>129 (74%)</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>50 (29%)</td>
<td>46 (26%)</td>
</tr>
<tr>
<td>Toilet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flush</td>
<td></td>
<td>118 (67%)</td>
<td>117 (67%)</td>
</tr>
<tr>
<td>None/latrine</td>
<td></td>
<td>57 (33%)</td>
<td>58 (33%)</td>
</tr>
<tr>
<td>Water piped in house</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>158 (90%)</td>
<td>154 (88%)</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>17 (10%)</td>
<td>21 (12%)</td>
</tr>
<tr>
<td>Maternal age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger than 20 years</td>
<td></td>
<td>52 (30%)</td>
<td>64 (37%)</td>
</tr>
<tr>
<td>20 years or older</td>
<td></td>
<td>123 (70%)</td>
<td>111 (63%)</td>
</tr>
<tr>
<td>Mother literate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>132 (75%)</td>
<td>131 (75%)</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>43 (25%)</td>
<td>44 (25%)</td>
</tr>
<tr>
<td>Antenatal care</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>162 (93%)</td>
<td>167 (95%)</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>13 (7%)</td>
<td>8 (5%)</td>
</tr>
<tr>
<td>Parity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One</td>
<td></td>
<td>64 (37%)</td>
<td>70 (40%)</td>
</tr>
<tr>
<td>Two or more</td>
<td></td>
<td>111 (63%)</td>
<td>105 (60%)</td>
</tr>
<tr>
<td>Birthweight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (&lt;2500g)</td>
<td></td>
<td>7 (4%)</td>
<td>9 (5%)</td>
</tr>
<tr>
<td>Delivery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaginal</td>
<td></td>
<td>123 (70%)</td>
<td>127 (73%)</td>
</tr>
<tr>
<td>Caesarean</td>
<td></td>
<td>52 (30%)</td>
<td>48 (27%)</td>
</tr>
</tbody>
</table>

*Equivalent to US$40 per month.

Table 2: Comparison of activities in the maternity wards to promote and support breastfeeding in 1998 (before intervention) and 2001 (after training), according to mothers and fathers.

<table>
<thead>
<tr>
<th>Activity</th>
<th>1998 (n=364)</th>
<th>2001 (n=349)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 4 Skin-to-skin contact in delivery room</td>
<td>94 (26%)</td>
<td>131 (38%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Helped to breastfeed in delivery room</td>
<td>21 (6%)</td>
<td>22 (6%)</td>
<td>0.89</td>
</tr>
<tr>
<td>Step 5 Shown how to breastfeed (positioning</td>
<td>35 (10%)</td>
<td>80 (23%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>and attachment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 6 Infant given only breastmilk</td>
<td>77 (21%)</td>
<td>244 (70%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Given no water/tea</td>
<td>102 (28%)</td>
<td>280 (80%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Given no other milk</td>
<td>250 (99%)</td>
<td>345 (99%)</td>
<td>1.00</td>
</tr>
<tr>
<td>Step 7 Roomed-in</td>
<td>High</td>
<td>307 (88%)</td>
<td></td>
</tr>
<tr>
<td>Step 8 Advised to breastfeed on demand</td>
<td>NA</td>
<td>41 (12%)</td>
<td></td>
</tr>
<tr>
<td>Step 9 Advised not to give pacifiers</td>
<td>NA</td>
<td>260 (75%)</td>
<td></td>
</tr>
<tr>
<td>Advised not to give bottles</td>
<td>NA</td>
<td>102 (29%)</td>
<td></td>
</tr>
</tbody>
</table>

Data for 1998 are from Marques and colleagues. NA=data not available. *Fisher’s exact test.

Figure 2: Proportions of infants exclusively breastfed from birth to 6 months when born in untrained hospitals (Before intervention, 1998) and after training (2001), with and without home visits.

In 2001 (after training), most practices had improved but delivery-room practices remained poor, and neither hospital attained baby-friendly status. Nevertheless, 70% of infants were exclusively breastfed in hospital compared with 21% in 1998 (p<0.0001).

Although the hospital-training intervention was associated with a significant increase in the proportion of infants exclusively breastfed in the maternity hospitals, the practice was not sustained and at 10 days of age only 53 (30%) of the 175 infants were exclusively breastfed (figure 2). By 30 days of age, the proportion had fallen to 26 (15%) of 168. When the patterns of exclusive breastfeeding in the two trial groups were compared for days 10–180, they differed significantly (p=0·0001), with a mean aggregated prevalence of 45% among the group assigned home visits compared with 13% for the group assigned none. The mean aggregated prevalence before the intervention was 7%. The difference in patterns for days 10–180 between the trial group assigned the hospital-training intervention but no home visits and the preintervention cohort was small but significant (p=0·0002).

The hospital-training intervention was associated with a significant increase in the proportion of infants exclusive breastfeeding (table 2).
breastfed at least partially in the maternity hospital (81% vs 70% before the intervention, p=0.009; figure 3) but the improvement was not sustained. The pattern of breastfeeding in the two trial groups for days 10–180 differed significantly (p<0.0001); the mean aggregated prevalence was 78% among the group assigned home visits compared with 62% for the group assigned none. When compared with the preintervention pattern, the hospital-training intervention was not associated with a significant difference (p=0.31), the mean aggregated prevalence before the intervention being 63%.

We investigated whether the hospital-training intervention, or hospital training plus home visits, had similar effects on rates of exclusive breastfeeding at 30 days among families below or above the poverty line and according to maternal educational attainment. After the hospital-training intervention, the proportions of better-off mothers (p=0.02) and better-educated mothers (p=0.01) who breastfed exclusively at 3 months increased, the proportion was 15%.

We found that the high rates achieved in hospital are very short-lived. Within 10 days, only 30% of infants were exclusively breastfed, and at 1 month the proportion was 15%.

Braun and colleagues16 found in Porto Alegre in southern Brazil that after BFHI implementation, exclusive breastfeeding rates in the first 6 months of life remained low; they concluded that the BFHI is insufficient to maintain the high rates achieved in hospital. In Italy, high rates of exclusive breastfeeding were achieved in eight hospitals after staff were trained, but again the benefit was not sustained.17 By contrast, the PROBIT randomised trial in Belarus of a training programme modelled on the BFHI found that 43% of infants were exclusively breastfed at 3 months compared to the preintervention pattern.

**Discussion**

The BFHI is the most widely promoted international programme to increase rates of exclusive breastfeeding and to extend breastfeeding duration. There are more than 18,000 baby-friendly hospitals worldwide, and Brazil has 289, more than any other country. The BFHI is based on *Ten Steps to Successful Breastfeeding,*18 and the evidence of effectiveness for each of the ten steps has been documented.19,20 Although the maternity hospitals in our study did not attain baby-friendly certification, the BFHI training programme was used and was associated with a striking improvement in exclusive breastfeeding in hospital, with 70% of infants exclusively breastfed compared with 21% previously. The significant effect of the BFHI on rates of exclusive breastfeeding while in hospital is well documented, but there have been few studies to test whether the benefit is sustained at home. We found that the high rates achieved in hospital are very short-lived. Within 10 days, only 30% of infants were exclusively breastfed, and at 1 month the proportion was 15%.

**Table 3:** Comparison of the proportions given water, tea, or other milk and using pacifier or bottle among infants born in hospitals in 2001 after training, with and without home visits

<table>
<thead>
<tr>
<th>Day</th>
<th>Sample responding</th>
<th>Water</th>
<th>Tea</th>
<th>Milk</th>
<th>Pacifier</th>
<th>Bottle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No visits (n=175)</td>
<td>Visits</td>
<td>No visits</td>
<td>Visits</td>
<td>No visits</td>
<td>Visits</td>
</tr>
<tr>
<td>10</td>
<td>175</td>
<td>174</td>
<td>89 (51%)</td>
<td>21 (12%)</td>
<td>98 (56%)</td>
<td>29 (17%)</td>
</tr>
<tr>
<td>30</td>
<td>168</td>
<td>174</td>
<td>115 (69%)</td>
<td>42 (24%)</td>
<td>83 (49%)</td>
<td>26 (15%)</td>
</tr>
<tr>
<td>60</td>
<td>167</td>
<td>171</td>
<td>128 (77%)</td>
<td>63 (37%)</td>
<td>61 (37%)</td>
<td>18 (11%)</td>
</tr>
<tr>
<td>90</td>
<td>166</td>
<td>167</td>
<td>138 (81%)</td>
<td>73 (44%)</td>
<td>42 (25%)</td>
<td>10 (6%)</td>
</tr>
<tr>
<td>120</td>
<td>164</td>
<td>166</td>
<td>133 (81%)</td>
<td>84 (51%)</td>
<td>36 (22%)</td>
<td>10 (6%)</td>
</tr>
<tr>
<td>150</td>
<td>160</td>
<td>166</td>
<td>130 (81%)</td>
<td>95 (57%)</td>
<td>24 (15%)</td>
<td>17 (10%)</td>
</tr>
<tr>
<td>180</td>
<td>161</td>
<td>169</td>
<td>134 (83%)</td>
<td>91 (54%)</td>
<td>21 (13%)</td>
<td>8 (5%)</td>
</tr>
</tbody>
</table>

*Proportion significantly lower among the group assigned home visits than in the group assigned none; χ² p<0.0001 at each timepoint except 150 days (p=0.0001), 150 days (p=0.0001), and 180 days (p=0.0006) for the proportions using pacifier.

Figure 3: Proportions of infants breastfed at least partially from birth to 6 months when born in untrained hospitals (Before intervention, 1998) and after training (2001), with and without home visits.
with 6% of infants in control sites (p<0·001). Unlike most other countries, mothers in Belarus normally stay in hospital for 6–7 days postpartum and have about 3 years’ obligatory maternity leave. Also for that trial, polyclinic staff were trained to provide postnatal support, and infants were seen routinely every month. Thus there was more than customary opportunity to establish successful lactation in hospital and for continuing postnatal support, and the quality and frequency of support might have differed little from that provided by home-based strategies. In our study setting, mothers stayed in hospital only 24–36 h or 48 h after caesarean deliveries. These short stays are typical of Latin America and countries where under-funded health systems struggle to meet demand, and thus the study has external validity. Early discharge precludes extended individual contact and support and might have contributed to the limited influence of the hospital-training intervention and the stronger effect of the home visits. Home support is likely to be especially important in countries where mothers stay in hospital for a short time.

In our randomised trial, postnatal visits helped to sustain the higher rates of exclusive breastfeeding associated with the hospital-based training programme. The strengths of the study are the randomisation of the home-visiting intervention and the prospective follow-up, which avoided recall bias. One possible limitation is that there were still shortfalls in helping mothers to breastfeed immediately after delivery. Although there was low adherence with advising breastfeeding on demand and not to give bottles, staff might believe that specific advice to breastfeed on demand is unnecessary in a population where this practice is the norm and might prefer to focus on advice not to give other liquids rather than specifically advising against bottles. One hospital was much more supportive than the other and was approaching baby-friendly certification, but the rate of exclusive breastfeeding from birth to 6 months was no higher among mothers who delivered in that hospital than among those who gave birth in the less supportive one (14% vs 12% at 6 months, both groups combined, p=0·73). This similarity is consistent with our finding that home support is more influential than support at hospital and is likely to be especially important in countries where mothers stay in hospital for a short time.

The challenge now is how to incorporate home visits, or some other effective means of postnatal support, into routine health-service delivery. In India, promotion of exclusive breastfeeding has been successfully integrated into existing primary health-care services by use of traditional birth attendants, village-based workers, auxiliary nurse midwives, and other health-care providers. In Brazil, the programme of community health agents is an option; in Recife, we are working in six city districts with the Municipal Health Secretariat to train more than 1400 community health agents, as well as maternity staff and doctors in 17 clinics. Training started in August, 2003, and is planned to end in December, 2005. Baseline breastfeeding data have been collected and the effects will be assessed. In our study in Palmares, four visits were made in the first month and the feasibility of such frequent visits can be questioned. In the scaled-up programme, this number has been reduced to two, giving a total of eight visits. This frequency is proving feasible, and preliminary results suggest that rates of exclusive breastfeeding in the scaled-up programme are similar to those in Palmares with ten visits. In Bangladesh, peer counsellors visited at least 15 times, before and after the infant was born, and 70% of visited infants were exclusively breastfed at 5 months compared with 6% of control infants. In Mexico, four postnatal visits were more effective than two postnatal visits. Further research is warranted to identify the minimum number of visits, and their timing, for programme effectiveness.

In terms of child survival, there is evidence, at least in Brazil, that the BFHI has limited influence as a strategy for achieving high rates of exclusive breastfeeding. Our findings add a further dimension to this concern because the hospital-training intervention was associated with inequity; it benefited the more affluent rather than the most disadvantaged. By contrast, home visits benefited all socioeconomic groups. We believe there is an urgent need to question reliance on the BFHI for breastfeeding promotion, especially in countries like Brazil where the postpartum hospital stay is short and there are strong traditions of giving water and tea from birth, and for early introduction of other milks and pacifiers. In the Ten Steps to Successful Breastfeeding, step 10 requires the establishment of support groups and is commonly the least emphasised. Moreover, Baby-Friendly certification can be awarded even if only rudimentary postnatal support is offered. Thus, although the BFHI was conceived as a hospital-based initiative with postnatal support, in practice the hospital component customarily stands alone.

In conclusion, we believe there could be a misplaced sense of security among international agencies and governments that the BFHI will sustain improved breastfeeding practices when mothers return home. Our results add weight to previous evidence that improvements are largely confined to the maternity hospital and are not sustained, and that a combination of systems (in the hospital and in the community) is...
needed so that mothers can receive continuing help locally, especially in the early weeks after the infant’s birth, when difficulties commonly arise. If the millennium development goal for reduction in child mortality is to be reached, delivery strategies for breastfeeding promotion, particularly reliance on the BFHI, need to be re-examined.

Contributors

All authors helped to conceive the research and design the study, analyse the results, and write the paper. S Bechara Coutinho trained the maternity staff and home visitors. P I Cabral de Lira and M de Carvalho Lima trained the fieldworkers and were responsible for project management and maintaining data quality.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgments

We thank the research team, hospital staff, home visitors, and mothers for their cooperation; the Paediatric Society of Pernambuco for their interest and administrative support; and Mike Kenward for the randomised-based statistical tests. The British Council facilitated interinstitutional collaboration, which we greatly appreciate.

The study was funded by the Superintendência para o Desenvolvimento do Nordeste (SUDENE), Brasil (Agency for the Development of the Northeast, Brazil).

References

Nitrous oxide and risk of surgical wound infection: a randomised trial

Edith Fleischmann, Rainer Lenhardt, Andrea Kurz, Friedrich Herbst, Béla Fülesdi, Robert Greif, Daniel I Sessler, Ozan Akça, on behalf of the Outcomes Research Group*

Summary

Background Nitrous oxide inactivates vitamin B12 and methionine synthase, thereby impairing DNA formation and, consequently, new cell formation. The gas also inhibits methionine production, which can reduce scar formation and depresses chemotactic migration by monocytes. Therefore, we assessed whether nitrous oxide increases the incidence of surgical wound infection.

Methods We recruited 418 patients aged 18–80 years, scheduled for colon resection that was expected to last more than 2 h, at three hospitals in Austria and Hungary. Patients were randomly assigned 65% intraoperative nitrous oxide (n=208) or nitrogen (n=206), with remifentanil and isoflurane. The primary outcome was the incidence of clinical postoperative wound infection, analysed by intention to treat.

Findings 206 patients in the nitrous oxide group and 202 in the nitrogen group were included in the final analysis. Duration of surgery was longer in the nitrogen group (3·4 h [1·5]) than in the nitrous oxide group (3·0 h [SD 1·3]) and arterial pressure (84 mm Hg [10] vs 81 mm Hg [9]), bispectral index values (53 [9] vs 44 [8]), and end-tidal isoflurane concentration (0·64% [0·14] vs 0·56% [0·13]) were greater in patients given nitrogen than in those given nitrous oxide. Infection rate was 15% (31/206) in patients given nitrous oxide and 20% (40/202) in those given nitrogen (p=0·205). Additionally, the ASEPSIS wound healing score, wound collagen deposition, number of patients admitted to critical care unit, time to first food ingestion, duration of hospital stay, and mortality did not differ between treatment groups.

Interpretation Nitrous oxide does not increase the incidence of surgical wound infection.

Introduction Wound infections are a frequent and serious complication of anaesthesia and surgery and can prolong hospital stay by 5–20 days per infection,1 substantially increasing the cost of care.2 Major factors affecting the incidence of surgical wound infections include site and complexity of surgery,3 underlying illness (including treatment with immunosuppressive drugs),4 timely administration of prophylactic antibiotics,5 intraoperative patient temperature,6 hypovolaemia,7 and tissue oxygen tension.8 Type of anaesthesia can also affect infection risk. Nitrous oxide has been used for more than a century and probably remains the most commonly used general anaesthetic. The gas has been given to several billion surgical patients. However, use of nitrous oxide has decreased in recent years, especially in Europe, and piped-in nitrous oxide is no longer provided in some new hospitals. Among the concerns about nitrous oxide are three properties that suggest that this gas might reduce resistance to surgical wound infection.

The first concern comes from in-vitro evidence indicating that exposure to nitrous oxide inactivates vitamin B12 and thus methionine synthase.9 Methionine synthase is the enzyme responsible for conversion of homocysteine to methionine and methyltetrahydrofolate to tetrahydrofolate, which are critical pathways for thymidine formation, which, in turn, is essential for DNA formation. Even after brief periods of nitrous oxide administration, DNA synthesis remains abnormal until the fourth postoperative day and does not return to normal until the sixth day.8 This change restricts formation of new cells, including haemopoietic cells critical for fighting infection. Inhibition of methionine synthase could explain the link between nitrous oxide exposure and spontaneous abortion.9

The second concern is that nitrous oxide inhibits methionine production, which in turn reduces protein formation.10 Without functioning methionine synthase, protein cannot be produced. Protein expression is a critical aspect of scar formation and tissue repair.11 Thus, nitrous oxide toxicity could impair healing. Consistent with this theory, nitrous oxide administration has been implicated in development of sepsis.11

The third disconcerting property of nitrous oxide is that the gas depresses chemotactic migration by monocytes, apparently by interfering with microtubules.10 By contrast, inhalation of 65% nitrous oxide for as little as 60 min significantly increases polymorphonuclear neutrophil chemotaxis.12 Chemotaxis is a key part of the bacterial killing process, which needs chemotaxis, phagocytosis, and killing.13 Which of these components dominates the process and the clinical consequences of these in-vitro observations remains unclear.

The effect of nitrous oxide on surgical wound infections has been investigated previously. However, the study involved patients at low risk of infection and
was thus underpowered for this particular outcome. Furthermore, oxygen concentration, which is an important confounder, was not controlled for. That this investigation did not identify an adverse effect is therefore insufficient reason to conclude that nitrous oxide has no effect on infection. We therefore assessed whether incidence of surgical wound infection is greater in patients given 65% nitrous oxide than in those given 65% nitrogen during elective colon surgery.

**Methods**

**Patients**

With the approval of the institutional review boards of each centre, we recruited 418 patients, with American Society of Anesthesiologists (ASA) physical status I–III, scheduled for elective colon resection expected to last more than 2 h, from two hospitals in Austria and one hospital in Hungary. The study was restricted to colon resection because the risk of wound infection is high in these patients. All patients gave written informed consent and were aged 18–80 years. Anaesthesia residents, Outcomes Research fellows, or attending anaesthetists enrolled patients; none of these doctors were involved in patient postoperative care or follow up.

Patients with acute bowel obstruction or those having minor colon surgery (eg, polypectomy, isolated colostomy) were excluded. However, we admitted patients undergoing restorative rectal resection and abdominoperineal excision of the rectum, which carries a particularly high risk of surgical-site infection. Patients in whom the surgeon did not anticipate primary wound closure were excluded, as were those with a history of fever or infection within 24 h of surgery.

**Procedures**

All patients received standard mechanical bowel preparation with an electrolyte solution the night before surgery. Intraluminal antibiotics were not used. Per surgical routine, cefuroxime (1·5 g) and metronidazole (1·5 g) were given intravenously during anaesthetic induction. Additional antibiotics (eg, to treat clinically suspected infections) were administered according to the judgment of the attending surgeon. Anaesthetic management was standardised. Thiopental sodium (3–5 mg/kg) or propofol (2–3 mg/kg), fentanyl (1–3 μg/kg), and vecuronium (0·1 mg/kg) or rocuronium (0·6 mg/kg) were used for induction; anaesthesia was maintained with isoflurane (0·6%) in 65% nitrous oxide (0·6 mg/kg) or propofol (2–3 mg/kg), fentanyl (1–3 μg/kg), and vecuronium or rocuronium. An infusion of remifentanil (0·2 μg kg⁻¹ min⁻¹) was subsequently started. After induction of anaesthesia, research fellows or attending anaesthetists, who were not involved in data collection, allocated patients to one of two groups. The assignments were based on computer-generated random numbers that were kept in sealed, sequentially numbered envelopes until used. Patients were not informed of their group assignments.

All patients were given 35% inspired oxygen during surgery, which was balanced by either 65% nitrous oxide or nitrogen, until immediately before extubation whereupon 100% oxygen was given. During the first hour of recovery, oxygen was given by nasal prongs at a rate of 2 L/min. Patients in both treatment groups subsequently breathed ambient air unless additional oxygen was needed to maintain oxygen saturation at 95% or more.

The anaesthetist was not blinded to treatment. However, great care was taken to prevent the surgeons from observing the administered gas mixture. Irrespective of group assignment, oxygen was given as necessary to maintain oxyhaemoglobin saturation (S$_O_2$) at 95% or more in all patients. Isoflurane and remifentanil doses were adjusted by the attending anaesthetist with the goal of maintaining mean arterial blood pressure at 90% of the pre-induction value. Ventilation was mechanically controlled to maintain end-tidal carbon dioxide tension near 35 mm Hg.

We gave around 12 mL kg⁻¹ h⁻¹ of crystalloid fluid throughout surgery. Additionally, blood loss was replaced with crystalloid fluid at a 4:1 ratio or colloid fluid at a 2:1 ratio. Crystalloid fluid was administered at 3·5 mL kg⁻¹ h⁻¹ for the first 24 h after surgery and at 2 mL kg⁻¹ h⁻¹ for the subsequent 24 h. Intraoperative core temperature was maintained near 36°C.

Target minimum haematocrit was determined prospectively based on the patient’s age and cardiovascular status. The target haematocrit was 26% in patients younger than 65 years without significant cardiovascular disease; 28% in patients either 65 years or older or those with significant cardiovascular disease; and 30% or more in patients 65 years and older with significant cardiovascular disease. Leucocyte-depleted allogeneic blood was given as necessary to maintain target haematocrit.

Patient-controlled analgesia with the opioid piritramide was provided postoperatively. Nausea and vomiting were treated with ondansetron (4 mg intravenously). After giving the report to the post-anesthesia care nurse, the attending anaesthetist sealed the anaesthesia record in an envelope marked with “Anaesthesia Record. Do not open unless necessary for clinical care until 16 days after surgery.” Thus, the surgeons and investigators were unable to determine group assignment.

Surgical wounds were examined daily by a physician unaware of group assignment. After discharge, patients were assessed during their 2-week clinic visits. Patients not returning to clinic were contacted by phone by an investigator unaware of group assignment. Preoperative morphometric characteristics, laboratory values, and historical factors that might have affected wound healing or resistance to infection were recorded. Core temperatures were measured in the distal oesophagus during surgery and orally on each subsequent hospital visit.
day. Heart rate and arterial blood pressure were recorded at 15-min intervals during anaesthesia, at 30-min intervals during the first postoperative hour, and then daily throughout the hospital stay. End-tidal isoflurane, nitrous oxide, and carbon dioxide concentrations were measured at 15-min intervals during anaesthesia. The bispectral index was recorded at 15-min intervals with an A1050 monitor (Aspect Medical Systems, Newton, MA, USA).

Patients rated their severity of nausea 2 h after surgery using a 100 mm long visual analogue scale. The number of emetic episodes was simultaneously recorded, along with the need for anti-emetic medication. Time to first flatus and first bowel movement, as well as time to first tolerated liquid and solid oral intake, were recorded. We assessed risk of infection using the Centers for Disease Control and Prevention (CDC) SENIC score, where one point each is assigned for three or more diagnoses, surgical duration 2 h or more, abdominal site of surgery, and the presence of a contaminated or dirty infected wound. We slightly modified the score by our use of admission, rather than discharge, diagnoses. Infection risk was further quantified using the national nosocomial infection surveillance system (NNISS), with which risk was predicted on the basis of type of surgery, ASA physical status rating, and surgical duration.

Primary outcome was the incidence of clinical postoperative wound infection. Wounds were suspected of being infected when pus could be expressed from the surgical incision or aspirated from a loculated mass inside the wound. Samples of pus were obtained and cultured for aerobic and anaerobic bacteria; wounds were deemed infected when the culture was positive for pathogenic bacteria. We also used a slight modification of the 1992 revision of the CDC criteria to diagnose wound infection. However, we modified the criteria by restricting the diagnostic period to 15 days rather than 30 days. Wounds were numerically scored for infection with the ASEPSIS system. This is an established and validated system derived from the weighted sum of points assigned for the following factors: 1) duration of antibiotic administration; 2) drainage of pus under local anaesthesia; 3) debridement of the wound under general anaesthesia; 4) serous discharge; 5) erythema; 6) purulent exudate; 7) separation of deep tissues; 8) isolation of bacteria from discharge; and 9) hospital stay exceeding 14 days.

In a subgroup of 52 patients at the University of Vienna (21 of whom were given 65% nitrous oxide), wound collagen and protein deposition were assessed. Near the end of surgery, a 7 cm long expanded polytetrafluoroethylene implant (Impra, International Polymer Engineering Inc, Tempe, AZ, USA) was inserted into the subcutaneous tissue a few centimetres lateral to the surgical incision. On the seventh postoperative day, the implants were removed and assayed for hydroxyproline and protein.

**Statistical analysis**

Logistic regression with data from a previous study of patients undergoing elective colon resection suggested that infection rate (as defined by pus and a positive culture) is about 9% in patients maintained at a core temperature of 36ºC. Our study was thus powered to detect a doubling of the infection rate to 18%. We calculated that 500 patients would provide 80% power to detect a statistically significant effect of nitrous oxide at an alpha level of 0.05. We therefore planned to enrol a total of 500 patients with the data and safety monitoring board to review the results after enrolment of 300 and 400 patients.

Morphometric and perioperative data were compared with unpaired, two-tailed, t tests for continuous data or χ² tests for categorical data. A p value less than 0.05 was regarded as statistically significant. Data are presented as number (%), median (quartiles), or mean (SD). Logistic regression was used to control for covariance that differed between the two treatment groups by p<0.25, and for blood transfusion, ASA, SENIC, and NNISS scores. Analyses were completed on an intention-to-treat basis.

**Role of the funding source**

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

---

**Figure: Trial profile**

418 enrolled in study
4 documentation misplaced
414 randomised

208 assigned to nitrous oxide group
206 assigned to nitrogen group

2 withdrawn (surgical complications)
1 withdrawn (attending physician request)

206 completed follow-up
205 completed follow-up

206 reached final analysis
202 reached final analysis

3 withdrawn (exclusion criterion present)
Results

Based on an analysis of the first 400 patients, the Data and Safety Monitoring Board stopped the study after concluding that additional enrolment was unlikely to alter the outcome. During their analysis, an additional 18 patients were enrolled. Under intention-to-treat rules, these patients were retained (figure). Of the 418 patients who consented, ten were not included in the data analysis for the following reasons. The data cover sheets were lost for four patients; thus, their group assignment was unknown. Surgical complications occurred in two patients in the nitrous oxide group that required stopping the study. Four patients in the nitrogen group were excluded from the analysis: one patient was excluded when the attending physician refused to allow the patient to participate; the other three patients that were excluded did not meet the inclusion criteria. Thus, 206 patients were included in the nitrous oxide group and 202 in the nitrogen group in the final data analysis; all these patients completed the trial. Patients were enrolled from Nov 10, 1998, until Nov 7, 2002.

The mean nitrous oxide concentration in patients assigned this gas was 64% (SD 5%). Intraoperative oxygen saturation averaged 98% (1%) in each group. Patient characteristics were closely similar in the two groups (table 1). The number of patients needing blood transfusions, the end-tidal partial pressure of carbon dioxide (PCO2), the average intraoperative core temperature, and the amount of fluids given during surgery did not differ between groups. The duration of surgery was slightly longer in the nitrogen group, whereas the end-tidal isoflurane concentration, mean arterial pressure, and bispectral index values were slightly less in the patients given 65% nitrous oxide than those given nitrogen. Although both groups consumed a similar amount of piritramide during the 2 h after surgery, the pain scores were greater in the nitrous oxide group than in the nitrogen group (table 2).

The incidence of infection, as determined by pus and a positive culture, was 8% in patients given 65% nitrous oxide and 11% in patients given nitrogen. The incidence of infection, as determined by CDC criteria, was also similar in the patients given nitrous oxide or nitrogen (table 3). Logistic regression analysis did not identify any differences in infection risk (table 4). With the combined criteria of infection diagnosed by either pus and a positive culture or CDC criteria, and given the numbers analysed at the end of the study (206 in the nitrous oxide group and 202 in the nitrogen treatment group), there was 80% power to detect a 13% point increase (nitrogen group 20% vs nitrous oxide group 33%) and 90% power to detect a 15% point increase (nitrogen group 20% vs nitrous oxide 35%). Additionally, the ASEPSIS score; wound collagen deposition; number of patients admitted to the intensive care unit; times to first solid food, flatus, and bowel movement; duration of hospital stay; and mortality were similar in each group (table 3).

Nausea or vomiting during the first 2 h after surgery was observed in 42% of patients given nitrous oxide and 43% of those given nitrogen (p=0.874). However, median VAS scores for nausea were significantly greater in the nitrous oxide group than in the nitrogen group (table 3).
Discussion

With the combined criteria of infection diagnosed by pus and a positive culture or CDC criteria, the incidence of infection did not differ significantly between patients given nitrous oxide and those given nitrogen. Additionally, the ASEPSIS wound healing score, wound collagen deposition, number of patients admitted to critical care units, time to first solid food, duration of hospitalisation, and mortality were similar in the two treatment groups.

Perioperative factors, including anaesthetic management, affect infection risk.21 For example, maintaining perioperative normothermia reduces infection risk three-fold,1 and supplemental oxygen halves this risk.4 Conversely, local administration of epinephrine, which reduces tissue perfusion, increases infection risk when given within several hours of contamination, but not when given later.22 There were thus reasons to believe that intraoperative exposure to nitrous oxide, a drug thought to impair wound healing and immune response via three mechanisms, might increase the incidence of postoperative infection. However, infection rates in our study did not differ between patients given intraoperative nitrous oxide and those given nitrogen, and this finding was true whether infection was defined by the presence of pus and positive culture or by CDC criteria. The study was adequately powered to detect a clinically important treatment effect, had one existed. We thus conclude that nitrous oxide administration does not substantially increase infection risk.

Our results are consistent with previous (underpowered) attempts to link nitrous oxide to infection risk,23,24 suggesting that any effect of nitrous oxide on wound infection risk is relatively small compared with other perioperative factors such as body temperature, oxygen administration, prophylactic antibiotics, glucose control, and surgical site preparation. Taken together, available data do not support a policy of avoiding nitrous oxide administration does not substantially increase infection risk.

Our results are consistent with previous (underpowered) attempts to link nitrous oxide to infection risk,23,24 suggesting that any effect of nitrous oxide on wound infection risk is relatively small compared with other perioperative factors such as body temperature, oxygen administration, prophylactic antibiotics, glucose control, and surgical site preparation. Taken together, available data do not support a policy of avoiding nitrous oxide administration does not substantially increase infection risk.

The 11% infection rate observed here, defined by expression of pus that was culture-positive for pathogenic bacteria in the patients given nitrogen is identical to the infection rate reported in a previous group of colon-resection patients given 30% oxygen in nitrogen.6 It was therefore imperative that our control patients be given nitrogen rather than oxygen. The institutional review boards at each participating hospital agreed that informed patients could continue to participate in the study using the original design.

We have previously reported in a subset of 344 patients that moderate-to-severe bowel distension occurs in 23% of patients given nitrous oxide, but in only 9% of those given nitrogen (p<0·001).26 Although our results are consistent with previous reports,27 they contrast with others.28 However, all the studies that did not identify a significant effect of nitrous oxide on bowel distension were seriously underpowered. Our results indicate that the number-needed-to-harm for a case of moderate or severe bowel distension from nitrous oxide was seven (95% CI 5–13). Although significant, this incidence of bowel distension was insufficient to make the investigators or surgeons aware of group assignment.

![Table 3: Principal results](https://www.thelancet.com/)

| Data are mean (SD), number of patients (%), or mean (95% CI). CDC=Centers for Disease Control and Prevention. §Superficial, deep, and peritoneal infections are not mutually exclusive. Infections as diagnosed by expression of culture-positive pus and by CDC criteria are not mutually exclusive. Consequently the total infection column is not the sum of each infection type. 1ASEPSIS is a wound-healing score. ‡Collagen deposition was measured in only 52 patients (21 in nitrous oxide group and 31 in nitrogen group). 5Measured 2 h after surgery on a visual analogue scale for nausea: 0 mm=no nausea; 100 mm=worst imaginable nausea. |

![Table 4: Multiple logistic regression results for surgical wound infection](https://www.thelancet.com/)

<table>
<thead>
<tr>
<th>Beta</th>
<th>Odds ratio</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas (nitrogen vs nitrous oxide)</td>
<td>0·11</td>
<td>1·26</td>
<td>(0·66–2·43)</td>
</tr>
<tr>
<td>Age</td>
<td>0·99</td>
<td>0·99</td>
<td>(0·96–1·01)</td>
</tr>
<tr>
<td>Smoking (no vs yes)</td>
<td>0·65</td>
<td>0·69</td>
<td>(0·49–1·52)</td>
</tr>
<tr>
<td>Sex (female vs male)</td>
<td>0·96</td>
<td>0·96</td>
<td>(0·47–1·99)</td>
</tr>
<tr>
<td>Weight</td>
<td>0·02</td>
<td>1·02</td>
<td>(1·00–1·05)</td>
</tr>
<tr>
<td>ASA</td>
<td>0·76</td>
<td>0·34</td>
<td>(0·07–1·66)</td>
</tr>
<tr>
<td>I vs III</td>
<td>0·44</td>
<td>1·13</td>
<td>(0·32–3·50)</td>
</tr>
<tr>
<td>SENIC</td>
<td>0·20</td>
<td>1·17</td>
<td>(0·28–4·85)</td>
</tr>
<tr>
<td>I vs 3</td>
<td>0·02</td>
<td>0·77</td>
<td>(0·30–2·43)</td>
</tr>
<tr>
<td>NNSS</td>
<td>0·36</td>
<td>0·36</td>
<td>(0·05–2·92)</td>
</tr>
<tr>
<td>I vs 3</td>
<td>0·07</td>
<td>0·24</td>
<td>(0·07–0·87)</td>
</tr>
<tr>
<td>Duration of surgery</td>
<td>0·04</td>
<td>0·97</td>
<td>(0·68–1·37)</td>
</tr>
<tr>
<td>Blood transfusion (yes vs no)</td>
<td>0·14</td>
<td>1·15</td>
<td>(0·91–1·45)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2·19</td>
<td>0·081</td>
<td></td>
</tr>
</tbody>
</table>
Nitrous oxide, unlike nitrogen, is an anaesthetic and provided about a 0.6 minimum alveolar concentration (MAC). MAC is the anaesthetic concentration that prevents movement in response to skin incision in 50% of patients and is a typical clinical dose. Despite assertions to the contrary, 29 MAC fractions of inhaled requirement and therefore the differences between the presumably, is that all patients were given an infusion of isoflurane, which would differ substantially in the two treatment groups. In fact, the difference was less than 0.1% isoflurane, would differ substantially in the two treatment groups.

All general anaesthetics, including isoflurane, impair immune function at least to some degree. However, isoflurane may do so less than other volatile anaesthetics, and immune effects have never been reported at low concentrations. Thus the tiny difference in isoflurane concentration between the groups (<10% MAC) is unlikely to have significantly confounded our results.

In the context of our findings, we suggest that nitrous oxide should not be avoided for fear of augmenting the risk of surgical wound infection.

Study investigators
Harnet Hopf and Heinz Scheunestuhl (University of California, San Francisco, California, USA); Franz Niedermayer, Klaus Krasser, and Valeria Rudel (Donauspital/ SMZ-Ost, Vienna, Austria); Tibor Bartok, Janos Hamza, Laszlo Orosz, and Zsolt Kanyari (University Medical School of Debrecen, Hungary); Thomas Scheck, Alexander Kober, Roman Fleischhackl, Tanja Treschan, Oliver Kimberger, Yuskeu Yuecel, Martin Freitag, Uli Schlechta, Karl Schebesta, Franz Lackner, Alois Werba, Michael Zimpfer, Helmut Hager, Barbara Kabon, Reinhold Fuegger, and Anton Stift (Vienna General Hospital, University of Vienna, Vienna, Austria).

Contributors
E Fleishmann, R Lenhardt, A Kurz, and O Ak ça participated in the organisation of the centre Vienna General Hospital, University of Vienna, Austria; E Fleishmann was involved in patient recruitment, perioperative study organisation, data management, and manuscript preparation; R Lenhardt was involved in data management and manuscript preparation; A Kurz undertook protocol preparation and manuscript preparation; F Herbst was involved in patient recruitment, postoperative study organisation (blinded wound assessments), and manuscript preparation; F Büdesi participated in the organisation of the centre in Debrecen, Hungary and undertook patient recruitment, organisation of data collection, and manuscript preparation; R Greif participated in the organisation of the centre SMZ-Ost Vienna, Austria, and was involved in patient recruitment, organisation of data collection, and manuscript preparation; D I Sessler was involved in protocol preparation, overall organisation of all centres, database preparation, data management, and manuscript preparation; and O Ak ça undertook protocol preparation, perioperative study organisation, data management, and manuscript preparation.

Conflict of interest statement
We declare that we have no conflict of interest.

Acknowledgments
We thank Marie Cote for suggestions; Gilbert Haugh for statistical assistance; Nancy Alisp for editorial contributions; Lale Ak ça for data management. This study was supported by NIH Grants GM 61655 and DE 14879-01A1 (Bethesda, MD, USA), the Joseph Drown Foundation (Los Angeles, CA, USA), the Oesterreichische Nationalbank Jubiläumsfonds, the Gheens Foundation (Louisville, KY, USA), and the Commonwealth of Kentucky Research Challenge Trust Fund (Louisville, KY, USA). O Ak ça is the recipient of a research training grant from the Foundation for Anesthesia Education and Research.

Tyco-Mallinckrodt Anesthesiology Products, Inc (St Louis, MO, USA) donated the thermocouples that we used.

References
Clinicians often reduce chemotherapy doses when treating obese patients because of concerns about overdosing. We assessed dose-response according to body-mass index (BMI) and oestrogen receptor (ER) expression of the primary tumour in premenopausal patients with node-positive breast cancer treated with classical CMF (cyclophosphamide, methotrexate, and 5-fluorouracil). Obese patients were significantly more likely to receive a lower chemotherapy dose (≤85% of expected dose) for the first course than were those with normal or intermediate BMI (39% vs 16%, p<0.001). For obese patients and for the total population, reducing the dose of chemotherapy was associated with a significantly worse outcome for the ER-negative cohort (total population hazards ratio ≥85% vs <85% 0·68 [95% CI 0·54–0·86] for disease-free survival; 0·72 [0·56–0·94] for overall survival) but not for the ER-positive cohort (1·16 [0·97–1·40] for disease-free survival; 1·16 [0·94–1·44] for overall survival) [interaction p values=0·0001 for disease-free survival and 0·0019 for overall survival]. Our findings suggest that for women with ER-absent or ER-low tumours, reduction in chemotherapy dose should be avoided.

Although long-term data for relapse-free survival and overall survival are available for many patients treated with adjuvant chemotherapy, there are conflicting results about the optimum dose intensity. Studies of the relation between obesity and breast cancer have described both an aetiological and negative prognostic effect for increased body-mass index (BMI). A possible explanation for this effect in overweight women receiving adjuvant systemic therapy for early-stage breast cancer is that although doses should be based on body surface area (BSA), suboptimum doses can be prescribed. Concerns of possible altered drug disposition in obese patients has resulted in empiric dose reductions (eg, reduction in the chemotherapy dose after final-dose calculation, or use of the ideal bodyweight to calculate BSA) to avoid excessive toxicity. Few data are available to lend support to these policies or to alternative descriptors of body size (eg, lean bodyweight). However, retrospective analyses suggest that toxicity is not excessive in the obese cancer patient who is fully dosed according to BSA, and that empirical dose reductions for the patient with raised BMI are not only unnecessary but might also result in shortened disease-free-survival.

To assess the relation between BMI, chemotherapy dose reduction, oestrogen receptor expression, and outcome, we reviewed the data from four randomised trials of the International Breast Cancer Study Group (IBCSG: formerly the Ludwig Group) assessing adjuvant classical CMF (cyclophosphamide, methotrexate, and 5-fluorouracil) in premenopausal women with node-positive breast cancer. These trials were done between 1978 and 1993 in many countries, and complied with ethics committee and informed consent requirements at the time for all localities. No endocrine therapy (eg, tamoxifen) was prescribed in these patients. Because clinical evidence and consensus guidelines accord with the hypothesis of differential treatment effects according to endocrine responsiveness of the primary tumour, all analyses were done separately for ER-negative and ER-positive cohorts. To reduce the risk that patient intolerance might contribute to the results, we defined dose reductions for patients with BMI >30 who were randomised to receive oophorectomy followed by CMFp × 12.

<table>
<thead>
<tr>
<th>Treatment group (cycles)</th>
<th>Median (range) years of follow up</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMF (12)</td>
<td>22 (6.5–25)</td>
</tr>
<tr>
<td>CMFp (12)</td>
<td></td>
</tr>
<tr>
<td>CMFp (12)</td>
<td>22 (1.6–24)</td>
</tr>
<tr>
<td>CMFp (6)</td>
<td>18 (1.6–22)</td>
</tr>
<tr>
<td>CMF (6) + 3 reint</td>
<td>12 (0.4–17)</td>
</tr>
<tr>
<td>CMF (3) + 3 reint</td>
<td></td>
</tr>
<tr>
<td>CMF (3) + 3 reint</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: IBCSG trials and treatment groups

C=cyclophosphamide, 100 mg/m² oral days 1–14 of each cycle. M=methotrexate, 40 mg/m² intravenously, days 1 and 8 of each cycle. F=5-fluorouracil, 600 mg/m² intravenously days 1 and 8 of each cycle. P=prednisone, 7.5 mg/day oral for 6 months or 12 months. PrCMF=perioperative CMF. Reint=reintroduction of 3 cycles of CMF. *Does not include 166 patients who were randomised to receive oophorectomy followed by CMFp × 12.
dose groups (<85% of the dose and ≥85%) for the first course and included only patients who received more than one course of CMF. A total of 2140 such patients who had known BMI and ER status were included in the data analyses (table 1). BMI groups were defined as: normal (<25 kg/m²); intermediate (25–29·9 kg/m²); and obese (≥30 kg/m²). The distribution of patients analysed in the three categories was 60% (1279), 28% (612), 12% (249), respectively.

A higher proportion of obese patients (39% [97 of 249]) received less than 85% of protocol specified dose during the first course of CMF compared with patients with normal and intermediate BMI (16% [38 of 235]; p=0·0001). Obese patients initially treated with expected doses of chemotherapy (≥85%) did not have more grade 3–4 toxicity than patients who received reduced (<85%) doses (14% [22 of 152] vs 12% [12 of 97] respectively; p=0·62).

For obese patients with ER-negative disease, those who received 85% or more of the first course dose had significantly better disease-free survival (DFS) and overall survival (OS) (table 2) than those who received less than 85%. Dose reduction for obese patients with ER-positive disease did not significantly compromise outcome (table 2). For the obese group, the interaction term (dose-level and ER-status) in the Cox model was significant for both DFS (p=0·0078) and OS (p=0·0104), indicating that the effect of dose on outcome differed significantly between the ER-negative and the ER-positive cohort.

Similar differences in the effect of dose reduction on outcome according to ER-status were seen for the intermediate and for the normal BMI groups (table 2). Considering all patients together in analyses stratified by BMI group, reduced dose in the first cycle compared with protocol specified dose was associated with a significantly worse outcome for the ER-negative cohort but with a non-significant better outcome for the ER positive cohort (table 2; DFS interaction term hazards ratio 1·77 [1·33–2·37]; p=0·0001; OS interaction 1·69 [1·34–2·02]; p=0·0019).

When the analyses were restricted to the 1563 (73%) patients included in trials of CMF alone (without low-dose prednisone) the size of the interaction between the dose reduction and ER cohort was larger than for the entire group, both overall and within BMI subgroups (DFS interaction term hazards ratio 2·01 [1·42–2·83]; p<0·0001; OS interaction 1·99 [1·34–2·95]; p=0·0007).

<table>
<thead>
<tr>
<th>ER-negative†</th>
<th>N</th>
<th>DFS</th>
<th>OS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10-year DFS% (SD)</td>
<td>Hazard ratio* (95% CI)</td>
</tr>
<tr>
<td>BMI obese</td>
<td></td>
<td>10-year DFS% (SD)</td>
<td>Hazard ratio* (95% CI)</td>
</tr>
<tr>
<td>Dose level ≥85%</td>
<td>49</td>
<td>49 (7)</td>
<td>0·55 (0·33–0·93)</td>
</tr>
<tr>
<td>Dose level &lt;85%</td>
<td>40</td>
<td>28 (7)</td>
<td>0·38 (0·20–0·74)</td>
</tr>
<tr>
<td>BMI intermediate</td>
<td>179</td>
<td>57 (4)</td>
<td>0·67 (0·43–1·03)</td>
</tr>
<tr>
<td>Dose level ≥85%</td>
<td>41</td>
<td>35 (8)</td>
<td>0·46 (0·28–0·75)</td>
</tr>
<tr>
<td>BMI normal</td>
<td>368</td>
<td>48 (3)</td>
<td>0·75 (0·54–1·05)</td>
</tr>
<tr>
<td>Dose level ≥85%</td>
<td>62</td>
<td>41 (6)</td>
<td>0·62 (0·38–1·00)</td>
</tr>
<tr>
<td>Total stratified by BMI group</td>
<td>596</td>
<td>49 (2)</td>
<td>0·68 (0·54–0·86)</td>
</tr>
<tr>
<td>Dose level ≥85%</td>
<td>143</td>
<td>36 (4)</td>
<td>0·76 (0·55–1·06)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ER-positive†</th>
<th>N</th>
<th>DFS</th>
<th>OS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10-year DFS% (SD)</td>
<td>Hazard ratio* (95% CI)</td>
</tr>
<tr>
<td>BMI obese</td>
<td></td>
<td>10-year DFS% (SD)</td>
<td>Hazard ratio* (95% CI)</td>
</tr>
<tr>
<td>Dose level ≥85%</td>
<td>103</td>
<td>37 (5)</td>
<td>1·20 (0·80–1·81)</td>
</tr>
<tr>
<td>Dose level &lt;85%</td>
<td>57</td>
<td>44 (7)</td>
<td>1·10 (0·72–1·65)</td>
</tr>
<tr>
<td>BMI intermediate</td>
<td>335</td>
<td>41 (3)</td>
<td>1·04 (0·73–1·49)</td>
</tr>
<tr>
<td>Dose level ≥85%</td>
<td>57</td>
<td>49 (7)</td>
<td>0·83 (0·55–1·26)</td>
</tr>
<tr>
<td>BMI normal</td>
<td>711</td>
<td>45 (2)</td>
<td>1·21 (0·94–1·55)</td>
</tr>
<tr>
<td>Dose level ≥85%</td>
<td>138</td>
<td>53 (4)</td>
<td>1·35 (1·06–1·72)</td>
</tr>
<tr>
<td>Total stratified by BMI group</td>
<td>1149</td>
<td>43 (1)</td>
<td>1·16 (0·97–1·40)</td>
</tr>
<tr>
<td>Dose level ≥85%</td>
<td>252</td>
<td>50 (3)</td>
<td>1·20 (0·94–1·51)</td>
</tr>
</tbody>
</table>

Table 2: Disease-free survival and overall survival comparing ≥85% dose level versus <85% dose level according to oestrogen-receptor (ER) cohort and body-mass index (BMI) group.
Our findings provide substantial additional evidence for the hypothesis that steroid hormone receptor status of the primary tumour defines distinct biological entities that need a differentiated approach to treatment and clinical trial investigations. Although we cannot exclude that dose-reduction was related to unrecorded comorbid conditions other than obesity, this possibility is unlikely because the patients included in our analysis all participated in randomised clinical trials that required documentation of adequate haematological, renal, and hepatic function before enrolment. In a retrospective assessment of the dose-response effect in obese patients given adjuvant chemotherapy, the adjusted failure risk ratio for obese patients who received more than 95% of the dose compared with lower doses was 0.73. We predict that a larger effect would be observed if the analysis were restricted to the population with endocrine non-responsive disease, and invite such analyses by other groups.

Our results indicate that for patients with endocrine non-responsive disease (ER-absent and ER-low tumour), a reduced dose during the first course of chemotherapy is detrimental. The absence of a clinically significant increase in the risk of grade 3/4 toxicity and the improved outcome in terms of DFS and OS for patients with endocrine non-responsive breast cancer who did not have doses reduced indicate that a priori reduction in the chemotherapy dose for these patients should be avoided.

Contributors
M Colleoni developed the hypothesis, defined the research parameters, and had primary responsibility for the background and discussion. S Li prepared the initial research project report and did all statistical analyses. R D Gelber was responsible for overseeing the statistical analysis and preparing the results for presentation. K Price oversaw the preparation of the research project report and the final version of the research letter. A Goldhirsch and M Castiglione are the study coordinators for the randomised trials included in the report, and both participated in the design and interpretation of this research letter. A Coates is the scientific co-chairman of the International Breast Cancer Study Group and was responsible for the final editing of this research letter. The article was revised and approved by all contributors.

Conflict of interest statement
We declare that we have no conflict of interest. The corresponding author had full access to all data in the study and final responsibility for the decision to submit for publication.

Acknowledgments
We thank the patients, physicians, nurses, and data managers who participated in the International Breast Cancer Study Group trials. We acknowledge the initial support provided by the Ludwig Institute for Cancer Research and the Cancer League of Ticino, and the continuing support for central coordination, data management, and statistics provided by the Swedish Cancer Society, The Cancer Council Australia, Australian New Zealand Breast Cancer Trials Group, the Frontier Science and Technology Research Foundation, the Swiss Group for Clinical Cancer Research (SAKK), the Swiss Cancer League, and the US National Cancer Institute (CA-75362). We also acknowledge support for the Cape Town participants from the Cancer Association of South Africa, and for the St Gallen participants from the Foundation for Clinical Cancer Research of Eastern Switzerland. The four clinical trials included in this report were funded through the contributions of the non-profit organisations and foundations listed above. None of these sources had any role in the conception, preparation, or analysis of this report.

References
Presbycusis

George A Gates, John H Mills

The inevitable deterioration in hearing ability that occurs with age—presbycusis—is a multifactorial process that can vary in severity from mild to substantial. Left untreated, presbycusis of a moderate or greater degree affects communication and can contribute to isolation, depression, and, possibly, dementia. These psychological effects are largely reversible with rehabilitative treatment. Comprehensive rehabilitation is widely available but underused because, in part, of social attitudes that undervalue hearing, in addition to the cost and stigma of hearing aids. Remediation of presbycusis is an important contributor to quality of life in geriatric medicine and can include education about communication effectiveness, hearing aids, assistive listening devices, and cochlear implants for severe hearing loss. Primary care physicians should screen and refer their elderly patients for assessment and remediation. Where hearing aids no longer provide benefit, cochlear implantation is the treatment of choice with excellent results even in octogenarians.

Introduction

Presbycusis, literally elder hearing, is the general term applied to age-related hearing loss. The term encompasses all conditions that lead to hearing loss in elderly people. The disorder is characterised by reduced hearing sensitivity and speech understanding in noisy environments, slowed central processing of acoustic information, and impaired localisation of sound sources. As a result, people with the disorder have difficulty, proportional to the degree of hearing impairment, in conversation, music appreciation, orientation to alarms, and participation in social activities. There are three classic types of the disorder—sensory, strial, and neural—that can occur alone or in combination. Each type has implications for treatment.

Because of the high prevalence of presbycusis, hearing difficulty is a common social and health problem. Overall, 10% of the population has a hearing loss great enough to impair communication, and this rate increases to 40% in the population older than 65 years. 80% of hearing loss cases occur in elderly people. Although hearing worsens with age, the severity of the hearing problem at any given age varies greatly. It is rare to find a person older than 70 years with no hearing impairment or whose hearing sensitivity has not declined from youthful levels. Hearing levels are poorer in industrialised societies than in isolated or agrarian societies. Thus, it is conceptually useful to regard presbycusis as a mixture of acquired auditory stresses, trauma, and otological diseases superimposed upon an intrinsic, genetically controlled, ageing process.

Presbycusis first reduces the ability to understand speech and, later, the ability to detect, identify, and localise sounds. The loss of hearing sensitivity begins in the highest frequencies, which has an adverse effect on understanding speech in noisy or reverberant places. Once the loss progresses to the 2-4 kHz range, which is important in understanding the voiceless consonants (t, p, k, f, s, and ch), speech understanding in any situation is affected. The most common complaint in presbycusis is that the patient cannot hear, but rather that they cannot understand what is being said. For example, people will confuse “mash”, “math”, “map”, and “mat”, or “Sunday” with “someday”. Even such minor misperceptions, left uncorrected, can lead to communication errors or worse. The hearing loss extends to the lower frequencies with time resulting in poor speech detection as well as poor speech understanding. High-frequency warning sounds, such as beepers, turn signals, and escaping steam, are not heard or localised, with potentially disastrous results.

Hearing loss affects an individual’s psychosocial situation. Untreated hearing impairment contributes to social isolation, depression, and loss of self-esteem. Hearing impairment has also been implicated as a cofactor in senile dementia, although others have contested this association. Many people regard presbycusis as an inevitable rite of passage into their senior years and are reluctant to seek help because of cost, vanity, and inconvenience. Others may ascribe their problem to mumbling speakers; for these people, recognition and acceptance of an age-related impairment is a difficult psychological hurdle.

Presbycusis management consumes an increasing portion of healthcare expenditures given the rising mean age of people in industrialised societies. Modern hearing aids are valuable aids to communication. Although they cannot restore lost sensory cells, they do provide acoustic power for declining metabolic function. On the horizon are investigations of treatment modalities that might correct the pathophysiological deficits of presbycusis and other hearing disorders. Such modalities will need more accurate diagnosis at the cellular level and greater medical involvement in their use than at present. The primary management of people with presbycusis in most communities is done by those who sell hearing aids. Only about 20% of people who might benefit from amplification have actually purchased an aid and 25-40% either underuse or abandon hearing aid use. Thus new treatment and management strategies need to be examined.

We review herein the structure and function of the ear, the altered functions in presbycusis, diagnostic considerations, and treatment options. We attempt to
clarify important issues related to the causes of presbycusis and highlight what the practising physician could do to advise patients about the diagnosis, prevention, and treatment of the disorder. Finally, we point out areas of controversy and misunderstanding of the disorder and its treatment, and mention emerging research that could alter the treatment landscape of this highly common problem. In general, knowledge about presbycusis comes from animal models, clinical experience, human temporal bone research, and epidemiological studies. We have drawn from all these sources and have relied upon our extensive personal reference databases for this review.

**Definitions**

Presbycusis (or presbyacusis) is a general term that refers to hearing loss in the elderly and, as such, represents the contributions of a lifetime of insults to the auditory system. Of these, ageing and noise damage are the chief factors, plus genetic susceptibility, otological disorders, and exposures to ototoxic agents. Some people refer to presbycusis as hearing loss due solely to ageing. Because it is very difficult to isolate age effects from other contributors to age-related hearing loss, we use presbycusis and age-related hearing loss synonymously. There is uncertainty about the mechanisms involved in auditory ageing. Furthermore, there is incomplete and contradictory evidence about the interactions between ageing and noise trauma. Both the peripheral and central auditory pathways are affected in presbycusis and the clinical findings often represent a mixture of abnormalities. Because of space considerations, this review will emphasise peripheral presbycusis, which is the most common problem, and mention central (neural) presbycusis only in passing.

**Background**

**Anatomy and physiology**

The external ear comprises the pinna and external auditory canal, which acts as a resonator and enhances sound transmission. The middle ear transforms air vibrations into the fluid-filled inner ear (cochlea) providing a pressure gain of 25–30 dB. The frequency range of human hearing is from 20 Hz to 20 kHz.
Although the structures of the middle ear undergo age-related changes, there is very little effect, if any, on their function, with the notable exception of cerumen production, which seems to increase with age. Whereas the middle ear is a passive device that is essentially unaffected by the ageing process, the cochlea is an active device, with non-linear characteristics, that is dramatically affected by ageing.

The cochlea is a coiled bony tube about 35 mm in length with three compartments: scala tympani, scala vestibuli, and scala media (figure 1). Scala media contains endolymph with a high concentration of potassium. A direct current resting potential (endolymphatic potential) of 80–90 mV is measurable in scala media. This large direct current resting potential arises from Na⁺ K⁺ ATPase pumps in the stria vascularis located on the lateral wall of the cochlea. The lateral wall (stria vascularis and spiral ligament) and the direct current resting potential are seriously affected by the ageing process and will be discussed in more detail later.

The auditory transducer is the organ of Corti, which contains sensory cells (three rows of outer hair cells and one row of inner hair cells). Deflection of stereocilia (hairs) of the sensory cells by a mechanical travelling wave initiates the transduction process. A travelling wave along the basilar membrane, moving from the base towards the apex of the cochlea, arises in response to the piston-like movements of the stapes in the middle ear. The travelling wave has a sharply tuned peak that is located basally for high-frequency sounds and progresses apically as the frequency is decreased. Deflection of stereocilia by the travelling wave opens and closes ion channels, resulting in a current flow (K⁺) into the sensory cell. The potassium flux arises from the positive 80–90 mV endocochlear potential of scala media added to the negative intracellular potential of outer and inner hair cells. The resulting depolarisation causes an enzyme cascade, releasing chemical transmitters and subsequently activating afferent nerve fibres. Although the notion of the cochlea and the organ of Corti as an active rather than passive organ is no longer debated, specific details of the cochlear amplifier and the biological basis of its operation are under investigation. This research is important because many forms of hearing loss involve the cochlear amplifier, whose non-linear properties allow the inner ear to respond to an extraordinarily wide range of sound intensities while maintaining excellent frequency specificity.

Each auditory nerve contains about 30 000 neurons connecting the sensory cells to the auditory brainstem (cochlear nucleus). The cell bodies of the auditory nerve are located in the central core (modiolus) of the cochlea. Each inner hair cell has ten to 20 dendritic connections. The auditory nerve has two types of fibres: 90–95% are type 1 fibres, which are large, myelinated, bipolar neurons that innervate inner hair cells; the remaining 5–10% are type 2 fibres, which are small, unmyelinated neurons that provide efferent synapses with outer hair cells.

**Cochlear ageing in animals**

Examination of the inner ears of animals raised in quiet environments provides important documentation of the pathophysiology of presbycusis. These experiments show that degeneration of the stria vascularis is the most prominent element. This degeneration usually originates in both the base and apex of the cochlea, extending to mid-cochlear regions as age increases. In some cases, the degeneration is patchy. In addition to age-related systematic degeneration of marginal and intermediate cells of the stria vascularis, there is a loss of Na⁺ K⁺ ATPase. Sometimes the loss of this important enzyme, which is detectable with immunohistochemical techniques, occurs when the stria vascularis appears normal or nearly normal under light microscopic examination. Accordingly, the prevalence of strial or metabolic presbycusis may prove to be substantially higher in human beings when the appropriate immunohistochemical techniques are applied to the study of temporal bones of older human beings.

The stria vascularis is heavily vascularised and has an extremely high metabolic rate. Histopathological studies of ageing gerbils have provided strong evidence for vascular involvement in age-related hearing loss. Morphometric analyses of lateral wall preparations stained to contrast blood vessels (figure 2) have shown losses of strial capillary area in aged animals. The vascular pathological changes first occurred as small focal lesions mainly in the apical and lower basal turns and progressed with age to encompass large regions at both ends of the cochlea. Remaining strial areas were highly correlated with normal microvasculature and with the endocochlear potential. Not surprisingly, areas of complete capillary loss invariably correlated with regions of strial atrophy. Subsequent ultrastructural analysis has revealed a significant thickening of the basement membrane, which is accompanied by an increase in the deposition of laminin and an abnormal accumulation of immunoglobulin as shown histochemically. Thus, considerable support exists for the major involvement of strial microvasculature in age-related degeneration of stria vascularis. It is important to note that laboratory studies in many animal species and histopathological studies of human temporal bones are consistent in showing age-related degeneration of the stria vascularis.

Age-related degeneration of the stria vascularis has a substantial effect on the basic physiology of the cochlea, especially on the endolymphatic potential, which provides a voltage to the cochlear amplifier. Thus, when the endolymphatic potential is reduced significantly, the operation of the cochlear amplifier is affected. Indeed,
when the endolymphatic potential decreases to values of 20 mV or lower the cochlear amplifier is deemed to be voltage starved with a maximum reduction in its gain. The gain ranges from 20 dB in the apex of the cochlea and increases to as much as 60 dB in the base. The audiogram resulting from a voltage-starved cochlear amplifier measured in laboratory animals with endolymphatic potentials less than 20 mV corresponds closely to audiograms recorded in ageing human patients (figure 3) with no history of exposure to noise. As much as 20–30% of the stria could degenerate with only a 20 mV reduction in the endolymphatic potential. As strial degeneration exceeds 50%, endolymphatic potential values drop substantially.

There is a unique feature of strial degeneration: in an ageing animal with a 40–50 dB hearing loss and an endolymphatic potential of 20 mV (normal=90 mV) hearing thresholds can be improved by 20–25 dB by introducing direct current into scala media, which raises the low endolymphatic potential. Direct current, introduced into scala media by the researcher, returns the endolymphatic potential to nearly normal values, reduces the extent of the hearing loss as indicated by the threshold of the compound action potential (CAP), and substantially increases the size of the CAP produced by moderate and high-level signals. Degeneration of the stria vascularis, which has been called the battery of the cochlea, and the resultant decline in the endolymphatic potential, has given rise to the dead battery theory of presbycusis. Indeed, the success of these current-injection experiments and the seeming validity of this theory has produced an engineering effort directed at the development of a direct current hearing aid.

In addition to the age-related decrease in the endolymphatic potential of scala media, there are other changes in the physiology of the ageing cochlea and auditory nerve. One such change is loss of auditory nerve function as indicated by increased thresholds of the CAP of the auditory nerve. Slopes of input-output functions of the CAP in ageing animals are decreased even when the loss of auditory thresholds is only 5–10 dB. In other words, as the signal intensity is increased the amplitude of the CAP increases by a fraction of that recorded in young animals. These shallow input-output functions of the CAP are also reflected in shallow input-output functions of physiological potentials arising from the auditory brainstem. Thus, what seems to be abnormal function of the auditory brainstem in older animals and human beings is only the abnormal output of the auditory nerve. The reduced amplitudes of action potentials recorded in ageing ears probably indicate asynchronous or poorly synchronised neural activity in the auditory nerve. The pathological basis of asynchronous activity in the auditory nerve is unknown, but probably involves the nature of the synapse between individual auditory nerve fibres and the attachment to the inner-hair cell, primary degeneration of spiral ganglion cells, and reductions in the endolymphatic potential. Malfunctions of the auditory nerve caused by a reduced endolymphatic potential remain difficult to differentiate from malfunctions caused by degeneration of spiral ganglion cells. Age-related asynchronous activity of the auditory nerve probably contributes to age-related declines in temporal resolving abilities, which are often attributed solely to age-related declines in the properties of the auditory central nervous system.

Figure 2: Surface preparations of lateral wall dissections from an old gerbil stained to contrast blood vessels
The strial capillary bed (between arrows) overlies vessels of the spiral ligament. A: focal area of strial capillary atrophy. B: complete loss of strial capillaries throughout the apical turn.

Figure 3: Threshold shifts in ageing in non-noise-exposed human patients
Reprinted with permission of The Society for Neuroscience.
Clinical pathophysiology

Loss of threshold sensitivity in the high-frequency region of the hearing spectrum is the first sign of presbycusis. Such changes can begin in young adulthood, but are initially evident at 60 years for the most part. Over time, the threshold elevation progresses to lower and lower frequency areas. This pattern of progression noted in quiet-aged gerbils (figure 4)\(^2\) is typical of declining metabolic function of the cochlea—ie, the stria vascularis\(^3\)—and is termed strial presbycusis.\(^2\) Many people also have loss of outer hair cells in the basal cochlea. Such loss of hair cells, termed sensory presbycusis, is more likely to be the result of specific disorders, usually damage from excessive noise. The audiometric pattern of sensory presbycusis is that of a steeply sloping high-frequency loss, often with a notch or dip in the 4 kHz region.

Other factors, such as genetic susceptibility, age-related diseases, and ototoxic medication, can add to the burden. Hearing loss accumulates from the effects of these various factors, but not necessarily in a linear manner. For example, age changes in hearing accelerate with time whereas the hearing loss from noise tends to decelerate with time. That societal noise contributes to presbycusis is lend support by Goycoolea and colleagues\(^4\) who showed that the mild hearing loss with age of people living their entire lives in the quiet of Easter Island was less than their kinsmen living on the noisier mainland for periods as short as 3–5 years.

Because of difficulties in controlling environmental ototoxicity in human beings and the notable similarities in the structure and function of the ear across mammalian species, animal studies have been vital in understanding the effects of noise and age on the inner ear. Mills and co-workers\(^2\) showed that gerbils raised in quiet have just as much, or more hearing loss with age than groups of noise-reared animals. The variability of the loss was far greater in the quiet-raised group than in the noise-reared group. The interaction of the effects of noise and ageing are not fully understood because, in part, noise and ageing both affect the high frequency regions of the cochlea first. However, noise damage is typified by threshold elevation in the 3–6 kHz range, whereas the earliest effects of age are seen at the highest test frequencies (usually 8 kHz) with a monotonic pattern of loss that is proportionately less at each lower test frequency.

Human pathology

Early knowledge about the pathology of human presbycusis came from the temporal bone laboratory of Schuknecht\(^5\) who described the light microscopic findings in the human inner ear and compared these to premortem hearing tests. Although the histological technique is crude by today’s standards, these examinations provided the opportunity to characterise the hearing in temporal bones with a single pattern of cellular loss. Of necessity, the cellular loss had to be far advanced to be detectable; thus, early lesions due to subtle dysfunction could not be detected at the light microscope level. We regard these archetypes as an important beginning in the study of the pathology of presbycusis.

Schuknecht classified presbycusis into four types: sensory (outer hair-cell loss), neural (ganglion-cell loss), metabolic (strial atrophy), and cochlear conductive (stiffness of the basilar membrane). The cochlear conductive subtype is theoretical and will not be discussed further as we believe such cases would be characterised today with modern histochemistry. Later, Schuknecht added two more categories: mixed and indeterminate, the latter accounting for 25% of cases.\(^2\) Others have noted that most cases exhibit a mixture of pathological changes.\(^2\) The audiometric pattern in Schuknecht’s sensory cases is typical of noise-induced hearing loss and the histology findings reveal missing hair cells in the area of the hearing loss. Indeed, most of these cases were men with histories of noise exposure. It is now known that ageing alone does not cause outer hair-cell loss.\(^2\) Thus, we emphasise that in human beings, and in many species of animals (excluding some genetic mouse models), sensory presbycusis probably has little to do with age and much to do with accumulated environmental noise toxicity.

We now know that the pure-tone threshold audiogram is a poor indicator of neural loss since any pattern from normal to anacousis can be seen. The sine qua non of neural lesions is reduced word recognition. Better clinical assessment techniques, such as auditory brainstem response audiometry and otoacoustic emission testing, provide additional detail about neural function. Primary dysfunction of the auditory nerve (auditory neuropathy)\(^6\) is more evident in children than in adults. In this condition, outer hair-cell function, as measured by otoacoustic emissions, is normal but the auditory brainstem responses are reduced or absent. The term auditory dyssynchrony is now used to describe the condition.
We are in agreement with the observations and conclusions of Schuknecht and Gacek* that age-related degeneration of the stria vascularis is the most prominent anatomical characteristic of age-related hearing loss. Additionally, the audiometric pattern resulting from strial degeneration in animals is also the most typical finding in cohort studies of elderly people.10,27–30 By contrast, the steeply sloping audiogram of people with confirmed noise-induced hearing loss31 differs greatly from the strial pattern and coincides with Schuknecht’s sensory presbycusis pattern,32 which can occur at any age. Indeed, many of his patients had clear histories of noise exposure. In the absence of any clinical method to measure the endocochlear potential in human beings, audiometric patterns are the only way to infer the probable pathophysiology.

In families, the strial pattern of hearing loss had a high heritability index (0·46).33 Heritability was significant in pairings of women (sisters 0·53, mother–daughter 0·36, siblings 0·38). Sensory presbycusis also had a significant heritability index for both sexes, but at lower correlations than strial presbycusis. Poor low–frequency hearing, which is typical of strial presbycusis, has also been shown to be associated with cardiovascular diseases (heart attacks, stroke, intermittent claudication) and to affect both men and women, but the effect is stronger in women.44

Central presbycusis
Given the complex neural connections through which auditory stimuli pass to be perceived in the auditory cortex, it is logical to expect that lesions of these pathways would affect hearing. Indeed, secondary degeneration of central pathways after loss of sensory cells in the cochlea, although slow, is a limiting factor in cochlear implantation in people with long-standing deafness. Primary neural lesions of the auditory pathways are distinctly uncommon, incompletely documented, and poorly understood. Whether primary degeneration and secondary degeneration can co-exist is not known. It is our view that secondary degeneration is the major cause of central presbycusis.

Age changes in the central auditory system are well known to affect speech perception. Such changes are now termed age-related auditory processing disorder. Other synonyms are central auditory dysfunction, neural presbycusis, central presbycusis, and nerve deafness. These changes typically affect speed of processing and result in poorer speech understanding in noise or with rapid or degraded speech. Given that the ultimate perception of speech is in the brain, it is not unexpected that age-related brain dysfunction would affect hearing. Fortunately, central presbycusis severe enough to limit rehabilitation is uncommon and few audiologists include central testing in the assessment of people for amplification candidacy because, in part, of complexity and uncertainty about the implications of the results.

Unfortunately, the term nerve deafness is firmly entrenched in the lay literature and by popular use even though it has led to the frequent misunderstanding that hearing aids will not help. Although people with central presbycusis do indeed have poorer speech understanding than expected from the audiogram, most people with poor speech understanding have outer hair-cell loss rather than neural lesions and are helped by amplification. Fortunately, isolated central presbycusis is uncommon and most cases with reduced speech discrimination also have loss of cochlear sensory cells. We could suspect central dysfunction in about 12% of the cases in the Framingham Heart Study Cohort using as a criterion a significant discordance between measured and predicted speech perception in quiet.35 Using multiple central auditory tests in the same cohort in an earlier study of 1026 people aged from 64 years to 92 years, we found one abnormal test in 23%, two abnormal tests in 4%, and three abnormal tests in none.36 None of these people had peripheral hearing loss severe enough to confound the central testing. The prevalence of abnormal test results increased significantly with age but age only accounted for 13% of the variance. We conclude that isolated central presbycusis is not common and that it is difficult to separate the effects of peripheral from central abnormalities.

Diagnosis
History
Hearing loss is often a silent disorder, characterised more by what is missed than what is heard. In many cases, family and friends are more aware of the problem than the patient. Nonetheless, most people with presbycusis will respond to a direct inquiry with a positive answer. People with depression and cognitive dysfunction should be assessed to exclude occult hearing loss as a contributing factor. The hearing loss is often accompanied by tinnitus—the perception of a ringing sound in the ears or head. Tinnitus is a sign of hearing loss and should be assessed to exclude treatable diseases (ie, acoustic neuroma) as a cause.

Risk factors for presbycusis are: 1) noise exposure; 2) smoking; 3) medication; 4) hypertension;7 and 5) family history. People with substantial exposure to workplace noise, recreational noise, and gun-shooting are more likely to have high-frequency hearing loss. Smoking is also associated with high-frequency loss.16 The use of aminoglycoside antibiotics, cisplatin, loop diuretics, or anti-inflammatory agents may contribute to hearing loss. Presbycusis does cluster in families.15

Physical examination
The physical examination is usually normal after removal of cerumen, which is a common problem in elderly people and a frequent cause of hearing loss and hearing aid malfunction. Topical sodium bicarbonate
solution, 10%, is an excellent cerumenolytic. Opacification of the normally translucent tympanic membrane is commonly seen. This has no effect on conduction of sound energy and is simply a manifestation of age.

Screening
Given the high prevalence of presbycusis in people of retirement age and the adverse effects of hearing loss on well-being, screening for hearing loss should be done at annual physical examinations or at the first visit for new patients over the age of 60 years. A single question on an intake form “do you have a hearing problem?” is a very cost effective and sensitive instrument to screen for presbycusis. The 10-item hearing handicap inventory for the elderly-short form (HHIE-S) is widely used for screening. However, the HHIE-S under-reports hearing loss because its sensitivity is lower than the single question “do you have a hearing problem”. The value of screening is justified by the effectiveness of remediation.

The use of clinical measures of hearing loss, such as spoken voice tests and finger friction tests, are imprecise and are not effective or reliable methods for screening. Screening audiometry administered by a trained office nurse or medical assistant is a practical and cost-effective method for detecting significant hearing loss. The equipment needed for screening audiometry is light weight, low cost, and well accepted by patients. The standard screening audiometer tests at 1 kHz, 2 kHz, and 3 kHz at intensity levels of 25 dB, 40 dB, and 60 dB. Failure at any one frequency at 25 dB for younger adults or 40 dB for retired individuals justifies a referral for definitive assessment.

Imaging is not done except where the loss of hearing is unilateral or significantly asymmetric, or where tinnitus is unexplained by the audiogram. A metabolic assessment might be indicated if the patient has not had a recent health examination. Diabetes, renal dysfunction, hypertension, and hyperlipidaemia should be excluded as cofactors.

Central auditory testing
A variety of central auditory tests are available but these are rarely done in the clinical setting unless there is a large discrepancy between the history and the results of standard peripheral auditory tests. For example, when a patient has reasonable speech understanding in quiet but has severe speech problems in noisy or difficult listening environments, central presbycusis is likely to be a factor. Unfortunately, contemporary rehabilitation measures—other than environmental manipulation—are ineffective in dealing with central problems. Moreover, few tests have any discrete localising value and are not done as part of the neurological workup. The most widely used tests assess speech recognition in relation to the pure-tone thresholds and to noise.

The simplest test is auditory rollover. Speech understanding normally improves as the level of the presentation of the phonetically balanced words increases until distortion due to overstimulation of the cochlea occurs. People with central presbycusis may exhibit a 20% or greater decrease in speech recognition at high signal intensities. Rollover was very uncommon (1% of patients) in our survey of the older members of the Framingham Heart Study Cohort. Some people with loudness recruitment, a common problem with cochlear hearing loss, have difficulty in performing the rollover test because at high presentation levels the apparent loudness of the sound increases substantially.

A widely available central auditory test is the synthetic sentence identification with either an ipsilateral competing message (SSI-ICM) or a contralateral competing message (SSI-CCM). The SSI-ICM seems to be more sensitive. In this test, the patient listens to a narrative about Davy Crockett and is instructed to listen for one of ten sentences superimposed on the narrative by the same speaker. Each sentence has no meaning even though each series of three words makes syntactical sense. The test is quite easy for most people even when the signal and the message are presented at the same intensity level, usually 40 dB above threshold. Since audibility is not an issue, the test requires the listener to attend to the message and ignore the narrative.

We have shown that poor performance on the SSI-ICM is common in people with probable Alzheimer’s disease and that very poor performance (<50% correct) in either ear might precede the clinical onset of dementia of the Alzheimer’s type by several years. We presume that frontal lobe executive control function is a key factor in the task because lesions of the central auditory pathway in Alzheimer’s disease are uncommon in early cases. With the possibility of treatments to delay the progression of Alzheimer’s disease, early identification assumes great importance. Referral to an otology centre can assess this possibility.

Many other tests are available. Two more widely available tests are the speech perception in noise (SPIN) test and the staggered spondaic words test (SSW). Both use carefully prepared speech materials to assess central auditory function.

Treatment
Hearing loss of all types affects not only communication but also quality of life. Mulrow and co-workers studied the effect of amplification on older patients with hearing loss and documented a positive effect of personal amplification (hearing aids) on quality of life. Therefore, treatment effects extend beyond communication. No treatment exists now to restore lost hearing. Research into hearing restoration is a growing scientific field.

Prevention
Although some degree of sensory presbycusis is inevitable, the deterioration can be reduced by avoidance of hazardous noise exposure or use of suitable hearing
Cardiovascular disease and its risk factors affect hearing to some extent. Stroke, myocardial infarction, claudication, hypertension, hyperlipidaemia, and diabetes mellitus have all been associated with excessive hearing loss. Therefore, that maintenance of good general health and fitness would reduce the risk of hearing loss due to systemic disease would seem logical. High lipid diets are associated with poor hearing. There is inconclusive evidence that low-caloric diets, which clearly prolong life in laboratory animals, have any effect on presbycusis. While free radical accumulation is implicated in presbycusis, anti-oxidant agents have not been shown to counter auditory ageing.

Rehabilitation

Communication courtesy
Communication is a two-way process; the burden of communication falls equally on the speaker and listener. Common courtesy holds that both should work at improving the communication environment as well as the process when one of them has a hearing problem. The speaker should: be face-to-face with the listener, speak clearly and unhurriedly, turn off competing sound sources (TV, radio), and make sure that the message was received. The hearing-impaired listener should also be serious about communicating and take steps to repeat what was heard so misunderstandings can be corrected. These principles apply to all degrees of hearing losses.

Environmental manipulation
Central nervous system processing of speech slows down with age. Older people have more difficulty in understanding speech than do younger people, even when cochlear sensitivity is similar. Degraded speech signals, such as those that occur in noise, in reverberant halls, or with rapid speakers, are more difficult to understand. Such problems are thought to indicate slowing of the central auditory system’s ability to integrate and synthesise speech sounds into meaningful language elements. Therefore, optimising the listening environment, often by simple means (turning off the radio, speaking slower), has important effects on speech comprehension.

Amplification
As a rule of thumb, when the average hearing thresholds reach 40 dB on the audiogram, amplification is indicated. People with smaller losses also receive benefit, especially when employment or educational needs are considered. The choice of types of hearing aids and models is staggering. Hearing aids may have analogue or digital circuitry, may be adjusted manually or automatically, and may have sophisticated directional microphones, noise suppression technology, telephone coils, and multiple programme modes (for quiet, noise, music). The smaller the physical size of the device, the greater the cost.

Although amplification technology has made great strides in recent years, controversy remains regarding the benefits of these various choices, especially in relation to cost. Yeuh and colleagues have shown a clear advantage for digital technology in hearing aids. Notable advances in directional microphones and noise-suppression circuitry in high-end hearing aids have greatly extended their benefit.

Hearing aids have many drawbacks. They do not restore normal hearing. They need a long learning and adjustment period in which the brain adapts to the new way things sound. They are uncomfortable, unsightly, and costly. Nonetheless, until such time in the distant future that a biological correction for presbycusis is developed, they are the mainstay of rehabilitation for moderate or worse hearing loss. The greater the loss, the more the individual relies on amplification. People with mild high-frequency losses are often annoyed with the amplification of sounds such as paper shuffling and their own breathing, and, as a consequence, use the aids intermittently. Such a usage pattern prevents central adaptation to amplified sounds and contributes to user dissatisfaction. Reliance on education and environmental manipulation is more likely to succeed in the early case. Education about reasonable expectations enhances hearing aid acceptance.

Speech reading
Aural communication is enhanced by viewing the speaker’s face. Given that speech is often redundant, facial expressions and lip contours provide assistance in filling the gaps resulting from unheard speech sounds. Formal speech reading classes are available at a few centres. Unfortunately, not enough centres are available to provide adequate speech reading training. Materials are now becoming available on the internet (search engine phrase “speech reading”).

Auditory training
For people with severe losses, auditory training is of benefit. The hearing impaired listener is trained to identify speech sounds and key words with amplification in place. Such training is tedious and worthwhile, but is seldom available outside of University centres.

Assistive listening devices
A wide range of assistance is available. Ranging from frequency-modulation transmitters, amplified telephones, teleconnectors to couple hearing aids with the telephone, captioning of televised or live programmes, infrared systems for home television, flashing alarms for
the doorbell, telephone, and smoke detectors. Assistive listening devices are important and increasingly available adjuncts for the severely hearing impaired person.

**Cochlear implants**

Cochlear implants are indicated at any age for people with bilateral severe hearing losses not materially helped by hearing aids. Current criteria include hearing no better than identifying 50% or fewer key words in test sentences in the best aided condition in the worst ear and 60% in the better ear. People of retirement age generally are excellent implant candidates since language skills are good and duration of deafness is short. Surgical morbidity is low and acceptance is high.

**Conclusions**

Presbycusis is common. The disorder deprives older people of key sensory input, which seriously affects their quality of life. Modern rehabilitation strategies are effective but underused. Primary physicians, especially those who care for elderly people, should consider the effect of presbycusis on health. Improvement in general health occurs after auditory rehabilitation. Recent evidence suggests that hearing loss may be an early sign of, as well as a contributor to, dementia. Screening for hearing loss in elderly people has health as well as hearing benefits. Modern amplification methods provide improved communication ability for most users. Regeneration biology is a vigorous new research field that is pursuing methods to restore hearing by regrowth of new hair cells and enhancement of stria dysfunction by chemical and prosthetic means.

These promising efforts for future treatments await further research and development.

**Conflict of interest statement**

G A Gates is a consultant to Advanced Cochlear Systems.

**References**


32. Schuknecht HF. Further observations on the pathology of presbycusis. *Arch Otolaryngol* 1964; 80: 369–82.


Genetic Epidemiology 3

Genetic association studies

Heather J Cordell, David G Clayton

We review the rationale behind and discuss methods of design and analysis of genetic association studies. There are similarities between genetic association studies and classic epidemiological studies of environmental risk factors but there are also issues that are specific to studies of genetic risk factors such as the use of particular family-based designs, the need to account for different underlying genetic mechanisms, and the effect of population history. Association differs from linkage (covered elsewhere in this series) in that the alleles of interest will be the same across the whole population. As with other types of genetic epidemiological study, issues of design, statistical analysis, and interpretation are very important.

Genetic association studies aim to detect association between one or more genetic polymorphisms and a trait, which might be some quantitative characteristic or a discrete attribute or disease. Association differs from linkage in that the same allele (or alleles) is associated with the trait in a similar manner across the whole population, while linkage allows different alleles to be associated with the trait in different families. However, genetic associations arise only because human populations share common ancestry and it has been argued that association studies are really just a special form of linkage study in which the extended family is the wider population. In linkage analysis, data from distantly related individuals are more powerful for detecting small effects than data from closely related individuals, but this advantage is offset by the fact that, owing to increased possibility for linkage to be destroyed by recombination, linkage extends over shorter distances in distantly related individuals, necessitating a greater density of markers. Association in apparently unrelated people represents the extreme of this effect: association analysis has greater power than linkage studies to detect small effects, but requires many more markers to be examined. The fact that association operates only over short distances in the genome has for long guaranteed association studies an important place in fine mapping genetic loci initially detected by linkage. More recently, it has been realised that genetic susceptibility to common complex disorders probably involves many genes, most of which have small effects. This fact, together with the identification of large numbers of single nucleotide polymorphisms (SNPs) throughout the genome and rapidly falling genotyping costs, has led to the importance of association studies in genetic epidemiology. Indeed, it is possible to envisage the search for disease susceptibility genes being done by screening large numbers of SNPs across the whole genome.12

Although family-based studies still have a place in the study of population association (in addition to linkage), such research has much more in common with classic epidemiological studies of environmental and behavioural risk factors than do linkage studies. Consequently, issues of study design and analysis have more in common with the rest of epidemiology. Parallels with classic epidemiology are also clear if we consider why association between a genetic polymorphism and a trait might exist in a given population: (1) the polymorphism has a causal role; (2) the polymorphism has no causal role but is associated with a nearby causal variant; or (3) the association is due to some underlying stratification or admixture of the population. In a mixed population in which strata have different environmental exposures or the founder populations entail different genetic risks, any locus whose allele frequencies differ between strata or founder populations will be associated with disease to some extent, whether or not it is near to a causal locus.

Direct association

The first of these forms of association is termed direct association, and studies of direct association target polymorphisms which are themselves putative causal variants. This type of study is the easiest to analyse and the most powerful, but the difficulty is the identification of candidate polymorphisms. A mutation in a codon which leads to an aminoacid change is a candidate causal variant. However, it is likely that many causal variants responsible for heritability of common complex disorders will be non-coding. For example, such variants may cause variation in gene regulation and expression, or differential splicing. We do not know enough to predict which variants may have such effects. Thus, direct association studies only have the potential to discover some of the genetic causes of disease and disease-related traits. However, some 10 000–15 000 aminoacid changing SNPs with minor allele frequency exceeding 1% in Europeans have been identified, and screening of these in whole genome studies is feasible.

Indirect association

In the second type of association, the polymorphism is a surrogate for the causal locus and this type of association allows us to search for causal genes in indirect...
association studies. However, indirect associations are even weaker than the direct associations they reflect, and it will usually be necessary to type several surrounding markers to have a high chance of picking up the indirect association. Indirect association studies are more difficult to analyse, and there is still debate as to the best methods. They are also less powerful than direct studies. Finally, by contrast with direct studies, until we can be sure that we have adequately charted the polymorphisms in a region, there cannot be a definitive negative result since we cannot exclude the possibility that a causal variant exists but is not picked up by the markers chosen. The next phase of the Human Genome Project—the International HapMap Project—aims to improve our knowledge in this respect. This project will be discussed in more detail in a later paper in this series. The imminent completion of the second phase of this study, plus rapid recent advances in high throughput genotyping technology, mean that screening of perhaps 80% of the genome for disease associations is becoming feasible, if costly. In the meantime, most indirect association studies concentrate on candidate genes identified either on the basis of their known function or from animal models. Even as whole genome studies are increasingly used, such candidate gene studies will continue to play an important part. Such studies will allow typing of markers more densely, not only to improve detection of true causal associations but also to increase confidence that negative findings represent true negatives.

Confounded association

The final type of association is that due to confounding by stratification and admixture (substructure) within the population. Confounding, as in the rest of epidemiology, raises the possibility both of generating false findings (positive confounding) or obscuring true causal associations (negative confounding). However, although the problem of unobserved confounding is intractable in classic epidemiology, dictating limits on the size of causal effect that can be safely inferred from observational studies, genetic epidemiology offers possibilities for circumventing the difficulty.

The most obvious way of avoiding this difficulty is to measure association in well-mixed, outbred populations. Failing this, any stratification and admixture effects could be reduced by matching (in the design or the analysis, or both) by geographical region and by any markers of ethnic origin. In this manner, comparisons can be made, as far as possible, within homogeneous subpopulations. It has been argued that such devices will avoid the small confounding effects expected to arise from stratification and admixture. However, this view has been questioned. Meta-analyses have indicated that causal variants for complex disease might, when looked at one at a time, have rather small effects and large studies will be necessary to detect them. Against this background, even modest confounding by stratification and admixture could have important repercussions. It is not yet known how serious this problem will be for association studies in populations of European origin, but it poses a grave difficulty in admixed populations such as African Americans or Afro-Caribbeans. Admixture does present opportunities for gene mapping by exploiting a back-crossing experiment of nature, but such studies are beyond the scope of this article.

The first method for dealing with confounding by population structure is matching by family; if comparisons are made between siblings with the same parents, confounding by population structure is excluded. However, such studies are not always very powerful and they are difficult, or even impossible, to undertake on a sufficiently large scale to detect genetic associations reliably. The role of such studies will probably be to confirm findings generated by less expensive methods and to answer more complex secondary questions.

The second method for dealing with the problem is to seek genetic markers for population substructure, or ancestry informative markers—loci whose allele frequencies differ between the founder populations. Inevitably there will be some loss of power due to the imperfect measurement of admixture proportions. This loss of power might be modest for populations in which founder populations are very different and there are good markers of substructure, such as African Americans, but it remains to be seen whether this method can be applied efficiently to control for the smaller differences which might exist, for example, within European populations.

The third approach is genomic control. Confounding is regarded as a random process, potentially affecting all loci, such that the effect of positive confounding is to increase the type I error (false positive) rate for association tests; although conventional tests for association are correct if regarded as tests for association within the population studied, they will have an inflated false positive rate when judged as tests of causal effects in the presence of stratification or admixture (or both). Another perspective (more intuitive to geneticists) is that, although people in a population-based association study can be regarded as having been independently sampled from the particular population studied, they are not independently sampled when regarded as a sample of all mankind; they are cryptically related because they have been drawn from the same population. As a result, when regarded as tests of the causal null hypothesis, conventional $\chi^2$ tests for association have greater variance than they should and use of conventional significance levels will lead to a higher false-positive rate.

Genomic control is less ambitious than other methods that control for confounding by substructure in that it seeks only to control the false positive rate by increasing
the threshold required for statistical significance. The factor by which the variance is inflated by confounding can be estimated by typing a large number of unselected markers across the genome and estimating the variance of association test statistics empirically. This method is simple to do. However, no attempt is made to deal with negative confounding, which increases the false negative rate and reduces power. Use of more stringent test criteria to control the false positive rate will accentuate loss of power. It also remains to be empirically tested whether the distribution of test statistics is inflated by the same multiple, irrespective of allele frequency and throughout the entire distribution.

It remains to be seen whether correcting for confounding by substructure by statistical modelling will be more powerful than accepting some degree of confounding and controlling the resultant type 1 error rate. Much will depend on how serious the problem turns out to be, and whether sufficiently informative markers will be identified for the former approach to work efficiently. However, the approaches could turn out to be complementary, with gross effects addressed by statistical models and surrogate measures of substructure and more subtle effects, such as those due to cryptic relatedness between cases and/or controls, left to genomic control.

**Direct association: patterns of genotype–phenotype relationship**

We shall consider a diallelic locus, directly related to either a quantitative trait or to a discrete trait such as presence (prevalence), or occurrence (incidence), of a disease. Multiallelic loci lead to more complicated scenarios and generate tests with many degrees of freedom. Even in the simplest diallelic case, different patterns for the genotype–phenotype relationship must be considered. Since there are three possible genotypes, which have a natural order (1/1, 1/2, and 2/2), the question of linearity of the relationship must be considered.

**Linear dose-response modelling**

In classic mendelian genetics of fully penetrant discrete traits, the description of an allele as dominant implies the linear relationship for quantitative traits. He defined absence of dominance to imply the linear relationship:

\[ \text{Mean trait value} = \alpha + \beta x \]

where \( x \) codes genotypes 1/1, 1/2, and 2/2 as 0, 1, and 2 respectively and \( \beta \) is the additive effect of each copy of the 2 allele. Since this model predicts that the trait mean for heterozygotes will lie precisely midway between the means for the two types of homozygote, it is easy to see why Fisher identified linearity with absence of dominance, but this idea is based on a stronger model than the earlier concept.

The importance of a simplifying model such as the linear dose-response model above is that the strength of genotype–phenotype relationship is expressed in a single parameter (\( \beta \)) and statistical tests for existence of such a relationship only have one degree of freedom. To extend the model to allow a quite general pattern of relationship we must introduce an additional parameter to measure deviation from linearity. For example, we might introduce a variable \( z \) coded as 0 for homozygotes and 1 for heterozygotes, to give the following model:

\[ \text{Mean trait value} = \alpha + \beta x + \gamma z \]

\( \gamma \) is then said to represent a dominance effect. In this extended model, all patterns of relationship between phenotype mean and the three genotypes are possible, but two parameters now code the association and statistical tests have two degrees of freedom. Consideration of this broader class of models inevitably carries the penalty of reduced power if the pattern of relationship truly is linear. Some have argued that in most cases we would wish to constrain the two-parameter model so that the trait mean for heterozygotes cannot lie outside the range delimited by the means for homozygotes. This approach leads to tests that are intermediate between conventional tests with one and two degrees of freedom.\(^{17}\) In any situation, the choice of the most powerful test depends on the pattern of association that actually exists and, unless we are simply doing confirmatory studies, this pattern is unknown a priori—a ubiquitous problem for statistical analysis. Perhaps for most complex disease genetics the model in which heterozygote risk is constrained to lie within the range defined by the two homozygote risks is the best compromise between generality and parsimony. However, such a model is little used, perhaps because of lack of software implementations.

To model gene effects on binary qualitative traits that are not fully penetrant, Wright\(^{16}\) introduced the notion of an underlying, unobserved, and normally distributed quantitative trait (liability) governed by Fisher’s linear model; the discrete trait is assumed to manifest when liability exceeds some threshold value. The predictions from Wright’s model are very close to those from the logistic regression model, which is the mainstay of statistical analysis in the rest of epidemiology.\(^{15,20}\) With this approach, absence of dominance means that the log odds of response for 1/2 heterozygotes is midway between that for 1/1 and 2/2 homozygotes, and so each allele contributes multiplicatively to the odds. For uncommon traits (as most diseases are), this model is nearly the same as the model of multiplicative effects of each allele on risk.
The multiplicative risk model could be argued to be the natural model for lack of dominance in this context. The multiplicative risk model has one particularly useful property. Hardy-Weinberg equilibrium is defined by genotype frequencies consistent with the two alleles being independently sampled from a population of alleles. Genotypes of controls, in a case-control study, should therefore be in Hardy-Weinberg equilibrium. But if disease risk is related to genotype multiplicatively, such that genotype risk can be decomposed into a product of effects of the two alleles, then genotypes of the cases of disease are also expected to be in Hardy-Weinberg equilibrium, with alleles being independently drawn from a population in which the frequency of high risk alleles is increased. This result justifies the common practice of counting alleles rather than genotypes in statistical analyses.

Epistasis
The general issue of dominance relates to the extent to which the joint effect of two alleles at a single autosomal locus might be different from the sum (or product in a multiplicative model) of the effects anticipated for each allele independently. A related issue is the degree to which the combined effect of alleles at two or more loci can reasonably be modelled by the individual locus contributions. The fact that inheritance of some traits could only be explained by joint action of two unlinked loci was first demonstrated by Bateson,22 who termed the effect epistasis. In these first examples, variation of phenotype with genotype at one locus was only apparent in those with certain genotypes at the second locus; others would show no effect. Thus epistasis was defined as one locus masking the effect of another. Fisher23 used a similar term, epistacy, to refer to a statistical interaction meaning deviation from additive effects of the two loci upon the trait mean. The term epistasy soon evolved into epistasis,24 and in modern genetics the two uses of the word coexist, often causing confusion.25

Epistasis in Fisher’s sense is dependent on scale and, in general, does not have a clear interpretation in terms of mechanism. The interpretation of the causal implications of statistical interaction in epidemiology has been vigorously debated over at least three decades.26,27 A similar debate continues in relation to interaction between genes and environmental risk factors. Some have argued that the interaction of genes and environment will become a major influence on the epidemiological study of disease causation and on public health interventions,28,29 whereas others have been more sceptical.30

If interpretation of statistical interaction between genes is problematic, an important reason to consider such interaction relates to our ability to discover the genes related to complex diseases in the first place. If such genes act together, epistatically, with several genes acting in the same pathway, the marginal effect of each gene on its own might be small, but might reflect much larger effects of collections of genes.31,32 Some have even postulated scenarios in which marginal effects are absent altogether.33 But this hypothesis requires one gene to reverse the direction of effect of another, which although possible, is perhaps unlikely to happen widely. Such arguments have led these same researchers to suggest that the analysis of association studies should move away from analysis of genes one at time, focusing instead on pairs or even larger constellations of genes. It is not yet clear whether the gains in effect size realised in practice by consideration of several genes at a time will be sufficient to compensate for the requirement for more stringent correction for the number of hypotheses to be tested.34 A further debate concerns the relative merits of recursive partitioning methods, which derive from the automatic interaction detection methods of Sonquist and Morgan,35 originating in the social sciences but now widely used in the computer science and bioinformatics communities, over more standard regression-based approaches.

Indirect association: patterns of linkage disequilibrium
The mapping of susceptibility genes for common complex disorders and genes for other common traits by the indirect method depends on the existence of association, at the population level, between the causal variants and nearby markers. Such association, because of the proximity of loci on the genome, is termed linkage disequilibrium. (Some use this term to describe any population-wide association between loci, whether due to proximity or to another reason such as population stratification and admixture. We prefer the term allelic association for this more general circumstance. The term gametic phase disequilibrium is also used to describe allelic association due to proximity). Success of this strategy depends upon some understanding of patterns of linkage disequilibrium and the forces that determine them—mutation, recombination, and population history.

The figure shows the genealogy of the same small segment of eight versions of the same chromosome. It is assumed that they will be descended from a common ancestor, and that the segment is so small that no recombination will have arisen within the segment. This latter assumption is necessary because the recombination in the sample history is an important complication: an adjacent segment separated by recombination will have an entirely different genealogy above this point. We assume that a mutant allele cannot revert back to wild type (lightning does not strike twice), and every copy of the mutant allele in the population is descended from the same ancestral mutation. The scale for the height of the genealogy is meioses (ie, generations). In this example there are four haplotypes. Labelling the initial allele at each locus as 1 and the new
allele created by mutation as 2 these are 111 (individuals 6, 7, and 8), 122 (individuals 4 and 5), 211 (individuals 1 and 2), and 121 (individual 3). Alleles that are common in the sample are older mutations, and the number of different haplotypes increases in direct proportion to the number of polymorphisms, unless some polymorphisms correspond to mutations on the same branch of the genealogy. Less obvious is that fact that, even under these simplifying assumptions, the pattern and strength of association between polymorphisms is very variable.

The situation in the figure represents complete linkage disequilibrium between the three loci. This fact is apparent when looking at loci two at a time; each pair of loci define only three haplotypes. Table 1 shows the two-locus haplotype frequencies as 2×2 contingency tables. Complete linkage disequilibrium between pairs of loci is evident because at least one cell of the corresponding table is zero, since this is the maximum degree of association possible given the row and column totals. Linkage disequilibrium decays for three reasons: (1) recombination(s) in the genealogy occurring at some point between the two loci; (2) recurrence of the same mutation; and (3) gene conversion (transfer of information between alleles or loci).37 The first reason is the most important for this decay. Since the probability of recombination increases with the distance between the loci, the strength of linkage disequilibrium is expected to decline with distance.

Various different measures of pairwise linkage disequilibrium have been proposed,38 including Lewontin’s D’39 which has also been termed the “association probability”.40 Lewontin’s D’ is an important measure for identification of regions in which there has been little recombination and, therefore, in which there is the potential to map causal loci by indirect association studies. However, this measure does not directly determine the power of indirect association studies. Formally, the power of tests for indirect association depends largely on the index r², the square of the conventional correlation coefficient between the allele at the typed locus, scored 1 or 2, and the allele at the causal locus, scored similarly. The dependence of power on r² rather than on any other measure of association is complete for quantitative traits in the absence of a dominance variance component due to the causal locus.41 The nature of the relationship between r² and the power to detect association is such that, if B is causal, we would require a sample size 2·8 times as large (0·56/0·2) to detect the indirect association with A than to detect the association with C. The panel shows that, even when loci are in complete disequilibrium (D’=1), the pairwise r² values can vary widely, because they are related to the allele frequencies and to the position of the corresponding mutations in the genealogy.

Linkage disequilibrium is also relevant to the more recent discussion of “haplotype blocks”.42 Genetic loci across large areas of the genome were suggested to divide into blocks characterised by little disequilibrium between blocks and limited haplotype diversity within blocks. These two aspects of blocks, physical extent and haplotype diversity are, in a sense, reflected by the measures D’ and r², respectively. However, they are not necessarily linked since they are determined by the different random processes of recombination and mutation; both the extent and haplotype diversity of blocks are extremely variable. Further, haplotypic diversity almost inevitably increases as more polymorphisms are discovered.
There has been some discussion as to whether blocks have clear boundaries, coincident with so-called recombination hot spots, or whether they arise as a result of purely random forces.44–46 Sperm-typing experiments show the existence of hot spots55 but, random forces undoubtedly also have an important role. This debate is important in relation to the stability of block structures across populations and to the sharpness of block boundaries. If the extent of haplotypes is determined by random recombination, then all haplotypes encompassing a given point in the genome will not be the same length and we should not be surprised to see a few high values of D’ extending well outside the main block of linkage disequilibrium.

The idea of haplotype blocks tends to be linked with the idea of haplotype tagging SNPs, largely because the ideas were published simultaneously. However, the idea of haplotype tagging SNPs arose from studies of candidate genes after it was noted that, after discovering large numbers of SNPs by a combination of searching databases and exon resequencing, there is usually substantial redundancy—a few haplotype tagging SNPs capture, in some sense, most of the polymorphism of the gene. Many different methods have been proposed for the choice of such SNPs.46–50 Consideration of the power to detect indirect association via haplotype tagging SNPs suggest that the important criterion is the coefficient of determination, a generalisation of r² to multiple regression models and usually denoted by R². This quantity measures the ability of a set of tag SNPs to predict another dimorphism.46,51 The R² values with which tag SNPs predict the remaining known polymorphisms provide an estimate of the likely ability to predict a causal variant, but, with our limited knowledge of human polymorphism, its accuracy cannot be guaranteed.

Study designs

Familiar epidemiological designs such as population-based case-control or cohort designs52,53 are often used for genetic association studies and the data are analysed much the same way too, risk factors such as smoking and obesity etc, being replaced by the presence or absence of a particular genetic polymorphism. Risk can be considered in terms of either a predisposing allele or genotype, or in terms of multiple categories of disease risk such as the risks associated with different alleles at a multiallelic genetic locus, or the risks associated with the three possible genotypes 1/1, 1/2, 2/2 at a single diallelic locus.

Other designs have been specifically proposed for genetic studies. Family-based designs such as the case-parent triad design,53,54 case-parent-grandparent design,55 or analysis of general pedigrees have been proposed to counteract confounding due to population stratification that can occur in case-control or other population-based designs.56 In family designs, alleles or genotypes

<table>
<thead>
<tr>
<th>Details</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Statistical analysis method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-sectional</td>
<td>Genotype and phenotype (ie, note disease status</td>
<td>Inexpensive. Provides estimate of disease prevalence</td>
<td>Logistic regression, χ² tests of association or linear regression</td>
</tr>
<tr>
<td></td>
<td>or quantitative trait value) a random sample from</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohort</td>
<td>Genotype subsection of population and follow</td>
<td>Provides estimate of disease incidence</td>
<td>Survival analysis methods</td>
</tr>
<tr>
<td></td>
<td>disease incidence for specified time period</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-control</td>
<td>Genotype specified number of affected (case) and</td>
<td>No need for follow-up</td>
<td>Logistic regression, χ² tests of association</td>
</tr>
<tr>
<td></td>
<td>unaffected (control) individuals. Cases usually</td>
<td>Provides estimates of exposure effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>obtained from family practitioners or disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>registries, controls obtained from random</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>population sample or convenience sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extreme values</td>
<td>Genotype individuals with extreme (high or low)</td>
<td>Genotype only most informative individuals hence save</td>
<td>Linear regression, non-parametric, or permutation</td>
</tr>
<tr>
<td></td>
<td>values of a quantitative trait, as established</td>
<td>on genotyping costs</td>
<td>approaches</td>
</tr>
<tr>
<td></td>
<td>from initial cross-sectional or cohort sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-parent triads</td>
<td>Genotype affected individuals plus their parents</td>
<td>Robust to population stratification. Can estimate</td>
<td>Transmission/disequilibrium test, conditional</td>
</tr>
<tr>
<td></td>
<td>(affected individuals determined from initial</td>
<td>maternal and</td>
<td>logistic regression or log-linear models</td>
</tr>
<tr>
<td></td>
<td>cross-sectional, cohort, or disease-outcome</td>
<td>imprinting effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>based sample)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-parent-</td>
<td>Genotype affected individuals plus their parents</td>
<td>Robust to population stratification. Can estimate</td>
<td>Log-linear models</td>
</tr>
<tr>
<td>grandparent septets</td>
<td>and grandparents</td>
<td>maternal and</td>
<td></td>
</tr>
<tr>
<td>General pedigrees</td>
<td>Genotype random sample or disease-outcome based</td>
<td>Higher power with large families Sample may already</td>
<td>Pedigree disequilibrium test, family-based</td>
</tr>
<tr>
<td></td>
<td>sample of families from general population</td>
<td>exist from linkage studies</td>
<td>association test, quantitative</td>
</tr>
<tr>
<td>Case-only</td>
<td>Genotype only affected individuals, obtained</td>
<td>Can estimate interaction effects. Very sensitive to</td>
<td>Logistic regression, χ² tests of association</td>
</tr>
<tr>
<td></td>
<td>from initial cross-sectional, cohort, or disease-</td>
<td>population stratification</td>
<td></td>
</tr>
<tr>
<td></td>
<td>outcome based sample)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA-pooling</td>
<td>Applies to variety of above designs, but</td>
<td>Potentially inexpensive compared with individual</td>
<td>Estimation of components of variance</td>
</tr>
<tr>
<td></td>
<td>genotyping is of pools of anywhere between two</td>
<td>genotyping (but technology still under development)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and 100 individuals, rather than on an individual</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Study designs for genetic association studies
transmitted to affected individuals are compared with untransmitted alleles or genotypes, providing a control sample that is inherently matched to the case sample with regard to population structure. An alternative approach is to use population-based studies and correct for population stratification. However, these methods involve typing of either a large number of unselected markers or a panel of markers chosen to be highly informative for the type of admixture in the study population. Such corrections will only be possible in large studies. The case-parent triad design typically requires the same number of triads (consisting of a case and two parents) to be typed as the number of cases required in a case-control design (assuming an equal number of controls), to give the same power. Thus a sample of 500 case-parent triads will have roughly the same power as 500 cases and 500 controls, but the case-parent triad design requires 1.5 times the amount of genotyping, and could also be more difficult to obtain (except when family samples had already been collected for previous linkage study). For this reason, many prefer the case-control approach; however, family-based approaches provide a useful complementary strategy because of their robustness to population stratification and because they allow estimates of effects due to direct maternal genotype or maternal-fetal interaction and parent-of-origin (imprinting) effects. Case-parent triad designs allow such effects to be estimated at the expense of rather weak assumptions concerning population distributions of parental genotypes. These assumptions may be avoided by use of the case-parent-grandparent design (but such families may be difficult to obtain in practice).

Some designs try to reduce the genotyping effort, for instance by only typing individuals at the extremes of the phenotype distribution. DNA pooling studies reduce the amount of genotyping by typing DNA pooled from a group of individuals as opposed to genotyping each person separately. With the haplotype tagging approach, genotypes from an initial sample (say 32 people) are used to select loci to genotype in the larger sample. This strategy essentially exploits the indirect association approach to gene mapping. Further cost savings and efficiency can be obtained by a staged strategy.

Various different designs are commonly used in genetic association studies (table 2). Methods and programs have been developed for power and sample size calculations (table 3). 

### Statistical analysis

The analysis of data depends crucially on the study design. In the simplest case, familiar methods such as logistic regression, $\chi^2$ tests of association, and odds ratios may be suitable. At a single marker, the issue arises as to whether to analyse on the basis of allele counts or genotype counts. Suppose we have case and control data for a single diallelic genetic locus (table 4). A simple $\chi^2$ test for independence has 2 degrees of freedom. Two odds ratios can be calculated: $af/be$ (for genotype 2/2 vs 1/1) and $cf/de$ (for 1/2 vs 1/1). Alternatively, if there is a reason to expect dominance or recessiveness in the effect of allele 2, we could group the top two rows or bottom two rows together to provide a $\chi^2$ test with 1 degree of freedom and an odds ratio of $(a+c)f/(b+d)e$ or $a(d+f)/b(c+e)$, respectively. Another approach might be a test of trend, with a dose-response effect in regard to the number of copies of the 2 allele. A similar test could be done by uncoupling the alleles within a genotype and constructing a test in terms of case and control chromosomes (table 5). A $\chi^2$ test of association with 1 degree of freedom on the data in table 4 assumes that chromosomes or alleles are independent units, which essentially means Hardy-Weinberg equilibrium (and, for estimation of effects under the alternative hypothesis, multiplicative effects of alleles). All of these tests can be done with statistical software.

Table 6 lists some sources for statistical methods commonly used in genetic association studies. Although many can be done in standard packages, some (particularly for family data) require special software. Some are designed to be simple, others require some specialist knowledge.

The simplest and most powerful statistical analyses arise in the direct association studies in which causal hypotheses, and hence analyses, are specific to single, typed polymorphisms. However, in indirect studies, which exploit linkage disequilibrium between typed markers and causal variants, analysis of marker loci one at a time might not be ideal—the $r^2$ between single

<table>
<thead>
<tr>
<th>Reference</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytical calculation</td>
<td><a href="http://hydra.usc.edu/gse">http://hydra.usc.edu/gse</a></td>
</tr>
<tr>
<td>QUANTO</td>
<td><a href="http://odin.mdacc.tmc.edu/anonftp/TDTASP">http://odin.mdacc.tmc.edu/anonftp/TDTASP</a></td>
</tr>
<tr>
<td>Genetic power calculator</td>
<td><a href="http://statgen.iop.kcl.ac.uk/gpc/index.html">http://statgen.iop.kcl.ac.uk/gpc/index.html</a></td>
</tr>
<tr>
<td>Stata power and sample size programs</td>
<td><a href="http://www.uni-bonn.de/~umt70e/soft.htm">http://www.uni-bonn.de/~umt70e/soft.htm</a></td>
</tr>
<tr>
<td>TDTPOWER</td>
<td><a href="http://lin.damacc.tnc.edu/anonftp/TDTPOWER">http://lin.damacc.tnc.edu/anonftp/TDTPOWER</a></td>
</tr>
<tr>
<td>TDTASP</td>
<td><a href="http://lin.damacc.tnc.edu/anonftp/TDTASP">http://lin.damacc.tnc.edu/anonftp/TDTASP</a></td>
</tr>
<tr>
<td>TDT-PC</td>
<td><a href="http://www.biostat.jhsph.edu/~wmchen/pc.html">http://www.biostat.jhsph.edu/~wmchen/pc.html</a></td>
</tr>
</tbody>
</table>

### Table 3: Resources for power and sample size calculations

<table>
<thead>
<tr>
<th>Chromosomes</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>1/2</td>
<td>e</td>
<td>f</td>
</tr>
<tr>
<td>2/2</td>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>

### Table 4: Counts of genotypes in case-control study

<table>
<thead>
<tr>
<th>Allele</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2a+c</td>
<td>2b+d</td>
</tr>
<tr>
<td>1</td>
<td>2e+c</td>
<td>2f+d</td>
</tr>
</tbody>
</table>

### Table 5: Counts of chromosomes in case-control study
markers and the causal locus could be much lower than the R² for prediction of the causal locus from a group of markers. For such studies, therefore, it will be preferable to use multilocus approaches to analysis. However, these methods are still at an early stage of development.

Multilocus approaches are generally assumed to involve consideration of haplotypes. Analysis at the haplotype level can reveal an effect marked by an ancestral haplotype but has two main drawbacks: since the number of haplotypes could be large, the potential gain might be offset by an excessive increase in the degrees of freedom in the test; and haplotype phase will often be uncertain.

There are several ways in which the first problem might be approached. The simplest, and most common is to pool rare haplotypes, which will certainly sacrifice some information. Instead some have proposed grouping strategies based on cladistic considerations. However, for markers in regions in which linkage disequilibrium is strong, it is questionable whether the use of any haplotype information is worth the increase in degrees of freedom since, in such circumstances, a simple multiple regression equation with one parameter per marker can achieve prediction of untyped loci with R² only slightly worse than haplotype-based predictions. This finding suggests testing for indirect associations either by regression of trait on marker loci, without inclusion of the interaction terms, or by an appropriate variant of Hotelling’s T² statistic. These analyses have the additional attraction of not requiring resolution of haplotype phase and can often be done with conventional statistical packages.

When linkage disequilibrium is less strong, haplotype analyses remain important, especially for fine mapping (eg, in a stepwise logistic regression strategy). Long haplotypes, spanning several blocks of linkage disequilibrium, are particularly important for identifying rare variants, although these would have to have large effects to be detectable. The resolution of phase can then present a serious practical problem. Various computational algorithms address the phase-estimation problem, both in unrelated individuals and in families. In a two-stage procedure (whereby haplotype scoring based on a first haplotype analysis stage are used in a second stage test of association), these algorithms can be satisfactory in population-based studies since significance can be assessed by permutation arguments. For estimation of relative risks, however, association and haplotype phase must be considered simultaneously. In family studies, too, association and haplotype phase must be considered simultaneously. In family studies, too, family-based controls are relevant to haplotype phase patterns, which are important for haplotype analysis, albeit with some loss of information.

### Table 6: Statistical methods for genetic association studies

<table>
<thead>
<tr>
<th>Approach</th>
<th>Reference</th>
<th>Software</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>Model log odds of disease as linear function of underlying genotype variables</td>
<td>20, 74, 20</td>
<td><a href="http://www.stata.com/">http://www.stata.com/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model quantitative trait as linear function of underlying genotype variables</td>
<td>See above</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model survivor function or hazard as function of underlying genotype variables</td>
<td>See above</td>
</tr>
<tr>
<td>Survival analysis</td>
<td></td>
<td>Model counts of genotype combinations for mother, father, and affected offspring</td>
<td>See above</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test for association or linkage between disease phenotypes and haplotypes by utilizing family-based controls</td>
<td>See above</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linkage disequilibrium analysis of quantitative and qualitative traits based on variance components</td>
<td><a href="http://www.sph.umich.edu/csg/abecasis/QTDT/">http://www.sph.umich.edu/csg/abecasis/QTDT/</a></td>
</tr>
<tr>
<td>DNA pooling</td>
<td></td>
<td>Test for differences in allele frequencies in different pooled samples while estimating components of variance due to experimental error</td>
<td>See above</td>
</tr>
</tbody>
</table>

markers and the causal locus could be much lower than the R² for prediction of the causal locus from a group of markers. For such studies, therefore, it will be preferable to use multilocus approaches to analysis. However, these methods are still at an early stage of development.

Multilocus approaches are generally assumed to involve consideration of haplotypes. Analysis at the haplotype level can reveal an effect marked by an ancestral haplotype but has two main drawbacks: since the number of haplotypes could be large, the potential gain might be offset by an excessive increase in the degrees of freedom in the test; and haplotype phase will often be uncertain.

There are several ways in which the first problem might be approached. The simplest, and most common is to pool rare haplotypes, which will certainly sacrifice some information. Instead some have proposed grouping strategies based on cladistic considerations. However, for markers in regions in which linkage disequilibrium is strong, it is questionable whether the use of any haplotype information is worth the increase in degrees of freedom since, in such circumstances, a simple multiple regression equation with one parameter per marker can achieve prediction of untyped loci with R² only slightly worse than haplotype-based predictions. This finding suggests testing for indirect associations either by regression of trait on marker loci, without inclusion of the interaction terms, or by an appropriate variant of Hotelling’s T² statistic. These analyses have the additional attraction of not requiring resolution of haplotype phase and can often be done with conventional statistical packages.

When linkage disequilibrium is less strong, haplotype analyses remain important, especially for fine mapping (eg, in a stepwise logistic regression strategy). Long haplotypes, spanning several blocks of linkage disequilibrium, are particularly important for identifying rare variants, although these would have to have large effects to be detectable. The resolution of phase can then present a serious practical problem. Various computational algorithms address the phase-estimation problem, both in unrelated individuals and in families. In a two-stage procedure (whereby haplotype scoring based on a first haplotype analysis stage are used in a second stage test of association), these algorithms can be satisfactory in population-based studies since significance can be assessed by permutation arguments. For estimation of relative risks, however, association and haplotype phase must be considered simultaneously. In family studies, too, family-based controls are relevant to haplotype phase patterns, which are important for haplotype analysis, albeit with some loss of information.

### Table 6: Statistical methods for genetic association studies

<table>
<thead>
<tr>
<th>Approach</th>
<th>Reference</th>
<th>Software</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>Model log odds of disease as linear function of underlying genotype variables</td>
<td>20, 74, 20</td>
<td><a href="http://www.stata.com/">http://www.stata.com/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model quantitative trait as linear function of underlying genotype variables</td>
<td>See above</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model survivor function or hazard as function of underlying genotype variables</td>
<td>See above</td>
</tr>
<tr>
<td>Survival analysis</td>
<td></td>
<td>Model counts of genotype combinations for mother, father, and affected offspring</td>
<td>See above</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test for association or linkage between disease phenotypes and haplotypes by utilizing family-based controls</td>
<td>See above</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linkage disequilibrium analysis of quantitative and qualitative traits based on variance components</td>
<td><a href="http://www.sph.umich.edu/csg/abecasis/QTDT/">http://www.sph.umich.edu/csg/abecasis/QTDT/</a></td>
</tr>
<tr>
<td>DNA pooling</td>
<td></td>
<td>Test for differences in allele frequencies in different pooled samples while estimating components of variance due to experimental error</td>
<td>See above</td>
</tr>
</tbody>
</table>
Alternatively, as in population–based studies, phase uncertainty can be accounted for explicitly in the association analysis. The problem of uncertain phase can be avoided altogether by use of molecular methods for haplotyping but such methods usually have lower throughput and are more expensive than those that yield only the diplotype.

In addition to associations between phenotypes and single genes, interaction effects between genes or between genes and environment can also be studied. After taking account of the vast increase in the number of potential tests, the expected power to detect interactions is low. However, power can be increased if we can safely assume independence between genes, or between a gene and an environmental exposure, within the population as a whole and, therefore, within controls. Evidence for statistical interaction can then be obtained from examination of cases only. However, as we have emphasised, the relationship between statistical and biological interactions (eg, functional interaction between proteins) is complex. Such analyses are more relevant to prediction of disease risk than to elucidation of the underlying trait pathogenesis.

Significance and importance

The standards of statistical proof that have become acceptable in the general biomedical literature are not appropriate for genetic association studies. Something akin to a multiple testing problem pervades the discipline, although there has been no clear consensus about how it should be dealt with. Approaches such as the Bonferroni correction are not appropriate because it is not the number of tests in any one investigation that is important. Rather it is that the vast majority of loci tested will not be associated, so that even a small false positive probability will mean that most positive results will turn out to be false. Thus, it is the a-priori probability of association that needs to be accounted for, rather than the number of tests. Thus, it has been suggested that Bayesian methods are more appropriate when prior probability of association is known, they allow calculation of the posterior probability that an association is genuine. However, the mathematics require knowledge not only of the prior probability of association but also of the distribution of the size of effects that will be encountered.

In gene expression array studies, so many tests are done simultaneously that these unknowns can be estimated within the experiment, and empirical Bayes methods can be used. When genome-wide association studies with many thousands of SNPs become feasible, such methods will become appropriate for association studies, but in the meantime, studies of candidate regions will dominate, and here the prior probabilities that determine appropriate standards of evidence remain largely subjective. However, such considerations show that, given the small a-priori probability that any genetic locus is associated with disease and the small effect sizes that seem to be typical and the inadequate study sizes that have also been typical, it should not be at all surprising that most findings judged positive with conventional levels of statistical significance have not been replicated.

Some would respond by pointing to the very low population attributable fractions that correspond to these small genetic effects and asking whether there is any utility in their discovery. However, no one would claim that the interventions that will follow from advances in genetic epidemiology will simply correct the less beneficial genetic variation. Instead, the important role of such studies will be the elucidation of mechanisms. In epidemiology, the role of genetic variation can be important in establishing the causal nature of environmental associations in which intervention could have major effects.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgments

Funding was provided by the Wellcome Trust and the Juvenile Diabetes Research Foundation.

References


www.thelancet.com Vol 366 September 24, 2005
End-of-life: a Catholic view

Hazel Markwell

Catholic bioethics is not a fully monolithic structure. Different theological and philosophical methodologies have resulted in differences of opinion on issues such as birth control, sterilisation, reproduction, and abortion. However, more common ground is found on end-of-life issues. The Catholic understanding of sickness, suffering, and death is grounded in a belief in Jesus Christ who, as the incarnation of God, suffered, died, and was resurrected. In light of this faith and hope for an afterlife, Catholics accept that although an effort must be made to eliminate sickness, suffering, and death, these things can also have a positive meaning. The belief that God participates in the human condition grounds Catholic values and positions on end-of-life issues. Catholic bioethics therefore has its source both in faith and in the ability of human reason to interpret scripture and, as Vatican II directed, to read the “signs of the times” in applying the teaching of the Church to contemporary situations. In other words, Catholics should remain attuned to the message of Christ through history in attempting to “do the right thing”.

In Catholic bioethics, two basic human values ground all others: human dignity, and the interconnectedness of every individual. The value of dignity of the individual arises from the belief that life has intrinsic worth because people are created in the image and likeness of God. Respect for human life results from this principle. Catholics believe that people are stewards, rather than owners, of their own bodies, and are accountable to God for the life that has been given to them, and for this reason life is said to be sacred. However, life is not an absolute good to be preserved at all costs, but is subordinated to the good of the whole person. Gaudium et Spes, one of the documents of Vatican II, makes the point that the dignity of the human person lies above all in the fact that he or she is called into a relationship with God. As social beings who are connected to each other in society, we also share a responsibility for one another.

All other values are derived from these two. The value of stewardship and creativity holds that we are accountable to God for the life that has been given to us. Belief in the value of the common good calls us to promote a just social order. This just order demands that we remain true to the value of charity or solidarity, in which we have a responsibility to respond to others in need, in particular the poor. This responsibility requires a commitment not only to the poor in our midst but also to those throughout the world. However, the huge gap between the rich and the poor of the world is widening, in large part due to the debt owed by developing countries. In 1960, the wealthiest fifth of the world’s population enjoyed an income that was 30 times greater than that of the poorest fifth, but a few years ago it was more than 80 times greater.

This situation has implications for all aspects of health care, including end-of-life care, research ethics, priority setting, women’s health, child health, mental health, and rehabilitation ethics. Sadly, unlike most of these issues, quality of end-of-life care has not been addressed at the global level. Solidarity with the poor and a commitment to social justice require that health systems engage in interventions that can lead to improvements in the global problem of end-of-life care. Some of these interventions could include “culturally specific educational programs for public health workers and the public; population based strategies to destigmatise death . . . and changes in social policies in relation to care for orphans”.

Caution is needed here to avoid the pitfall of applying the perspectives of developed countries to developing nations. Rather, any approach that attempts to address end-of-life care in developing countries must include indigenous people and a sensitivity to their own ethnic and medical systems.

The impetus and foundation for this commitment to the poor lies in a notion of justice that is grounded in love and an adherence to the message in the Gospel of Matthew (25:40): “. . . in so far as you did this to one of the least of these brothers of mine, you did it to me”.

Belief in the values of human dignity and interconnectedness has implications for decisions at the end of life. The value of interconnectedness implies a relationship between physician and patient that is covenantal rather than contractual. It suggests a process of decision making that is multifaceted and goes beyond a strict adherence to an individualistic autonomy, while ensuring the patient’s ongoing participation in making choices that affect his or her life. The value of human dignity and therefore respect for human life has implications relevant to the alleviation of pain and to the issues of withholding or withdrawing of treatment. The two values of the common good and charity require an awareness of the needs of people other than the patient when addressing the issues of patients’ demands. The focus of this article will be on how the values of human dignity and interconnectedness affect end-of-life care. I will begin with a discussion of how interconnectedness affects the physician-patient relationship, and argue that the notion of covenant best describes the ideal relationship between physician and patient. Second, I will discuss how the values of human dignity and respect for human life affect the questions of pain and suffering and the withholding or withdrawing of treatment. I will conclude with some brief suggestions for resolution of conflicts in areas where patients’ demands cannot be met.
The notion of covenant

Medicine today has become focused on the rational and intellectual, with emphasis on outcomes and evidence. Bioethics over the past 25 years has focused on the ordering of principles and the development of rules, but such development is inadequate, in that it has “not offered much insight into those ordeals confronting patients (and sometimes practitioners) that do not wholly admit of solution”.8 Writers such as William F May believe that these problems need to be faced, rather than solved, since “Moral reflection about such events does not simply trace back to a brace of sometimes conflicting principles; it forces meditation on the human condition; it probes one’s deepest convictions; it may even unsettle one’s habits; it asks of the agent the mobilization of resources, some of them already in place but untested; others, as yet, unbidden”.8

Whereas the notion of covenant takes its roots in the biblical context, it also figures prominently in the Hippocratic tradition in which the physician has, first, a duty to his or her patients; second, a “covenantal obligation to one’s teacher and . . . family; and third, sets both within the context of an oath to the gods”.9

Respect for the value of interconnectedness and a focus on covenant demands that the physician and patient enter into a relationship of trust with each other. The patient’s trust of the physician is in fact an act of faith in the good intentions of the physician. An ethic of trust calls the physician to enter into a relationship with the patient in order that he or she might begin to understand what the patient’s wishes might be. This interaction clearly transcends the limits of a contractual model of physician/patient relationship; rather, it demands a covenantal relationship between both participants, who see each other as a “gift”, with the ensuing obligations that this implies.

One of the necessary conditions that must be met for a covenantal relationship, which holds true to the value of interconnectedness, is trust. However, trust between physician and patient is complex. Patients since the time of Hippocrates have been asked to trust their doctors, but only recently, perhaps due to the arrival of the legal doctrine of informed consent, have physicians been asked to trust their patients by having conversations with them about their treatment options. For conversation to be meaningful, authors like Jay Katz in his book The Silent World of Doctor and Patient9 propose that we must differentiate between trust that is blind and trust that is earned following an acknowledgment that one person does not hold all of the answers. This understanding is crucially important in discussions around end-of-life issues, in which prognosis is often uncertain. Katz suggests that the proponents of patient self-determination have not appreciated or fully understood the difficulties in expression of uncertainty. In fact, as he notes, the only specific advice on conversation in the Hippocratic Oath “speaks against disclosure”.10 Similarly, Katz holds the opinion that physicians of ancient Greece would have found the notion of shared decision making unnecessary “because they viewed doctor and patient as united through philia, friendship, which made their objectives one and the same”.10 In covenantal relationships, on the other hand, the objectives of each party are not viewed as identical. Catholic bioethics places a strong focus on the informed and voluntary wishes of the capable patient in determining which treatments should be given or withheld. In view of the belief that human beings are created in the image of God, each person has an intrinsic tendency towards the good. However, every individual is also endowed with freedom and therefore can make choices not only for the good but also for things that are not good. The dignity of the human person, as one who is oriented toward God, requires him or her to make choices for the good within a free and informed conscience. In view of the proposed covenantal approach to the physician-patient relationship, the physician, capable patient, and anyone who the patient wishes to participate, together should make the decision as to which treatment plan is best aligned with the values of the patient. If the patient is not capable of making these decisions in partnership with their physician, then family members or substitute decision makers are called upon to make treatment decisions in the best interests of the patient.

Withholding or withdrawing treatment

Since at least the 16th century,11 Catholic theologians have made a distinction between ordinary and extraordinary measures. This position holds that while patients are obliged to choose ordinary methods for preserving life, they have the choice as to whether or not to accept extraordinary methods. A common definition of these terms is one proposed by Gerald Kelly: “Ordinary means of preserving life are all medicines, treatments, and operations which offer a reasonable hope of benefit for the patient and which can be obtained and used without excessive expense, pain or other inconvenience . . . . Extraordinary means of preserving life . . . . mean all medicines, treatments, and operations, which cannot be obtained without excessive expense, pain or other inconvenience, or which, if used, would not offer a reasonable hope of benefit.”12

It seems that the term “ordinary” was originally used to mean “what is medically customary”. However, in today’s medical practice, in which many measures such as cardiopulmonary resuscitation are routinely used in dying patients, many extraordinary measures are in danger of becoming customary. Again, the free and informed choice of the patient and family in collaboration with the treating team and the medical indications should inform and guide the process.

It has also become clear that the expressions cannot be defined in terms of categories of treatment; there is not one list of ordinary procedures, and another list of extraordinary procedures.13 Mechanical ventilation, for
example, could be ordinary at one stage in an illness, and extraordinary at a later stage as the illness advances. Although the physician has a right and an obligation to provide the patient and family with information about what is medically possible, what is medically indicated, and which treatment provides the best outcome in terms of a risk-benefit analysis, it is primarily the patient and family who have the right to determine what is or is not ordinary or extraordinary from an ethical point of view. For patients, issues of pain and suffering play a crucial role in determining whether or not a procedure should be used, not the fact that the procedure has become routine.

Pain and suffering
The issue of pain and suffering is important to Catholic bioethics. However, it is prudent at the outset to make the distinction between the two. Pain is the physical discomfort that often accompanies illness, whereas suffering refers to the existential anguish experienced by patients when they come face to face with the loss of all that they have hoped for in the future. However, the two issues are connected. Authors like Ira Brock have pointed out that untreated physical pain can lead to an increase in suffering. Patients whose pain is untreated often experience feelings of abandonment, which in turn increases their suffering. While Catholic bioethics believes that the experience of pain and suffering is not without meaning, this belief does not imply that pain relief should be withheld in order that a patient might come to understand the redemptive nature of suffering. Control of physical pain is a patient’s right and not a privilege that is meted out to those who we feel deserve it. Although personal growth may occur through suffering, the Catholic tradition does not present pain and suffering as goods in themselves. As early as the 1950s, a group of anaesthesiologists asked Pope Pius XII whether or not pain relief should be offered to a patient, if in so doing the patient’s life might be unintentionally shortened. The Pope replied that painkillers should be offered if no other means existed, even if this led to unconsciousness and the inability to fulfil one’s moral duties and family obligations. This judgment reflects the principle of double effect, which has a critical role in the care of the dying and specifies that “An action with 2 possible effects, one good and one bad, is morally permitted if the action: (1) is not in itself immoral, (2) is undertaken only with the intention of achieving the possible good effect, without intending the possible bad effect even though it may be foreseen, (3) does not bring about the possible good effect by means of the possible bad effect, and (4) is undertaken for a proportionately grave reason”.10

In Pellegrino and Thomasma’s book, Helping and Healing: Religious Commitment in Health Care, the authors point out that when patients suffer, they experience a sense of their own vulnerability and finitude, as well as a disruption and fracture of their own person and sense of community. As a result, while the experience of suffering can be an opportunity to experience God, this experience occurs through an encounter with another person. Since human beings are interconnected, human flourishing comes to fruition in community and not in isolation, especially in the experience of illness. However, in illness, the patient has specific needs that can only be fulfilled by the healer. As a result, treating pain, holding a patient’s hand, administering chemotherapy, and performing surgery have the potential to become moments of opportunity for the experience of God, according to Pellegrino and Thomasma. As such, the practice of medicine takes on the nature of sacrament, the visible sign of the invisible presence of God. According to Pellegrino and Thomasma, the sense of finitude, vulnerability, loss of self, and destruction of a person’s normal life that is experienced in illness can be transformed when a sacramental approach to medicine is taken. They note that the art of medicine is a human endeavour that imitates the beauty and creativity of God. In experiencing this art, the patient experiences transcendence rather than despair. In this way finitude is transformed. When physicians treat patients regardless of how the patient might have contributed to their own illness, and when they respond to the cry of the patient for help, vulnerability is overcome. When an attempt is made to treat the whole person, rather than focusing on bodily functions, personhood is restored. Finally, when the focus of medicine is to place the patient, even though he or she may be dying, back into the community where he or she has the opportunity to experience the love of those around him or her, the sense of disruption of life is attended to. All these components of healing function as signs of God’s grace. However, “the clinical event first and foremost becomes a saving event if the intention of the healer is to imitate what Christ did as a sign of sacrificial love, a love that joins the participants together with Christ.”

Respect for the dignity of the human person reminds us that physical health is only one good among many and, in itself, is not the highest good. The Catholic tradition believes that God has created the human person within the context of a destiny that lies beyond the earthly condition. As a result, for Catholics, the process of dying is more than a medical crisis. While dying may provoke feelings of fear and abandonment it is also a time for remembering both the joyful and the painful moments of one’s life. It is therefore an opportunity for celebration as well as forgiveness and reconciliation. As a result, spiritual support is crucial. The presence of the chaplain and priest should be offered to patients as part of their ongoing care during the process of dying. The sacraments become of particular importance at this time because of the need of patients to be nourished and strengthened in their faith. In particular the “sacrament of the sick”, or “extreme unction”, is not only a sacrament for patients who are about to die, but also for those who are perhaps proceeding to the end of their lives due to illness.
Sadly, for some people the journey through illness is more an experience of terror than one of entering into the mystery of life and death. When this happens, many families and patients respond by demanding treatments that not only are medically futile but in fact are not in keeping with their own personal values or religious beliefs. Although respect for human dignity requires an understanding of the patient’s values and the fostering of free and informed choice, it does not necessitate blind obedience to demands that are not beneficial or may even be harmful. However, a response that focuses on legal rights and obligations is a pitfall to be avoided. Medical practice must operate within the law, but it should not be reduced to the law. Christian justice does not operate within a set of abstract principles or apart from the human condition. Neither does it focus on what one is owed. Rather, it offers a way of love shown to us by Christ. This provides direction in situations of conflict.

A covenantal approach to care of the dying requires an ongoing commitment of each party to come to an understanding of the other’s position. With dying patients who are Catholic, a priest or chaplain can be of invaluable assistance. Unreasonable demands on the part of patients or families often arise more out of a deep sense of anxiety, grief, and unfinished business rather than a philosophical position on autonomy and justice. As Henri Nouwen writes, the deepest fear experienced by the dying person and family is rejection: “Indeed we can be healed from our fear of death, not by a miraculous event that prevents us from dying, but by the healing experience of being a brother or sister of all humans—past, present, and future—who share with us the fragility of our existence.”19

In conclusion, this discussion has pointed out some of the main components of Catholic bioethics that affect end-of-life care (panel). The value of interconnectedness results in a relationship between physician and patient that is like kinship and demands ongoing conversation and trust. The value of human dignity results in an approach to pain and suffering that requires appropriate medical intervention, participation in the patient’s journey, and an awareness that the process of dying is more than a medical event concentrated at the end of life. Here the importance of the priest or chaplain in addressing the spiritual and religious needs of the Catholic patient has been raised. Of crucial importance is an acceptance that physicians are called both to cure and to care. For Catholic bioethics, care involves a recognition of the fragility and vulnerability of every human being as one who has first been loved by God and therefore deserves our total commitment.

References
5 Editor’s choice: “the champagne glass of world poverty”. BMJ 1999; 318.
18 Vatican II. Sacrosanctum Concilium: 22.
A 21-year-old girl with recurrent abdominal pain after a robbery

Albert Selva-O’Callaghan, Jose Gómez-Acha, Pere Munne, Miquel Vilardell-Tarrés

In July, 2003, a 21-year-old girl, born in Ecuador, was seen in our department for recurrent abdominal pain. 4 years earlier, in June, 1999, she had experienced a violent robbery and was shot by a firearm; exploratory laparotomy and gastric suture were done at that time. Since then, she had periodically (once per month) experienced cramps and colicky abdominal pain, for which she had presented to the emergency department in her local hospital in Ecuador. She received symptomatic relief with analgesic drugs, but a diagnosis was not established; she had been diagnosed with gastritis, cholecystitis, and pancreatitis in previous hospital admissions. 1 year later she came to Spain with her family. 1 month before admission to our hospital, she went to the emergency department with another episode of abdominal pain and was diagnosed with intestinal intussusception on the basis of CT findings; this was not confirmed by laparotomy.

When seen by us, she had no gingival Burton’s lines, and there were no signs of neuritis of the radial nerve or other upper limb motor nerves. An abdominal radiograph (figure) showed foreign bodies with a metal density, suggestive of firearm pellets. Laboratory tests showed that her haemoglobin was 80 g/L, with low haptoglobin levels, high reticulocyte count, and basophilic stippling. Her blood lead concentration was 7·5 μmol/L (accepted, <0·5 μmol/L), and urinary δ-aminolaevulinic acid was 26·4 mmol/mol creatinine (normal, 0–5). We diagnosed the patient with lead poisoning and treated her with EDTA (edetic acid) chelation for 10 days. Total urinary lead excretion was 85 μmol/day at the start of chelation, and had dropped to 4 μmol/day on the last day of chelation. The pellets were surgically removed from the abdominal and peritoneal sites. At final follow-up in May, 2005, the patient was symptom-free, and had no abdominal pain; her blood lead concentration was 0·7 μmol/L, and anaemia with basophilic stippling had resolved.

The signs and symptoms of lead poisoning may be difficult to recognise, and with a low level of suspicion and without appropriate testing, the diagnosis can be difficult. Lead exposure in adults occurs in numerous work settings, such as the manufacture or use of batteries, solder, paint, and ceramics. Retained bullets after firearm injuries is probably an underestimated source. Lead poisoning is a rare, yet well-documented, complication of gunshot injuries, and patients with retained lead fragments in a joint are at higher risk for the development of lead toxicity. Prospective studies in patients with retention of projectile fragments have shown that blood lead tends to increase with time after the injury, and that the increase depends to some extent on the presence of a bone fracture caused by the gunshot, or on exposure of the lead to synovial fluid. Our patient did not have a broken bone or pellets in the intra-articular space. Nevertheless, exposure of the pellets to serous membranes with a wide absorption surface, in this case the peritoneum, was undoubtedly the factor causing lead toxicity in our patient. Irrespective of their location, surgical removal of the bullets and chelation therapy are the mainstay measures for treating lead poisoning due to gunshot injuries.

References