1. Ensuring drug safety: lessons from the thiazolidinediones
   Page 1101
   The Lancet

2. A Framework Convention on Alcohol Control
   Page 1102
   The Lancet

3. The traditional white coat: goodbye, or au revoir?
   Page 1102
   The Lancet

4. Thiazolidinediones, deadly sins, surrogates, and elephants
   Pages 1103-1104
   John GF Cleland and Stephen L Atkin

5. Patient-important outcomes in diabetes—time for consensus
   Pages 1104-1106
   Victor M Montori, Gunjan Y Gandhi and Gordon H Guyatt

6. Pathology of human H5N1 infection: new findings
   Pages 1106-1108
   Wai Fu Ng and Ka Fai To

7. Supported employment for people with severe mental illness
   Pages 1108-1109
   Paul B Gold and Geoff Waghorn

8. Mental health and global movement of people
   Pages 1109-1111
   Dinesh Bhugra and Iraklis Harry Minas

9. The built environment and health
   Pages 1111-1113
   Mala Rao, Sunand Prasad, Fiona Adshead and Hasitha Tissera
10. Learning from new initiatives in maternal and child health
   Pages 1113-1114
   Cesar G Victora, Robert E Black and Jennifer Bryce

11. Community workers key to improving Africa's primary care
    Pages 1115-1117
    Hannah Brown

12. Doctors to pay for patients' medicine in Germany
    Page 1118
    Rob Hyde

13. An impressive dictionary of medical lives
    Pages 1119-1120
    David Weatherall

14. Vaccine maker
    Page 1120
    Anne Harding

15. Packaging disease
    Page 1120
    Katherine Nightingale

    Page 1121

17. Abbas Ali Mansour
    Page 1121

18. Marguerite M Vogt
    Page 1122
    Ivan Oransky

19. Efficacy of ivermectin against Onchocerca volvulus in Ghana
    Page 1123
    Ed Cupp, Frank Richards, Patrick Lammie and Mark Eberhard

20. Efficacy of ivermectin against Onchocerca volvulus in Ghana
    Page 1123
    Charles D Mackenzie
21. Efficacy of ivermectin against Onchocerca volvulus in Ghana
   Pages 1123-1124
   Jan HF Remme, Uche Amazigo, Dirk Engels, Andriamahefazafy Barryson and
   Laurent Yameogo

22. Efficacy of ivermectin against Onchocerca volvulus in Ghana – Authors' reply
   Pages 1124-1125
   Mike Osei-Atweneboana, Jeffrey Eng, Daniel Boakye, John Gyapong and Roger
   Prichard

23. Efficacy of ivermectin against Onchocerca volvulus in Ghana
   Page 1125
   Gilbert Burnham

24. Coeliac disease and lymphocytic hypophysitis
   Page 1125
   Weekitt Kittisupamongkol

25. Coeliac disease and lymphocytic hypophysitis – Authors' reply
   Pages 1125-1126
   Ami Schattner and Taiba Zornitzki

26. Dual inhibition of the renin system by aliskiren and valsartan
   Pages 1126-1127
   Suzanne Oparil, Steven A Yarows, Samir Patel, Jack Zhang and Andrew Satlin

27. Global Fund: harmonisation and good governance vital
   Page 1127
   Nicolaus Lorenz and Kaspar Wyss

28. Chronic myeloid leukaemia in India
   Page 1127
   Pankaj Malhotra and Subhash Varma

29. Prescribing in elderly people
   Page 1128
   Marie-Laure Laroche, Jean-Pierre Charmes and Louis Merle

30. Who said that?
    Page 1128
    Judith A Whitworth

31. Department of Error
    Page 1128
32. Congestive heart failure and cardiovascular death in patients with prediabetes and type 2 diabetes given thiazolidinediones: a meta-analysis of randomised clinical trials
   Pages 1129-1136
   Rodrigo M Lago, Premranjan P Singh and Richard W Nesto

33. H5N1 infection of the respiratory tract and beyond: a molecular pathology study
   Pages 1137-1145
   Jiang Gu, Zhigang Xie, Zhancheng Gao, Jinhua Liu, Christine Korteweg, Juxiang Ye, Lok Ting Lau, Jie Lu, Zifen Gao, Bo Zhang, Michael A McNutt, Min Lu, Virginia M Anderson, Encong Gong, Albert Cheung Hoi Yu and W Ian Lipkin

34. The effectiveness of supported employment for people with severe mental illness: a randomised controlled trial
   Pages 1146-1152
   Tom Burns, Jocelyn Catty, Thomas Becker, Robert E Drake, Angelo Fioritti, Martin Knapp, Christoph Lauber, Wulf Rössler, Toma Tomov, Jooske van Busschbach, Sarah White and Durk Wiersma

35. Achieving health equity: from root causes to fair outcomes
   Pages 1153-1163
   Michael Marmot

36. Barriers to improvement of mental health services in low-income and middle-income countries
   Pages 1164-1174
   Benedetto Saraceno, Mark van Ommeren, Rajaie Batniji, Alex Cohen, Oye Gureje, John Mahoney, Devi Sridhar and Chris Underhill

37. Energy, energy efficiency, and the built environment
   Pages 1175-1187
   Paul Wilkinson, Kirk R Smith, Sean Beevers, Cathryn Tonne and Tadj Oreszczyn

38. 4-week headache after 60 pints of beer
   Page 1188
   Zia I Carrim, Jane MacPhillimy and Ravi Jampana
Ensuring drug safety: lessons from the thiazolidinediones

In today’s Lancet, Rodrigo Lago and colleagues add to the ongoing analysis of the safety of thiazolidinediones for control of hyperglycaemia. Initial concerns about the cardiovascular risks of one of the thiazolidinediones, rosiglitazone, erupted in a firestorm of controversy when a meta-analysis by Nissen and Wolski was published online in the New England Journal of Medicine in May. Nissen and Wolski found that rosiglitazone (Avandia, GSK) was associated with a significantly increased risk of myocardial infarction, and a risk of death from cardiovascular causes that was of borderline statistical significance. The authors acknowledged that their analysis was limited by the public availability of trial results and a lack of access to patient-level data, and that the pooled studies were not designed to assess cardiovascular outcomes. But they concluded that the risks of rosiglitazone use in diabetes should be carefully considered by both doctors and their patients.

Response to these findings was swift. An editorial that accompanied the paper, while acknowledging the findings’ “fragility”, called for urgent regulatory action; GSK vigorously defended its product, saying that studies show Avandia’s cardiovascular profile to be comparable to other oral antidiabetes agents; and this journal counselled against a rush to judgment, advising patience until the final results of RECORD, a phase III trial designed to assess cardiovascular outcomes, were available. Media coverage was extensive. A US Congressional hearing was hastily called, as was a Food and Drug Administration (FDA) Advisory Committee meeting. The Congressional hearing was a spectacular display of partisan agendas and bipartisan ignorance. The FDA meeting resulted in a 22 to 1 vote to keep Avandia on the market, and to add a “black box” warning on the label of the risks of the drug’s use in patients with congestive heart failure. If this sequence of events sounds familiar, it is because a nearly identical path was trodden when the safety of selective cyclo-oxygenase 2 (COX-2) inhibitors was called into


For GSK’s defence see http://www.gsk.com/media/press-kits/avandia-21may2007.pdf

For The Lancet’s view on rosiglitazone see Editorial Lancet 2007; 369: 1814


agent in the thiazolidinedione class (Actos, Takeda), and one assessing long-term cardiovascular risk with rosiglitazone. Here, both drugs were associated with an increased risk of heart failure, though without an associated increase in mortality. This is the same conclusion reached by Lago and colleagues: patients taking thiazolidinediones seem to have a higher risk of congestive heart failure, but do not have a higher risk of death from cardiovascular causes.

Is there then a bottom line to all these bits of evidence? What should doctors and patients do? Is there in fact enough good evidence on which to decide anything? It seems that the jury is still out for the thiazolidinediones as a class. But there are many interim take-home messages. Some of these are highlighted by Comments in today’s issue. First, it must be remembered that meta-analysis is a technique with important limitations. And the studies on which the thiazolidinedione meta-analyses are based have thus far all involved surrogate markers; the studies were not designed to assess cardiovascular outcomes, but rather improved glycaemic control. This outcome, as Victor Montori and colleagues note, is not a patient-centred one. The current clinical emphasis on glucose control (as measured by HbA1c) skirts the outcomes that matter most to patients—microvascular and macrovascular complications, quality of life, and survival.

These commentators highlight issues that must be taken into account in the ongoing debate about thiazolidinediones. Future trials ought to be designed with these issues firmly in mind. Further, it is no secret that the regulatory system is also in urgent need of repair. Manufacturers must do—in a timely fashion—postmarketing studies that assess the long-term safety of their drugs, and regulatory agencies must hold manufacturers’ feet to the fire to ensure that these are performed, performed properly, thoroughly evaluated, and made available to guide decisions about prescribing. Agencies like the US FDA must have the resources and authority to close what is now a potentially dangerous gap. Unless limitations on the understanding, analysis, and communication of drug safety issues are addressed, the thiazolidinediones might simply become the latest in a series of preventable drug disasters. ■ The Lancet
The traditional white coat: goodbye, or au revoir?

Last week, UK Health Secretary Alan Johnson outlined new measures to prevent hospital-acquired infection, including hand washing and a “bare below the elbows” dress code for clinical staff in hospitals, to start in January, 2008. The new policy means short sleeves. The policy bans the traditional long-sleeved white coat. Also banned are wristwatches, presumably leaving doctors to gaze expectantly for a convenient wall-clock when timing an event or taking a pulse.

Cuffs on long sleeves can become contaminated with microorganisms, but do cuffs transmit infection? The Health Secretary’s working group stated that “there is no conclusive evidence that uniforms (or other work clothes) pose a significant hazard in terms of spreading infection”. So, on what basis did the Health Secretary make his recommendations? The working group resorted to “informed common sense”—a level of evidence just above guesswork.

What will be the effect of these new, possibly ineffectual, guidelines about short sleeves? Johnson’s guess is that they will help to ensure better washing of hands and wrists. The reforms do signal that the National Health Service is taking hospital-acquired infection seriously, especially because the public perceives that uniforms present an infection risk, according to the working group.

Johnson’s boss has also grasped the wrong end of the evidence stick. This week, UK Prime Minister Gordon Brown called for ward-by-ward cleaning in hospitals to stop the spread of infection. But disinfection of high-touch surfaces is what is needed, more so than removing visible dirt. The public understandably wants clean wards and crisp uniforms, but politicians must stop pandering to populism about hospital cleanliness and listen to the evidence. Brown also plans to double the number of hospital matrons, to check on ward cleaning, and accost doctors wearing long sleeves. They would be better employed making sure doctors, nurses, and visitors wash their hands properly, the proven way to stop hospital-acquired infections. ■ The Lancet

A Framework Convention on Alcohol Control

International conventions exist to control narcotics, psychotropic substances, tobacco, and doping in sport. But when it comes to alcohol—the drug that causes 4% of global deaths and disability—nearly as much as tobacco (4·1%) and five times the burden of illicit drugs—no similar legally binding agreement has been sought.

In many cultures, alcohol has long been considered in a more salubrious, socially acceptable light than other drugs. This attitude might be because of the putative health benefits associated with moderate drinking. Or perhaps it is due to the popularity of alcohol among higher socioeconomic groups, who, in many countries, are more likely to drink and more likely to drink regularly, while lower socioeconomic groups are more likely to smoke. Public-health advocates have also been lenient on the alcoholic beverage industry compared with big tobacco.

These lax attitudes along with free trade and competition rules, which treat alcohol as any other commodity, have undermined effective alcohol control measures, such as increasing taxes or restricting the hours or days of sale. Instead, at country level, ineffective control strategies (warning labels, education in schools) have been adopted. Internationally, resolutions on alcohol control exist but are non-binding and easily flouted.

For alcohol-control measures to be taken more seriously by governments, an international treaty modelled on the Framework Convention on Tobacco Control (FCTC) is needed. One of WHO’s greatest achievements, the FCTC aims to counter the increase in tobacco consumption by making it a legal requirement for countries to introduce certain tobacco-control strategies.

Momentum is already gathering for a Framework Convention on Alcohol Control (FCAC). In the past couple of years the World Medical Association and the American Public Health Association have been among those who have voiced their support for such a move.

From its initial inception, the FCTC took 10 years to become a reality. The road to an FCAC is likely to be similarly long. Next year’s World Health Assembly provides a crucial opportunity for WHO and member states to make those first steps towards a global treaty to reduce alcohol-related harm. ■ The Lancet
Thiazolidinediones, deadly sins, surrogates, and elephants

The heated debate about the safety of thiazolidinediones has made the scientific community and the public uneasy, which is good because complacency is a deadly sin, akin to sloth. This debate was triggered by a meta-analysis of 42, mainly short-term, trials that investigated the effects of rosiglitazone on glycaemic control in more than 27 000 patients, most of whom had type 2 diabetes mellitus.1 The analysis suggested that rosiglitazone increased the relative risk of myocardial infarction or death from cardiac causes by about 50%, although this risk represented an absolute difference of less than 0·15%. Had no substantial outcome study been underway, this meta-analysis would have been an important stimulus to do one; however, one such study is underway.2 The scare caused by the meta-analysis was widely aired in the medical and lay press. Such publicity could have damaged the conduct of the trial and rendered the outcome neutral, which would provide no proof of safety because stopping rosiglitazone would mean both treatment groups would end up on the same treatment. Accordingly, the steering committee and sponsor of this trial had little choice but to publish an interim analysis. Rates of myocardial infarction and death were low and much the same in patients assigned to rosiglitazone or comparator drug, but there were too few events to prove the absence of excess risk, and the incidence of heart failure doubled in the rosiglitazone group. These findings were similar to those of a large trial that compared pioglitazone and placebo.3

In today’s Lancet, Rodrigo Lago and colleagues report a meta-analysis of seven trials of rosiglitazone and pioglitazone in more than 20 000 patients with, or at high risk of developing, diabetes.4 This analysis also noted a doubling in the incidence of heart failure but suggested that neither agent is associated with increased cardiovascular mortality, although the confidence interval cannot exclude a 25% increase.

Lago and colleagues’ report included far fewer trials than the previous meta-analysis of rosiglitazone.1 It is not certain which set of trial selection criteria was least prone to bias. When presented with small trials, journals have a bias against publishing negative results but will often publish positive ones. Any resulting meta-analysis will reflect the same bias.

The outcome measures assessed in a meta-analysis should be chosen with care. Myocardial infarction is not necessarily a definitive outcome measure because only a few patients who have myocardial scars consistent with infarction report any event.1 Avoidance of the pain and worry of myocardial infarction is an appropriate goal of treatment, but subsequent disability and mortality can be measured directly. Ankle swelling is usually not due to heart failure and is certainly not a valid endpoint. Thiazolidinediones can cause fluid retention that might result in peripheral and pulmonary oedema but there is little evidence that such drugs affect cardiac function adversely.5 Patients with oedema caused by thiazolidinediones do not seem to have an unfavourable prognosis,3 although symptoms and signs might only resolve when these drugs are stopped. Why should fluid retention be blamed on the heart rather than the kidneys? The diagnosis of heart failure was generally not robust in studies of thiazolidinediones and much of the adverse prognosis of “diastolic” heart failure may be due to non-cardiovascular comorbidity rather than cardiac dysfunction.7

However, all the meta-analyses consistently fail to spot the elephant in the room. Treatments should be effective rather than merely innocuous. Improved glycaemic control is not a surrogate for effective care of patients who have diabetes, which should be to reduce disability and increase lifespan. Surrogate markers of treatment effects have not done well in many cardiovascular disease areas, including hyperlipidaemia,8 hypertension,9 or heart failure.11 In the RECORD study,7 more than 90% of patients had no important cardiovascular event in 3·75 years of follow-up. Having no symptoms and low rates of events leads to difficulties in designing trials to show real benefits

Retinal maculopathy in diabetes
to patients. However, such difficulty suggests the absence of an important role for that intervention in present medical practice.

Government should bear the brunt of the criticism for the approval of pointless drugs. Patients and clinicians have always emphasised improvement of symptoms, reduction of disability, and delaying of death as much as safety. The regulatory authorities need greater emphasis on ensuring that drugs have effects that are clinically relevant, both in their actions and extent, without stifling innovation in an industry that is valuable to society. Short patent-life contributes to the development of expensive low-efficacy drugs as companies scramble to make money for profit and to invest in further research. Longer patents, akin to that provided by copyright for an author or a song, would encourage long-term investment in high-quality clinical research. Short-term trials with large numbers of patients designed to show statistically significant but clinically spurious differences would be replaced by trials with fewer patients designed to show substantial and clinically relevant long-term benefits. Trials should also show that such benefits wane if treatment is stopped, to ensure that long-term treatment is effective for chronic relatively low morbidity, such as diabetes and hypertension. The main problem with this arena of debate is the use of surrogate markers and surrogate forms of proof, such as meta-analysis. Let us ensure that the sins of our fathers are not meted out to future generations.

*John G F Cleland, Stephen L Atkin
Department of Cardiology, Castle Hill Hospital, University of Hull, Hull HU16 5JQ, UK (JGFC); and Department of Diabetes, Hull York Medical School, Hull, UK (SLA)
j.g.cleland@hull.ac.uk

Patient-important outcomes in diabetes—time for consensus

The epidemic of type 2 diabetes and its resulting complications—cardiovascular, renal, ophthalmological, and neurological—has generated widespread alarm. Despite the alarm, the resultant flurry of investigation, and an interval of 50 years since the introduction of the first oral agent for the treatment of diabetes, we remain uncertain if any antihyperglycaemic drug can favourably affect key patient-important outcomes, including morbidity, mortality, and quality of life. Why this profound ignorance? One key reason is that diabetes trials have focused on the effect of interventions on glucose control (eg, HbA1c) rather than on patient-important outcomes.1

Measurement of HbA1c purportedly captures the effect of therapy on diabetes complications: the lower the HbA1c achieved, the lower the risk of diabetes-related complications. The Diabetes Control and Complications Trial (DCCT) offers the best evidence in support of this putative causal link.2 DCCT established that, in patients with type 1 diabetes, the near-physiological replacement of insulin led to lower HbA1c concentrations and reductions

5 Cleland JG, Coletta AP, Nikitin NP, Clark AL. Clinical trials update from the American College of Cardiology: Darbepoeitin alfa, ASTEROID, UNIVERSE, paediatric carvedilol, UNLOAD and ICELAND. Eur J Heart Fail 2006; 8: 326–29.
in the risk of microvascular complications (retinopathy, neuropathy, and nephropathy) and, with less certainty, reductions in the risk of macrovascular complications. In this context, HbA1c may be a credible surrogate endpoint.

Unfortunately, HbA1c loses its validity as a surrogate marker when patients have a constellation of metabolic abnormalities, when the most common complications are macrovascular, and when the treatments have multiple poorly understood effects. This situation characterises type 2 diabetes.

In the UK Prospective Diabetes Study, insulin and sulfonylureas achieved similar HbA1c reductions as did metformin, but failed to achieve the same reduction in the risk of myocardial infarction and other important diabetes complications. Indeed, sulfonylureas and glitazars might reduce HbA1c and increase cardiovascular risk. Recent evidence that treatment of diabetes with rosiglitazone increases cardiovascular risk has further reduced the credibility of HbA1c as an adequate surrogate.

Available antihyperglycaemic agents have effects that go beyond glycaemic control. Thus revelations of a paradox between lower HbA1c concentrations and increased risk of coronary events should not surprise. Both clinicians and experts must acknowledge that it matters how (ie, with what agent or agents) we achieve glucose-control targets.

Medical therapeutic history is littered with instances in which reliance on surrogate outcomes has provided misleading results. Encainide and flecainide, anti-arrhythmic drugs that suppressed malignant-looking arrhythmias, increased mortality. Oestrogen-replacement therapy favourably affected HDL-cholesterol and carotid plaque, but did not prevent cardiovascular events. Torcetrapib increased HDL-cholesterol and modified vascular morphology, but increased mortality. Despite these lessons, trialists continue to err in relying on surrogate outcomes to substitute for effects on patient-important outcomes.

The direct measurement of patient-important outcomes in diabetes trials remains uncommon. Only one in five randomised trials in diabetes published in top general and specialty journals was powered to measure such outcomes. The future seems even grimmer. Clinical trial registries reveal that only 14% of randomised trials in diabetes will assess patient-important outcomes as primary endpoints. Even large trials miss the opportunity to measure these outcomes: the ADOPT trial followed up over 4000 patients for 4 years but was powered to assess when patients will need to add further therapy to maintain glycaemic control, rather than whether therapy improves patients’ wellbeing, limits their signs and symptoms, or prolongs their lives.

The medical community is increasingly aware of the need to engage patients with chronic conditions in clinical decisions. A conscientious patient with diabetes would like to choose drugs that maximise benefit (reduce complications) and minimise burden (route of administration, need for self-monitoring, cost), side-effects (weight gain), and efficacy failure (hypoglycaemia). But if an informed asymptomatic patient asks whether they will be better off if they follow a regimen that will reduce HbA1c by 0.5%, the lack of high-quality evidence leaves us unable to provide a satisfactory response. Clinicians and policymakers who view quality-of-care criteria that focus on reductions in unsubstantiated surrogate markers should share patients’ well-warranted incredulity.

What is the way forward? Surrogate outcomes allow smaller, shorter, and cheaper trials, provide a faster offering of more choices to patients and clinicians, save research money, and allow new drugs more rapid access to market. The apparent benefits are, however, a mirage and the apparent savings represent false economy. Any savings are quickly overwhelmed by costs associated with potentially ineffective or even harmful (yet heavily advertised) expensive therapies, and the incremental costs of treating the harms these interventions might cause. Patients and society may end up paying dearly for drugs that cause more harm than good.
The medical community should insist that we invest the resources needed to do trials that ascertain the effect of interventions on patient-important outcomes. This policy will prevent the premature dissemination of therapies that ultimately prove harmful, facilitate patients’ participation in decisionmaking, and speed the day when we can confidently offer treatments that will provide long-term benefit to patients with diabetes.

"Victor M Montori, Gunjan Y Gandhi, Gordon H Guyatt
Knowledge and Encounter Research Unit, Division of Endocrinology, Diabetes, Metabolism, Nutrition, and Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN 55905, USA (VMM, GYG); and CLARITY Research Group, Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada (GHG)
Montori.victor@mayo.edu
We declare that we have no conflict of interest.

1 Guyatt G, Montori V, Devereaux PJ, Schunemann H, Bhandari M. Patients at the center: in our practice, and in our use of language. ACP J Club 2004; 140: A11–12.

Pathology of human H5N1 infection: new findings

Human cases of the highly pathogenic avian influenza virus H5N1 were first documented in Hong Kong in 1997. From late 2003, this disease became endemic in Asia, often but not invariably associated with an outbreak in poultry. According to WHO, the total number of cases reported from 12 countries until July 25, 2007, was 319, with a fatality rate of 60%.1 Sporadic cases of person-to-person transmission have been reported and a pandemic outbreak poses a serious threat. The table summarises the results of the few postmortem studies after human H5N1 infection. In today’s Lancet, Jiang Gu and colleagues report new findings from two postmortem studies in people who had H5N1, including a rare case of a pregnant woman and her fetus.1

In birds, H5N1 affects multiple organs. In human beings, H5N1 infection mainly affects the lower respiratory tract, causing diffuse alveolar damage and respiratory failure, by contrast with human influenza infection, which mainly affects the upper respiratory tract. Diarrhoea is also a common presenting feature of H5N1, occurring in up to 70% of patients. Viral RNA was detected in seven of nine faecal samples tested.3 Positive and negative strands of viral RNA in the intestine in case 5 (table) and in Gu and colleagues’ two cases suggest viral replication occurred in the intestine. This finding could have important implications for infection control. The virus was also cultured from cerebrospinal fluid and faecal, throat, and serum specimens from a child who presented with diarrhoea before developing coma.10 Cases 1 and 5, both children, had brain lesions, although the role of H5N1 in these lesions remained unclear. Gu detected viral genomes in the neurons of an adult brain without pathological changes. The infection is interesting. With the development of antibodies in the mother and their transplacental crossing into the fetus, pathological lesions in the fetus may result. Thus data are accumulating in support of extrapulmonary infection at many sites in the body.

In the respiratory tract, the receptor for human-adapted influenza viruses, α2,6-linked sialic acid, is mostly expressed in the upper airways; the cells in the alveoli and terminal bronchiole express α2,3-linked sialic acid, the receptor for avian influenza viruses. Viral replication has been detected in type II pneumocytes in the lung, but also in ciliated and non-ciliated epithelial cells of the trachea in Gu and colleagues’ study, which contrasts with a previous report (case 5). Receptor affinity is believed to be a major factor that prevents efficient person-to-person transmission. Successful infection of the epithelial cells in the trachea by H5N1 virus has two implications. First, other mechanisms that mediate virus entry might exist. Second, the virus might develop mechanisms to overcome respiratory-tract defences. However, Gu and co-workers implied that infection of the epithelial cells in the trachea by H5N1 virus has been detected in type II pneumocytes in the lung, which contrasts with the severe and widespread histopathological changes of diffuse alveolar damage, which is consistent with a late phase of viral eradication in an immunocompetent host.

<table>
<thead>
<tr>
<th>Case: year, place</th>
<th>Sex (age)</th>
<th>Time from onset of disease to death (days)</th>
<th>Main extrapulmonary findings</th>
<th>H5N1</th>
<th>Influenza virus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: 1997, Hong Kong(^*)</td>
<td>Female (13 years)</td>
<td>20</td>
<td>Reactive haemophagocytic syndrome, lymphoid depletion, microglial nodule in cerebral white matter</td>
<td>Negative: bone marrow, brain, heart, intestine, liver, lung, lymph node, spleen, kidney*(^\dagger)</td>
<td>Negative: tissues as listed on left (H5 antigen)</td>
</tr>
<tr>
<td>2: 1997, Hong Kong(^*)</td>
<td>Male (25 years)</td>
<td>30</td>
<td>Reactive haemophagocytic syndrome, lymphoid depletion</td>
<td>Negative: bone marrow, brain, heart, intestine, liver, lung, lymph node, spleen, kidney*(^\dagger)</td>
<td>Negative: tissues as listed on left (H5 antigen)</td>
</tr>
<tr>
<td>3: 2003, Hong Kong(^*)</td>
<td>Male (22 years)</td>
<td>10</td>
<td>Reactive haemophagocytic syndrome, lymphoid depletion</td>
<td>Positive: lung*</td>
<td>Negative: bone marrow, brain, kidney, liver, spleen*</td>
</tr>
<tr>
<td>4: 2004, Thailand(^*)</td>
<td>Male (26 years)</td>
<td>10</td>
<td>Reactive haemophagocytic syndrome, lymphoid depletion</td>
<td>Positive: lung*</td>
<td>Positive: epithelial cells of lung (influenza A nucleoprotein)</td>
</tr>
<tr>
<td>5: 2004, Thailand(^*)</td>
<td>Male (6 years)</td>
<td>17</td>
<td>No reactive haemophagocytic syndrome, small foci of necrosis in brain</td>
<td>Positive: lung, small and large intestine (positive and negative strands of RNA)<em>, spleen (negative strands of RNA)</em></td>
<td>Positive: pneumocytes (influenza A nucleoprotein)</td>
</tr>
<tr>
<td>6–8: 2004, Thailand(^*)</td>
<td>..</td>
<td>..</td>
<td>Examination of lung and spleen only: atypical lymphocytes in spleen</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td>9: 2005, China(^*)</td>
<td>Pregnant female (24 years)</td>
<td>9</td>
<td>Reactive haemophagocytic syndrome, lymphoid depletion</td>
<td>Positive: lung, intestines, heart, spleen, liver, kidney, placenta*</td>
<td>Positive: trachea, pneumocytes, cytotoxic lymphocytes, Hofbauer cells (nucleoprotein and haemagglutinin), lymph node (T cells), small intestine, negative: bronchi, spleen, heart, endothelial cells, hepatocytes, Kupffer cells, kidney, syncytotrophoblasts, circulating mononuclear cells (small intestine)*</td>
</tr>
<tr>
<td>9: 2005, China(^*)</td>
<td>Fetus (4 months)</td>
<td>..</td>
<td>No specific changes</td>
<td>Positive: lung, liver*(^\dagger)</td>
<td>Positive: bronchi, pneumocytes, circulating mononuclear cells, Kupffer cells, heart, endothelial cells, hepatocytes, Kupffer cells, kidney, small intestine*</td>
</tr>
<tr>
<td>10: 2005, China(^*)</td>
<td>Male (35 years)</td>
<td>27</td>
<td>Reactive haemophagocytic syndrome, lymphoid depletion</td>
<td>Positive: lung, trachea, intestines, heart, spleen, liver, kidneys*</td>
<td>Positive: trachea, lymph node (T-cells), brain, small intestine*</td>
</tr>
</tbody>
</table>

All cases had diffuse alveolar damage. For cases 9 and 10, see reference for strand-specific reverse-transcriptase PCR results. *Detected by reverse-transcriptase PCR. **Detected by culture. §Detected by in-situ hybridisation. ¶Formalin-fixed tissue PCR. **Real-time reverse-transcriptase PCR.

Table: Summary of postmortem case-reports on H5N1 by year, location, and distribution of viral genomes and antigens
Gu and colleagues’ successful use of newly developed molecular techniques on paraffin-embedded tissues enables broad use outside level III biological laboratories, and also makes review of previous material possible. These molecular techniques have pitfalls, including cross-contamination, operator dependence, and other technical issues. Correlation with viral culture to confirm productive viral replication is needed and is absent from Gu’s report. Reproduction of these studies, including experimental models, is awaited.

*Wai Fu Ng, Ka Fai To*  
Department of Pathology, Princess Margaret Hospital, Hong Kong, China (WFN); Department of Pathology, Yan Chai Hospital, Hong Kong, China (WFN); and Department of Anatomical and Cellular Pathology, State Key Laboratory in Oncology in South China, Li Ka Shing Institute of Health Science, Chinese University of Hong Kong, Hong Kong, China (KFT)

ngwaufl@cuhk.edu.hk

We declare that we have no conflict of interest.

---

**Supported employment for people with severe mental illness**

All countries can do more to improve the employment of people with severe mental illness, and facilitating access to the competitive labour market offers one way to achieve this goal. In today’s *Lancet,* Tom Burns and colleagues report a randomised trial in six European countries of supported employment (an evidence-based job-placement programme) versus the typical and dominant alternative vocational rehabilitation service available locally for people with severe mental illness. The investigators looked at competitive employment and clinical outcomes. Supported employment programmes assist adults with severe mental illness to enter jobs that meet their personal preferences, thereby achieving social inclusion, relief from poverty, and diminished reliance on governmental welfare assistance.

Over an 18-month intervention period, Burns and colleagues observed that more participants in the supported employment programme (55%) obtained competitive employment compared with participants in traditional services (28%), without increased admissions to hospital for illness relapse. These outcomes closely resemble those of randomised trials in the USA and Canada, confirming an effective way to improve employment prospects of people with severe mental illness across widely differing cultural, health, welfare, and labour-market contexts.

In the USA, most researchers and policymakers describe competitive employment in terms of jobs open to anyone, located in typical business environments, and staffed by workers recruited on the basis of qualifications, not disabilities. In Burns and colleagues’ study, most jobs obtained by study participants were in unskilled or support positions (eg, warehouse or catering work). The primary study outcome was the number of participants who typically obtain unskilled and semi-skilled entry-level positions (eg, warehouse or catering work). The primary study outcome was the number of participants who obtained competitive jobs, paying wages at or near the local minimum and earning an average US$3000–5000 over 10–20 full-time-equivalent weeks per year. These low wages rarely lead to economic independence, do not markedly reduce reliance on governmental income support, and do not lead to career development. Moreover, few competitive jobs acquired by people in supported

employment programmes provide on-the-job training to prepare participants for more sophisticated jobs. We believe that increasing individuals’ career prospects requires formal opportunities for higher education, combined with enhanced on-the-job training, to help people learn new ways to complete essential tasks more efficiently, communicate with others more effectively, organise multiple activities, adjust to changing demands from the job, and learn new work skills relevant to the evolving needs of employers.

Few higher-education institutions around the world accommodate the unique impairments and learning needs of people with severe mental illness. Advanced learning programmes tailored for people with mental illness, such as supported education, have neither been widely tested nor widely adopted in the USA and elsewhere. Perhaps the time has come for studies on the career development of people with severe mental illness. Such studies could aim to restore social inclusion and citizenship rights through higher education and career development.

Every nation faces the challenge of developing human capital, yet people with severe mental illness have not generally been regarded as an asset, even though several studies show a capacity to contribute. People with mental illness strive for self-determination through work just as passionately as people in good health. We believe that developing an empowered, educated, healthy, and skilled labour force could reduce the economic and social marginalisation of people with mental illness. At present we know little about how to apply evidence-based practices in supported employment to countries with developing market economies. Similarly, in more regulated economies than in the USA, such as Australia, Canada, and Europe, we know little about the extent to which established evidence-based practices can be implemented with high fidelity, and whether the active ingredients need further development in local contexts. Despite international differences in labour markets, health and welfare systems, employment outcomes from Burns and colleagues’ study warrant optimism that high rates of competitive employment could be achievable in more regulated economies and lead to increasingly viable career prospects for people with histories of lengthy unemployment, such as those with severe mental illness.

Further progress will inevitably require new international partnerships, funding from a wide variety of sources, different research designs, a long-term focus to track vocational recovery, and inclusive communities prepared to restore equal rights of citizenship and value human strengths over deficits. Equally important will be the need for researchers to produce evidence for immediate use in developing policy and in sponsoring local evidence-based programmes.

Paul B Gold, *Geoff Waghorn*
Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, SC, USA (PBG); and School of Population Health, University of Queensland, and the Queensland Centre for Mental Health Research, Park Centre for Mental Health, Richlands, QLD 4076, Australia (GW) geo7_waghorn@qcmhr.uq.edu.au

We declare that we have no conflict of interest.


Mental health and global movement of people

Migration is a key part of globalisation, and the social, cultural, economic, and political forces of globalisation have substantially changed the determinants and consequences of migration. 170 million people live outside their country of origin, and every year more than 700 million people cross national boundaries. The public-health importance of this massive movement of people is apparent for communicable diseases, and, although less visible, is no less important for mental health.

Although globalisation promises universal economic and social benefits, it leads to increased marginalisation, unemployment, erosion of job security, increased poverty, reduced access to health care and education, and reduced social provision for people who are ill or unemployed. Many factors that lead to permanent and temporary migration are products of globalisation, and are determinants of mental health and illness. Rapidly emerging market economies need cheap labour, which might involve women from rural areas with poor education who have been separated from their family and who have access to few legal protections. High-income countries (ie, those in Europe, North America, and Australia and New Zealand) with declining and ageing populations need immigrants, but are often ambivalent about them when they come. Complex emergencies and human-rights abuses produce large flows of asylum seekers and refugees, mostly into neighbouring low-income countries that have little capacity to receive and to care for them. Poverty fuels the deadly trade of people-trafficking, and is the major engine for undocumented immigration. The decline of rural economies everywhere and rapidly escalating global ecological problems will substantially increase the pressure on people to move.

The encounter between people and cultures is mediated by a global flow of commodities, information, and, increasingly, direct contact between people. Most migration occurs from poor, generally sociocentric (ie, collectivist) cultures to those that are richer and egocentric (ie, individualist). Acculturation, which occurs when different cultural groups are in sustained contact with one another, might lead to distress and dysfunction in some individuals and to tension between cultural groups. The power of institutional or individual racism over the mental health of immigrants must not be ignored. Fragmentation and erosion of identity, the loss associated with displacement from familiar contexts and support networks, the difficulties of settlement, and the pressures on accustomed family structures and relationships can increase vulnerability to mental illness. Sociocentric individuals who migrate from sociocentric societies to those that are egocentric are likely to develop distress—especially if they do not have access to a community of people with similar backgrounds.
An additional source of stress on immigrants is that host societies have generally failed to respond effectively to the reality of ethnic, cultural, and linguistic diversity. The effect of movement of trained and professional individuals and their families around the world cannot be underestimated. The cost of training a doctor or an engineer in a low-income country may be proportionately low, but the brain drain can cost millions of dollars, especially in countries that can ill afford to lose such skills. Therefore the treatment gap in low-income countries will probably widen, and there have been recent calls for developed countries to refund donor countries. The money that migrants send back to their country of origin may compensate, but this is not the full story. The mental health of people who are left behind and who have poor resources, the potential resulting resentment, and the role of acculturation on the health of migrants need urgent wider discussion.

*Dinesh Bhugra, Iraklis Harry Minas

The built environment and health

In a recent British Medical Journal poll, the sanitary revolution that introduced clean water and sewage disposal was voted the most important medical advance since the journal was first published. That result reminds us of the crucial part played by urban planners and engineers in health improvement 150 years ago. As the squalor and decay of that time gave way to improvements in the environment, together with substantial advances in medicine and improved life expectancy, the 20th century saw a corresponding decrease in interest in the effect of the built environment on health. The automobile was pivotal in the planning of communities, with unforeseen results, such as urban sprawl and changes in lifestyles, health, and wellbeing.

The damaging social results of urban planning that ignore lessons from real life have been eloquently described in a seminal study of American cities. Crime, absence of social cohesion, noise, air pollution, and road-traffic accidents are some of the negative characteristics that have persisted or even increased, despite the efforts of well-meaning planners, architects, and transportation experts in past decades. Related to this changed landscape, the causes of disease that were prevalent a century ago have been replaced by more chronic disorders, such as asthma, obesity, and diabetes. Against this background, a collective recognition of the crucial relation between the built environment and health, particularly within the neighbourhood as an essential health setting, is only just beginning to re-emerge (figure). Furthermore, climate change is leading to an urgent focus on sustainable building, which has the potential not only to minimise environmental degradation but also to improve health.

Researchers define the built environment as encompassing all buildings, spaces, and products that are created or modified by people. The built environment affects indoor and outdoor physical environments, social environments, and subsequently health and...
quality of life. It includes urban design, transportation systems, and land-use planning and policies that affect communities in urban, rural, and suburban areas. Causal relations between the built environment and health have been difficult to establish, but there is a growing body of evidence showing the pathways and mechanisms by which the built environment affects health and factors associated with specific aspects of physical and mental health (webfigure).

Access to green open space can increase physical activity and mental wellbeing, because most sustained exercise is incorporated into daily routine activities. Putting health and wellbeing centre stage would result in streets, green spaces, and neighbourhoods that encourage more walking and cycling and opportunities for informal social contact and interaction. Noise and light pollution, which cause stress and inhibit communication, could be addressed through a skilful and balanced application of legislation and planning. Interior environments would play their part both in modification of behaviour (eg, more attractive and prominently positioned stairs) and by reduction of stress (eg, clear directions and good acoustics and natural light).

There is a consensus that until now the consideration of health and wellbeing has had little effect in the creation of the built environment, with socioeconomically disadvantaged communities being worst affected. To make a difference, public-health and built-environment professionals need “to learn from each other how best to address the needs of the communities they serve, to determine what answers each has that the other needs, to create a common language, and to initiate the opportunities to use it.”

Two recent developments might strengthen this collaboration in England. The Strategic Environmental Assessment (SEA) Directive, which came into force across the EU in 2004, requires the likely significant effects on the environment of implementation of spatial plans, including the effects on population and human health, to be considered. Moreover, sustainability appraisals, which are mandatory in England and which incorporate the SEA, need spatial plans to be assessed for their economic, social, and environmental effects, with health being relevant to all three aspects of the assessment. Getting theory into practice has nevertheless been challenging for two reasons—the urgent need to establish greater and more routine collaboration between planners and health professionals, and to strengthen competence to undertake meaningful health assessments. Therefore the Department of Health in collaboration with its partners has recently issued draft guidance to help authorities assess the health impacts of their spatial plans effectively.

The second development is the establishment of Teaching Public Health Networks across England by the Department of Health. The Networks aim to encourage higher and further education as well as mainstream public-health learning across a diverse range of curricula. With strong support from the Department for Innovation, Universities and Skills, the development is already stimulating interdisciplinary collaboration for learning, which should create more interagency research and practice. For example, the University of the West of England (Bristol, UK) is leading the work to integrate health into built-environment curricula (Grant M, University of the West of England, Bristol, UK; personal communication). Furthermore, organisations such as the Commission for Architecture and the Built Environment, the Royal Town Planning...
Institute, and the Royal Institute of British Architects are promoting greater understanding of health effects among the built-environment professions.

We now need a paradigm shift in the way professions work together to translate our growing understanding of the link between the built environment and health and wellbeing into real and effective action.  

We echo the points made by Chris Murray and colleagues.  

1 Ferriman A. BMJ readers choose the “sanitary revolution” as the greatest medical advance since 1840. BMJ 2007; 334: 111.


5 Barton H, Grant M. A health map for the local human habitat. J R Soc Prom Health 2006; 126: 252-53.


Learning from new initiatives in maternal and child health

In fact, our views were accepted as a Comment in The Lancet early in July, and were waiting for a publication slot.

Most low-income countries are making slow progress in improving child and maternal survival—too slow to achieve the fourth and fifth Millennium Development Goals (MDGs) by 2015.  

Countries, donors, and development agencies are responding to the situation by redoubling efforts to stimulate and support country efforts, particularly in Africa. More than 100 countries are working with WHO, UNICEF, and others to improve health workers’ performance and family and community behaviours that are essential for child survival, and to strengthen health systems through the Integrated Management of Childhood Illness strategy.  

UNICEF has gained experience in the Accelerated Child Survival and Development initiative in west African countries, with particular emphasis in Benin, Ghana, Mali, and Senegal, and is now moving ahead to apply the lessons learned to additional countries. The Partnership for Maternal, Newborn & Child Health has selected Burkina Faso, Malawi, and Mozambique as jump-start countries for coordinated action to support country plans for the rapid scale-up of high-impact interventions.

Bilateral agencies from the USA, UK, Canada, Sweden, and several other high-income countries are supporting these initiatives, and extending them to address broader health-system supports and a wider range of countries, including those in south Asia. Under the leadership of the Norwegian Government, a Global Business Plan to accelerate progress towards MDGs 4 and 5 has been formed and will explore performance-based dispersement mechanisms on the basis of documented progress in maternal and child health. Canada is spearheading the Catalytic Initiative to Save a Million Lives, and the UK has recently announced an ambitious International Health Partnership.

We have a unique opportunity to learn from these initiatives. The world needs to know, a few years down the line, whether these different approaches had an effect.
on maternal and child health—and why. Although some of these strategies have been rigorously assessed,1 others have not. Even for initiatives that have been or will be properly assessed, results are often difficult to compare because of the use of different frameworks, methods, and indicators.

Valid information about what is being implemented at country level, how, at what levels of quality, at what cost, and with what outcomes and effects is essential as a basis for sustained improvements in programmes. Existing mechanisms for data collection, such as the Demographic and Health Surveys2 and Multiple Indicator Cluster Survey,3 could be tapped for population-based information on coverage, but additional data for process indicators, contextual factors, and health effects will be needed.

The methodological challenges for the assessment of large-scale interventions are formidable.4 The initiatives described above aim for scaling up at national level, so there will be no control areas within each country that can be used for comparison. Implementation under real conditions is always less than perfect, and thus the actual effect is consistently smaller than that predicted from efficacy studies. The timeline for such assessment is necessarily long: a couple of years are usually needed to reach high coverage, and time must then be allowed for the biological effect of the intervention to take place.

Additionally, methods for measuring mortality are typically retrospective, so that more time is required to detect effects. 5–7 years is not unusual, which is substantially longer than the political timespan of most funding agencies. Furthermore, mortality levels are already declining in many countries, albeit slowly and variably, so simple before-and-after comparisons of national trends can mislead. Contextual factors affecting mortality trends should be measured and taken into account when trying to attribute any progress with the intervention under study.

With the imminent launch of initiatives that use performance-based disbursement of funds, demands for frequent (real-time) monitoring and evaluation will be even greater. Continued support of country activities on the basis of documented evidence of improvements in maternal and child health will bring additional challenges about how to measure outputs, outcomes, and effects in a timely, reliable, and independent manner.

Building political commitment and creating demand for action require valid data that are updated frequently, to show progress and identify bottlenecks that must be addressed. Although country-level assessments can provide important insights, a systematic multicountry effort is needed to supplement and systematise these efforts. We propose that, for the common good, agencies and governments implementing these different initiatives should coordinate their evaluation efforts in different countries, with a common framework and consistent indicators of implementation, costs, coverage, and effect. Results from these coordinated evaluations will then be comparable, and can be used to inform the global community about approaches that can achieve rapid, sustained, and equitable improvement in the health of mothers and children in specific contexts. Capacity in programme monitoring and evaluation can best be developed through collaborative fieldwork; therefore the involvement of scientists and managers from the low-income countries where the assessments are done is essential.

The worst-case scenario is business as usual.3 Without a common framework, 10 years down the road we will have some initiatives that were not evaluated at all and others that were poorly evaluated. Among the few that will be properly assessed, comparability of findings will be hampered by the use of different indicators and incompatible designs. Once again, we will have lost the opportunity to learn from our successes and failures.

4 Cesar G Victora, Robert E Black, Jennifer Bryce

Universidade Federal de Pelotas, Pelotas, RS 96001, Brazil (CGV); and Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA (REB, JB)

cvictora@terra.com.br

We are involved in the evaluation of the IMCI, ACSD, and PMNCH initiatives.

Community workers key to improving Africa’s primary care

In parts of rural Africa, where conflict and neglect have destroyed any remnants of a functioning health system, there is one long-running public-health programme that is not only surviving but thriving—by capitalising on communities’ desires to help themselves. Hannah Brown reports.

Standing at the dark doorway of a large wattle-and-daub hut, Michel Mazogo, a health worker in a remote village of northern Democratic Republic of Congo (DRC), proudly shows off his charge: a makeshift health centre, built by the villagers themselves, to serve ten neighbouring communities.

The building’s cramped interior houses four individual rooms, but little in the way of equipment: a small trolley displaying a few tired surgical instruments sits under a window; a fragile-looking maternity bed doubles as a bike stand in one room; and some empty shelves line the walls of another. Mazogo, who trained for 3 years at technical college for this post, says he has just aspirin, paracetamol, chloroquine, quinine, adrenaline, and iron to distribute to patients who visit. If he writes a prescription for other treatments, or if patients present with symptoms he cannot address, then they must travel 50 km on foot along a dirt track, to obtain the drugs from the regional town of Buta, an hour and a half’s flight north over dense rainforest from DRC’s second city, Kisangani.

Mazogo points out crumbling patches of wall in what he calls the recovery room for new mothers. Termites are slowly eating away at the building, and in time it will collapse completely. What Mazogo wants is some government money for corrugated iron to protect the roof from this inevitable damage. But in DRC, whose infrastructure was largely shattered during the brutal 5-year civil war that came to an end in 2002, financial support for this kind of basic health-system strengthening is little more than a pipe dream.

The difficulties faced by Mazogo’s community echo those of many rural populations in sub-Saharan Africa, where health systems have been decimated by conflict, political neglect, and health-worker migration, while being overrun by increasing burdens of disease. However, thanks to a public-health programme that has been heralded as one of Africa’s most successful, communities in DRC and 15 other African countries have been taking their health into their own hands—and achieving impressive reductions in the burden of common tropical diseases as a result.

What has given communities the opportunity to bypass government-run health structures is an innovative strategy adopted by the African Programme for Onchocerciasis Control (APOC), in which volunteers from each village are trained to deliver treatment, watch for side-effects, and report back to local health workers. The system was designed as a sustainable way of ensuring that distribution of a drug to cure river blindness consistently reached those in need, even in countries affected by war.

But the enthusiasm of communities for helping themselves means this strategy is being extended beyond river blindness. Now, these voluntary workers, known as community drug distributors (CDDs), are taking on vitamin A distribution, deworming, and other simple interventions. By doing so, this army of unpaid health workers, which numbers over 400 000, is helping to supplement and reinforce the traditional health system—and the knock-on effects have been so successful that WHO Director General Margaret Chan has publicly called for more governments to consider this strategy as a way of strengthening primary care in Africa.

For APOC, involving communities in their disease-control efforts was a simple question of necessity, explains the programme’s director, parasitologist Uche Amazigo. Established in 1995 on a shoe-string budget, APOC was set up as a successor to the remarkably successful Onchocerciasis Control Programme (OCP), which since 1974 had been working to stamp out river blindness in 11 west African nations.

The strategy used by OCP was based initially on pesticide spraying to kill the...
onchocerciasis parasite’s vector, the blackfly, which breeds near fast flowing rivers and can travel up to 400 km by air (compared with the mosquito’s 50 km flight capacity). However, in 1987, a new drug called ivermectin, which if given as a single annual dose effectively rids infected individuals of the onchocerciasis microfilariae (precursors of the symptom-causing adult worms), was incorporated into OCP’s disease-control efforts, thereby providing hope of interrupting transmission of the parasites in endemic communities. According to Amazigo, the availability of the new drug treatment raised the possibility that onchocerciasis control might be extended to cover the many other African countries in which spraying of blackfly breeding grounds for vector control was not a realistic option.

An unprecedented pledge by the drug’s manufacturer, Merck and Co, to provide an unlimited supply of the agent free of charge to all those in need for as long as necessary put the onus on WHO to come up with a way to make sure affected communities could get the drug. Scientists at WHO, the World Bank, and a UNDP-sponsored Special Programme for Research and Training in Tropical Diseases (TDR) recruited a team of African scientists, of which Amazigo was one, to work out a solution.

There were some basic requirements for the distribution system: it had to be cheap, sustainable—to break transmission of the parasite, scientists estimated that the drug had to be administered for 16–18 years—and to work in some of the most difficult conditions in Africa. “We asked: how can we deliver this drug for a long period?” recalls Amazigo. “Can the health system do it? No. Can donors support it? No. Can NGOs do it? No. So after meetings and workshops we decided to see if communities could do it. And, if yes, would their performance be better than that of the health system?”

These deliberations led to a strategy termed community-directed treatment, which took community involvement in public health to a level no previous programme had done. Although many other programmes attempt to consult with communities on decisions, their interventions are usually community-based—with health workers leading projects—and imposed instructions on communities.

After some studies in which TDR tested the efficiency of the strategy, and managed to show that distributing drugs via community volunteers was in fact a more efficient mechanism than when the traditional health system took control, the newly established APOC adopted the strategy. It now runs 108 projects that span 16 endemic countries in sub-Saharan Africa and has maintained OCP’s achievements in ten of 11 west African nations (the exception is Sierra Leone, where conflict has disrupted surveillance, allowing the return of the blackfly). “What the community has done is to share the cost of ivermectin distribution with us and with the health system. Imagine asking health staff in countries to treat 41 million people every year just for oncho—that would not be feasible—but it is now being done because communities are doing it”, says Amazigo.

Mass administration of the drug ivermectin with this community-directed strategy has, according to a recent health impact assessment done by a group from Erasmus University Medical Centre in Rotterdam, reduced the prevalence of infection by about 73% compared with pre-APOC levels, and reductions in the symptoms of blindness and the disabling unrelenting itching were even greater. However, Amazigo asserts that an additional achievement of the programme has been to build up a network of community volunteers who are trained and primed to deliver drugs, help educate their peers about health issues, and participate actively in other disease control measures. In effect, APOC’s projects have provided a mechanism through which primary health care in general can be improved.

The results are already showing. In Mazogo’s village, a group of around 20 or so newly trained community distributors are crammed into a round mud hut, balancing on boxes, stools, and benches, and earnestly clutching exercise books while staring intently at a large blackboard propped up on a chair against the back wall of the
hut. Listening to instructions, they are taught how to add vitamin A supplements to their usual distribution of ivermectin and how to add this information into record books, which are returned to local health centres from where APOC can keep track of who has received what.

The group has amassed from surrounding villages up to 15 km away, travelling to the training session on foot the night before. Although vitamin A was added in this region only last year, coverage is already at amazingly high levels: 92% of children under 5 now get the supplement. There are additional plans for distributors to help dole out condoms, the deworming drug mebendazole, and to help look for signs of tuberculosis among their communities. Treatments for lymphatic filariasis and schistosomiasis may soon be added to the CDDs list of tasks as well. These are activities that simply were not being done through the conventional health system, says Amazigo.

Virgile Kikaya, a public health doctor who supervises onchocerciasis activities from DRC’s capital Kinshasa, says establishing an effective public-health programme in his country’s remote areas has been very tough. “The health system is very weak, we do not have roads, and the government has many difficulties in supporting health system strengthening in terms of financial contributions. But people want to work and so the community-directed system works”, he says. “People were saying to us ‘how are you going to do it during the conflict?’ But we said we could try. And through strong partnerships we have made it work.”

Now that the programme has been running for 5 years, it is starting to produce reciprocal benefits for the health system. Communities that were completely out of reach of the health system are now getting treated. The records collected by CDDs have provided valuable population data in the absence of a census. And Kikaya says the need to watch for drug-related side-effects has strengthened the links between communities and health workers. “We try to reinforce the health structure in very remote areas so even if there are side-effects, there is the capacity to deal with them”, he says. What is more, the money APOC has invested in computers, transport vehicles, advocacy, and research has provided skills for ministry of health staff that would not have existed.

But, there are some challenges with the CDD system. These workers cannot do everything the health system should: some drugs require complicated dosing calculations, and proper technical training for administration and follow-up. And, according to Kikaya, although the motivation of CDDs to learn and take control of drug distribution is generally strong, the system can be vulnerable when other health programmes offer incentives that seem more attractive than working for free. “We have problems with immunisation, malaria, and AIDS programmes because they come with so much money that the distributors get poached”, he explains.

However, the biggest issue is yet to come. APOC’s mandate is due to run out in 2015, and unless the ministries of health in project areas step in to replace the funding for CDD training by this date, the network of community distributors—which number over 400,000 in all APOC countries—is at risk. Ensuring a smooth transition is what Amazigo believes must happen to ensure that APOC’s achievements in onchocerciasis control are maintained, and also that governments reap the biggest benefits from the CDD network. “The challenge is how to get governments to provide regular financial support to the programme to maintain their own achievements so the flies will have no parasite and the free areas will remain free?” she asks.

There is a positive example. Uganda, which was one of the first countries to receive APOC support in the mid-1990s, has already made the successful transition from APOC to government support, while maintaining good onchocerciasis control. In fact, the ministry of health was so impressed with the results of the community-directed strategy as a cost-effective way of delivering primary care that it has reorganised its district health services around the model, and CDDs are now being used to deliver home malaria treatment, bednets, and several other drugs, in addition to ivermectin.

But because central budgetary funds are tight, training sessions—like the one APOC supported in Mazigo’s village—where the volunteers not only learn the necessary information to carry out their tasks, but also receive feedback on the coverage rates they have achieved in their communities and get the opportunity to converse with others doing a similar role, are lapsing. And because it is this feedback that helps keep the distributors motivated to do the work for free year after year, Uganda could be storing up problems for the future. “Community-directed treatment needs patience, time, and commitment”, says Amazigo. “And if you do not have it, implementing the process is difficult. All the communities want is for you to come and train their own chosen people and they will do the job that the health system should do.”

Hannah Brown
Doctors to pay for patients’ medicine in Germany

Physicians’ groups in Berlin have slammed an aspect of the government’s health reform which has left around 20% of doctors having to pay for their patients’ medicine out of their own pocket, with some also facing fines for their prescribing practices. Rob Hyde reports.

Health legislation forcing 20% of Berlin doctors to pay for their patients’ medicine has been slammed by a leading German medical group. The Association of Statutory Physicians (KV) in Berlin says 1200 of the city’s 6200 general practitioners now face an average fine of €90 (£62) for spending above their budget on prescriptions between January and April, 2007.

Spokesperson Annette Kurth revealed that one Berlin doctor had even received a €2700 (£1880) fine. She said that the situation for the city’s doctors is appalling. “They are being punished for doing their job properly. It has created a real climate of fear amongst doctors. What I hate is that doctors are now having to justify themselves for having simply made sure patients get the medicine they need.”

Cost-cutting is a key part of the government’s continuing health reform plans. The openly declared aim of the Pharmaceutical Prescribing Efficiency Act (Arzneimittel-Versorgungs-Wirtschafts-Gesetz) in May, 2006, was therefore to reduce expenditure on drugs for patients. By January, 2007, however, the new Bonus-Penalty Ruling (Bonus-Malus-Regelung) placed even more pressure on doctors to spend less on medicine.

The Bonus-Penalty Ruling drew on the daily defined dosage (DDD) notion pioneered by WHO. Though WHO originally introduced DDD solely as a means to compare prescription demands in different countries, the Bonus-Penalty Ruling takes the notion of a DDD and attaches a fixed daily budget to it.

As a result doctors in Germany are now allowed to prescribe only a set daily dosage at a set daily rate. The final part of the ruling, however, is that doctors prescribing over 10% above the set rate, are fined. “What is really wrong is that even doctors who are being very, very economical with their budget are being affected”, says Kurth.

Under terms of the Bonus Penalty Ruling, set daily rates apply for drugs used to treat high blood pressure, depression, migraines, prostate illnesses, and osteoporosis. Drugs, for example, used to treat depression, migraines, and blood pressure are only allowed to cost 37 cents (around 25 pence) per day.

The Berlin Ministry for Health, Environment and Consumer Protection, says regulating costs for these drugs, and those to treat prostate illnesses and osteoporosis, is a necessary reaction. Spokesperson Regina Kneiding said extremely high levels of spending on medicines before the Bonus-Penalty System, combined with massive price rises within the pharmaceutical sector, meant the government had been forced to take action. “Before the ruling there was simply exorbitant spending on health care because of rocketing costs for pharmaceutical products. With the ruling there is no question of quality being compromised here, it is simply a necessary measure to try and control exorbitant spending.”

“Part of this whole development is also encouraging preventive treatment. The facts are simply that Germans tend to take more medicine than in other countries, and we are just ensuring that there is more control in the system.”

The potential benefit of the legislation for doctors is that if they spend less than their budget, their regional KV will award them with a form of “credit” allowing them to spend more on prescribing drugs thereafter. For this to happen all doctors within one regional KV must together spend less than a collective budget. If, however, this collective budget is exceeded, the individual doctors who spent less than their individual budget, go unrewarded.

Although drafted as national legislation, in eight of Germany’s 16 federal states the regional KVs and health insurance providers (Krankenkassen) have worked out an alternative agreement. A few other states have negotiated a regional amendment to the Bonus-Penalty Ruling, but the states of Hessen, Mecklenburg-Vorpommern, and Berlin remain wholly bound by it.

“The ruling is bad for Berlin because people are simply not as healthy here as in areas such as Baden-Württemberg or Bavaria”, says Kurth. “More people have chronic illnesses requiring highly-expensive treatment. One in five people with HIV [in Germany] comes from Berlin and receives treatment here.”

Berlin and Brandenburg are the only states so far whose KVs have received bills for over-prescribing during the first quarter of 2007. In Brandenburg, 10% of doctors face fines.
In his biography of Samuel Johnson, John Wain includes a memorable quotation about Johnson’s feelings on completing A Dictionary of the English Language, published in 1755: “I have protracted my work till most of those whom I wish to please have sunk into the grave, and success and miscarriage are empty sounds. I therefore dismiss it with frigid tranquillity, having little fear or hope from censure or from praise.” Although clearly a labour of love, and aided by e-mail and other means of communication that were not available to Johnson, the editors of the Dictionary of Medical Biography surely must have experienced similar feelings of intellectual burnout as this remarkable work went to press.

The raw statistics alone are staggering: the lives of doctors stretching over three millennia; five volumes; an international board of 22 medical historians; 379 contributors; and 1143 entries. In addition, there are six extended essays on the evolution of medical tradition in different parts of the world and separate lists of individuals by country, fields of activity, and dates. At least the editors, William and Helen Bynum, can rest assured that nobody is likely to take on a similar task in a hurry.

As emphasised in the Introduction, the study of medicine’s past has changed dramatically in recent years; the older “Great Doctors” approach has been replaced by a broader vision of health and disease set against changing environments and social conditions. In short, the discipline has become the “Social History of Medicine”. This change in outlook is certainly reflected in the Dictionary of Medical Biography which, as well as a conventional list of the good and the great (and a few of the not so good), includes those who have contributed to the arts as well as fields ranging from acupuncture and alchemy, through craniology and Daoism, to mesmerism, sexology, Tibetan medicine, and women’s rights. All this makes for hours of fascinating browsing, interrupted only occasionally by finding that one of one’s great heroes, in my case the 18th-century polymath Thomas Young, has missed the boat. And there is no shortage of surprises. The inclusion of Maurice Pappworth, that chronic thorn in the flesh of the UK’s Royal College of Physicians, may cause some uneasy rustlings in Regent’s Park, although his writings on human experimentation undoubtedly influenced the development of improved guidelines for the ethics of medical research. In the latter context, some readers may find the inclusion of a group of Nazi doctors rather disturbing, even though the name of Josef Mengele, who committed such ghastly atrocities in Auschwitz in the name of eugenics, does not appear.

From years of traumatic experience trying to advise the editors of The Dictionary of National Biography about names to be included in my own field, I am well aware of the difficulties involved. Considering the much greater scope and timescale of the Dictionary of Medical Biography, this must have been a major headache. Overall, however, the balance is excellent and it is only in the choice of entries for those who lived in the recent past that there is more scope for debate.

Perhaps it is not surprising that medical historians often seem to be uneasy when they survey the 20th century; there may simply not have been enough time to put the life and work of more recent figures into its historical perspective, a problem intensified by the huge expansion in the number and activities of those involved. For example, it must have been very difficult to decide whether the names of those who trained in medicine but spent most of their lives in scientific research are better placed in a dictionary of science or medical biography, or both. Otto Warburg and Hans Krebs, both Nobel Prize winners, were among the founders of modern biochemistry. They were both medically qualified, yet Warburg’s name is included whereas Krebs’ is not, despite the fact that he discovered the ornithine cycle of urea synthesis at the same time as he had full clinical responsibility for a ward of 22 patients in Freiburg.

The increasing amalgamation of medicine and science in the 20th century raises many related problems for medical biographers, particularly if their aim is to restrict their coverage to practising doctors. Since the introductory essay entitled Western Medical Tradition points out that genetics had an increasingly important role in medical research and practice in the second half of the 20th century, the reader might wonder how this specialty developed and who were its most important figures. Yet under this subject heading there are only two names: Archibald Garrod, the undisputed father of medical genetics, and Cornelia De Lange, a Dutch paediatrician who, apparently, gave her name to the Brachmann-De Lange syndrome. Although it is true that many of those who developed human genetics in the first half of the 20th century—J B S Haldane and Ronald Fisher, for example—were not medical scientists, the evolution of medical genetics relied on the work of distinguished medical men like Lionel...
In brief

Book  Vaccine maker

Maurice Hilleman was intimidating, driven, and gruff. He was also a loving family man and the inventor of nine vaccines that save millions of lives every year. Paul Offit interviewed this little-known hero during the last 6 months of his life, when Hilleman was dying of lung cancer.

Few people have heard of the man who created vaccines against measles, mumps, rubella, chickenpox, hepatitis A and B, pneumococcus, meningococcus, and Haemophilus influenzae. Offit suggests some reasons why. For one, Hilleman worked in industry; for another, he was no self-promoter.

He grew up on the harsh eastern plains of Montana, helping out on the family farm. Rebell ing against his father’s strict Lutheran faith, Hilleman rejected the church and embraced Darwin, and science. After his degree in microbiology, he went into industry. As he told Offit: “I came off a farm. We had to do marketing. We had to do sales. I wanted to do something. I wanted to make things!”

Hilleman went on to Squibb, where he learned to mass-produce influenza vaccine, and then to the Walter Reed Army Institute. He next brought his “committee-of-one” approach to Merck, where he spent the rest of his career. He liked to keep shrunken heads of employees he’d fired (made by his daughters from dried apples) behind his desk, and inspired fear—and fierce loyalty—in his staff.

This vivid portrait of a unique man is interwoven with a thorough account of the science and politics of vaccination. Offit makes it clear why Hilleman couldn’t have accomplished what he did in today’s regulatory environment, and why we are so fortunate that he was who he was, and did what he did when he did it.

Anne Harding
anne_harding@yahoo.com

Book  Packaging disease

Most people know what the pink ribbon symbolises, but what of the blue ribbon? Its status reflects the low-profile of prostate cancer. In Cancer Activism, Karen Kedrowski and Marilyn Stine Sarow analyse the breast and prostate cancer movements in the USA. They find that, by maintaining a high profile in the media, using advocates in Congress, corporate sponsorship, and social marketing, the breast cancer movement has achieved great success. By contrast, prostate cancer came later to the advocacy game and is smaller and quieter, leading to one of the book’s paradoxes: men who can’t find their voice while women stalk the corridors of power. Cancer Activism highlights a cynical race for attention and money, one the authors lament affects many other disease movements.

Katherine Nightingale
kathnightingale@yahoo.com
### Ten most wanted

**July, 2007**

1. **White-collar worker’s brain** *(Clinical Picture, July 21)*

2. **Cannabis use and risk of psychosis** *(Comment, July 28)*

3. **Cannabis and psychosis** *(Articles, July 28)*

4. **HIV entry inhibitors** *(New Drug Class, July 7)*

5. **Inflammatory bowel disease** *(Series, May 12)*

6. **HPV vaccine** *(Articles, June 30)*

7. **VALIDD study on valsartan** *(Articles, June 23)*

8. **HDL cholesterol and atherosclerosis** *(Comment, July 14)*

9. **Meningococcus** *(Seminar June 30)*

10. **Psychosis and cannabis** *(Editorial, July 28)*


### Lifeline

Abbas Ali Mansour is assistant professor of medicine at the Department of Medicine, Basrah College of Medicine, Iraq. His research specialises in diabetes, metabolism, and endocrine disorders. He has taught postgraduate and undergraduate students in Basrah, southern Iraq, since 1993.

**What has been the greatest achievement of your career?**
Studying the epidemiology of diabetes in southern Iraq.

**And the greatest embarrassment?**
Absence of soluble insulin in a major hospital while we have two patients with diabetes ketoacidosis.

**What do you think is the most over-hyped field of medicine at the moment?**
HIV/AIDS.

**And the most neglected?**
Diabetes.

**Which research study has had most effect on your work?**
The UK Prospective Diabetes Study, which revolutionised the management of diabetes.

**Who inspires you?**
My wife: she has stayed at home helping me for 16 years.

**What apart from your family is the passion of your life?**
My country, Iraq.

**Who was your most influential teacher, and why?**
Ihssan Al-Shama, who steadfastly attended the hospital for 15 years.

**What would be your advice to a newly qualified doctor?**
Start to build yourself gradually, do not forget your teachers.

**How do you relax?**
By forgetting the political problems of Iraq.

**What items do you always carry with you?**
A digital camera.

**What is your greatest fear?**
Civil war in Iraq.

**What are you currently reading?**
Texts on screening for diabetes.

**What was your first experiment as a child?**
Swimming.

**What keeps you awake at night?**
Erratic electricity.

**Which would you choose, money or power?**
Power.

**If you knew you had a week to live, how would you live those days?**
Helping patients with diabetes.
Marguerite M Vogt

Pioneering virologist. Born on Feb 19, 1913, in Berlin, Germany, she died on July 6, 2007, in San Diego, CA, USA, aged 94 years.

Marguerite Vogt was only 14 years old when she published her first scientific paper, using the fly to study developmental genetics in her native Germany. The precociousness was the result of some direction from her parents, both prominent neuroscientists, and Vogt went on to earn a medical degree from the University of Berlin in 1937.

It was only after emigrating to the USA after World War II, however, that Vogt switched to the virology work for which she would become known. She initially worked with Max Delbruck, but in the early 1950s moved to Renato Dulbecco’s laboratory at the California Institute of Technology, Pasadena, CA, USA. She and Dulbecco worked together studying tissue culture and cell biology, and created a way to grow the poliovirus and purify it using the plaque assay. “It was a fantastic combination”, Dulbecco said, “and at the time, these were fantastic results”. The plaque assay “made virology into a quantitative science in the sense that you could assess the infectivity of viruses without having to go through animals”, said Walter Eckhart, who joined Dulbecco’s laboratory as a postdoctoral student in the 1960s.

By the time Dulbecco and Vogt moved to the Salk Institute in La Jolla, CA, USA, in 1962, they had begun studying the role of viruses in cancer, starting with the recently discovered polyomavirus. At the time, the only way to test the ability of viruses to transform cells into cancerous cells was with serial dilutions in animals. “When I started with the polyomavirus, we needed an assay method for the virus”, Dulbecco said. “The virus can do two things—kill cells and transform them. The killing effect could be easily assayed. What was difficult was assaying the transforming effect. In general we used mouse cells, but these are killed. Marguerite found that if we used hamster cells, they are not killed, but are transformed.”

Dulbecco would use molecular biology techniques to move virology forward, and would share the 1975 Nobel Prize in Physiology or Medicine for work on the interaction between tumour viruses and the genetic material of the cell. “But tissue culture techniques were a prerequisite for that”, said Eckhart, now a Salk professor. “She and Dulbecco used them to figure out what was going on at the cellular level.” Vogt went on to study senescence and telomeres, publishing her last paper in 1998 but remaining active in her Salk laboratory until just a few years ago.

Lee Hartwell, who shared the 2001 Nobel Prize in Physiology or Medicine, joined Dulbecco’s laboratory as a postdoctoral student as he and Vogt were joining the Salk. The group published two papers based on the discovery that polyomavirus induces the synthesis of cellular DNA in stationary cells. “People expected it to be making viral DNA, but it was turning on the cell cycle, probably to create a milieu to replicate its own DNA”, said Hartwell, now President and Director of the Fred Hutchinson Cancer Research Center, Seattle. “It was a very exciting finding.”

Vogt “was just this unusual person, just so passionate about science, and enthusiastic”, Hartwell said. “Every time a new postdoc would come to the lab, she would sort of attach herself to them and pick up whatever new techniques they could bring. Given the lab was so small, we worked quite closely together. She taught me all about cell culture. We used to meet in the morning and have crumpets and tea and just talk about science.”

Vogt was also known for throwing large parties at her home. She “played piano in a magnificent way”, Dulbecco said. Despite her strong links to so many Nobel Prize-winning scientists, Vogt’s work went largely unrecognised outside of her laboratory. When others asked whether she was frustrated by that, she would say “No, that just gets in the way”, said Susan Forsburg, who met Vogt in 1993 when she joined the Salk faculty. Colleagues remember a time she ran out of the Salk library, smiling and waving a journal article. She’d been scooped, but instead of being upset, she said “Isn’t it wonderful? They got the answer, we can move ahead now.” She was just that generous, said Forsburg, now a professor at the University of Southern California in Los Angeles. “The process of doing science was paramount.” Vogt never married and leaves no survivors. “Science was my milk”, she told The New York Times in 2001, saying that she felt that she had to choose between being a scientist and being a wife and mother.

Ivan Oransky
ivan-oransky@erols.com
Efficacy of ivermectin against *Onchocerca volvulus* in Ghana

In their Article on possible resistance to ivermectin by *Onchocerca volvulus* in Ghana (June 16, p 212),1 Mike Osei-Atweneboana and colleagues note that the typical microfilaricidal effect continued after treatment with ivermectin, but suggest that a “rapid repopulation” of skin with microfilariae was likely owing to a resistance mechanism developed by female adult *O volvulus* worms.

Such a resistance mechanism would have to be different from any previously described in helminths. We note, however, the earlier reports from Ghana of “suboptimal responders” in patients with onchocerciasis and the presence of B-tubulin-associated benzimidazole resistance in lymphatic filariasis. In neither situation was resistance conclusively documented.

Substantially more data are needed to substantiate ivermectin resistance. Moreover, a more parsimonious explanation is inadequate treatment coverage. Taking ivermectin on an annual basis provides no prophylactic benefit, and studies indicate that people thus treated begin infecting vector blackflies only 4–6 months later. Therefore, annual ivermectin treatments provide only a 4–6-month window of interrupted transmission even when community coverage is high, which was not the case here. Thus, the people being skin-snipped were living in an environment where active transmission was going on for 6–8 months per year, during which young worms were constantly being reintroduced into the population.

Osei-Atweneboana and colleagues might simply have measured the reinfection rate (not resistance), and the rapid repopulation of the skin might have represented the progeny of newly maturing worms, rather than the resumption of microfilarial production by ivermectin-treated worms. If our hypothesis is correct, the prevalence of the “resistant” phenotype should decline if the frequency of ivermectin, and level of ivermectin coverage, were increased.

We declare that we have no conflict of interest.

Ed Copp, *Frank Richards, Patrick Lammie, Mark Eberhard
fxr1@cdc.gov
Auburn University, Auburn, MS, USA (EC); The Carter Center, Atlanta, GA, USA (PR); and Centers for Disease Control and Prevention, Atlanta, GA (PL, ME)


Mike Osei-Atweneboana and colleagues’ conclusion3 has serious implications for the Mectizan Donation Program. It might be correct, but the report has some weaknesses.

Many of the data come from communities in which ivermectin coverage levels are only 50%. Moreover, the study is based on a parasitological technique of limited reliability, especially when parasite loads are very low, as in this case. The data are not expressed as worm counts per mg skin, and since the amount of skin taken with corneoscleral punches can vary substantially, this could have significantly affected the results. We cannot determine from the numerical data presented whether or not the statistical analyses control for variations due to technique. Additionally, there is no comparison with pretreatment levels and no analysis of changes within individuals over time.

There are at least two phases in the action of ivermectin: an initial phase of direct parasite killing, which involves mechanisms where resistance has been described in intestinal nematodes, and a second phase of recrudescence, the mechanisms of which could involve host factors as well as parasite factors, including parasite-specific immune responses. The data in this paper indicate that the first phase is normal in these patients, suggesting a lack of ivermectin resistance as it is currently understood. The second phase is the one that Osei-Atweneboana and colleagues seem to believe is compromised.

A definitive finding of *Onchocerca volvulus* resistance to ivermectin will have profound effects on control programmes in Africa and Latin America. Changes, although costly, could be made if necessary. First, however, the existence of true drug resistance must be confirmed and the resulting effects better understood.

I declare that I have no conflict of interest.

Charles D Mackenzie
mackenz8@msu.edu
Filarial Diseases Laboratory, Michigan State University, East Lansing, MI 48824, USA


Mike Osei-Atweneboana and colleagues conclude that ivermectin-resistant parasite populations are emerging in some areas. We are not convinced that their results provide conclusive evidence.

Their study shows that ivermectin retains its full microfilaricidal effectiveness, but Osei-Atweneboana and...
colleagues base their conclusion on the microfilarial repopulation dynamics after treatment. In four villages, repopulation was faster than in five other villages, and Osei-Atweneboana and colleagues took this as “evidence that ivermectin resistance is developing”. However, comparable repopulation dynamics have also been seen in situations where incomplete control led to renewed transmission and the introduction of young worms into an ageing worm population.

The critical question is why these four villages had relatively high infection levels after 12–17 treatment rounds. One possible explanation is insufficient treatment coverage. Reports available to us are incomplete but suggest that treatment coverage has been highly variable, sometimes as low as 20–40%, and with no treatment reported for some years. Osei-Atweneboana and colleagues even report on a hyperendemic village that never received treatment. Such untreated villages, and the untreated members of villages with poor treatment coverage, will contribute to transmission and infections with new onchocercal worms, also in neighbouring villages and among those who received treatment.

We believe that inadequate coverage provides a plausible explanation for the observations. We do not dismiss the possibility of resistance and the African Programme for Onchocerciasis Control will coordinate further investigations. Furthermore, WHO is working with partners towards the establishment of comprehensive systems for monitoring drug use and efficacy in large-scale anthelmintic treatment programmes.

We declare that we have no conflict of interest.

*Jan H F Remme, Uche Amazigo, Dirk Engels, Andriamahefazafy Barryson, Laurent Yameogo Remmelj@who.int

UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, World Health Organization, 20 Avenue Appia, 1211 Geneva, Switzerland (DE); and Other Tropical Diseases, Regional Office for Africa, World Health Organization, Brazzaville, Congo (AB)


**Authors’ reply**

Ed Cupp and colleagues suggest that an ivermectin resistance mechanism involving reduced suppression of reproduction by adult female *Onchocerca volvulus* would be unique. Although ivermectin resistance in *trichostongylids* of livestock is manifested by the survival of the parasitic stages, including adult worms, the action of ivermectin against adult filariae is unique in that ivermectin mainly suppresses reproduction rather than killing adult filariae. Nevertheless, it is of interest that *Caenorhabditis elegans unc-9* mutants share phenotypes of ivermectin resistance and an egg-laying defect.

It has been suggested that low treatment coverage might explain the observed rapid skin repopulation with microfilariae, and Charles Mackenzie and Jan Remme and colleagues suggest treatment coverage of less than 50%. In fact, the average treatment coverage in the 5 years before the study, in the communities that showed poor parasitological responses, were: Jagbenendo 73.5%, New Longoro and Kyingakrom 66.5% each, and Wia 68.7%. These levels of coverage are in excess of the target of 65% set out by the African Programme for Onchocerciasis Control. Furthermore, the almost complete removal of skin microfilariae by 30 days after treatment confirms that treatment coverage was 100% for all patients in our study. A low treatment coverage cannot explain the rapid repopulation of skin microfilariae in these patients in the communities that showed poor parasitological response to ivermectin.

Remme and colleagues point to the ivermectin-naive community and suggest that incoming infections could cause newly maturing adult worms to rapidly repopulate the skin with microfilariae. Ivermectin-naive communities are rare in Ghana and the naive community found in this study was in the East Gonja district. This area is remote from Kyingakrom and New Longoro (Kintampo district), two of the communities in which skin repopulation with microfilariae was alarming.

Cupp and colleagues also suggest that the rapid repopulation of the skin with microfilariae could be due to newly maturing adult parasites which had established in the 6–8 months when ivermectin might not fully suppress parasite reproduction. This explanation seems unlikely for several reasons. The infective larvae take at least 1 year to become reproductively active adults. At 30 days after treatment, virtually all microfilariae were removed from the skin. Yet by day 90 after treatment there were already significant differences between the communities that responded suboptimally and the ivermectin-naive community or the previously treated communities that responded as expected. This would leave only a 60-day window for newly maturing adult parasites to begin microfilaria production, while not being affected at the time of treatment. Furthermore, in the ivermectin-naive community where transmission could be expected to be highest, the percentage repopulation in comparison with pretreatment microfilaria levels (2.89%) was less (as were absolute microfilaria counts) than in the communities that showed the suboptimal responses (as high as 21.12% in Kyingakrom).

Mackenzie also suggests that variability in microfilaria estimates in skin snips could explain the differences between the communities that responded poorly and those that responded as expected. Each community had between 12 and 68 study participants, with two skin snips taken at each sampling time. Individuals
were followed up longitudinally, with the data being presented as geometric means for each community. Any sampling variability would be random and be accounted for in the statistical analyses. Comparisons were made with pretreatment microfilaria counts and analyses were done on individuals over time. Furthermore, in a subsequent unpublished study on these communities, adult worms, embryograms, and larval stages in the vector have been examined and the conclusions from these additional results are consistent with those reported.

Our results cannot be explained by the alternative hypotheses suggested, which strengthens the concern that ivermectin resistance, manifested as an increase in the rate of microfilaria repopulation by adult worms, is developing. As suggested by many of the correspondents, this suggestion warrants increased monitoring for ivermectin resistance and an acceleration of efforts to develop better tools for such monitoring and for new means of onchocerciasis control.

We declare that we have no conflict of interest.

**Coeliac disease and lymphocytic hypophysitis**

Ami Schattner and Taiba Zornitzki describe a man with anaemia, arthralgia, and gastrointestinal involvement (June 30, p 2214). Many ancillary tests were done, resulting in an eventual diagnosis of corticotropin deficiency possibly caused by lymphocytic hypophysitis. One test that was not done was for coeliac disease. This disease is very common and has protean manifestations matching most findings in Schattner and Zornitzki’s patient. Steroids are also used in its treatment, and the patient’s recovery might therefore be due to their effect on coeliac disease, which can occur with lymphocytic hypophysitis. Finally, as well as heeding Occam’s razor, “When you hear hoofbeats, don’t think zebras” should also be kept in mind.

I declare that I have no conflict of interest.

**Weekitt Kittisupamongkol**

weekitti@gmail.com

Surin Hospital, Surin 32000, Thailand

1 Schattner A, Zornitzki T. When the whole-body scan shows no abnormality. Lancet 2007; 369: 2214.
3 Ciclitira PJ, King AL, Fraser JS. AGA technical review on celiac sprue. Gastroenterology 2001; 120: 3526–40.

**Authors’ reply**

We applaud the brilliant suggestion by Weekitt Kittisupamongkol that our patient’s isolated corticotropin deficiency could have been due to associated unrecognised coeliac disease. As many as 3–8% of blood donors in Israel have positive serology for coeliac disease, and when they undergo
intestinal biopsies a prevalence of at least one in 157 is found.1

With such a high prevalence in Israel and other countries, one should be wary of associations that might be no more than coincidental. However, as an autoimmune disease, coeliac disease is often associated with other endocrine and non-endocrine autoimmune disorders. Only the associations with type 1 diabetes and with thyroid disease are well established by large controlled studies (about 6% and 5% of patients with coeliac disease, respectively), whereas concomitant occurrence of Addison’s disease or autoimmune hypophysitis in coeliac disease is limited to case reports.2

Impaired growth hormone was identified in five of seven children with coeliac disease whose growth did not catch up after more than a year of gluten-free diet; in four of them, high titres of antipituitary antibodies were discovered and the MRI was normal.1 Low titres of these autoantibodies were also present in three of 125 children with coeliac disease and normal growth after treatment. Thus, clinically significant autoimmune hypophysitis seems to be linked to coeliac disease. Although we are not aware of any case of central cortisol deficiency associated with coeliac disease, we concur that it should have been examined in our patient.

Notwithstanding this possibility, our patient’s work-up did include endoscopy and biopsies of the duodenum; these were normal and therefore rule out coeliac disease. The patient also fully recovered on a gluten-rich diet. Additionally, we have now done a search for serological markers. A stored, frozen sample of the patient’s serum was negative for antibodies against gliadin and endomysial and tissue transglutaminases.

In our medical school we often use the aphorism “When you hear hoofbeats think horses, not zebras”. However, although coeliac disease might perhaps be regarded as a “horse”, coeliac disease presenting as adrenal insufficiency secondary to isolated corticotropin deficiency caused by autoimmune lymphocytic hypophysitis is not just a “zebra” but one with shining pink stripes.

We declare that we have no conflict of interest.

*Ami Schattner, Taiba Zornitzki
amimd@clalit.org.il
Hadassah Medical School, Jerusalem, Israel


Dual inhibition of the renin system by aliskiren and valsartan

We would like to clarify some aspects of our study (July 21, p 221) in light of the selective reporting of our findings in the Comment by Willem Birkenhager and Jan Staessen.2

Although drugs that inhibit the renin system are known to lead to modest increases in potassium concentrations,3 the Comment gives an exaggerated picture of the risks associated with combined aliskiren/valsartan. As we report, most potassium increases to above 5·5 mmol/L were transient, with concentrations returning to normal at study end without any interruptions to treatment. Potassium concentrations of 6·0 mmol/L or more, the danger level referenced in the Comment, were more common with placebo (six patients; 1%) than combination treatment (two patients; 0·5%). In a previous study,4 only two of 118 patients who received the aliskiren/valsartan combination had potassium concentrations above 5·5 mmol/L; none had concentrations above 6·0 mmol/L.

We also disagree with Birkenhager and Staessen that our findings have limited generalisability. Although diastolic blood pressure was used for patients’ selection, the mean systolic blood pressure in all treatment groups at baseline ranged from 152·8 to 154·1 mm Hg, with 86·6% of patients having systolic blood pressure of 140 mm Hg or greater. 8-h ambulatory blood pressure monitoring was used to exclude those with white-coat hypertension. Other baseline characteristics, including age, ethnic origin, duration of hypertension, and the proportion of patients with obesity and metabolic syndrome, indicate that our study population was representative of hypertensive patients seen in clinical practice. Both diastolic and systolic blood pressure were assessed as efficacy endpoints in the study.

Finally, the claim that this combination “is unlikely to make it to general practice or even to primary prevention in specialist care” vastly overstates the issue of hyperkalaemia. The requirement for biochemical monitoring with the aliskiren/valsartan combination is no greater than that needed for patients receiving moderate diuretic doses in combination with other classes of antihypertensive drugs.

SO is the recipient of grants-in-aid from Abbot Laboratories, AstraZeneca, Aventis, Biovail, Boehringer Ingelheim, Bristol Myers Squibb, Forest Laboratories, GlaxoSmithKline, Novartis, Merck, Pfizer, Sankyo Pharma, Sanofi Synthelabo, and Schering-Plough. She is a consultant for Bristol Myers-Squibb, Daiichi Sankyo, Merck, Novartis, Pfizer, Sanofi Aventis, and the Salt Institute. She is a member of the Board of Directors for Encysive Pharmaceuticals. SAY has served as a speaker for Novartis. SP, HF, JZ, and AS are employees of Novartis Pharmaceuticals and are therefore eligible for Novartis stock and stock options. We thank Hui Fang for her contribution to this letter.

*Suzanne Oparil, Steven A Yarows, Samir Patel, Jack Zhang, Andrew Satlin
soparil@uab.edu
UAB Vascular Biology and Hypertension Program, Birmingham, AL, USA (SO); Chelsea Internal Medicine/ITHA, Chelsea, MI, USA (SAY); and Novartis Pharmaceuticals, East Hanover, NJ, USA (SP, HF, JZ, AS)
Global Fund: harmonisation and good governance vital

Michael McCarthy’s World Report (July 28, p 308)1 nicely summarises the success story of the Global Fund to Fight AIDS, Tuberculosis and Malaria in the past 5 years. However, we think that two important points have not received adequate attention.

First, the Global Fund has signed the Paris Declaration on aid effectiveness.2 Meetings of the Global Fund leadership are necessary, but will not suffice to achieve harmonisation, which also requires attention at the country level. In many countries malaria and HIV/AIDS control programmes are still executed in parallel, often by the same body, without even a minimum dialogue across programmes. Country coordinating mechanisms required by the Global Fund could have a more prominent role here and should be actively encouraged to pay attention to this aspect.

Second, McCarthy briefly mentions the need for transparency. Indeed, the Global Fund has been instrumental in promoting transparency and practices of good governance. Global Fund support to Uganda and Ukraine, for example, was interrupted partly because of poor governance and oversight from the country coordinating mechanisms.3 This is not a problem specific to the Global Fund, but concerns many other global health initiatives4 and new financial management mechanisms such as sector-wide approaches and budget support.

One way forward could be a mechanism of critical incident reporting as used in clinical practice.1 In this way, individuals and institutions would have the opportunity to report anonymously critical observations and claims of mismanagement, corruption, or fraud. This mechanism would serve as a complement to the Early Alert and Response System that the Global Fund has already put into place. Reported claims would need to be verified by an independent body or committee and could provide a safeguard for the Global Fund to maintain its good reputation as an innovative and efficient funding mechanism, which is vital for attracting and maintaining financial support.

The Swiss Centre for International Health of the Swiss Tropical Institute is currently acting as Local Fund Agent for the Global Fund in 11 countries.

*Nicolaus Lorenz, Kaspar Wyss
Swiss Centre for International Health, Swiss Tropical Institute, 4002 Basel, Switzerland

Chronic myeloid leukaemia in India

In their Seminar, Rüdiger Hehlmann and colleagues (July 28, p 342)1 suggest that chronic myeloid leukaemia (CML) is a rare disease with no obvious geographical or ethnic differences among different populations. However, available data suggest that the epidemiology of CML is different in the Indian subcontinent and in other developing countries from that of the rest of the world.2,3 CML is regarded as the most common form of adult leukaemia in this population,3 Onset is at a younger age than that expressed in the Seminar. The median age at onset varies between 30 and 40 years—almost half the median age of 65 years mentioned in the Seminar. The reasons for the younger age at onset in this region are not very clear. Shorter life expectancy, underdiagnosis in older people, or a higher prevalence of chronic infections in this population might be some of the reasons.3

The Seminar also did not mention homoharringtonine as being one of the upcoming treatment options either alone or in combination with other drugs for patients with imatinib-refractory disease. Homoharringtonine is a cephalotaxus alkaloid that induces cell differentiation and apoptosis by inhibiting protein synthesis.4 Available data suggest that it might induce a haematological as well as cytogenetic response in the chronic phase as well as the blastic phase of CML.5 A case report of its effect in imatinib-resistant CML harbouring a T315I mutation in BCR-ABL is very encouraging.1

We declare that we have no conflict of interest.

*Pankaj Malhotra, Subhash Varma
drpankaj@hotmail.com
Department of Internal Medicine, Post Graduate Institute of Medical Education and Research, Chandigarh 160012, India
Prescribing in elderly people

The review by Anne Spinewine and colleagues on appropriate prescribing in elderly people (July 14, p 173) is a welcome addition to the literature. We wish to add the following comments.

A list of inappropriate medications in elderly people (those aged 75 years and older) has recently been drawn up in France by 15 experts from different backgrounds and geographical origins using the Delphi method; this list mirrors French general practice. The list contains 34 criteria: 29 on medications or medication classes applicable to all elderly patients and five on medications that should be avoided by elderly patients with specific conditions. The list is a quality indicator of drug prescription in elderly people and should be regarded as a help when prescribing.

In a previous study involving 2018 patients admitted to an acute care geriatric unit, we showed that 66% of the patients were using at least one inappropriate medication according to the 1997 Beers list.1 The prevalence of adverse drug events was 19%; 6% of the patients experienced an adverse drug event attributable to an inappropriate medication. When relating inappropriate medications to the occurrence of adverse drug events, we calculated a 20% sensitivity, a 23% specificity, a 6% positive predictive value, and a 55% negative predictive value.

We do not agree with Spinewine and colleagues when they state that the inappropriateness of drugs is a relatively minor problem compared with inappropriate prescribing. Both points are important and some drugs—usually old drugs—are to be avoided whatever the circumstances. So, although inappropriate medications are only part of inappropriate prescribing, tackling this point is useful and beneficial for patients; fewer medications; fewer adverse drug events; and a smaller economic impact would ensue.

Who said that?

Contemporary technology allows frequent citation updates so that it is possible at the click of a mouse for a researcher to ascertain who in the last week has had the good judgment to quote his or her papers. The natural tendency is to admire the perspicacity and good taste of these citing authors. Recently, however, my natural sympathy for these discerning and临床 trials (one) were excluded.

For each of these 10 publications, I undertook to examine the scientific accuracy with which a series of the author’s papers were cited. (I considered that a review of the frequency with which I myself have sinned in this regard is beyond the scope of the present work.) From ISI Web of Knowledge at Aug 29, 2006, I used a citing reference search by date. The first 10 articles cited 10 or more times, including self-citations, were included, but clinical guidelines (four) and clinical trials (one) were excluded. For each of these 10 publications, I reviewed the first 10 citing papers for factual accuracy with respect to the article cited (as opposed to scientific disagreement).

The 10 articles were published between 1999 and 2002 and, excluding self-citations, were cited 10–29 times, median 13. Of the 100 citing papers (10×10) examined, 96 were in English, two German, one Spanish, and one Russian. Results are shown in the table. 18% of citing references, nearly one in five, were either partly or wholly factually incorrect. Only two of my papers were correctly cited by all 10 citing papers, seven were partly incorrectly cited by at least one citing paper, and five were incorrectly cited by one or more citing paper. The German, Spanish, and Russian papers all correctly cited the work in question.

As Oscar Wilde probably said, to misquote once may be regarded as a misfortune, to misquote more often looks like carelessness.

I declare that I have no conflict of interest.

Judith A Whitworth
Judith.Whitworth@anu.edu.au
John Curtin School of Medical Research, Australian National University, P O Box 234, Canberra, ACT 2601, Australia

Table: Accuracy with which a series of the author’s papers were cited by 10 consecutive other publications

<table>
<thead>
<tr>
<th>Paper</th>
<th>Correct</th>
<th>Partially correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>82</td>
<td>11</td>
<td>7</td>
</tr>
</tbody>
</table>

Who said that?

Contemporary technology allows frequent citation updates so that it is possible at the click of a mouse for a researcher to ascertain who in the last week has had the good judgment to quote his or her papers. The natural tendency is to admire the perspicacity and good taste of these citing authors. Recently, however, my natural sympathy for these discerning and临床 trials (one) were excluded.

For each of these 10 publications, I undertook to examine the scientific accuracy with which a series of the author’s papers were cited. (I considered that a review of the frequency with which I myself have sinned in this regard is beyond the scope of the present work.) From ISI Web of Knowledge at Aug 29, 2006, I used a citing reference search by date. The first 10 articles cited 10 or more times, including self-citations, were included, but clinical guidelines (four) and clinical trials (one) were excluded. For each of these 10 publications, I reviewed the first 10 citing papers for factual accuracy with respect to the article cited (as opposed to scientific disagreement).

The 10 articles were published between 1999 and 2002 and, excluding self-citations, were cited 10–29 times, median 13. Of the 100 citing papers (10×10) examined, 96 were in English, two German, one Spanish, and one Russian. Results are shown in the table. 18% of citing references, nearly one in five, were either partly or wholly factually incorrect. Only two of my papers were correctly cited by all 10 citing papers, seven were partly incorrectly cited by at least one citing paper, and five were incorrectly cited by one or more citing paper. The German, Spanish, and Russian papers all correctly cited the work in question.

As Oscar Wilde probably said, to misquote once may be regarded as a misfortune, to misquote more often looks like carelessness.

I declare that I have no conflict of interest.

Judith A Whitworth
Judith.Whitworth@anu.edu.au
John Curtin School of Medical Research, Australian National University, P O Box 234, Canberra, ACT 2601, Australia

Table: Accuracy with which a series of the author’s papers were cited by 10 consecutive other publications

<table>
<thead>
<tr>
<th>Paper</th>
<th>Correct</th>
<th>Partially correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>82</td>
<td>11</td>
<td>7</td>
</tr>
</tbody>
</table>

Table: Accuracy with which a series of the author’s papers were cited by 10 consecutive other publications

<table>
<thead>
<tr>
<th>Paper</th>
<th>Correct</th>
<th>Partially correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>82</td>
<td>11</td>
<td>7</td>
</tr>
</tbody>
</table>
Congestive heart failure and cardiovascular death in patients with prediabetes and type 2 diabetes given thiazolidinediones: a meta-analysis of randomised clinical trials

Rodrigo M Lago, Premranjan P Singh, Richard W Nesto

Summary

Background The overall clinical benefit of thiazolidinediones (TZDs) as a treatment for hyperglycaemia can be difficult to assess because of the risk of congestive heart failure due to TZD-related fluid retention. Since prediabetic and diabetic patients are at high cardiovascular risk, the outcome and natural history of such risks need to be better understood. We aimed to examine the risk of congestive heart failure and of cardiac death in patients given TZDs.

Methods We used a search strategy to identify 3048 studies. 3041 were excluded, and we did a systematic review and meta-analysis of the seven remaining randomised double-blind clinical trials of drug-related congestive heart failure in patients given TZDs (either rosiglitazone or pioglitazone). We calculated pooled random-effects estimates of the risk ratios for development of congestive heart failure in patients given TZDs compared with controls. The main outcome measures were development of congestive heart failure and the risk of cardiovascular death.

Findings 360 of 20191 patients who had either prediabetes or type 2 diabetes had congestive heart failure events (214 with TZDs and 146 with comparators). Results showed no heterogeneity of effects across studies ($I^2=22.8\%$; $p$ for interaction=$0.26$), which indicated a class effect for TZDs. Compared with controls, patients given TZDs had increased risk for development of congestive heart failure across a wide background of cardiac risk (relative risk [RR] 1.72, 95% CI 1.21–2.42, $p=0.002$). By contrast, the risk of cardiovascular death was not increased with either of the two TZDs (0.93, 0.67–1.29, $p=0.68$).

Interpretation Congestive heart failure in patients given TZDs might not carry the risk that is usually associated with congestive heart failure that is caused by progressive systolic or diastolic dysfunction of the left ventricle. Longer follow-up and better characterisation of such patients is needed to determine the effect of TZDs on overall cardiovascular outcome.

Introduction The prevalence of type 2 diabetes is increasing, and mortality from cardiovascular disease is two-fold to eight-fold higher in people with diabetes than in those without.¹ Thiazolidinediones (TZDs) are synthetic selective ligands of the nuclear transcription factor, peroxisome-proliferator-activated receptor γ (PPARγ). TZDs enhance insulin sensitivity¹¹ and are effective agents for control of glycaemia in patients with type 2 diabetes. Both rosiglitazone and pioglitazone have been shown to have a positive cardiometabolic profile that is independent of their effects on glycaemia.¹² One trial¹ showed that pioglitazone reduced major cardiovascular events in patients with type 2 diabetes; another¹ that pioglitazone slowed progression of carotid artery intima-media thickening. However, a recent meta-analysis has questioned the cardiac safety of rosiglitazone.⁶

The clinical use of TZDs has been limited by the fact that they cause fluid retention, and therefore could potentially lead to development of congestive heart failure in patients with or without pre-existing left ventricular systolic or diastolic dysfunction. However, TZDs do not directly affect left ventricular systolic or diastolic function.⁷ In 2003, the American Heart Association and American Diabetes Association (AHAADA) issued a consensus statement on the issue of congestive heart failure and provided guidelines for use of TZDs in patients with type 2 diabetes, with or without coexisting cardiovascular disease.⁶ The overall clinical benefit from use of TZDs in randomised clinical trials might be difficult to gauge because of the risk of congestive heart failure. Moreover, the outcome and natural history of congestive heart failure that is caused by TZD-related fluid retention has not been defined. TZD-related congestive heart failure could either be a drug-related adverse event or a negative outcome that might ultimately affect the survival of patients who receive TZDs. We did a systematic review and meta-analysis of pooled data from randomised trials of TZDs in subjects with prediabetes or type 2 diabetes to assess the risk of development of heart failure and death from cardiovascular causes in patients given TZDs.

Methods

Search strategy Randomised, double-blind, controlled trials of TZDs were eligible for inclusion in our meta-analysis if they reported risk estimates or frequency data for congestive heart failure and cardiovascular death. We excluded non-randomised clinical trials and those in which outcomes were not reported.

---

See Editorial page 1101
See Comment pages 1103 and 1104
Lahey Clinic Medical Center, Burlington, MA, USA
(R M Lago MD, P P Singh MD, R W Nesto MD)
Correspondence to: Richard W Nesto, Lahey Clinic Medical Center, 41 Mall Road, Burlington, MA 01805, USA
Richard.W.Nesto@lahey.org
We did a systematic review of Embase, MEDLINE, Database of Abstracts of Reviews of Effects (DARE), and the Cochrane Library (from January, 1998, to March, 2007). Figure 1 summarises the flowchart of article selection and inclusion. We searched for the following MeSH terms: “heart failure, congestive”, “mortality”, “cardiovascular system”, OR “edema” OR the text words “cardiovascular”, “CHF”, “edema”, “cardiac”, “heart”, “death”, “mortality”, OR “congestive heart failure”, OR “adverse” [all fields] AND “events” [all fields] AND (the MeSH term “thiazolidinediones” OR the text word “thiazolidinediones”) OR “TZD” [all fields] OR (“rosiglitazone” [substance name] OR “rosiglitazone” [text word]) OR (“pioglitazone” [substance name] OR “pioglitazone” [text word]). We selected only randomised controlled trials with male human patients that were written in English. We also searched the databases of the European Society of Cardiology, American Heart Association, American College of Cardiology, and American Diabetes Association by hand to identify full publications that had not yet been indexed. We reviewed bibliographies of retrieved publications to further increase the yield of potentially relevant articles.

Data collection and quality assessment
Two independent investigators extracted and tabulated data in a standardised data-extraction form. Discrepancies were resolved by group discussions and by reference to the original reports. The standard form included first author, publication year, mean age of participants, sex proportion, trial duration, type of TZD agent, type of control or controls, number of participants in drug and control groups, and number of events of interest (congestive heart failure and cardiovascular death) in the drug and control groups.

<table>
<thead>
<tr>
<th>Daily TZD dosage (mg)</th>
<th>Participants</th>
<th>CHF definition criteria</th>
<th>Trial duration (months)</th>
<th>Mean (SD)</th>
<th>Sex (men)</th>
<th>BMI (kg/m²)</th>
<th>Baseline HbA1c (%)</th>
<th>Baseline medical history</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone vs placebo</td>
<td>8</td>
<td>200</td>
<td>Investigator-reported new-onset heart failure</td>
<td>12</td>
<td>59.4</td>
<td>160 (80%)</td>
<td>32.7 (5.0)</td>
<td>5.8 (0.6)</td>
</tr>
<tr>
<td>Rosiglitazone vs metformin and sulfonylurea</td>
<td>4 to 8</td>
<td>4447</td>
<td>Adjudicated CHF that required admission to hospital</td>
<td>45</td>
<td>58.4</td>
<td>2290 (51.5%)</td>
<td>31.5 (4.8)</td>
<td>7.9 (0.7)</td>
</tr>
<tr>
<td>Rosiglitazone vs placebo</td>
<td>4 to 8</td>
<td>224</td>
<td>Adjudicated adverse event</td>
<td>13</td>
<td>64.0</td>
<td>184 (82.0%)</td>
<td>29.7 (3.6)</td>
<td>7.8 (1.3)</td>
</tr>
<tr>
<td>Rosiglitazone vs metformin/ Rosiglitazone vs glimebamide</td>
<td>4 to 8</td>
<td>4351</td>
<td>Serious CHF events, as reviewed by panel of independent cardiologists</td>
<td>48</td>
<td>56.3</td>
<td>2511 (57.7%)</td>
<td>32.2 (6.4)</td>
<td>7.4 (0.9)</td>
</tr>
<tr>
<td>Rosiglitazone vs placebo</td>
<td>4 to 8</td>
<td>5269</td>
<td>Adjudicated CHF</td>
<td>36</td>
<td>54.7</td>
<td>3214 (61.0%)</td>
<td>30.5 (5.6)</td>
<td>–</td>
</tr>
<tr>
<td>Pioglitazone vs glimepiride</td>
<td>15 to 45</td>
<td>462</td>
<td>Investigator-reported adverse event</td>
<td>18</td>
<td>59.6</td>
<td>289 (62.5%)</td>
<td>32.0 (5.0)</td>
<td>7.4 (1.0)</td>
</tr>
<tr>
<td>Pioglitazone vs placebo</td>
<td>15 to 45</td>
<td>5238</td>
<td>Adjudicated CHF</td>
<td>36</td>
<td>61.8</td>
<td>3866 (73.8%)</td>
<td>30.9 (4.8)</td>
<td>8.1 (1.4)</td>
</tr>
<tr>
<td>Overall</td>
<td>20 191</td>
<td></td>
<td></td>
<td>29.7 (15.2)</td>
<td>59.2 (3.1)</td>
<td>13508 (66.9%)</td>
<td>31.5 (5.0)</td>
<td>7.72 (1.1)</td>
</tr>
</tbody>
</table>

Data are mean (SD) or number (%), unless otherwise specified. TZD=thiazolidinedione. BMI=body mass index. HTN=hypertension. HLD=hyperlipidaemia. CVD/CAD=cardiovascular disease/coronary artery disease. HbA1c=glycosylated haemoglobin. CHF=congestive heart failure. CKD=chronic kidney disease. *This study included only individuals with prediabetes (with impaired fasting glycaemia (fasting plasma glucose 5.6 mmol/L to 6.9 mmol/L) and impaired glucose tolerance (plasma glucose level after a 2-h 75 mg oral glucose tolerance test of less than 11.0 mmol/L). †Only stable patients in NYHA class I-II CHF were included. All patients were required to have a left ventricular ejection fraction of 45% or less. ‡This study excluded patients with class II, III, and IV heart failure, and included those with class I heart failure.

Table 2: Characteristics of trials and participants

Figure 1: Search strategy profile
We measured the quality of these trials on the basis of internal validity and control of selection bias, detection bias, and attrition bias. We assessed selection bias according to incorporation of age, gender, and cardiac-disease history in risk estimates, when applicable. For control of attrition bias, we assessed the extent of loss to follow-up, represented as a proportion of the total initial study population. We calculated the loss–event ratio, from the number of patients lost to follow-up and the number of outcome events in the study, to assess the importance of loss to follow-up risk estimates in each study. We arbitrarily regarded studies with loss–event ratios of less than 10% as having satisfactory control of attrition bias.

Statistical analysis
We allocated the results of each randomised control trial as dichotomous frequency data. We did separate meta-analyses with the DerSimonian and Laird\textsuperscript{13} random-effects models to obtain pooled relative risks (risk ratios, RR) and associated 95% CIs for outcomes, with an intention-to-treat approach as a measure of association. Natural log transformations were done on the RR calculations. All analyses were initially done with a fixed-effects model, and were repeated if we detected heterogeneity across studies with a random-effects model, which included a measure of variance in the calculation of pooled results. Conventional random-effects weighting was used in all analyses. All p values were two-sided and p values of less than 0·05 were regarded as significant. We also did subgroup analyses with meta-regression to assess potential effect modification by type of control. To avoid statistical duplication of data, multiple control groups in a trial were collapsed as one independent control group for comparison to a specific TZD agent.

For studies that reported no events in a trial group, we applied Haldane’s correction with the classic half-integer correction to calculate the RRs and variances. We did

<table>
<thead>
<tr>
<th>A</th>
<th>Weight</th>
<th>Risk ratio (95% CI)</th>
<th>Risk ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosiglitazone vs control\textsuperscript{19}</td>
<td>12.9%</td>
<td>1.49 (0.62, 3.53)</td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone vs placebo\textsuperscript{10}</td>
<td>7.3%</td>
<td>1.81 (0.55, 6.02)</td>
<td></td>
</tr>
<tr>
<td>Pioglitazone vs glimepiride\textsuperscript{8}</td>
<td>11%</td>
<td>2.97 (0.12, 72.63)</td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone vs placebo\textsuperscript{21}</td>
<td>5.0%</td>
<td>7.00 (1.59, 30.76)</td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone vs placebo\textsuperscript{18}</td>
<td>12.2%</td>
<td>2.88 (0.12, 69.94)</td>
<td></td>
</tr>
<tr>
<td>Pioglitazone vs placebo\textsuperscript{7}</td>
<td>49.0%</td>
<td>1.31 (0.03, 1.67)</td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone vs metformin and sulfonylurea\textsuperscript{20}</td>
<td>23.5%</td>
<td>2.24 (1.27, 3.96)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>1.72 (1.21, 2.42)</td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity: $\chi^2=7.77$, df=6 (p=0.26), $I^2=22\%$
Test for overall effect: $Z=3.06$ (p=0.002)

<table>
<thead>
<tr>
<th>B</th>
<th>Weight</th>
<th>Risk ratio (95% CI)</th>
<th>Risk ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosiglitazone vs control\textsuperscript{13}</td>
<td>23.7%</td>
<td>1.49 (0.62, 3.53)</td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone vs placebo\textsuperscript{21}</td>
<td>12.2%</td>
<td>1.81 (0.55, 6.02)</td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone vs placebo\textsuperscript{30}</td>
<td>8.0%</td>
<td>7.00 (1.59, 30.76)</td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone vs placebo\textsuperscript{31}</td>
<td>11.7%</td>
<td>2.88 (0.12, 69.94)</td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone vs metformin and sulfonylurea\textsuperscript{20}</td>
<td>54.4%</td>
<td>2.24 (1.27, 3.96)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>2.18 (1.44, 3.32)</td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity: $\chi^2=3.33$, df=4 (p=0.50), $I^2=0\%$
Test for overall effect: $Z=3.65$ (p=0.0003)

<table>
<thead>
<tr>
<th>C</th>
<th>Weight</th>
<th>Risk ratio (95% CI)</th>
<th>Risk ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pioglitazone vs glimepiride\textsuperscript{8}</td>
<td>0.6%</td>
<td>2.97 (0.12, 72.63)</td>
<td></td>
</tr>
<tr>
<td>Pioglitazone vs placebo\textsuperscript{2}</td>
<td>99.4%</td>
<td>1.31 (0.03, 1.67)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>1.32 (1.04, 1.68)</td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity: $\chi^2=0.25$, df=1 (p=0.62), $I^2=0\%$
Test for overall effect: $Z=2.24$ (p=0.02)

Figure 2: Overall risk for congestive heart failure with (A) TZDs; (B) rosiglitazone; and (C) pioglitazone
TZD=thiazolidinediones. Forest plot of risk ratios (RR) on a logarithmic scale for trials pooled with Mantel-Haenszel weighting. The area of each square is proportional to the weight of the corresponding study, measured as the inverse of the estimated variance of the log risk ratio. The diamond represents the pooled relative risk, and its width represents its 95% CI. A horizontal line represents each study, with its effect size and 95% CIs. The solid vertical line corresponds to no risk. df=degrees of freedom.
further sensitivity analyses of data with Mantel–Haenszel weighted pooling of trials. We used the $I^2$ test to assess the percentage of total variation across studies that was due to heterogeneity rather than chance; to quantify inconsistency across studies; and to check that results from individual studies could be pooled.\textsuperscript{14,15}
To check for potential publication bias we used both Egger’s test and Begg’s tests. Accordingly, we examined for relative symmetry of individual study estimates around the overall estimate with Begg funnel plots in which log RRs were plotted against their corresponding standard errors. All primary analyses were done with Review Manager statistical software (version 5.0; The Cochrane Collaboration, Oxford, England; 2007), and additional analyses with Comprehensive Meta Analysis (version 2, Biostat, Englewood, NJ, USA).17

Role of funding source

No funding source was involved in the conception or development of the study. The first author and corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

We identified 3048 publications from our systematic review. A preliminary review of these studies led us to reject 2187 of them: 634 because they were not original and 1753 because they were not relevant to our aim. Of the 661 remaining studies, 654 were excluded because they either did not randomise patients in their study design (n=394) or did not measure or report outcome data for congestive heart failure or death (n=260). Ultimately, the meta-analysis included seven studies including one with two control groups. Table 1 summarises the characteristics of these studies and their participants. All studies had been published since 2005 and followed up patients for between 12 and 48 months, with a mean of 29·7 months. Trial populations ranged from 200 to 5269 participants, with a median of 4351.

The definition of congestive heart failure varied between trials, but in general consisted of investigator-reported or adjudicated congestive heart failure events that required admission to hospital (table 1). Although one study also reported events that did not require admission to hospital, we chose to exclude those events from analysis, to be consistent with the level of certainty of the diagnosis and severity of congestive heart failure reported in most other studies.

In the seven randomised clinical trials, 360 congestive heart failure events were reported in 20191 participants with prediabetes or type 2 diabetes (214 in the 9360 patients who were given TZDs, and 146/10831 were given comparators). The age of the participants ranged from 54·7 to 64 years, with a mean of 59·4; 83% were white, and 64–8% were men.

Results showed no heterogeneity of effects across studies (I²=22–8%; p=0·26), which indicated a class effect for TZDs. The observed inconsistency suggested that most of the variability between studies was due to chance. Figure 2 shows that, compared with controls, patients given TZDs had increased risk of congestive heart failure (RR 1·72, 95% CI 1·21–2·42, p=0·002). Figure 3 shows that, by contrast, the risk of cardiovascular death was not significantly increased with the use of either rosiglitazone or pioglitazone, compared with controls (0·93, 95% CI 0·67–1·29, p=0·68).

The overall event rate for congestive heart failure was 2·3% in patients in the TZD group and 1·4% in the comparator group. The overall event rate for cardiovascular death was 0·7% in both groups. Furthermore, the overall attributable event rate for congestive heart failure was 0·9% per year in the TZD group and 0·5% per year in the comparator group. The overall estimated number-
needed-to-harm for congestive heart failure was 107 over the 29.7-month mean follow-up. Table 2 shows that the number-needed-to-harm varied across studies from 35 in one trial to 491 in another.

Sequential exclusion of each trial from the analysis of congestive heart failure and cardiovascular death did not affect the overall relative risks. To account for heterogeneity in the studied populations, we excluded trials of patients with prediabetes and with metabolic syndrome without type 2 diabetes; this did not affect the RR for congestive heart failure (1.46, 1.19–1.78, p=0.0003) or cardiovascular death (0.91, 0.63–1.30, p=0.59). Neither did exclusion of a trial that enrolled a diabetic population with established and treated heart failure affect the RR for congestive heart failure (1.74, 1.21–2.50, p=0.03) or that for cardiovascular death (0.91, 0.65–1.28, p=0.60). The overall rate of congestive heart failure was 1.8% (360 of 20,191 patients). If only trials with placebo–control groups were included, the RR for congestive heart failure was 1.97 (95% CI 0.94–4.13, p=0.07), and the RR for cardiovascular death was 1.08 (0.66–1.76, p=0.77).

Pooled results from each of the TZD trials showed an increased risk for congestive heart failure was associated with both rosiglitazone and pioglitazone (table 3, figure 2). The pooled RR for development of congestive heart failure was 2.18 (95% CI 1.44–3.32, p=0.0003) in the five trials of rosiglitazone, and 1.32 (1.04–1.68, p=0.02) in two studies with pioglitazone. Figure 3 shows that the pooled RR for cardiovascular death was not higher than in controls with either rosiglitazone (RR 0.91, 95% CI 0.63–1.32, p=0.63) or pioglitazone (1.01, 0.51–2.01, p=0.98).

Table 2 summarises the raw number of events in the trials we assessed and the number of events per year. The risk for congestive heart failure did not differ for rosiglitazone and pioglitazone (1.74, 0.97–3.14, p=0.07). The risk of cardiovascular death did not differ between both drug groups (1.01, 0.73–1.40, p=0.96) (figure 4). The funnel plots were asymmetric, which suggested a publication bias, but this was not sufficiently large to affect our results or interpretations in a meaningful way.

Discussion
In 20,191 patients with prediabetes or type 2 diabetes in seven randomised trials, the risk of congestive heart failure was higher in patients given TZDs than in controls. However, despite the higher incidence of congestive heart failure in patients given TZDs, these patients did not have a higher rate of cardiovascular death. Even with the expected heterogeneity of baseline risk for congestive heart failure in these patients, the final outcome, measured for either congestive heart failure or cardiovascular death, was not affected by exclusion of any specific trial or group of trials in which patients’ characteristics were different.

The relative risk for congestive heart failure was increased across a wide background of cardiovascular risk in these trials: patients with prediabetes; those with type 2 diabetes without cardiovascular disease; and established cardiovascular disease without congestive heart failure; and with type 2 diabetes with documented congestive heart failure (NYHA class I and II) and an ejection fraction of less than 40%. However, although the relative risk for congestive heart failure was similar across the trials, the absolute risk varied according to the severity of the glucometabolic state and the presence or absence and degree of cardiovascular disease at baseline. Except in one trial, congestive heart failure occurred in the absence of any history or clinical evidence of congestive heart failure or systolic dysfunction at the time of enrolment.

Since drug exposures in these trials were short, and patients did not have previous histories of congestive heart failure or evidence of left ventricular dysfunction at entry, the excess of congestive heart failure event rates related to TZDs was probably the result of TZD-related fluid retention and diastolic dysfunction in susceptible patients. The natural history of congestive heart failure when caused by TZD-related fluid retention is unknown. A recent study suggested that either amiloride or spironolactone could induce diuresis in rosiglitazone-induced fluid retention; this effect is consistent with activity of the PPARγ agonists in the distal collecting duct.

Several factors can affect left ventricular diastolic function in type 2 diabetes such that a small increase in plasma volume (as noted in patients given TZDs) could precipitate pulmonary oedema because of an increase in left atrial and pulmonary venous pressures.

Diabetes is regarded to be commonly associated with diastolic dysfunction. Devereux and colleagues showed that type 2 diabetes might be associated with an increased prevalence of diastolic dysfunction, especially in those with hypertension. In most studies, patients who have an episode of heart failure due to diastolic dysfunction have a similarly poor prognosis to patients with systolic heart failure, at least in older patients with comorbidities. However, these epidemiological studies have not addressed whether an episode of heart failure that might be unmasked by fluid retention would have a different outcome when compared with the occurrence of diastolic heart failure in the general population.

This meta-analysis has shown that the increased event rate of congestive heart failure with TZDs was not associated with an increased risk of cardiovascular death in these trials. One observational study reported that, ambulatory patients with congestive heart failure given TZDs did not have a higher risk of admission to hospital for congestive heart failure or death than those who did not receive TZDs. Another observational study showed that, when compared with other antidiabetic drugs, insulin-sensitisers did not change the risk of death within a year after acute myocardial infarction, although the risk of readmission to hospital for heart failure was higher.

A meta-analysis of three randomised clinical trials (one with prediabetic patients and two with type 2 diabetic patients with a history of either cardiovascular disease or congestive heart failure) and four observational studies
showed that the risk of congestive heart failure doubled with use of TZDs. Our meta-analysis extended this by including four additional randomised trials with patients across a wider background of cardiac risk, that compared TZDs both with placebo and with active treatments. Our meta-analysis also assessed cardiovascular death in these trials.

A recent meta-analysis showed that patients given rosiglitazone had a higher risk of myocardial infarction than controls; they also had a higher risk of cardiovascular death, although this was not significant. By contrast, our meta-analysis did not show an increased hazard of cardiovascular death for either rosiglitazone or pioglitazone, despite the increased risk of congestive heart failure, even though the occurrence of congestive heart failure confers a poor prognosis for patients with type 2 diabetes. We did not include the smaller trials available for either rosiglitazone or pioglitazone, since they might not have had long enough observation times to accurately measure the risk for congestive heart failure and cardiovascular death. Furthermore, in these smaller studies, congestive heart failure events were defined in variable terms, investigator-reported, and not adjudicated.

Whether TZD-related fluid retention is a more benign cause of congestive heart failure than other causes cannot be confirmed without a comparison of outcomes between patients who developed congestive heart failure in the treatment and control groups. In one trial, mortality due to heart failure was similar in both the pioglitazone and placebo groups. In a more recent analysis of this trial, more cases of serious heart failure were associated with pioglitazone than with placebo, but the number of primary and main secondary events were similar in each group. One interpretation of these findings could be that congestive heart failure in the pioglitazone group (which was probably the result of fluid retention in many patients) carried a more favourable overall prognosis than did congestive heart failure in the placebo group (which was caused by progressive cardiac dysfunction). Another explanation could be that despite the potential for more adverse cardiovascular events associated with congestive heart failure, pioglitazone could have a cardioprotective effect compared with placebo.

We did not have data about the outcomes of patients in whom congestive heart failure was due to fluid retention, to enable comparison with those in whom it was due to other causes. Moreover, the overall benefit for patients given pioglitazone in the main secondary endpoint compared with controls, could be explained by pioglitazone-related directional changes in blood pressure, high-density lipoproteins, triglycerides, and glycaemia, which are factors associated with a lower cardiac risk. Any intrinsic negative effect of pioglitazone related to the risk of congestive heart failure or coronary ischaemia could have been counterbalanced or negated by beneficial changes in these cardiovascular risk factors. The true risk–benefit profile of a TZD when compared with another treatment for diabetes should be assessed when glycaemia and other cardiovascular risk factors are similar in the two treatment groups. Despite the glucose-lowering effect of TZDs, our data indicate that these drugs should not be used in patients with heart failure and should be cautiously used for glycaemic control in patients with cardiovascular disease who do not have heart failure. In patients with type 2 diabetes without cardiovascular disease in whom the absolute risk for congestive heart failure is much lower, the use of TZDs should be weighed against the risks and benefits of other antidiabetic medications.

Limitations of this study included the inherent assumptions made for any meta-analysis, and the use of aggregated data either as reported or as provided by individual study authors. Individual patient data and original data were not available, which limited our ability to do more detailed time-to-event analyses. The meta-analytic approach used might not have detected methodological problems in the primary studies. Moreover, definitions of heart failure differed between the included trials. Another potential limitation was that the comparator populations included both placebo and active treatments. The incidence of heart failure across trials was also heterogeneous across trials. Moreover, information about patient-specific data did not include the outcomes of congestive heart failure in the TZD and the control groups. Insufficient follow-up durations could have affected our conclusions about the association between congestive heart failure and cardiovascular mortality. We also did not have sufficient data to assess whether the risk of congestive heart failure differed between the two TZDs. We need longer follow-up and better characterisation of patients in whom congestive heart failure develops because of fluid retention to determine the effect of TZDs on overall cardiovascular outcome and whether congestive heart failure should be regarded as an adverse event or a characteristic cardiovascular endpoint.

Contributors
RN conceived and coordinated the review, wrote to authors of papers for additional information; provided additional data from papers; obtained and screened data on unpublished study; provided a clinical perspective; provided general advice on the review; and performed previous work that was the foundation of the current study. RN and RL designed the review and appraised the quality of papers. RL performed searches; screened search results; organised retrieval of papers; screened retrieved papers against inclusion criteria; extracted data from papers; managed data for the review; did analysis and interpretation of data; provided a methodological perspective; and wrote the review. PPS did searches and retrieved papers. All authors have read and approved the final version of the manuscript.

Conflict of interest statement
RWN is on the Speaker’s Bureau and is involved in research funded by Glaxo Smith Kline and Takeda. RML and PPS declare no conflicts of interest.

References


H5N1 infection of the respiratory tract and beyond: a molecular pathology study

Jiang Gu,* Zhigang Xie,* Zhancheng Gao,* Jinhua Liu,* Christine Korteweg,* Juxiang Ye, Lok Ting Lau, Jie Lu, Zifen Gao, Bo Zhang, Michael A McNutt, Min Lu, Virginia M Anderson, Encong Gong, Albert Cheung Hoi Yu, W Ian Lipkin

Summary

Background Human infection with avian influenza H5N1 is an emerging infectious disease characterised by respiratory symptoms and a high fatality rate. Previous studies have shown that the human infection with avian influenza H5N1 could also target organs apart from the lungs.

Methods We studied post-mortem tissues of two adults (one man and one pregnant woman) infected with H5N1 influenza virus, and a fetus carried by the woman. In-situ hybridisation (with sense and antisense probes to haemagglutinin and nucleoprotein) and immunohistochemistry (with monoclonal antibodies to haemagglutinin and influenza virus, and a fetus carried by the woman. In-situ hybridisation (with sense and antisense probes to haemagglutinin and nucleoprotein) and immunohistochemistry (with monoclonal antibodies to haemagglutinin and nucleoprotein) were done on selected tissues. Reverse-transcriptase (RT) PCR, real-time RT-PCR, strand-specific RT-PCR, and nucleic acid sequence-based amplification (NASBA) detection assays were also undertaken to detect viral RNA in organ tissue samples.

Findings We detected viral genomic sequences and antigens in type II epithelial cells of the lungs, ciliated and non-ciliated epithelial cells of the trachea, T cells of the lymph node, neurons of the brain, and Hofbauer cells and cytrophoblasts of the placenta. Viral genomic sequences (but no viral antigens) were detected in the intestinal mucosa. In the fetus, we found viral sequences and antigens in the lungs, circulating mononuclear cells, and macrophages of the liver. The presence of viral sequences in the organs and the fetus was also confirmed by RT-PCR, strand-specific RT-PCR, real-time RT-PCR, and NASBA.

Interpretation In addition to the lungs, H5N1 influenza virus infects the trachea and disseminates to other organs including the brain. The virus could also be transmitted from mother to fetus across the placenta.

Introduction A pandemic outbreak of human infection with avian influenza H5N1 currently poses a potentially serious health threat worldwide. Since the outbreak of infection with avian influenza H5N1 virus in 2003, WHO has reported 277 laboratory confirmed cases in ten countries with a mortality rate of about 60%.1 So far the virus has spread only from animals to human beings. However, human-to-human transmission potentiated by viral genomic mutation and reassortment of genomic subunits could be imminent. Recently, the first cases of probable human-to-human transmission have been reported.2,3 The H5N1 influenza A virus is a negative-stranded RNA virus in which the genome consists of eight segments encoding ten viral proteins including haemagglutinin, neuraminidase, polymerase proteins, and nucleoprotein.4

Little is known about the specific effects in organs and cells targeted by the virus. The infection initially seemed to be restricted to the lungs, but later reports5–8 have suggested that influenza A H5N1 could disseminate beyond the lungs. For various reasons (eg, religion), full autopsies of H5N1-infected human cases cannot often be obtained. Accordingly, only a few reports9–11 have described histopathological and virus distribution in H5N1 cases. Studies using in-situ hybridisation to detect viral genomic sequences in target cells have not been reported thus far.

We present clinicopathological data from H5N1 autopsies of two unrelated Chinese cases, as well as the histopathological changes and pattern of infection in the placenta and fetus from one of the patients, who was pregnant at the time of death. To gain further insight into the tissue tropism of influenza A H5N1 virus, we used in-situ hybridisation and immunohistochemistry to analyse viral localisation in various organs. Reverse-transcription (RT) PCR, real-time RT-PCR, and nucleic acid sequence-based amplification (NASBA) H5 detection assays were also done to detect viral RNA in tissue samples, as well as strand-specific RT-PCR.

Methods

Patients The clinical data of patient 1 have previously been published in detail.10 A 24-year-old Chinese woman from China’s Anhui province who was 4 months pregnant presented with a 6-day history of fever, cough, and dyspnoea. 2 weeks before admission, she had handled birds, several of which had died. On admission, she was lymphopenic, confused, and irritable, had bilateral infiltration on chest radiograph, and substantially reduced oxygen saturation. She was placed on a ventilator and treated with antibiotics, corticosteroids (hydrocortisone 400 mg on day 6 and day 7, and methylprednisolone 160 mg on day 8 and 240 mg on day 9), and fluids, but died 62 h after admission, 9 days after the onset of symptoms. No antiviral treatment was given.

www.thelancet.com Vol 370 September 29, 2007 1137
Patient 2 was a 35-year-old Chinese man from China’s Jiangxi province who had a 6-day history of fever and productive cough. He had participated in selling birds, of which several had died. On admission, chest radiographs showed evidence of pneumonia, and laboratory results showed abnormally increased hepatic-associated and cardiac-associated enzymes and lymphopenia. He was treated with antibiotics. Corticosteroids were given initially with methylprednisolone 40 mg on day 10 and then 120 mg per day for 17 days. Antiviral treatment was given, including rimantadine 100 mg twice daily on days 10 and 11 and then oseltamivir 150 mg per day from day 11 for 10 days. After admission, he became increasingly irritable and confused, followed by lowered consciousness. He developed respiratory failure and mechanical ventilation was initiated. In a sputum culture, several gram-positive microorganisms and Candida albicans were isolated. Antifungal drugs were added to his treatment. He developed multiple organ failure and died 27 days after the onset of symptoms.

The Chinese Centre for Disease Control and Prevention confirmed human infection with avian influenza H5N1 in both patients. RT-PCR detected H5N1 viral sequences in nasal swabs and nasopharyngeal aspirates, which were obtained on day 6 of illness for patient 1 and day 10 for patient 2. Viruses were isolated from the nasopharyngeal aspirate cultures, and designated as influenza A/Anhui/1/2005 virus in patient 1 and A/Jiangxi/1/2005 virus in patient 2. The haemagglutinin genes of viruses in patient 1 (GenBank accession number: DQ371928)\(^6\) and patient 2 (webappendix)\(^6\) were sequenced. The receptor-binding sites of both viruses were identical to those of previous H5N1 isolates.\(^9\) H5N1 viruses isolated from both patients were susceptible to both the M2 inhibitors amantadine and rimantadine, and the neuraminidase inhibitors oseltamivir and zanamivir.

Both cadavers were stored at 4°C and underwent autopsy about 18–20 h after death. The autopsies were done following conventional protocols and strict safety procedures.\(^1\) Tissue samples from all major organs and tissues were taken and fixed in 4% formalin. The brain of patient 1 was not available for investigation.

**Immunohistochemistry**

Immunohistochemistry was done on the basis of the technique of Lin and colleagues, with antigen retrieval by a standard technique.\(^6,10\) To detect viral antigen, tissue slides of 4 μm thickness were incubated with mouse monoclonal antibodies to nucleoproteins and haemagglutinin. Furthermore, monoclonal antibodies to the following cell markers were used: CD68 (for macrophages), CD3 (T lymphocytes), CD20 (B lymphocytes), CD8 (cytotoxic T cells), S100 (dendritic cells), cytokeratin AE1/AE3 (epithelial cells), surfactant protein A (type II pneumocytes), tubulin-β (ciliated epithelial cells), placental alkaline phosphate (PLAP, syncytiotrophoblasts), E-cadherin (cytotrophoblasts), neurofilament (neurons), neuron-specific enolase (neurons), and factor VIII (endothelial cells, webtable).\(^7\)

For controls, we used unrelated antibodies in place of the primary antibody.

**In-situ hybridisation**

For the development of probes, we used haemagglutinin (GenBank accession number DQ100556)\(^6\) and nucleoprotein gene sequences (DQ100560) of the H5N1 A/black-headed gull/Qinghai/1/2005 virus, which was recently isolated from a migratory bird at China’s Qinghai lake.\(^11\) Plasmids were generated by cloning of the full haemagglutinin gene (1779 bp) and full nucleoprotein gene (1565 bp) into a plasmid vector PGEM-T (Promega, Madison, WI, USA) yielding pGEM-HA for haemagglutinin and pGEM-NP for nucleoprotein. Both plasmids were linearised with appropriate restriction enzymes. Two sense and two antisense RNA probes were prepared by in-vitro transcription with T7 and Sp6 RNA polymerase (Promega) in the presence of digoxigenin-UTP (Roche Diagnostics, Penzberg, Germany). Since H5N1 is a negative-stranded RNA virus, sense probes were defined as the probes that detect the viral RNA (negative-stranded), whereas antisense probes detected mRNA and complement RNA (cRNA), which are both positive-stranded.

Briefly, before hybridisation, all solutions were prepared with diethyl pyrocarbonate (DEPC)-treated water.\(^12\) After deparaffinisation and rehydration, tissue sections of 4 μm thickness were treated with proteinase K digestion or microwave heating. Tissue sections were then incubated with a hybridisation cocktail containing 50 μg/mL of one of the four sense and antisense probes at 55°C for 16 h. All sense and antisense probes were applied separately on consecutive tissue sections. After blocking with horse serum (1:100), sections were incubated with alkaline phosphatase-labelled digoxigenin antibody (1:500, Roche Diagnostics, Penzberg, Germany) for 1 h, and the reaction products were colourised with nitroblue tetrazolium/5-bromo-4-choloro-3-indolyl phosphate (Promega).

As a positive control, we used brain tissue samples of a black-headed gull, for which H5N1 infection of the brain was confirmed by viral isolation.\(^6\) We used lung tissues from a mouse infected with H9N2 influenza virus as a negative control. Negative controls also included an unrelated antisense probe against the fragment of the polymerase gene (R1AB) of the severe acute respiratory syndrome-associated coronavirus (SARS-CoV),\(^20\) as well as H5N1 in-situ hybridisation probes to tissues (including lung and tracheal) obtained from seven adults who died from infectious lung diseases other than H5N1 influenza (four, SARS; one, purulent bronchitis; two, pneumonia), one adult who died from a non-infectious disease (gastric ulcer), one pregnant woman who died from an amniotic embolism, and one aborted fetus.

We identified the cell types infected by the virus with double labelling by combining in-situ hybridisation for
viral genomic sequences and immunohistochemistry for one of the cell-associated markers. To identify placental cells and cerebral neurons containing viral sequences or antigens, consecutive sections adjacent to the sections used for in-situ hybridisation or immunohistochemistry were immunostained with CD68, E-cadherin, PLAP, neurofilament, or neuron-specific enolase. Tissues from the pharynx, nose and paranasal sinuses, and lymph nodes other than hilar nodes were not available for investigation.

On preliminary investigation of lung tissues, we noted a contrast between the extent of histological damage and the limited number of positive pneumocytes (patient 1) or absence of positive pneumocytes (patient 2) shown by in-situ hybridisation and immunohistochemistry studies. Therefore, we did extensive pulmonary sampling for further histological assessment to ensure that the results were representative.

**Expected localisation of signals from in-situ hybridisation and immunohistochemistry**

After receptor-mediated endocytosis of the virus, the polymerase proteins, nucleoprotein, and encapsidated RNA segments migrate to the nucleus of the infected cell. In the nucleus, RNA segments are transcribed into mRNA and cRNA. mRNA is subsequently translated to viral proteins (eg, haemagglutinin and nucleoprotein) in the cytoplasm. Newly synthesised nucleoprotein is transported back to the nucleus. cRNAs function as antigenomic templates for the production of progeny RNA segments in the nucleus. Assembly of progeny gene segments and proteins occurs in the cytoplasm. Therefore, sense and antisense signals after in-situ hybridisation for nucleoprotein and haemagglutinin could be seen in both the nucleus and cytoplasm of infected cells, as well as immunohistochemical signals for nucleoprotein in the nucleus and for haemagglutinin in the cytoplasm. Since antisense probes hybridise to mRNA and cRNA, a positive signal would probably indicate active viral replication.

**RT-PCR, real-time RT-PCR, and NASBA**

RNA was extracted from deparaffinised tissue samples, or directly extracted from formalin-fixed tissues after overnight incubation with proteinase K (10 µg/µL, Ameresco, Cleveland, OH, USA). We did RNA extraction with Trizol (Invitrogen, CA, USA), and PCR as previously described. The haemagglutinin gene of the H5N1 virus was detected with H5for as the forward primer and H5rev as the reverse primer for H5 gene amplification (panel). Reamplification was done for specific samples with the same set of primers if no band was seen in gel electrophoresis. We obtained negative controls for RT-PCR from uninfected tissues (eg, lung, brain, placenta, and intestine) from three patients who died from non-infectious diseases. RT-PCR for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was assessed in parallel as an internal control.

To detect viral RNA in fetal tissues, real-time RT-PCR (H5 avian influenza virus Nucleic Acid Amplification Fluorescent Quantitative Detection Kit, PG Biotech, Shenzhen, China) was used as recommended by the manufacturer and the national standards of the People’s Republic of China.

<table>
<thead>
<tr>
<th>Patient 1</th>
<th>Patient 2</th>
<th>Fetus</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISH*</td>
<td>IH*</td>
<td>ISH</td>
</tr>
<tr>
<td>Trachea</td>
<td>+/−</td>
<td>+/−</td>
</tr>
<tr>
<td>Bronchi</td>
<td>−/−</td>
<td>−/−</td>
</tr>
<tr>
<td>Alveolar pneumocytes</td>
<td>+/+</td>
<td>+/+</td>
</tr>
<tr>
<td>Lymph nodes</td>
<td>+/+</td>
<td>+/+</td>
</tr>
<tr>
<td>Spleen</td>
<td>−/−</td>
<td>−/−</td>
</tr>
<tr>
<td>Heart</td>
<td>−/−</td>
<td>−/−</td>
</tr>
<tr>
<td>Endothelial cells</td>
<td>−/−</td>
<td>−/−</td>
</tr>
<tr>
<td>Hepatocytes</td>
<td>−/−</td>
<td>−/−</td>
</tr>
<tr>
<td>Kidneys</td>
<td>−/−</td>
<td>−/−</td>
</tr>
<tr>
<td>Small intestine</td>
<td>+/+</td>
<td>+/+</td>
</tr>
<tr>
<td>Brain</td>
<td>−/−</td>
<td>−/−</td>
</tr>
<tr>
<td>Syncytiotrophoblasts</td>
<td>−/−</td>
<td>−/−</td>
</tr>
<tr>
<td>Cytotrophoblasts</td>
<td>+/+</td>
<td>+/+</td>
</tr>
<tr>
<td>Hofbauer cells</td>
<td>+/+</td>
<td>+/+</td>
</tr>
<tr>
<td>Circulating mononuclear cells</td>
<td>−/−</td>
<td>−/−</td>
</tr>
<tr>
<td>ISH=in-situ hybridisation. IH=immunohistochemistry. n/a=not applicable. Plus sign=positive. Minus sign=negative. *Results presented for sense/antisense probes (identical results for nucleoprotein and haemagglutinin). †Results presented for nucleoprotein/haemagglutinin signals (nucleoprotein mainly detected in nucleus and haemagglutinin in cytoplasm). ‡Includes putative T lymphocytes.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Panel: Nucleotide sequences of primers used in RT-PCR, strand-specific RT-PCR, and NASBA**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>H5for</td>
<td>5’ gTg AYA ATg gYA Tgg AA 3’</td>
</tr>
<tr>
<td>H5rev</td>
<td>5’ CCA iAa YAa ACC AgC TA 3’</td>
</tr>
<tr>
<td>Tag-H5for</td>
<td>5’ TCT Aag Tgg CqA gTA gT gAAT gAA TgY ATg gAA 3’</td>
</tr>
<tr>
<td>Tag-H5rev</td>
<td>5’ TCT Aag Tgg CqA gTA ACC AIA AAg AYA gAC CAg CTA 3’</td>
</tr>
<tr>
<td>Tag</td>
<td>5’ TCT Aag Tgg CqA gTA A 3’</td>
</tr>
</tbody>
</table>

NASBA capture probe
DIG-GCRATTCYCTAGCAGCAAT
NASBA detection probe
Biotin-GATGCAAGGTCCGATATGAG

DIG=digoxigenin.
Figure 1: In-situ hybridisation locating gene sequences of H5N1 viral antigens (nucleoprotein and haemagglutinin).

Signals seen with nitroblue tetrazolium/5-bromo-4-chloro-3-indoly phosphate (purple-blue) and immunohistochemical signals with 3-amino-9-ethylcarbazole (double labelling, brown-red). No counterstaining done unless stated otherwise. (A) Lung tissue showing severe damage, hyaline membrane formation, and oedema (by haematoxylin and eosin staining), by contrast with limited number of cells positive for in-situ hybridisation in lung tissue of patient 1 (figure 1B). (B) Positive signals (with nucleoprotein sense probe) in nuclei of isolated pneumocytes (arrows). (C) Double-labelling of in-situ hybridisation (with nucleoprotein antisense probe) and immunohistochemistry (with tubulin-β antibody, brown, arrowheads) showing dark-blue viral genomic sequences (arrows) in cytoplasm of tubulin-negative non-ciliated cell (arrow 1) and tubulin-positive ciliated cell (arrow 2) in trachea. No signal seen in nuclei (lightly counterstained with methyl-green). (D) Positive signals (with nucleoprotein sense probe) in several mononuclear cells in lymph node (arrows, lightly counterstained with methyl-green). (E) Positive signals (with nucleoprotein antisense probe) in mucosal epithelial cells of small intestine (arrows, lightly counterstained with methyl-green). (F) Positive signals (with nucleoprotein sense probe) in brain cells (arrows) from left parietal lobe (lightly counterstained with methyl-green), mainly located in cytoplasm and confirmed to be neurons from immunostaining for neurofilament or neuron-specific enolase (webfigures 1D and 1F). (G) Positive signals (with nucleoprotein antisense probe) in large mononuclear cells with morphological features of cytotrophoblasts in periphery of chorionic villus (arrows). Cells confirmed as cytotrophoblasts (webfigure 4D). No positive signals noted in any syncytiotrophoblasts. (H) Positive signals (with nucleoprotein sense probe) in fetal liver cells (arrows), confirmed as Kupffer cells (webfigure 1F). (I) Positive signals (with haemagglutinin sense probe) in bronchial epithelial cells in fetal lung tissue (arrows).

Role of the funding source

The sponsor 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.

Results

The microscopic features of both patients were similar, apart from more extensive fibroproliferative changes in the lungs of patient 2. Both lungs of patient 1 showed features of diffuse alveolar damage (figure 1A) and focal desquamation of epithelial cells into alveolar spaces without evidence of type II pneumocyte hyperplasia. In the lungs of patient 2, patchy foci of consolidated bronchopneumonia and areas of fibrosis were seen. We found variable numbers of macrophages in the alveoli (especially in patient 2), and moderate numbers of scattered neutrophils and rare lymphocytes in the interstitial spaces. Both patients had substantially depleted lymphoid tissue in the spleen, lymph nodes, and mucosal lymphoid tissue in the gastrointestinal tract. The liver in both patients had spotty necrosis. In both patients, we detected very low numbers of macrophages with haemophagocytosis in the spleen, lymph nodes, and liver. The kidneys showed extensive tubular necrosis. Other organ systems showed no pronounced histological changes, apart from hypertrophy in the thymus of patient 2.

The placenta showed development appropriate for the length of gestation. We saw scattered foci of syncytiotrophoblast necrosis, sometimes with associated dystrophic calcification. Whether this finding was induced virally or was the sole result of maternal shock is unclear. Acute necrotising deciduitis was detected focally, and regarded as consistent with maternal shock.

Fetal tissues mostly showed no specific histopathological findings, and development was also consistent with gestational age. However, sections of fetal lung showed oedema and very small numbers of scattered interstitial neutrophils, which raised the possibility of mild acute interstitial pneumonitis, although this appearance was notably less severe than that seen in patients 1 and 2 (webfigure 1).

Strand-specific RT-PCR was undertaken on the basis of the technique of Yue and co-workers, with minor modification. Briefly, two-part reactions were used. First, the RT reaction was done in the presence of tagged primer, tag-H5for or tag-H5rev (panel). A third of cDNA products then underwent PCR with primers Tag/H5rev or Tag/H5for. We reamplified under the same PCR conditions for a specific sample if no band was seen after gel analysis of the first PCR reaction.

Virus-specific RNA was also detected with the avian influenza virus (H5 subtype) NASBA diagnostic test kit (MP version, Hong Kong DNA Chips, Hong Kong). The test was done as previously described with H5-specific capture and detection probes (panel). Absorbance of the amplified product was measured at 405 nm by an ELISA plate reader (Bio-Rad, Hercules, CA, USA). A negative control from the test kit was also included for H5 detection. The study was approved by the internal review board and ethics committee of the Peking University Health Science Centre.

Sense and antisense probes for in-situ hybridisation detected viral genomes focally in tissue samples from various organs (table 1). The two sets of probes (for...
haemagglutinin and nucleoprotein) generated identical staining results. In both autopsies, the trachea, small intestines, and lymph nodes showed positive signals of in-situ hybridisation, although pneumocytes were positive only in patient 1. In the lungs, sense probes hybridised in the nuclei of pneumocytes (figure 1B), whereas antisense probes hybridised in both the cytoplasm and nuclei. In all other organs with positive signals, both sense and antisense were present mainly in the cytoplasm of infected cells.

In the respiratory tract, we detected positive signals in tracheal epithelial cells and alveolar epithelial cells (figures 1B and 1C, webfigure 1). However, only an estimated 10–20% of epithelial cells in the trachea and about 5% of epithelial cells in the alveoli showed positive signals. Both bronchi and bronchioles were negative. Double labelling combining in-situ hybridisation and immunohistochemistry with tubulin-β antibody showed that both ciliated (tubulin-β-positive) cells and non-ciliated (tubulin-β-negative) cells of the trachea had viral sequences (figure 1C, webfigure 2). We also found putative basal cells to be infected. Double labelling with antibodies for cytokeratin and surfactant protein A showed that the positive alveolar cells were type II pneumocytes (webfigures 1C and 1D). In-situ hybridisation showed no positive staining in endothelial cells, macrophages, lymphocytes, fibroblasts, or any other cell type in the lungs or blood. However, we found positive viral signals in the cytoplasm of mononuclear cells in hilar lymph nodes (figure 1D). Double labelling and consecutive sections showed that cells positive for in-situ hybridisation were T lymphocytes (ie, positive for CD3 and negative for CD68, CD20, and S100, webfigure 3). Additionally, positive signals of intracytoplasmic viral sequences were present in mucosal epithelial cells of the small intestine (>50% in some intestinal segments, figure 1E). We also detected viral sequences in the cytoplasm (and to a much lesser extent in the nuclei) of brain cells from patient 2 (figure 1F, webfigure 1). Double labelling with neural markers neurofilament or neuron-specific enolase showed that these H5N1-positive cells were neurons (webfigures 1O and 1P). Table 2 shows the topographic distribution of cells with positive signals from in-situ hybridisation. No positive signals were seen in the heart, liver, spleen, kidneys, oesophagus, bone marrow, or stomach.

Placental tissue samples showed a large number of infected cells in chorionic villi (webfigure 4). Most positive cells were localised in the connective tissue core of these villi. These cells were confirmed (by labelling with monoclonal antibodies to CD68 and PLAP in consecutive sections) to be Hofbauer cells but not syncytiotrophoblasts. A subgroup of the cells with positive signals was found in the periphery of chorionic villi (figure 1G), which morphologically resembled cytotrophoblastic cells and were confirmed by immunostaining with E-cadherin antibody on consecutive sections. No positive signal was seen in syncytiotrophoblasts.

In the fetus, in-situ hybridisation identified viral sequences in the lungs, circulating mononuclear cells (webfigure 1), and mononuclear cells in the liver (figure 1H). The latter cells were identified as macrophages (Kupffer cells) by double labelling with antibody to CD68 (webfigure 1). Both sense and antisense probes were

<table>
<thead>
<tr>
<th>ISH*</th>
<th>RT-PCR</th>
<th>NASBA-based H5 detection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sense</td>
<td>Antisense</td>
</tr>
<tr>
<td></td>
<td>Optical density (405 nm)</td>
<td>Result</td>
</tr>
<tr>
<td>Parietal lobe</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Frontal lobe</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Occipital lobe</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Internal capsule</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Temporal lobe</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Amygdala</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Thalamus</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Mesencephalon</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Pons</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Medulla oblongata</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

*ISH=in-situ hybridisation. + Signals obtained with sense and antisense probes were mainly located in cytoplasm of infected neurons.

Table 2: Results of in-situ hybridisation, RT-PCR, and NASBA for H5 detection assay in selected brain regions of patient 2.

Figure 2: Immunohistochemical staining for H5N1 viral antigens (nucleoprotein, haemagglutinin) Positive signals seen with diaminobenzidine (brown, Zymed laboratories, San Francisco, CA, USA) or 3-amino-9-ethylcarbazole (red-brown, Sigma, St Louis, MO, USA). Slides counterstained with haematoxylin. (A) Positive staining (with nucleoprotein antibody) in nuclei and cytoplasm of pneumocytes (arrows). (B) Positive signals (with nucleoprotein antibody) in nuclei of mononuclear cells (arrows) with morphological features of macrophages in core of chorionic villi. Immunostaining in adjacent sections indicate cells to be Hofbauer cells (webfigure 4). (C) Positive signals (with nucleoprotein antibody) in nuclei of epithelial cells of trachea (arrows). (D) Positive staining (with nucleoprotein antibody) in cytoplasm and nuclei of neurons (arrows) from hippocampus. (E) Positive staining (with haemagglutinin antibody) in mononuclear cell in lymph node (arrow). (F) Positive staining (with haemagglutinin antibody) in cytoplasm of pneumocytes (arrows) in fetal lung tissue.
positive, mainly in the nuclei of pneumocytes and in a few detached epithelial cells of the bronchi (figure 1I). In Kupff er cells and circulating mononuclear cells, sense and antisense probes showed positive signals in both the cytoplasm and nucleus. The fetal lungs appeared to have more positively stained cells than did the maternal lungs.

The specificity of in-situ hybridisation was established by the results with the negative and positive controls (webfigure 1). The brain sections of the black-headed gull showed almost all neurons with positive signals for in-situ hybridisation, indicating the high detecting sensitivity of this technique.

Distribution of immunohistochemical staining (table 1) was consistent with that of in-situ hybridisation, apart from the absence of viral antigens in the intestines. Positive staining for nucleoprotein and haemagglutinin was detected in pneumocytes (figure 2A) and cytotrophoblasts and Hofbauer cells in the placenta (figure 2B) in patient 1; as well as in tracheal epithelial cells (figure 2C), the brain (figure 2D), and T lymphocytes in hilar lymph-node tissue (figure 2E) in patient 2. The fetus showed positive staining in bronchial epithelial cells, pneumocytes (figure 2F), and circulating mononuclear cells. Nucleoprotein was mainly detected in the nucleus and haemagglutinin in the cytoplasm. Negative controls validated the specificity of the immunohistochemistry protocol (webfigure 5).

All organs tested showed positive RT-PCR results, apart from the lymph nodes of patient 2 (tables 2 and 3, figure 3). All non-paraffin-embedded samples were positive for H5 expression. However, paraffin-embedded tissues only showed H5 expression after reamplification of the RT-PCR products. NASBA showed positive results on both types of samples without the need of reamplification.

With real-time RT-PCR, viral RNA was detected in the lungs and liver of the fetus (table 3). GAPDH was an internal control for successful RNA extraction in these assays. Positive-stranded RNA was detected in the heart and placenta of patient 1 and in the lungs, trachea, intestines, and brain of patient 2 (table 3, figure 3). The specificities of the RT-PCR and NASBA were further confirmed by use of the negative controls.

### Discussion

Our comprehensive investigation of the tissue tropism of H5N1 influenza virus, based on two adult autopsies and one fetal autopsy, focuses on the localisation of viral genomic sequences and antigens. We present evidence suggesting that the virus disseminates beyond the respiratory system. In addition to the lungs, viral sequences and antigens were found in the cerebral neurons and lymphocytes.

Presence of viral sequences and antigens in the CNS is consistent with the recent isolation of H5N1 virus from cerebrospinal fluid of a boy who died from encephalitis with neurological symptoms commonly seen in patients with H5N1 influenza (Gao Zh, unpublished), including the two cases in this study. Brain neurons were found to be infected by the virus. We also saw regional variations in positive neuronal distribution and negative neurons next to positive neurons. Possible reasons might include differing densities of the avian influenza virus receptor in human beings, differences in blood supply pathways and nerve connections that allow virus-target cell contact, and differing viral loads and viral replication stages. The detection of positive-stranded RNA by RT-PCR and in-situ hybridisation could indicate active viral replication in the brain. The virus could reach the CNS by penetrating the blood-brain barrier or by invading the afferent fibres of the olfactory, vagal, trigeminal, and sympathetic nerves after replicating in the respiratory mucosa, as has been shown in animals.14

---

### Table 3: Results of RT-PCR, strand-specific RT-PCR, real-time RT-PCR, and NASBA for H5 detection in specific tissue samples

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Patient 1 RT-PCR</th>
<th>Patient 2 RT-PCR</th>
<th>Fetus RT-PCR</th>
<th>Patient 1 NASBA</th>
<th>Patient 2 NASBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lungs</td>
<td>Y N Y N N Y Y</td>
<td>Y N Y N N Y Y</td>
<td>Y Y</td>
<td>2.579</td>
<td>0.961</td>
</tr>
<tr>
<td>Trachea</td>
<td>Y Y Y Y</td>
<td>n/a</td>
<td>1.236</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Intestines</td>
<td>Y Y N Y Y N Y N</td>
<td>2.277</td>
<td>2.112</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td>Y Y Y Y</td>
<td>n/a</td>
<td>2.424</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Heart</td>
<td>Y N Y N Y N N N</td>
<td>2.415</td>
<td>2.555</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Spleen</td>
<td>Y N Y N N N</td>
<td>1.326</td>
<td>1.442</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>Y N Y N N N Y N</td>
<td>2.176</td>
<td>2.225</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Kidneys</td>
<td>Y Y N Y N N N</td>
<td>2.371</td>
<td>1.587</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Lymph node</td>
<td>Y Y N N N N N</td>
<td>n/a</td>
<td>1.107</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Placenta</td>
<td>Y Y Y</td>
<td>n/a</td>
<td>1.653</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

+/-, + represent total, negative-stranded, and positive-stranded RNA, respectively. *Samples regarded as H5-positive if absorbance higher than 0.45.
Our data imply that the virus also infects and actively replicates in the small intestines, which is consistent with previous studies. The origin of infection in the intestines could be blood-borne, which is lent support by previous studies isolating live H5N1 virus from the serum and plasma. However, infection of infected respiratory secretion cannot be excluded as a possible route of infection, since H5N1 influenza viruses maintain sialidase activity despite the low pH in the upper digestive tract. Although in-situ hybridisation, NASBA, and RT-PCR detected viral RNA in the intestines, immunohistochemistry for viral antigens was negative. This discrepancy is consistent with the findings of Uiprasertkul and colleagues, although the reason is still unclear.

In-situ hybridisation and immunohistochemistry detected viral sequences and antigens in lymphocytes in the lymph nodes, and fetal macrophages in the placenta. Circulating mononuclear cells in the fetus and macrophages in the liver were found to harbour viral sequences. Previous in-vitro experiments have shown infection of macrophages by H5N1, and ex-vivo experiments have shown that the virus attaches to alveolar macrophages in human lung tissue. In addition to viraemia, infected immune cells could also carry the virus to extrapulmonary organs, which has been thought to participate in the pathogenesis of SARS.

The virus localised to type II pneumocytes in the respiratory tract, which has also been reported previously. However, with double labelling, we found viral sequences and antigens in both ciliated and non-ciliated epithelial cells of the trachea (figure 1C), contrasting with previous in-vivo and ex-vivo studies. In cultures of human tracheobronchial epithelial cells, H5N1 influenza viruses have been reported to infect mainly ciliated cells, which express mainly avian influenza virus receptors (α-3,3-linked sialic acids), although a limited number of non-ciliated cells (<20% of all infected cells) have also been reported to be infected.

Some studies have detected only human influenza virus receptors (α-2,3-linked sialic acids) on non-ciliated cells, whereas others also have found avian influenza virus receptors in these cells, albeit to a lesser extent. Changes in receptor-binding properties of A/Anhui/1/2005 and A/Jiangxi/1/2005 viruses could, in theory, also account for the infection of non-ciliated cells. However, preliminary tests have not revealed any substantial changes in the receptor-binding sites of either virus, compared with previous H5N1 isolates.

Notably, only a few scattered epithelial cells in the lungs were found to harbour the virus, contrasting with the severe and widespread histopathological changes in the lungs. Since this contrast was unexpected, lung tissue was sampled and analysed extensively, but the number of cells with viral localisation was consistently low in both patients. With the technique’s very high detection sensitivity (close to 100%), the percentages of positive epithelial cells recorded in this study could be reasonable estimates of H5N1-infected cells. In view of the low number of infected cells in patient 1 and the absence of cells with positive signals after in-situ-hybridisation in patient 2, direct viral injury to the epithelial cells of the respiratory tract is, in our view, unlikely to cause such severe pathological changes. The lack of histopathological changes in the brain, despite our findings indicating active viral replication in the region, also suggests that viral replication might not be specifically pathogenic. Recent in-vitro and in-vivo studies have indicated that hyperinduction of cytokines and chemokines could take part in the pathogenesis of H5N1 influenza.

Despite the high number of infected cells in the fetal respiratory system, we saw no evidence of severe damage to the fetal lungs, which greatly contrasts with the extensive damage found in the adult lungs. The absence of severe pulmonary damage (ie, high numbers of infected cells) in the fetus probably indicated an immunological naive status, which would be expected to result in low concentrations of the cytokines or chemokines to which the fetal lung tissues were exposed, and thus reduce or eliminate their induction of tissue damage. This theory is supported by in-vitro experiments showing that H5N1-infected neonatal macrophages express much lower amounts of chemokines than H5N1-infected adult macrophages.

Although the intracellular distribution pattern of immunohistochemical signals conformed to our expectations, it did not for signals from in-situ hybridisation. Probes hybridised mainly in the nuclei of pneumocytes and in the cytoplasm of other organs. In mice infected with H5N1 influenza virus, nucleoprotein sense and antisense probes have also hybridised mainly in the cytoplasm of infected cells, although the reason for this finding is unclear.

RNA analysis with RT-PCR and NASBA assays showed that H5-specific RNA was present in all tissues examined apart from the lymph nodes of patient 2, for which only NASBA showed positive result. This result could be due to the higher sensitivity of NASBA than that of RT-PCR. In fact, RT-PCR needed reamplification of the PCR products on the paraffin-embedded samples, which indicated a lower detecting sensitivity than NASBA.

RT-PCR and NASBA results were generally consistent with those of in-situ hybridisation and immunohistochemistry. However, viral RNA was also seen in viscera and some regions of the brain that showed negative results for both in-situ hybridisation and immunohistochemistry. A similar discrepancy has also been reported in a SARS study, which was attributed either to very low copy numbers of RNA and protein in these organs that might not be detectable or to false-positive RT-PCR results. False-positive results might be caused by the presence of virus in blood perfusing the organs without actual viral replication in the tissue parenchyma. Detection of positive-stranded...
RNA in the lung, heart, intestines, placenta, brain, and trachea in our study could imply that viral replication occurs in these organs. The absence of corresponding negative-stranded RNA in the lung and heart could be due to a lower detecting sensitivity of RT-PCR for negative strands than for positive strands.

This study has shown the capacity for human vertical transmission of the H5N1 virus. Transplacental transmission of the H5N1 virus warrants careful investigation, since maternal infections with common human influenza virus are generally thought not to affect the fetus. A sero-epidemiological study showed no evidence of transplacental transmission in pregnant women with human influenza infection. Our placenta autopsy showed viral genomic sequences in cytotrophoblasts and resident macrophages; furthermore, the virus infected the fetus. Viraemia has been reported in avian influenza virus infections, which is by contrast with the rare occurrence of viraemia in human influenza virus infections. Therefore, the likelihood of virus reaching the uterus and placenta is probably higher in avian influenza than in human influenza.

The vertical transmission route of avian influenza virus could be similar to that of human cytomegalovirus, which also targets cytotrophoblasts and Hofbauer cells. Two possible routes of transplacental transmission have been suggested: transcytosis across syncytiotrophoblasts to cytotrophoblasts in chorionic villi, or via infection of invasive cytotrophoblasts in the uterine wall (which could be infected after contact with maternal blood). These infected cells subsequently transmit the virus to the anchoring chorionic villi and could then be transmitted to Hofbauer cells that enter the fetal circulation. The presence of viral sequences and antigens in cytotrophoblasts, Hofbauer cells, and circulating fetal mononuclear cells supports this theory.

We detected viral sequences in cytotrophoblasts of chorionic villi but not in syncytiotrophoblasts. Differences in virus receptor expression could explain why cytotrophoblasts are susceptible to avian influenza virus infections but not human influenza virus infections. Both syncytiotrophoblasts and cytotrophoblasts have been found to lack α-2,6-linked sialic acids, but whether the placenta expresses α-2,3-linked sialic acids is unknown. The relative number of infected cells in the fetal lungs, as detected by in-situ hybridisation and immunohistochemistry, was substantially higher than in the two adults, which could be explained by the dominance of avian-influenza-virus receptors over human-influenza-virus receptors in the bronchial and alveolar epithelia during the pseudo-glandular stage of lung histogenesis (up to the 20th gestational week).

Despite the long duration of the disease and antiviral treatment in patient 2, viral sequences and antigens were detected in the post-mortem tissues. This finding is different from a previous study. The delayed clearance of viral antigens and sequences could be due to the immunosuppressive effect of the high-dose corticosteroids with which the patient had been treated for a long period before death. Since viral cultures were not done on post-mortem tissues, whether the detection of antigens and sequences indicates active viral replication is unclear. Positive results with in-situ hybridisation and RT-PCR have been reported in patients with SARS who died late in the course of disease. However, these positive RT-PCR results have been ascribed to the presence of small amounts of residual genome, rather than to active viral replication.

We have shown that the H5N1 virus spreads beyond the lungs, infecting both ciliated and non-ciliated epithelial cells of the trachea, the placenta, T lymphocytes in lymph nodes, and cerebral neurons. We also report evidence of transplacental transmission, resulting in infection of fetal organs. These newly obtained data are important in the clinical, pathological, and epidemiological investigation of human H5N1 infection, and have implications for public-health and health-care providers.

**Contributors**

JG initiated, designed, and coordinated the study, analysed the results, and took part in the writing of the manuscript. ZX took part in the autopsies, tissue collection, and routine pathology. ZiG was responsible for the clinical management and clinical data analysis. JLu took part in the probe-design molecular biology, viral test, and the writing of the manuscript. JY did the immunohistochemistry and in-situ hybridisation. CK took part in the study design, result and literature analysis, and writing of the manuscript. LTL took part in the RT-PCR and NASBA, result analysis, and writing of the manuscript. JL did the RT-PCR and NASBA. ZiG took part in the autopsies, tissue collection, routine pathology, and clinical data analysis. BZ did the molecular pathology, in-situ hybridisation, and probe design. MAM took part in the routine pathology, clinical data analysis, and writing of the manuscript. ML took part in the autopsies, tissue collection, and routine pathology. VMA did the fetoplacental pathology and molecular pathology. Eg took part in routine pathology and tissue processing. ACHY designed and coordinated the RT-PCR and NASBA study, analysed the results, and took part in the writing of the manuscript. WIL had overall responsibility for the study design, and took part in the writing of the manuscript.

**Conflict of interest statement**

We declare that we have no conflict of interest.

**Acknowledgments**

We thank Hongquan Shao and Ning Li for their assistance with the autopsies; Lu Yao, Ruishu Deng, and Ruiqi Xue for their assistance in the experiments; and Ting Zhang for helping with the photos. This study is supported partly by the Lifu Educational Foundation, National Basic Research Program (973) of China (grant no 2005CB523003), National Natural Science Foundation of China (grant no 30599431), and awards from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. CK is supported by grants from the Prins Bernhard Cultuurfonds (Wassink-Hesp Fonds and Kuitse Fonds), the Netherlands.

**References**


The effectiveness of supported employment for people with severe mental illness: a randomised controlled trial

Tom Burns, Jocelyn Catty, Thomas Becker, Robert E Drake, Angelo Fioritti, Martin Knapp, Christoph Lauber, Wulf Rössler, Toma Tomov, Jooske van Busschbach, Sarah White, Durk Wiersma, for the EQOLISE Group

Summary

Background The value of the individual placement and support (IPS) programme in helping people with severe mental illness gain open employment is unknown in Europe. Our aim was to assess the effectiveness of IPS, and to examine whether its effect is modified by local labour markets and welfare systems.

Methods 312 patients with severe mental illness were randomly assigned in six European centres to receive IPS (n=156) or vocational services (n=156). Patients were followed up for 18 months. The primary outcome was the difference between the proportions of people entering competitive employment in the two groups. The heterogeneity of IPS effectiveness was explored with prospective meta-analyses to establish the effect of local welfare systems and labour markets. Analysis was by intention to treat. This study is confirmed with ClinicalTrials.gov, with the number NCT00461318.

Findings IPS was more effective than vocational services for every vocational outcome, with 85 (55%) patients assigned to IPS working for at least 1 day compared with 43 (28%) patients assigned to vocational services (difference 26·9%, 95% CI 16·4–37·4). Patients assigned to vocational services were significantly more likely to drop out of the service and to be readmitted to hospital than were those assigned to IPS (drop-out 70 [45%] vs 20 [13%]; difference –32·1% [95% CI –41·5 to –22·7]; readmission 42 [31%] vs 28 [20%]; difference –11·2% [–21·5 to –0·90]). Local unemployment rates accounted for a substantial amount of the heterogeneity in IPS effectiveness.

Interpretation Our demonstration of the effectiveness of IPS in widely differing labour market and welfare contexts confirms this service to be an effective approach for vocational rehabilitation in mental health that deserves investment and further investigation.

Introduction

Unemployment for people with mental-health disorders is very high, with rates of up to 95% for those with severe mental illness.1 In the UK, the contribution of mental-health problems to absence from work due to sickness has substantially increased over the past decade,2 and people with mental-health disorders represent the largest group (40%) who claim incapacity benefit.3 A European study4 reported that mental-health problems are a rising cause of sickness, absenteeism, and work disability pensions. Traditional rehabilitation, including referred to as the train-and-place model, has addressed the mental health of individuals with severe mental illness.10 However, the successful placement of patients with severe mental illness has been limited. The train-and-place model, which emphasises rapid rehabilitation in mental health that deserves investment and further investigation.

The most intensively studied place-and-train or supported employment intervention is individual placement and support (IPS), which emphasises rapid job search on the basis of patient preference and training in job skills to prepare patients for a return to employment. This approach remains the most widespread but has had very little success, and many patients obtain employment only in sheltered workshops.1 Developments in the USA emphasise direct job placements, often in simple entry-level occupations, plus support to patient and employer. This model is called place-and-train.

The most intensively studied place-and-train or supported employment intervention is individual placement and support (IPS), which emphasises rapid job search on the basis of patient preference and continuing support to patient and employer from an employment specialist working as an integral member of the mental-health service contributing to treatment planning and delivery.1 Results from several randomised trials and two meta-analyses2–3 have shown the effectiveness of the programme in the USA, where this intervention is now the recommended evidence-based practice.9 There are almost 20 experimental and quasi-experimental studies of IPS. Several of these studies investigated combined interventions (eg, IPS and assertive community treatment) or examined specific aspects of the intervention (such as degree of IPS integration for agencies, teams, and individual providers). Results from randomised trials11–16 have shown that rates for competitive employment on the open job market for patients using IPS were more than doubled, and a large scale implementation trial in eight sites with locally-determined supported employment and control services noted much the same degree of clinical effectiveness.7

Europe differs greatly from the USA in both its employment practices (varying amounts of employment protection compared with a hire and fire culture in the USA) and in having more generous welfare systems. Such systems might generate a benefit trap, in which there could be perceived or real financial disincentives to returning to work—eg, loss of housing benefits or high disability payments.20 Differences in both labour markets and welfare systems might reduce the effectiveness of IPS. Moreover, welfare systems and job markets vary...
considerably across Europe,\textsuperscript{79} and there are substantial differences in unemployment rates.

Our aim was to assess the effectiveness of IPS compared with existing good quality rehabilitation and vocational services for people with severe mental illness in terms of open employment outcomes (in the competitive labour market), and to examine its effectiveness in different European welfare systems and labour markets.

**Methods**

**Study design**

We undertook a randomised trial in six European centres—London (UK), Ulm-Guenzburg (Germany), Rimini (Italy), Zürich (Switzerland), Groningen (Netherlands), and Sofia (Bulgaria). Patients were included if they were diagnosed with severe mental illness (psychotic illness, including bipolar disorder), were aged between 18 years and local retirement age (ie, between 60 and 65 years), had been ill and had major role dysfunction for at least 2 years, were living in the community at baseline, had not been in competitive employment in the preceding year, and wished to enter competitive employment. They were randomly allocated to either IPS or vocational service (control service). Since the effect of sex and work history on vocational outcomes needed to be considered,\textsuperscript{20} service allocation was stratified by centre, sex, and work history (more or less than 1 month’s competitive employment in the 5 years before baseline). Recruitment took place between April 1, 2003, and May 30, 2004, with follow-up ending on Nov 30, 2005. Randomisation was done centrally with MINIM (version 1.5). A researcher at every centre recruited patients and submitted their details to the statistician for randomisation, and researchers were notified of allocation by email. The allocation sequence was concealed until the services had been assigned, but patients, professionals, and researchers could not be blinded to service allocation thereafter.

The primary hypothesis was that patients assigned to IPS would be more likely to obtain open employment than would control service patients. Secondary hypotheses were that they would be in open employment for longer than would control patients, and they would not spend more time in hospital. The secondary outcomes included drop-out from service and admission to hospital. All analyses, apart from that of job tenure, were undertaken on an intention-to-treat basis with the entire sample, and then repeated for every centre alone.

**Interventions**

IPS was provided by one or two IPS workers at every centre, who were trained in the model. The IPS model consists of identification of patients who want to work in the competitive labour market, and helps them develop realistic goals and seek appropriate employment directly; there is no training phase. The IPS worker builds up a network of employers willing to accept patients, with whom the IPS worker continues contact, supporting both patient and employer. This support is open ended (in our study until the end of the 18-month follow-up), and the IPS worker had a maximum caseload of 25 patients. When the local services operated a community mental health team system, all IPS workers were located within such a team, providing a service to study patients recruited from that team and liaising with team staff.

The vocational service at every centre was chosen on the basis that it was the best alternative vocational rehabilitation service available locally, and it was the typical and dominant service in the area. All services provided high quality vocational rehabilitation according to the train-and-place model. This rehabilitation consisted of an assessment of the patient’s rehabilitation needs, and the provision of a structured training programme aimed at combating deficits related to illness and training in appropriate work skills (eg, reintroduction of a daily routine for attending...
the centre, time management, or information technology skills). The structured programme usually occupied most of the week and was generally at a day centre, although in Ulm it involved mostly residential care. Every vocational service had to make a commitment to take patients into the service within 2 months of randomisation.

All IPS workers were novices who undertook an equal amount of training at a project conference at the start of the study, by the originator of IPS, Deborah Becker, and a London vocational rehabilitation specialist, who then continued to supervise workers by telephone conference every 2 weeks. The fidelity of the IPS workers to the model was assessed with the IPS fidelity scale at three time points. The scale distinguishes successfully between supported employment and other vocational interventions.

The characteristics of the vocational services were assessed at two time points with a vocational services questionnaire developed for the study. This questionnaire was a data collection method developed from narrative accounts from all centres of the vocational services being used, and was designed to capture the nature of the service offered and its distinctiveness from IPS. Both the IPS fidelity scale and the vocational services questionnaire were administered to both IPS workers and vocational services at every centre, to measure systematically the differences between them.

### Procedures

Patients were followed up for 18 months, with interviews at baseline and 6, 12, and 18 months. Data were obtained on vocational outcomes, hospital admission, and service use by interview, on job satisfaction and hours worked by questionnaire at the start and end of each job obtained, and on job status by vocational staff. Clinical and social functioning, quality of life, and needs for care were assessed at all interviews with validated, structured assessments and will be reported separately. Researchers were trained centrally in administration of all measures, and inter-rater reliability was assessed periodically with videotaped interviews. The clinical diagnosis of psychosis was confirmed from case-notes by OPCRIT—a validated structured assessment by independent research staff who were clinically trained.

### Statistical analysis

The study was funded to provide one full-time equivalent IPS worker for every centre, which meant a maximum sample size of around 300 patients in total (50 people at all six centres, divided equally between treatment groups). With Drake and colleagues finding that 9% of the control group entered competitive employment, and on the basis of a total sample size of 300, our study had 90% power to detect differences approaching those found by Drake and colleagues.

95% CIs were calculated for primary and other binary outcomes. Continuous vocational outcomes (number of hours worked, number of days employed, and job tenure) and time in hospital were analysed by presenting bootstrapped estimates for both the differences in

### Table 1: Patient characteristics at baseline

<table>
<thead>
<tr>
<th>IPS (n=156)</th>
<th>Vocational service (n=156)</th>
<th>Total (n=312)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>37.3 (9.80)</td>
<td>38.3 (9.94)</td>
</tr>
<tr>
<td>Men</td>
<td>93 (60%)</td>
<td>95 (61%)</td>
</tr>
<tr>
<td>Age at first psychiatric contact (years)</td>
<td>26.8 (8.36)</td>
<td>26.5 (8.54)</td>
</tr>
<tr>
<td>Number of admissions in lifetime</td>
<td>0</td>
<td>13 (8%)</td>
</tr>
<tr>
<td>1-5</td>
<td>117 (75%)</td>
<td>105 (68%)</td>
</tr>
<tr>
<td>6-10</td>
<td>16 (10%)</td>
<td>21 (14%)</td>
</tr>
<tr>
<td>&gt;11</td>
<td>10 (6%)</td>
<td>11 (7%)</td>
</tr>
<tr>
<td>Clinical diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia / schizoaffective disorder</td>
<td>122 (79%)</td>
<td>126 (82%)</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>28 (18%)</td>
<td>23 (15%)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (3%)</td>
<td>5 (3%)</td>
</tr>
<tr>
<td>Work history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;1 month in past 5 years</td>
<td>88 (56%)</td>
<td>86 (55%)</td>
</tr>
<tr>
<td>≤1 month in past 5 years</td>
<td>68 (44%)</td>
<td>70 (45%)</td>
</tr>
<tr>
<td>Number of years in education</td>
<td>12.1 (3.83)</td>
<td>11.6 (3.09)</td>
</tr>
<tr>
<td>Living situation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone</td>
<td>51 (33%)</td>
<td>54 (35%)</td>
</tr>
<tr>
<td>With friends / relatives</td>
<td>85 (55%)</td>
<td>77 (49%)</td>
</tr>
<tr>
<td>Sheltered accommodation</td>
<td>20 (13%)</td>
<td>25 (16%)</td>
</tr>
<tr>
<td>Born in country of residence</td>
<td>135 (87%)</td>
<td>147 (94%)</td>
</tr>
</tbody>
</table>

Data are mean (SD) or number (%).

### Table 2: Patient characteristics at baseline

<table>
<thead>
<tr>
<th>IPS</th>
<th>Vocational service</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worked for at least 1 day</td>
<td>85 (55%)</td>
<td>156</td>
</tr>
<tr>
<td>Number of hours worked*</td>
<td>428 (706.77)</td>
<td>143</td>
</tr>
<tr>
<td>Number of days employed*</td>
<td>130 (174.12)</td>
<td>154</td>
</tr>
<tr>
<td>Job tenure (days)*</td>
<td>213.6 (159.42)</td>
<td>83</td>
</tr>
<tr>
<td>Drop-out from service</td>
<td>20 (13%)</td>
<td>156</td>
</tr>
<tr>
<td>Admission</td>
<td>28 (20%)</td>
<td>148</td>
</tr>
<tr>
<td>Percentage of time spent in hospital*</td>
<td>4.6 (12.56)</td>
<td>148</td>
</tr>
</tbody>
</table>

Data are mean (%) or mean (SD). *Data for hours worked were not available for all patients, since not all patients completed follow-up interviews or were able to supply this information. Data for days employed were collected outside interview. Job tenure data were only calculated for the subgroup of patients who worked. Data for hospital use were missing for 23 patients. †Bootstrapped estimates of difference between means and bias corrected and accelerated 95% CIs presented.

### Table 2: Vocational, admission, and drop-out outcomes

<table>
<thead>
<tr>
<th>IPS (n=156)</th>
<th>Vocational service (n=156)</th>
<th>Total (n=312)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admission 28 (20%)</td>
<td>148</td>
<td></td>
</tr>
<tr>
<td>Percentage of time spent in hospital*</td>
<td>4.6 (12.56)</td>
<td>148</td>
</tr>
</tbody>
</table>

Data are mean (%) or mean (SD). *Data for hours worked were not available for all patients, since not all patients completed follow-up interviews or were able to supply this information. Data for days employed were collected outside interview. Job tenure data were only calculated for the subgroup of patients who worked. Data for hospital use were missing for 23 patients. †Bootstrapped estimates of difference between means and bias corrected and accelerated 95% CIs presented.

### Table 2: Vocational, admission, and drop-out outcomes

<table>
<thead>
<tr>
<th>IPS (n=156)</th>
<th>Vocational service (n=156)</th>
<th>Total (n=312)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of admissions in lifetime</td>
<td>0</td>
<td>13 (8%)</td>
</tr>
<tr>
<td>1-5</td>
<td>117 (75%)</td>
<td>105 (68%)</td>
</tr>
<tr>
<td>6-10</td>
<td>16 (10%)</td>
<td>21 (14%)</td>
</tr>
<tr>
<td>&gt;11</td>
<td>10 (6%)</td>
<td>11 (7%)</td>
</tr>
<tr>
<td>Clinical diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia / schizoaffective disorder</td>
<td>122 (79%)</td>
<td>126 (82%)</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>28 (18%)</td>
<td>23 (15%)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (3%)</td>
<td>5 (3%)</td>
</tr>
<tr>
<td>Work history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;1 month in past 5 years</td>
<td>88 (56%)</td>
<td>86 (55%)</td>
</tr>
<tr>
<td>≤1 month in past 5 years</td>
<td>68 (44%)</td>
<td>70 (45%)</td>
</tr>
<tr>
<td>Number of years in education</td>
<td>12.1 (3.83)</td>
<td>11.6 (3.09)</td>
</tr>
<tr>
<td>Living situation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone</td>
<td>51 (33%)</td>
<td>54 (35%)</td>
</tr>
<tr>
<td>With friends / relatives</td>
<td>85 (55%)</td>
<td>77 (49%)</td>
</tr>
<tr>
<td>Sheltered accommodation</td>
<td>20 (13%)</td>
<td>25 (16%)</td>
</tr>
<tr>
<td>Born in country of residence</td>
<td>135 (87%)</td>
<td>147 (94%)</td>
</tr>
</tbody>
</table>

Data are mean (SD) or number (%).
means and their 95% CIs since they were positively skewed. The analysis of job tenure was based only on patients who had been in competitive employment for at least 1 day. Missing secondary vocational data were handled with a conservative approach. Secondary vocational data (hours worked and days employed) for patients who had worked for at least 1 day were scored as missing, whereas patients who had not worked were scored as zero.

Prospective meta-analyses27 were used to explore the possible effect of labour market and welfare system factors on the heterogeneity of the effectiveness of IPS and returning to competitive employment. These analyses were done to account for variability in the primary employment outcome across the six centres that might have been due to factors other than the interventions or characteristics of the study sample. The factors considered were identified through a detailed consultation exercise, through review of published work, analysis of international data sets, and use of a semistructured questionnaire to IPS workers associated with the study. The factors identified were local unemployment rate, percentage change in gross domestic product (GDP), long-term national unemployment rate as a proxy for social exclusion, benefit trap, and indirect income redistribution. Cochran’s Q test24 was used to examine whether there was significant heterogeneity in outcomes between the centres. The Cochran’s Q test was then used to examine whether the factors of interest explained a significant amount of the variation between centres. To confirm the appropriateness of the assumptions made by the prospective meta-analyses, we used an alternative method, logistic regression analysis, for the categorical variable benefit trap only. The logistic regressions against both IPS effect sizes and getting a job irrespective of service, including centre as a random effect, were fitted with R (version 2.4).

Likelihood of benefit trap was assessed by asking IPS workers whether they considered their client group to be at risk of having their income reduced if they took a job, and centres were categorised as high, low, and no risk. The factors were tested first against IPS effect sizes and then against finding a job irrespective of service.

Analyses were done with SPSS for Windows (version 12.0), except for the meta-analyses (Comprehensive Meta-Analysis [version 2.0]) and the bootstrapping analyses (Stata for Unix [version 8.1]). This study is registered with ClinicalTrials.gov, with the number NCT00461318.

Role of the funding source

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the corresponding 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.

Results

Figure 1 shows the trial profile. Table 1 shows baseline sociodemographic characteristics of the two groups. Data for the primary outcome measure (in competitive employment for at least 1 day) were available for the whole sample. Of these patients, 252 (81%) completed the final follow-up interview. There were no systematic differences in any baseline characteristics between those who dropped out (did not complete the final interview) and those who remained, nor between the number dropping out of the study between IPS and vocational service groups (difference 7.7 percentage points, 95% CI –1.01 to 16.4, p=0.085). Interview data were supplemented by data from questionnaires and vocational workers. Five people (three IPS and two vocational service patients) died from natural causes during the study.

All IPS workers maintained good or fair levels of IPS fidelity throughout the study (median 65, min–max 61–70 of 75). By contrast, no vocational services achieved 56 (the lower cutoff for fair) at any time-point (median 31,

![Figure 2: Proportions assigned to IPS or VS who worked at least 1 day within centres](Image)

Proportions assigned to IPS or VS who worked at least 1 day within centres. Error bars=95% CIs. IPS=individual placement and support. VS=vocational service.

Table 3: Socioeconomic sources of heterogeneity

<table>
<thead>
<tr>
<th>Source</th>
<th>IPS effect size</th>
<th>Getting a job</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local unemployment rates*</td>
<td>5.82</td>
<td>0.006</td>
</tr>
<tr>
<td>GDP per head growth (annual %) 2003†</td>
<td>1.66</td>
<td>0.198</td>
</tr>
<tr>
<td>% GDP spent on health 2002</td>
<td>0.229</td>
<td>0.632</td>
</tr>
<tr>
<td>Long-term unemployment rate (1999)</td>
<td>0.322</td>
<td>0.456</td>
</tr>
<tr>
<td>Benefit trap? (2004-05)¶</td>
<td>1.62</td>
<td>0.445</td>
</tr>
</tbody>
</table>

These socioeconomic variables should not be compared with each other, since the data are from different years and different sources. *Ranges from 3.6 in Zürich and Sofia to 8.1 in Groningen. Information provided by authors adjusted using ratio of national rates (EIU 2004 database accessed online via the Economist Intelligence Unit Market Indicators and Forecasts website) and ratio applied to local rates. †Ranges from -1.4 in Groningen to 4.9 in Sofia. Information from World Development Indicators Online database, accessed via the Economic and Social Data Services (ESDS) website. ¶Ranges from 7.3 in Sofia to 11.2 in Zurich. Information from World Development Indicators Online database, accessed via ESDS website. §Persons unemployed for a period of 1 year or more as a percentage of the labour force. Ranges from 1.2 in Zurich to 2.3 in Sofia. Information from ESDS website. ¶High risk centres: London, Groningen, low risk: Ulm, Zurich; no risk: Rimini, Sofia. GDP=gross domestic product.
min–max 24–40), confirming that the services they were delivering were not classifiable as IPS. Vocational service questionnaire data also confirmed that the IPS and vocational services at all centres differed substantively in their aims and scope, with vocational services working to the train-and-place model.

In all six centres, IPS was more effective than were vocational services for every vocational outcome, with 85 (55%) patients assigned to IPS working for at least 1 day compared with 43 (28%) patients assigned to vocational services (difference 26·9%, 16·4–37·4). Patients assigned to IPS worked for more hours than did vocational service patients during the 18 months of follow-up, and they were employed for more days (table 2). Of those who worked for at least 1 day, patients allocated to IPS maintained their jobs for longer periods than did those assigned to vocational services (214 days vs 108 days).

Vocational service patients were significantly more likely to drop out of the service than were IPS patients (table 2). Vocational service patients were also more likely to be admitted during the study period than were those assigned to IPS, and they spent on average twice as much time in hospital (table 2).

IPS was significantly more effective than was the vocational service in terms of vocational outcomes in London (difference 32·0%, 7·7–56·3), Rimini (30·8%, 5·0–56·5), Zürich (38·5%, 16·7–60·2%), and Sofia (40·7%, 17·1–64·4), but there was no difference in Ulm (11·5%, –15·4 to 38·5) or Groningen (7·7%, –17·3 to 32·7). In Ulm, the number of patients assigned to IPS working for at least 1 day (n=14) was close to that of the other centres, but more vocational service patients (n=11) at this centre also worked for at least 1 day than did those at the other five centres (figure 2). There was no statistically significant heterogeneity in variation in effectiveness of IPS (Q=6·12, p=0·295).

Only local unemployment rates explained a significant amount of the heterogeneity in effectiveness of IPS (Q= 5·82, p=0·016), whereas increased GDP growth per head, long-term unemployment rate, and risk of a benefit trap (as assessed by IPS workers) accounted for a significant amount of heterogeneity in getting a job, irrespective of service (table 3). Figure 3 shows the effect sizes of the centres grouped by the variable risk of benefit trap. Where benefits were deemed likely to be higher than salary (a greater risk of benefit trap), this was associated with a lower risk difference (a measure
of the effectiveness of IPS). A higher risk of benefit trap was associated with a lower event rate (of getting a job). Logistic regression analyses confirmed these findings (data not shown).

Discussion
This study clearly shows the effectiveness of IPS, since the rate of obtaining competitive employment for people with severe mental illness who were motivated to work was doubled compared with usual, high-quality, vocational rehabilitation. Not only did patients assigned to IPS obtain competitive employment more often than did those assigned to vocational services, but they also kept their jobs for longer and worked for more hours. We noted that a high rate of employment did not have a detrimental effect on clinical wellbeing and relapse, which would have been indicated by an increased number of psychiatric admissions. This result is important confirming for many of the clinicians we approached, who were concerned about the potential stress that working in the competitive labour market might cause their patients. Indeed, the finding of a reduced rate of admission with IPS is not reported in US studies, and could relate to the generally greater degree of integration of health and social care in Europe.

We have shown that IPS is effective in Europe, despite very different economies and labour markets from the USA, where previous IPS studies have largely been done. Although the heterogeneity of effect size between the six centres was not statistically significant, the test for heterogeneity is known to have low power, especially when the number of sites is small as in this case. It was still valid to explore the sources of clinical heterogeneity.

Unlike the US trials, our study showed that socioeconomic context did affect IPS effectiveness, especially local unemployment rates, which accords with a non-randomised US study and could relate to the generally greater degree of integration of health and social care in Europe.

We have shown that IPS is effective in Europe, despite very different economies and labour markets from the USA, where previous IPS studies have largely been done. Although the heterogeneity of effect size between the six centres was not statistically significant, the test for heterogeneity is known to have low power, especially when the number of sites is small as in this case. It was still valid to explore the sources of clinical heterogeneity.

Unlike the US trials, our study showed that socioeconomic context did affect IPS effectiveness, especially local unemployment rates, which accords with a non-randomised US study and could relate to the generally greater degree of integration of health and social care in Europe.

Overall, more patients obtained jobs when the country's economy was growing and job creation was increased than they did when the economy was slow. High amounts of social exclusion were also associated with more patients obtaining jobs; this counter-intuitive finding might have been because these countries offered less welfare support, thus providing greater incentives to work in the competitive employment market. Furthermore, the benefit trap (in which there could be a perceived or real financial disincentive to returning to competitive employment) was shown to be a demonstrable impediment to successful vocational rehabilitation overall in this group, although its association with IPS effect size was not significant. Recruitment to the study was especially difficult in two countries with a substantial benefit trap (the UK and Netherlands). The prospective meta-analysis was exploratory, however, and the findings should be treated with some caution.

To ensure comparability across very different mental-health-care systems, we restricted our study to patients with severe mental illness who had been unemployed for at least 1 year. Our inclusion criteria were close to many of the US studies and, like patients in those studies, our patients had very limited work history and work skills, as well as several longstanding role impairments. Most of the US studies, however, included non-psychotic patients; yet we showed equal IPS effectiveness. We therefore believe that the IPS approach would be at least as effective in Europe as it has been in the USA. The accumulated evidence for IPS in North America, plus our findings of its effectiveness in widely differing labour market and welfare contexts, should confirm this service as an effective approach for vocational rehabilitation in mental health that deserves investment and further investigation.

Contributors
TB designed the study with WR and AF. All authors were involved in the conduct of the study, interpreting the results, and in revising and correcting the paper, which was drafted by TB and JC. The analyses were led by SW. All authors read and approved the final version of the manuscript.

The EQOLISE Group
Tom Burns, Jocelyn Catty, Connie Geyer, Marsha Koletsi, Pascale Lissouha, Miles Rinaldi, Sarah White (London), Thomas Becker, Ulrike Elhiosun, Rana Kalkan, Reinhold Kilian (Ulm), Angelo Fioritti, Denise Manchini (Rimini), Astrid Niersman, Jooske van Busschbach, Durk Wiersma (Groningen), Christoph Lauber, Wulf Rösler, Ingeborg Warnke (Zürich), Dimitar Germanov, Toma Tomov (Sofia), Adelina Comas, Claire Curran, Martin Knapp, Anita Patel (LSE).

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

Acknowledgments
This study was funded by a grant from the European Union Quality of Life and Management of Living Resources Programme QLRT 2001-00683. We thank Greg McHugo for methodological advice, Deborah R Becker and Miles Rinaldi for training the IPS Workers, and the IPS workers themselves: Alison Lewis (London), Wulf Dorn and Eva Marischka (Ulm), Donato Piegari (Rimini), Bettina Barsch and Patric Meyer (Zürich), Anne Mieke Epema, Laureen Jansen, and Bea Hummel (Groningen), Petar Karaginev (Sofia).

References
7 Crowther RE, Marshall M, Bond GR, Huxley P. Helping people with severe mental illness to obtain work: systematic review. BMJ 2001; 322: 204–08.
17 Cook JA, Leff HS, Blyler CR, et al. Results of a multisite randomized trial of supported employment interventions for individuals with severe mental illness. Arch Gen Psychiatry 2005; 62: 505–12.
Achieving health equity: from root causes to fair outcomes

Michael Marmot, on behalf of the Commission on Social Determinants of Health

Health is a universal human aspiration and a basic human need. The development of society, rich or poor, can be judged by the quality of its population’s health, how fairly health is distributed across the social spectrum, and the degree of protection provided from disadvantage due to ill-health. Health equity is central to this premise and to the work of the Commission on Social Determinants of Health. Strengthening health equity—globally and within countries—means going beyond contemporary concentration on the immediate causes of disease. More than any other global health endeavour, the Commission focuses on the “causes of the causes”—the fundamental structures of social hierarchy and the socially determined conditions these create in which people grow, live, work, and age. The time for action is now, not just because better health makes economic sense, but because it is right and just. The outcry against inequity has been intensifying for many years from country to country around the world. These cries are forming a global movement. The Commission on Social Determinants of Health places action to ensure fair health at the head and the heart of that movement.

Health inequality, inequity, and social determinants of health

Consider three children: one African, one south Asian, and one European. At birth each, representing the average for their country, has life expectancy of less than 50 years. The African and south Asian figures come from 1901, the European figure from 1901. Over the past century, life expectancy for the European child increased by about 30 years, and is still rising.1 Between 1970 and 2000, the south Asian child’s life expectancy rose by 13 years, whereas for the child in sub-Saharan Africa, during the same period, life expectancy rose by 4 months.2

The improvement in health in 20th century Europe, North America, and the other countries that now make up the Organisation for Economic Cooperation and Development, is a major societal achievement. Although there is no certainty as to what accounted for the improvement in Europe, it is most likely a combination of improvement in the conditions in which people live and work and, more recently, advances in medical care.

The health achievements that Europe has enjoyed have already started happening in south Asia and other regions (figure 1)—but have considerable distance still to go—and could happen in sub-Saharan Africa. No country or region should have to live with levels of ill-health that are avoidable. The lack of improvement in health in the countries of central and eastern Europe and the former Soviet Union is of concern, as are the other differences shown in figure 1. Improvements in living and working conditions, and finding a way to deliver known medical solutions, would lead to dramatic reductions in these global inequalities in health.

Such inequalities in health should not be tolerated. In many poor countries, maternal mortality ratios exceed 500 per 100000 livebirths. In Sweden the ratio is two per 100000.

Inequalities within countries and the social gradient

There is a second problem of inequalities in health: the dramatic differences within countries. These differences in health occur along several axes of social stratification including socioeconomic, political, ethnic, and cultural. One way of describing the magnitude of inequalities is the gap between top and bottom socioeconomic groups. In El Salvador, for example, if mothers have no education their babies have a one in ten chance of dying in the first year of life; if mothers have at least secondary education the infant death rate is a quarter of that.3

Such striking inequalities in health within countries are seen in rich countries, too. In Glasgow, UK, life expectancy of men in one of the most deprived areas was 54 years, compared with 82 years in the most affluent.4 Thus the poorest in Glasgow have lower life expectancy than the Indian average. Men with the lowest life expectancy in the USA (1997–2001) had lower life expectancy than the Pakistan average (1995–2000).4 In every instance, indigenous peoples of the world have life expectancies lower than the national average.5

But focusing on the gap between top and bottom fails to draw attention to a pervasive finding: the social gradient in health (figure 2).4 With few exceptions, the evidence shows that the lower an individual’s socioeconomic position the worse their health. There is a social gradient in health that runs from top to bottom of the socioeconomic range. The gradient can be obvious or subtle. In general,

---

1. This paper is an abridged version of the Interim Statement of the Commission on Social Determinants of Health
2. Correspondence to: Professor Sir Michael Marmot, International Institute for Society and Health, University College London, London WC1E 6BT, UK m.marmot@ucl.ac.uk

---

**Figure 1**: Life expectancy at birth by region, 1970–75 and 2000–05

Source: Human Development Report 2005.2
people second from the bottom have worse health than those above them but better health than those below. In Sweden, adults with a PhD have lower mortality than those with a professional qualification or Master’s degree. The gradient is a worldwide occurrence, seen in low-income, middle-income, and high-income countries.

The gradient in health should not deflect attention from the plight of people at the bottom of the gradient, the poorest of the poor. Rather, the social gradient in health means that we are all implicated.

Inequalities in health within and between countries arise from inequalities within and between societies: in social and economic conditions and their effects on people’s lives that determine their risk of illness, and the actions taken to prevent or treat illness when it occurs. Such inequalities are not inevitable or immutable. For example, we see increasing differences in Russia in life expectancy by level of education among both men and women (figure 3). There is also evidence that conditions can be changed for the better (figure 4).

A central aim of the Commission on Social Determinants of Health is to assemble the evidence, particularly of what will make a difference, to lay the basis for action to reduce inequalities in health within and between countries. Where such evidence is lacking the Commission will make recommendations on how to redress the gaps.

Justice, inequality, and inequity
All societies have social hierarchies in which economic and social resources, including power and prestige, are distributed unequally. The unequal distribution of resources affects people’s freedom to lead lives they have reason to value, which in turn has a powerful effect on health and its distribution in society. The Commission takes issue with the unequal distribution of social conditions when health suffers as a consequence.

Not all health inequalities are unjust or inequitable. If good health were simply unattainable, this would be unfortunate but not unjust. Where inequalities in health are avoidable, yet are not avoided, they are inequitable. This distinction can be illustrated by the difference in
men’s and women’s health. Women, in general, live longer than men. This difference is likely due to biological sex differences, and is not, therefore, inequitable. However, in cases where women have the same or lower life expectancy as men—that is, where social conditions act to reduce their apparently natural longevity advantage—inequality is a mark of inequity.15 The injustice that the Commission seeks to address comes from failure to achieve levels of health that, but for lack of action, should be attainable.

The right to the highest attainable level of health is enshrined in the charter of WHO and many international treaties.16 This right obliges governments and others to act—to take steps that increase all individuals’ chances of obtaining good health. The realisation of this right, however, will take not just access to health care but action on the social determinants of health.

Although we see health as having intrinsic value—health as an end in itself—the Commission also recognises its instrumentality. Good health enables people to participate in society, with potentially positive consequences for economic performance.17,18 Addressing the social determinants of health will yield greater, and sustainable, returns to existing efforts to improve global health.

Empowerment and freedom

At the heart of the concern with social determinants of health, and health inequity, is concern for people without the freedom to lead flourishing lives.19 To make a fundamental improvement in health equity, technical and medical solutions such as disease control and medical care are, without doubt, necessary. But they are insufficient. There will need to be empowerment of individuals, communities, and whole countries.

We see empowerment operating along three interconnected dimensions: material, psychosocial, and political. People need the basic material requisites for a decent life; they need to have control over their lives, and they need political voice and participation in decision making processes. Although individuals are at the heart of empowerment, achieving a fairer distribution of power requires collective social action—the empowerment of nations, institutions, and communities.

The differential status of men and women in almost every society is perhaps the most pervasive and entrenched inequity. As such, the relation between the sexes represents as pressing a societal issue for health as the social gradient itself. Indeed the feminisation of the catastrophic AIDS epidemic in southern Africa clearly shows the lack of power of women to enjoy fundamental social freedoms.20,21 This marked health inequity encapsulates disempowerment at many levels—government and institutional incapacity to act on evidence of gender effect, and the unequal participation of women in political institutions from village to international levels; unequal access to and control over property, economic assets, and inheritance; unequal restrictions on physical mobility, reproduction, and sexuality; sanctioned violation of women’s and girls’ bodily integrity; and accepted codes of social conduct that condone and even reward sexual violence against women. It is not enough to focus on delivering antiretrovirals to women with AIDS in southern Africa while doing little to deal with their profound disempowerment.

Panel 1: SEWA, the Self-Employed Women’s Association, India

Many Indians, both urban and rural, experience severe disadvantage as a result of low social status: the combined effect of caste, education, and income. They have poor housing, with restricted access to clean water and sanitary facilities. They have had little in the way of financial resources and have difficulty pursuing their rightful livelihoods. Their children have had little opportunity for development and education, especially where they forego schooling to work with their parents. When ill, they have little access to health care, frequently only available for a fee.

In Ahmedabad, there are around 100 000 street vendors, forming a sizable proportion of the informal employment sector in the city. They sell fruit, vegetables, flowers, fish, clothes, vessels, toys, footwear, and many other items for daily and household use. Most vendors have been selling in the city’s markets and streets for generations.

Like other poor self-employed women, the vegetable sellers of Ahmedabad live in poor parts of the city. They start work at dawn, buying their wares from merchants in the wholesale markets. They frequently need to borrow money, incurring very high rates of interest, and routinely face harassment and eviction from their vending sites by local authorities. The Self-Employed Women’s Association (SEWA) is a striking example of collective action by these women and others like them, to challenge and change these conditions.

To strengthen control over their livelihoods, vegetable sellers and growers (all SEWA members) linked together to set up their own wholesale vegetable shop, cutting out exploitative middlemen. As a result, both growers and sellers have seen improved incomes through better prices for their produce. SEWA also organises child care, running centres for infants and young children, and campaigns at the state and national level for child care as an entitlement for all women workers. And SEWA members are improving their living conditions through slum upgrading programmes to provide basic infrastructure such as water and sanitation. This happens in partnerships with government, people’s organisations, and the corporate sector.

To address the problem of access to credit, the SEWA Bank provides small loans and banking facilities to poor self-employed women, like the vegetable sellers, avoiding the interest rates demanded by private loan agents. The Bank is owned by its members, and its policies are formulated by an elected Board of women workers.

In times of health crisis, poor families not only lose work and income, but often also have to sell assets to secure the wherewithal to pay for treatment. Poor informal sector workers and their families are pushed further into the cycle of poverty and indebtedness. With SEWA, however, when the vegetable sellers or their family members fall ill, collectively organised health insurance can be used to pay for health care costs. SEWA has started an integrated insurance scheme for women in times of crisis.

Frequently harassed by local authorities, the vegetable sellers campaigned with SEWA to strengthen their status, through formal recognition in the form of licenses and identity cards, and representation on the urban Boards which govern market activities and urban development. That campaign, started within Gujarat, subsequently went all the way to the India Supreme Court, and inspired international attention and alliances.

Source: SEWA website http://www.sewa.org/services/bank.asp
Conditional Income Transfer (Bolsa Familia), Brazil

In many ways, Brazil of recent years is a good example of managed growth and commitment to poverty reduction. However, even though the government of President Lula Da Silva has set a course to address the high rates of inequality in the country, chronic poverty in parts of Brazil mean that the poorest households continue to suffer from multiple forms of disadvantage. They are frequently unable to secure adequate nutrition for the family, and in rural areas can be highly vulnerable to environmental hazards such as drought and flood. The poorest urban households are connected neither to the water nor the sewage system, and poor communities have no trash collection services. Poor access to education leads to relatively high rates of illiteracy, compromising employment opportunities for young men and women.

The period of so-called redemocratisation from the mid-1980s brought with it important changes in Brazil’s approach to governance, social policies, and poverty reduction. A key component of this new policy environment is the Family Stipend Programme, or Bolsa Familia, a form of conditional cash transfer targeted at poor and extremely poor families to address key aspects of extreme poverty and reduce inequality.

The Bolsa Familia, launched in October, 2003, unified four federal programmes designed to address key aspects of household well being among the poorest families. These were: the school stipend, food stipend, food card, and fuel support programmes. Conditionalities stipulated that children between 7 and 15 years should be regularly attending school, and that growth, nutrition, development and immunisation status of children from 0 to 6 years should be regularly monitored. The programme also included prenatal care for pregnant women.

Complementary interventions, designed to safeguard household income and promote further poverty reduction, included adult literacy classes, aid to family-based agriculture, access to micro-credit, and professional or vocational training. At the federal level, the programme was coordinated through an Inter-Ministerial Management Committee. Originally, the Bolsa Familia secretariat was directly linked to the President’s office. While municipalities were responsible for registering eligible families, the legislation enacting Bolsa Familia established local councils, including the participation of civil society organisations, to monitor interventions.

The Bolsa Familia represents a holistic approach to social welfare, poverty reduction and addressing the interconnected conditions that lead to poor and inequitable health. Coordinated cross-sectorally, through inter-ministerial management, the programme acts on key aspects of wellbeing at the family and household level—from child development through stimulating uptake of health and education services, through nutrition for children and mothers, to living conditions with the fuel subsidy, and employment through vocational training, support to family agriculture, and micro-credit.

Although the share of total income represented by the conditional cash transfers has been relatively small, the programme’s outstanding targeting (using a unified registry) has resulted in an impressive equalising effect, responsible for about 21% of the fall in the Brazilian Gini index (a measure of inequality of income distribution).

The effect of these processes of disempowerment is most pronounced in indigenous populations, among the most marginalised and disenfranchised peoples in the world, many of whom have experienced profound dispossession of land and erosion of culture. Their crisis is reflected in “wide disparities between the health status of indigenous peoples and non-indigenous peoples within the same country”.

In emphasising the need for both empowerment and technical solutions, we draw the parallel with contemporary models of development. An increase in national income, by itself, does not capture development in its fullest sense. At the least, education and health should be included. To achieve development in this fuller sense, economic growth is insufficient—it needs to proceed hand in hand with empowerment.

A social determinants of health approach has several advantages. It bridges the artificial distinction between technical and social interventions, and shows how both are necessary aspects of action. It seeks to redress the imbalance between curative and preventive action and individualised and population-based interventions. And, by acting on structural conditions in society, a social determinants approach offers a better hope for sustainable and equitable outcomes.

Social determinants of health and health equity

There is not a great deal of mystery as to why poor people in low-income countries suffer from high rates of illness, particularly infectious disease and malnutrition: little food, unclean water, low levels of sanitation and shelter, failure to deal with the environments that lead to high exposure to infectious agents, and lack of appropriate medical care. Similarly, we have a great deal of knowledge of the causes of non-communicable diseases that represent the major burden of disease for people at the lower end of the social gradient in middle-income and high-income countries. The WHO Global Burden of Disease study identified underweight, overweight, smoking, alcohol, hypertension, and sexual behaviour as major causes of morbidity and mortality. For both groups of causes the question is how they and their inequitable distribution come about. That is, what are the causes of the causes?

The Commission believes that these health inequities are the result of a complex system operating at global, national, and local levels which shapes the way society, at national and local level, organises its affairs and embodies different forms of social position and hierarchy. The place people occupy on the social hierarchy affects their level of exposure to health-damaging factors, their vulnerability to ill health, and the consequences of ill health.

Putting all these levels in context is the natural environment, and the macro-level to micro-level effects of environmental change. Risks to health include heatwaves and other extreme weather events, changes in infectious disease patterns, effects on local food yields and freshwater supplies, impaired vitality of ecosystems, and loss of livelihoods. If present trends continue the adverse health effects from human-induced environmental changes will be distributed unequally. The poor, the geographically vulnerable, the politically weak, and other disadvantaged groups will be most affected.

Addressing the intersection between social determinants of environmental change and the effect of environmental change on health inequities will benefit sustainable ecological and population health alike.
To translate this conceptual understanding into action on the social determinants of health, the Commission convened nine thematic knowledge networks: globalisation, health systems, urban settings, employment conditions, early child development, social exclusion, women and gender equity, measurement and evidence, and priority public health conditions. Each network is reviewing evidence of what is likely to work and why. Other key factors such as violence and conflict, food and nutrition, and the environment were investigated. Special consideration is being given to the increase in the world’s ageing population and its implications. Recommendations based on a comprehensive analysis of this work will be reported in the Commission’s Final Report in 2008.

Case studies from low-income, middle-income, and high-income countries are described in panels 1–3. These case studies show the range of social determinants of health, the causes of the causes, and illustrate types of action that can be taken to tackle health determinants—from structural conditions of society to more immediate influences, at all levels from worldwide to local, across government.

The Commission sees action as a truly multi-stakeholder process, including government and non-government organisations, civil society more broadly, including trades unions, political parties, popular movements and alliances, private sector organisations and, crucially, health practitioners themselves. Key to multilevel, multisector action is coherence.

The three panels indicate how a combination of environments—home, school, work, neighbourhood, and health-care system—can unequally expose different groups to factors that damage health. None of them captures all elements of the ideal comprehensive strategy necessary to tackle health inequities. Rather, they illustrate various approaches and show how action on the conditions within the environments can improve people’s material conditions, psychosocial resources, and behavioural opportunities.

Growing, living, and working
The tragedy of infant and child deaths in poor countries is that most are preventable. Child mortality shows a clear social gradient (figure 5).6 Child survival is crucial. But so is the quality of children’s development. More than 200 million children worldwide are not reaching their development potential.7

The Commission’s Early Child Development knowledge network stresses the need for a balanced approach to children’s development, consisting of physical, cognitive and language, and social and emotional components. In addition to economic circumstance, each component of child development is dependent on the nature of the environments in which children exist. A child’s early environment has a vital effect on the way their brain develops. The more stimulating the environment the more connections are formed in the brain and the better the

Panel 3: Multilevel intersectoral action for health—Sweden
Sweden is, in general, a healthy place to live; life expectancy is among the highest in the world and infant mortality among the lowest.8 Comparing absolute levels of mortality for manual and non-manual workers, Sweden has lower health inequities than other European countries.9 Health in Sweden is contextualised by a stable, wealthy democracy with strongly developed social welfare policies broadly based on equal treatment.10 The changing global context, in combination with an economic recession in the early 1990s, is, however, affecting the way work and life are organised. Although health is improving for all groups, health inequalities are growing.

Structural intervention
Norrbotten, an area in the north of Sweden, is characterised by traditional livelihoods in logging and mining. The region has started to see effects of globalisation in the increasing segmentation of traditional sectors, and increasingly precarious forms of employment—indicated by high and rising rates of sickness absence. The region has among the lowest rates of disposable income per person in the country. There are higher rates of death by cardiovascular diseases, suicide and alcohol-related diseases, especially among men. Norrbotten’s unemployment rates are higher and education levels are lower than the national averages. The FRISK Initiative by the governor of Norrbotten is aimed at structural drivers in the field of employment and working conditions. Although concerned initially with sickness absence it now takes an integrated approach to (1) management training with a focus on positive health effects and health promotion; (2) improving the work environment and increasing worker safety; (3) providing information resources for the expansion of professional networks; and (4) supporting the rehabilitation of individuals who have been long-term unemployed.11

Community intervention
A more disease-oriented approach, combining individual and population-level efforts involving multiple sectors, is the Västerbotten Intervention Program. Västerbotten, a county in northern Sweden, had the highest cardiovascular mortality in Sweden. A long-term prevention programme was initiated in 1985 to address this problem. Especially the community intervention in Norsjö has been followed carefully and offers valuable experience for other communities. Contrary to other models, the health sector and its primary health care providers took an active role in the work, including health counselling and food labelling. In the 10-year assessment, the intervention area had a significantly larger decline in cholesterol, systolic blood pressure, and predicted coronary disease mortality.12 People with low education seemed to benefit the most from the prevention programme, suggesting that the reduction of health inequity is possible through this type of programme.
child thrives in all aspects of their life: physical development, emotional and social development, and their ability to express themselves and acquire knowledge. Although fundamentally important for childhood health, early child development also has far-reaching societal effects, with implications for health inequities in adult life.

In both India and Brazil, the approach of the Self-Employed Women’s Association (SEWA, panel 1) and the policy orientation of the Bolsa Familia (panel 2) empower households to break intergenerational poverty through enhanced and more equitable support in childhood.

Brazil and India, like much of the developing world, are undergoing rapid urbanisation. In 2007, 1 billion of the 3 billion people who live in urban settings live in slums. The scale of the urban problem might seem vast and unmanageable. However, urban areas can provide a healthy living environment. Better housing and living conditions, access to safe water and good sanitation, efficient waste management systems, safer neighbourhoods, food security, and access to services such as education, health, welfare, public transportation, and child care are social determinants of health that can be addressed through good urban local governance. SEWA, supported by the World Bank, shows the value of community-driven improvements to the living environment of the urban poor.

For most people in the world, living conditions are largely determined by economic opportunity afforded through the labour market. A major challenge to health is the conditions under which people work. This challenge applies both to working conditions and to the nature of employment contracts and the availability of work itself. In high-income countries, much action has been taken on physical and chemical hazards in the workplace. But with greater segmentation of the labour market precarious employment has become more prevalent. The example from Sweden (panel 3) shows how changing employment conditions towards less job security and control, are affecting people’s wellbeing and health in a high-income country.

In low-income countries, persistent physical and chemical hazards are compounded by high rates of informal employment with negligible labour protection. Employment conditions provide a fertile area for major improvements in conditions of the physical and social environment. In India more than 80% of workers are outside the formal employment sector, excluded both from the protection afforded by labour standards and from whatever social security provisions are linked to formal employment. SEWA represents a good example of collective action among informal workers including collective bargaining and health and social insurance schemes. Producing goods for export, for example in textiles and clothing, provides employment for people in low-income countries. This benefit should not be at the cost of substandard employment conditions that damage health. The price of apparently cheap consumer goods for people in high-income countries should not be poor health in low-income countries.

**Contextualising behaviour**

Contemporary public-health interventions have often given primary emphasis to the role of individuals and their behaviours. The Commission recognises the important role of these factors, but sets them in the wider social context to illustrate that behaviour and its social
patterning, as shown in figures 6 and 7, is largely determined by social factors. Figure 6 shows how cirrhosis associated with heavy drinking is more common in lower socioeconomic groups. Countries with more restrictive alcohol policies tend to have lower levels of alcohol consumption, lower levels of mortality from liver cirrhosis, lower levels of other alcohol-related mortality, and fewer social problems due to alcohol. National tobacco control efforts show the responsiveness of health-damaging behaviours to intersectoral action.

We believe that unless action also addresses the structural drivers of inequity in behaviour, it will not tackle the contribution of these behaviours to health inequities.

A new global trend is the so-called nutrition transition—increasing consumption of fats, sweeteners, energy-dense foods, and highly processed foods. The world now faces a double burden of malnutrition—under-nutrition and over-nutrition, both of which are socially patterned. Addressing nutrition inequities requires action on the structural drivers of food availability, accessibility, and acceptability at the global and national levels.

### Health systems

Although inequities in health result from the social conditions that lead to illness, the high burden of illness particularly among socially disadvantaged populations, creates a pressing need to make health systems responsive to population needs. International, national, and local systems of disease control and health services provision are both a determinant of health inequities and a powerful mechanism for empowerment. Central within these systems is the role of primary health care, as indicated by the community-based programme in Sweden (panel 3).

In some instances, health systems perpetuate injustice and social stratification. In low-income and middle-income countries, public money for health-care tends to go to services that wealthy people use more than poor people. Reforms that lead to charging at the point of use are a disincentive to use of health care. Out of pocket expenditures for health care deter poorer people from using services, leading to untreated morbidity. Such expenditures can also lead to further impoverishment or bankruptcy. The larger the proportion of health care that is paid out of pocket, the larger the proportion of households that are faced with catastrophic health expenditures.

SEWA illustrates an organisational model that provides a safety net for some groups unable to meet acute health-care costs. The conditional cash transfer model of Bolsa Família (panel 2) stimulates uptake of health services that typically do not get to poor communities. While financial support to improve access to and use of health services among the poor is crucial in the short term, the underlying issue for policy intervention is the need to reduce and remove financial barriers to such services. National health systems are pivotal in addressing health inequities; they need to be adequately resourced, function well, and be accessible to all. Appropriately configured and managed health systems provide a vehicle to improve people’s lives, protecting them from the vulnerability of sickness, generating a sense of security, and building social cohesion within society; they can ensure that all groups benefit from socioeconomic development and they can generate the political support needed to sustain them.

Current efforts to revitalise primary health care worldwide should go hand-in-hand with attention to the social determinants of health. Just as a social determinants approach to improving health equity must involve health care so must programmes to control priority public health conditions include attention to the social determinants of health. Such action has to involve multiple sectors in addition to the health-care sector. It is not sufficient, for example, to provide treatment for people with diabetes in middle-income countries and not deal with the drivers of the obesity epidemic; to be concerned with childhood illness and not education of women who will become mothers; to deliver health education to individuals and not be concerned with their poverty; to deal with stress-related illness and ignore the conditions in which people live and work that gave rise to it. Lasting control of tuberculosis requires the combination of treatment and preventive action taking into account biological, health-behavioural, and socioeconomic factors.

### The shape of society

All societies are stratified along lines of ethnicity, race, gender, education, occupation, income, and class. Health inequities result from unequal distribution of power, prestige, and resources among groups in society. We see this very clearly in each of the case studies from India, Brazil, and Sweden. Although at very different stages of economic development, the differentiation of certain groups—be it by gender, caste, education, place, or income—is key to the way health inequity is generated.
At the core of gender health inequity are social norms and structures which support and perpetuate bias. Women account for only 17% of parliamentarians worldwide. The marginalisation of working women in India is substantial. While supporting its members’ material circumstances and working arrangements, SEWA also takes action to challenge the Indian legal system. The emphasis in conditional cash transfer programmes, such as that in Brazil and Mexico, on channelling resources through female household members shows the importance policy places on supporting their role in protecting children’s development and promoting family health.

**The social context**

Economic and social policies affect the distribution of the social determinants of health, including resources for education, health, and financial security. It is clear therefore why the relation between the Ministries of Health and Finance is so crucial to a social determinants view of health. Indeed, recognition of the importance of social determinants of health means that government social policy, not just health policy, is fundamentally important for health equity.

Pro-health equity policies seem to rely in many cases on the State in providing security via welfare programmes and a universal social safety net (figure 8). Sweden has, for much of the post-war period, maintained very strong State-led welfare policies. Redemocratisation in Brazil is associated with a strong commitment to address both poverty and inequity. In India, the material conditions of the vegetable sellers of Ahmedabad can be improved in the short term through local forms of collective action and empowerment. But a more sustained empowerment for workers comes from action at the structural level: through the state and national legislature, and improved access to credit.

Globalisation, with its remarkable acceleration of trade, knowledge, and resource flow, offers unprecedented promise for improving human health. Yet, to date, many feel that this promise has been disappointing. The emphasis placed on globalisation as an engine of economic growth has overlooked or under-estimated the initial conditions of inequality between rich and poor countries, and within them. Where the social institutions through which people share resources are relatively strong and fair, moderate inequality can be constructive, driving the efforts and risk-taking at the micro level that underpin economic success. But where institutions governing the distribution of societal resources are weak, corrupted or structurally inequitable, as they are both within many countries and between the rich and poor regions, inequality can act destructively, suppressing local enterprise and perpetuating impoverishment. So far, the benefits of globalisation have been largely asymmetrical, creating among countries and within populations winners and losers, with knock-on effects on health.

There is great benefit from increased trade openness, increasing inter-dependence among nations, and from an ability to deal with the major issues—environment,
health, security—at a global level. But there is something profoundly wrong in the assumption that all countries come to these new global fora equally equipped. Long historical trajectories bring countries together under globalisation at dramatically differing levels of institutional capacity and strength. A globalisation that does not provide for institutional capacity building among the developing nations is liable to foster and even increase inequity.

The poorest countries of the world, notably in sub-Saharan Africa, receive only small portions of global financial flows. In fact, net flows are increasingly from developing economies to high-income countries (figure 9). As a result, low-income countries rely heavily on official development assistance to finance their health systems and investments in social determinants of health and require more extensive forms of debt cancellation. Official development assistance continues to be important as a source of financing, complemented by more extensive forms of debt cancellation. Aid has the potential to lift as many as 30 million people out of absolute poverty each year, although its effectiveness is undoubtedly affected by issues of delivery. Strengthened social security systems would in the longer term act as a buffer against detrimental health effects of those benefiting less from trade liberalisation.

Although the Commission recognises the contribution that economic growth can make to the availability of resources for reducing health inequities, growth per se is not a sufficient prescription for equitable improvements in population health. Nor is growth with inequality a simple or automatic trade-off. Rather, action within and between countries to mitigate and remove structural, destructive inequality is the necessary counterpart to worldwide growth itself and the policies that aim to support it.

Time for action
We are at a turning point. 60 years ago, in 1948, the establishment of WHO embodied a new global vision, emerging from the ashes of conflict, of universal health at the highest attainable level. 30 years later, in 1978, the community of nations came together again in Alma Ata to call for a new approach to health, founded on a holistic understanding of local primary health-care needs, across the social determinants, and of people-centred action.

In 2008, the end of the Commission as a formal entity will, we believe, be the launch of a global movement, one that perceives equitable health as a societal good, at the heart of which lies social action, and a field in which countries and people, rich and poor, can unite in common cause.

Proponents of health for all have been numerous and vocal around the world. The primary health care movement, though sometimes overshadowed by disease-specific concerns, never died. Indeed primary health care, once again has a central role in WHO’s current agenda. The 1986 Ottawa Charter on Health Promotion, and its renewal in Bangkok, embraced a truly global vision of public-health action. The Latin American social medicine movement and the People’s Health Movement, the General Comment on the Right to Health, and the broad social vision of the Millennium Development Goals, all reaffirm the central importance of health, the need for social and participatory action on health, and the core human value of equity in health.

Building on these efforts, the Commission represents a unique opportunity for action. Whereas in the past, efforts have been fragmented, the Commission for the first time brings together at a global scale actors, experiences, and evidence concerned with social determinants of health and health equity. At the global level, we now understand, better than ever before, how social factors affect health and health equity. And although the need for better evidence remains, we have the knowledge to guide effective action. By linking our understanding of poverty and the social gradient, we now assert the common issues underlying health inequity. By recognising the nature and scale of both non-communicable and communicable diseases, we show the inextricable linkages between countries, rich and poor.

As processes of globalisation bring us closer together as peoples and nations, we begin to see the interdependence of our aspirations: for human security, including protection against poverty and exclusion; and for human freedom, not just to grow and flourish as individuals but to grow and flourish together. And in these aspirations, we recognise the interconnectedness of the causes of health inequity, and the imperative of action which is global, social, and collective.

CSDH Commissioners
Michael Marmot (Commission Chair), Frances Baum, Monique Bégin, Giovanni Berlinger, Mirai Chatterjee, William H Foege, Yan Guo, Kiyoshi Kurokawa, Ricardo Lagos Escobar, Alireza Marandi, Pascoal Mocumbi, Ndiroro Ndiaye, Charity Kaluki Nglu, Hoda Rashad, David Satcher, Amartya Sen, Anna Tibaijuka, Denny Vagero, Gail Wilensky.

CSDH Secretariat

Conflict of interest statement
I declare that I have no conflict of interest.

Acknowledgments
This paper is an abridged version of the Interim Statement of the Commission on Social Determinants of Health (CSDH), available from: http://www.who.int/social_determinants/en. It draws on a large body of work done by CSDH partners. I would like to acknowledge comments and fruitful discussions held with CSDH Commissioners, CSDH

Rasanathan, Gabrielle Ross*, Ritu Sadana, Orielle Solar, Sarah
partners, the CSDH secretariat, and the participants invited to the 8th Commission meeting in Vancouver in June. 2007. The Commission on Social Determinants of Health was set up in March, 2005, by the late Dr Lee Jong-Wook then Director-General of WHO. The endorsement of WHO has been carried forward with the support of the present Director-General Dr Margaret Chan. Valuable financial support has been received from partner countries including Canada, UK, Sweden, Chile, Egypt, India, Kenya, Mexico, Brazil, China, and Japan. This publication contains the collective views of the Commission on Social Determinants of Health and does not necessarily represent the decisions or the stated policy of WHO.

References


51 Gottlieb S. Medical bills account for 40% of bankruptcies. BMJ 2000; 320: 1295.


Barriers to improvement of mental health services in low-income and middle-income countries

Benedetto Saraceno, Mark van Ommeren, Rajaie Batniji, Alex Cohen, Oye Gureje, John Mahoney, Devi Sridhar, Chris Underhill

Despite the publication of high-profile reports and promising activities in several countries, progress in mental health service development has been slow in most low-income and middle-income countries. We reviewed barriers to mental health service development through a qualitative survey of international mental health experts and leaders. Barriers include the prevailing public-health priority agenda and its effect on funding; the complexity of and resistance to decentralisation of mental health services; challenges to implementation of mental health care in primary-care settings; the low numbers and few types of workers who are trained and supervised in mental health care; and the frequent scarcity of public-health perspectives in mental health leadership. Many of the barriers to progress in improvement and development of mental health services can be overcome by generation of political will for the organisation of accessible and humane mental health care. Advocates for people with mental disorders will need to clarify and collaborate on their messages. Resistance to decentralisation of resources must be overcome, especially in many mental health professionals and hospital workers. Mental health investments in primary care are important but are unlikely to be sustained unless they are preceded or accompanied by the development of community mental health services, to allow for training, supervision, and continuous support for primary care workers. Mobilisation and recognition of non-formal resources in the community must be stepped up. Community members without formal professional training and people who have mental disorders and their family members, need to partake in advocacy and service delivery. Population-wide progress in access to humane mental health care will depend on substantially more attention to politics, leadership, planning, advocacy, and participation.

Introduction

International public-health concerns for mental health have been accelerated by the World Development Report 1993 and the subsequent Global Burden of Disease report, which compared health conditions based on combined disability and mortality statistics. Although they did not make any explicit policy recommendations on mental health services, these reports showed, to the surprise and disbelief of many in the international public-health arena, the huge burden of disease imposed by mental disorders, not only in rich countries but also in low-income and middle-income countries.

Data on the global burden of disease prompted three high-profile international reports (table 1), and many important regional and national reports. Notably, one regional report was signed by all European ministers of health, including those from 27 low-income and middle-income countries in Eastern Europe. They committed to implementation of a detailed plan for service development, prevention of mental problems, and promotion of wellbeing. The global, high-level reports were largely concerned with the wellbeing of people affected by mental disorders: they called on decisionmakers to do everything in their

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Extend and improve care</td>
<td>Upgrade the quality of mental health services</td>
<td>Give care in the community</td>
</tr>
<tr>
<td>Improve mental health services for children and adolescents</td>
<td>Improve mental health services for children and adolescents</td>
<td>Provide treatment in primary care</td>
</tr>
<tr>
<td>Develop effective treatment and demand-reduction programmes for substance abuse</td>
<td>Develop effective treatment and demand-reduction programmes for substance abuse</td>
<td>Make psychotropic drugs available</td>
</tr>
<tr>
<td>Strengthen the workforce to provide care</td>
<td>Upgrade amount and quality of training for all health workers</td>
<td>Link with other sectors</td>
</tr>
<tr>
<td>Increase public and professional awareness; reduce stigma and discrimination</td>
<td>Increase public and professional awareness; reduce stigma and discrimination</td>
<td>Develop human resources</td>
</tr>
<tr>
<td>Strengthen other mental health system components that enable provision of care</td>
<td>Create national centres for training (and research) on brain disorders, linked with institutions in high-income countries</td>
<td>Establish national policies, programmes, and legislation</td>
</tr>
<tr>
<td>WMH=World Mental Health, IOM=Institution of Medicine, WHO=World Health Report</td>
<td>This is a summary of the recommendations in the books, rather than of the available summary documents. In addition to these 17 recommendations, the three reports together contain nine further recommendations on social determinants, prevention of mental disorders, and research.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Mental health recommendations from three high-level reports
power to organise care for those with mental disorders. One included extensive analysis of social factors in mental disorders, and mental wellbeing; another emphasised epidemiology and interventions for specific brain disorders; and the third focused especially on mental health policies and services. Despite these differences, the mental health service recommendations offered in these three reports are largely consistent (table 1). Their recommendations on services are of two kinds: direct recommendations to increase the availability of care, and recommendations on mental health system components to enable provision of more or better care, such as development of human resources or changes to mental health policy.

This Lancet Series has reviewed epidemiological evidence, availability of mental health resources, evidence for interventions in mental health, and the status of mental health system development in countries, including the extent of national-level progress in mental health service development. The Series reviewed the availability of mental health resources in 152 low-income and middle-income countries by analysis of data collected in 2001 and 2004. The review suggests that improvements in the availability of mental health resources between 2001 (when two of the high-profile reports were published) and 2004 were only slight. National-level successes (for example, case studies on Brazil, Chile, and Sri Lanka in panel 1) have occurred, though in most countries mental health service development continues to be fragmented and slow.

Despite the wide dissemination of high-level reports and evidence for the range of mental health interventions reviewed earlier in this Series, progress in scaling-up has not been as hoped. What hinders progress? Is it simply insufficient donor interest? If so, why? Or do other political or technical barriers exist? We aim to address these complex questions here to inform the Lancet’s call for action. Many of the barriers and lessons identified in this review will be common sense to experienced public-health experts, but the aim is to make them explicit, so that they become powerful tools for public mental health action.

Key barriers to service development

To understand the challenges to progress in the improvement of the quality and availability of mental health care in most low-income and middle-income countries, we surveyed a range of international experts and leaders. Our methods and results are reported in detail elsewhere, and panel 2 summarises the methods. We discuss the prevailing public-health priority agenda and its effect on funding; the complexity of and resistance to decentralisation of mental health services; challenges in implementation of mental health care in primary-care settings; the limited numbers and types of workers who are trained and supervised in mental health care; and the general scarcity

Panel 1: Sri Lanka: political will for mental health after a major disaster

Disasters can have devastating social and mental health effects. Professional understanding of these effects, paired with post-emergency mental health interest by the public (including media and politicians), can provide enormous opportunities for mental health system development. Before the 2004 tsunami, Sri Lanka’s formal mental health resources were mainly invested in mental hospitals in and around the capital, Colombo. Despite many efforts and important initiatives, mental health advocates (who were often at odds with one another) struggled to put mental health on the development agenda.

The political interest in and priorities related to mental health dramatically changed after the tsunami, which killed 35,322 Sri Lankans and displaced about 1 million people. A presidential committee was set up immediately, to provide support for mental health relief. During the following months, aid agencies offered various short-term mental health and social supports to Sri Lankans. Immediately after the emergency, the Ministry of Health, with sustained support from WHO, took steps to maintain the interest in mental health by initiation of a policy-development process. 10 months after the disaster, after consultation with a broad range of health-sector stakeholders, the Sri Lankan government approved a new, consensus-based National Mental Health Policy. The new policy guides efforts to strengthen the governance, management, and administration of mental health services and to reconfigure the organisation of mental health services so that care will be locally available in all districts. The policy emphasises human-resource development by outlining the appointment of different types of staff to all districts. The national policy’s vision includes acute inpatient care by a multidisciplinary team in each district. With the acute unit as a basis, the vision encompasses both fixed and mobile outpatient clinics throughout each district, and training and supervision of workers in mental health care, as was already happening in a few districts before the tsunami.

By the end of 2007, 18 out of 27 districts (67%) will have functioning acute inpatient units within general hospital settings, compared with 10 out of 27 (37%) before the tsunami (figure). The political will for mental health in Sri Lanka continues with the Ministry of Health’s active interest in new mental health legislation.

Figure: Construction of a mental health unit in Kalmunai, Sri Lanka

Staff organise acute inpatient care and outreach clinics across Kalmunai, which is the district that has been most severely affected by the tsunami.
Many challenges remain. First, innovative measures will be needed to employ qualified professionals in districts far away from Colombo. Such measures are being identified and implemented. The Ministry of Health is working with the national College of Psychiatrists to develop a new cadre of diploma holders, who will have 1 year of postgraduate training in psychiatry. The ministry will appoint these trained diploma holders to work in all districts to coordinate and provide mental health care. The idea is an extension of another innovative Sri Lankan invention, medical officers for mental health, who have 3 months of training in psychiatry. These personnel have been key in the provision of psychiatric care in many districts in recent years. Their numbers will increase from the present 57 to 257. A second major challenge is Sri Lanka’s colonial legacy—ie, about four of every five psychiatry beds in the country are in large mental hospitals near Colombo. A WHO survey that used the Community Placement Questionnaire showed that 978 (58%) of 1678 long-stay patients in mental hospitals could immediately leave the hospital and lead normal lives outside these facilities if they had places to stay and sufficient support in the community (Mahoney J; unpublished data, 2006). Negotiations with and recruitment of donors and local non-governmental organisations to provide such support is underway. In 2006, a nurse from every ward in one of the large hospitals in Sri Lanka was trained in rehabilitation in Bangalore, India, which substantially improved the quality of care. Other large psychiatric hospitals will soon be included in similar programmes. The active nurturing of opportunities in the aftermath of a major tragedy has resulted in substantial changes in the Sri Lankan mental health system.

Panel 2: Search strategy on opinion of experts and leaders

We surveyed a select group of international experts and leaders—all with experience and knowledge of low-income and middle-income countries. Our qualitative survey included seven open-ended questions about barriers and facilitating factors for mental health funding and development of mental health services in low-income and middle-income countries. Of 60 people we asked to participate, 50 (83·3%) responded. Seven additional individuals spontaneously submitted responses to the survey after it was shared with them by their invited colleagues. Our analyses cover these 57 responses. Accounting for multiple affiliations, we gathered the opinions of 12 current or previous senior national-government decisionmakers on mental health, eight civil-society leaders without training in mental health, 13 civil-society specialists or leaders in mental health services, three general public-health leaders, 20 associate or full professors, and 20 current or previous international health advisers or consultants on mental health services. These international advisers or consultants on mental health services included one current and four previous WHO Regional Mental Health Advisers. At the time of survey, respondents were based in 30 countries, including 18 low-income and middle-income countries. In total, we recorded 90 848 words of responses.

Two data analysts (RB and DS), who were unfamiliar with the international discourse on mental health services, analysed the text thematically. To reduce the risk of bias, these two analysts were unaware of the identities and affiliations of respondents during this initial stage. They read the texts separately, then conferred and agreed on themes that they had each independently identified. They then reread all responses to categorise text that related to the agreed themes. Since this analysis focused on broad themes, it only encompasses a subset of all issues raised by respondents. Further in-depth analyses would probably identify further barriers and facilitating factors.

Limitations of our analyses include the fact that the views expressed are those of senior experts and not of grass-roots workers, consumer groups, or other stakeholders. Also, the analysis does not cover the broad historical, geopolitical, and sociodemographic contexts in different countries. Finally, the review is focused on development of mental health services for people with mental disorders, and thus does not cover barriers to prevention of disorders, or to protection and promotion of psychosocial wellbeing in the general population.

(Continued from previous page)

The public-health priority agenda and its implications for funding

In response to a question about available funds for mental health services, respondents said that mental health had a low position on public-health agendas at national and international levels. They were concerned that mental health was named neither as a Millennium Development Goal (MDG) nor as an MDG-related target, despite established links between mental disorders and MDGs. Absence from the international public-health agenda can block progress even when investment in mental health has been agreed at the national level, as was the case in Rwanda:

“Rwanda, recognising the impact of the 1994 genocide as well as the rising rates of HIV infection, included mental health in the 2002 Poverty Reduction Strategy Paper document. However, when it came time to determine what will be financed within the Poverty Reduction Strategy Credit, mental health was not included, since it is not explicitly mentioned as an MDG. The result is that the Rwandan Ministry of Health cannot finance mental health services out of the World Bank loan/credit funds, even if mental health is an expressed need, an observed need, and a mental health strategy exists.”

F Baingana (formerly of the World Bank)

A raised profile on national and international agendas is not only essential for augmentation of funds but also for generation of the political support needed for the difficult decisions that are often part of mental health services reform.

Respondents identified a range of factors to explain the inadequacy of funding for mental health, which we take as a proxy indicator for a low position on the agenda. First, advocates for mental health might have different perspectives, which leads to contradictory messages. One observer, who was from outside the mental health profession, P Salama (Health Section, UNICEF), noted that “the field has suffered from a real and perceived lack of consensus among leading experts. This turns donors and policymakers off.” Because there are many types of mental health problems, advocates for mental health often lobby against one another to draw attention to different mental health problems (eg, severe mental disorders vs trauma-induced disorders vs lack of wellbeing), each of which might need different public mental health solutions. Yet, even when advocates agree on problem definition (eg, a focus on severe mental disorders), they too often offer competing views on the type of actions needed to address the problem, despite the evidence base that exists for specific interventions. Respondents commented that fragmentation in advocacy by different stakeholders—including governmental and non-governmental organisation service-providers,
consumers, family members, professional associations, leaders in mental health from non-governmental organisations, academics, and staff of international agencies—has prevented progress in many countries. R Jenkins (Institute of Psychiatry, London, UK) notes that senior psychiatrists in low-income and middle-income countries often prioritise increased expenditure on specialist services, which can be perceived by donors as “special pleading, not a priority, and not serving population needs.”

Second, respondents argued that advocacy and other communications by mental health practitioners might not be clearly understood by others. The concepts in mental health discourse are complex, diverse, often theoretical, and not clearly communicated to decisionmakers. D do Nascimento, former Director of mental health of Brazil, noted that “Mental health professionals have a hermetic discourse, difficult to understand by their colleagues in other sectors of health care. Some effort is needed to simplify this discourse.” Many advocacy efforts could have failed because they were not sufficiently clear.

Third, respondents pointed to the perception that mental health indicators were not sufficiently strong. Despite the measurability of mental disorders and the various components of mental health systems, respondents noted that mental health does not use indicators that are as tangible and convincing as, for example, mortality or vaccination coverage. This Lancet Series recommends that a set of simple, consensus-based indicators be monitored to track the progress of countries towards attainment of specific targets.

Fourth, respondents argued that advocacy for mental health has been weak because people with mental health problems and their families are too often invisible, voiceless, or at the margins of society. People with mental disorders and their families in low-income and middle-income countries are only rarely mobilised to form powerful constituencies, and to press for the availability of effective and humane mental health care. However, there are notable exceptions, such as in Zambia.

“The formation of a consumer movement in 2000 and involvement of family members played a very important part in mental health reforms. With clients and family members as stakeholders in mental health, there was demand that medical treatment should be a basic right for persons with mental health problems. Furthermore, in providing treatment and in protecting patients, we demanded that basic human rights must be protected. These demands led to some reforms in mental health as the government strived to respond to our demands.”

C S Katontoka (Mental Health Users Network of Zambia)

Although few low-income and middle-income countries have powerful consumer movements, this example shows the unmet potential represented by mobilisation of service users to ensure that their concerns are heard by decisionmakers.

Fifth, the general public’s interest in the well-being of those with mental disorders was reported to be low. L Vijaykumar, of the Indian non-governmental organisation, Sneha, noted the absence of a “ground swell of public opinion on mental health issues which will force the governments to allocate more funds for mental health”. On the contrary, respondents argued that stigma—which is common in the general population and in the health sector—is a barrier to progress.

Sixth, respondents suggested that advocacy might fail because decisionmakers often have the incorrect perception that mental health care is not cost effective relative to care of many other conditions.

“In Afghanistan...the national health authorities defined mental health as a priority, but the donor community had huge hesitations to fund service delivery. The main reason stated for this reluctance of the donors was the unavailability of clear studies about the cost-effectiveness of public mental health interventions. The institution charged with ‘costing’ the Basic Package of Health Services...said they could not provide the government and donors with data on the estimated costs and benefits.”

P Ventevogel (HealthNet-TPO)

Perceptions of insufficient gains from investment in mental health are not only common among international donors but also among national-level decisionmakers, who might “come to the World Bank and express the need for funding for mental health, but are not willing to take a loan for these activities. Policymakers [...] believe that mental health care is a ‘charity’ issue. They do not believe that there will be a return on investment” (F Baingana).

Cost-effectiveness, for which data are increasingly available, varies for different mental disorders and problems. For example, antiretroviral treatment for HIV/AIDS, which is firmly on the international public-health agenda, is as cost effective as treatment for depression delivered in primary care, which has not been widely implemented. Decisionmakers usually do not have up-to-date knowledge about the cost-effectiveness of mental health care and, thus, they often direct funding toward less cost-effective care. Respondents to our survey noted that, in many countries, scarce mental health funds are spent on long-term institutional care at mental hospitals and on new, patented, pharmaceuticals which, in general, are much less cost effective than community-based care and generic essential medicines.

**Organisation of services**

The way in which mental health services are organised affects treatment coverage for people with diverse mental disorders. Respondents were asked how low-income and middle-income countries should invest their scarce resources for mental health care. Many respondents discussed decentralisation of tertiary-care institutions,
development of community-based rehabilitation services, psychiatric care in general hospitals, and mental health care in primary-health care and other health-care settings.

The centralised location of most mental health resources (staff, budgets, and beds) in or near large cities was described by respondents as a key barrier to progress. Respondents reaffirmed the long-standing public-health and public mental health recommendation that resources for care need to be geographically decentralised so that care is available and accessible in the community.18

Most respondents were critical of large psychiatric institutions; some argued that progress was hindered in countries with such institutions. The respondents' main arguments were that institutions tend to consume a large proportion of scarce mental health resources (budgets, beds, and staff); have higher costs than care in the community; isolate people from support systems in their families and communities; and are associated with undignified life conditions, human-rights violations, and stigma. Many respondents recommended that large institutions should be downsized, or even closed.

Barriers to decentralisation and deinstitutionalisation are sizeable: despite long-standing recommendations,16–19 four of five psychiatry beds in low-income and middle-income countries are still in mental hospitals.18 Challenges to downsizing mental hospitals tend to be intertwined with challenges to development of community mental health services. Developing such services requires access to mental health resources that are mostly allocated to large psychiatric institutions. Yet, to avoid homelessness and neglect, many long-stay patients can only leave hospitals when care and support have been made available in local communities. Typically, additional funding will be needed during the transition to community care.

The vested interests of mental health professionals and hospital workers might be one of the most pervasive barriers to decentralisation. Hospital directors might fear that deinstitutionalisation threatens their power base. Mental health workers might fear forced relocation to rural areas, and might ask their trade unions to protest against policies that favour community-based care. According to A Minoletti (Ministry of Health, Chile):

“The main barrier for downsizing psychiatric hospitals is the high political cost that this entails, due to the pressure from the trade unions of hospital workers and organisations of mental health professionals (who should learn new skills for community care and may also lose some of their present privileges). In relation to the above, there are no professionals appropriately trained to be leaders of the process of downsizing psychiatric hospitals and face its technical challenges and social and political barriers.”

D Puras (Vilnius University, Lithuania) describes his experience in Eastern Europe and Central Asia:

“Over many decades the ineffective self-feeding system of centralised psychiatric institutions has developed sophisticated skills of survival and resistance...The system is controlled by a powerful lobby of administration of psychiatric institutions, which have good relations with the political and academic establishment. Ideologically, the system is supported by a still prevailing culture of paternalism and dependence, which is based on the presumption that mentally ill people are not capable to make independent decisions, so psychiatrists and other staff need to take care of them in a very paternalistic way... Even service users and family organisations are often on the side of the traditional system, because they do not know about alternatives or get financially dependent on organisations or institutions lobbying for institutional care and the biomedical paradigm.”

Moreover, division within government systems can hinder decentralisation. Two respondents, reflecting on experiences in Pakistan and South Africa, stated that successes in reform of policy or legislation at the national level do not necessarily translate to improvements in services in provinces or districts. Authorities at these levels of government, who were responsible for implementation of national policy and legislation, continued to allocate insufficient resources to develop mental health services. Similarly, respondents noted that, in some countries, competition between the government branches responsible for hospital services and community health have hindered transference of human and financial resources to community care. Respondents argued for international technical assistance, because decentralisation of resources is seen as technically complex, and countries might be able to learn a lot from others that have successfully implemented community care.

Broad agreement emerged among respondents that a mixed model of services that prominently included mental health in primary health care would best serve the millions of people with mental disorders. H Whiteford (formerly of the World Bank) described a systems approach, which was the most common viewpoint among respondents:

“I would argue for both an upskilling of the primary health care workforce in mental health and the expansion of specialist community mental health services. There will never be sufficient community mental health services to treat all people with mental illness so there cannot be a sole emphasis on these services. However primary care services cannot adequately diagnose and treat patients with serious mental illness without the support of specialised services. I believe the mutual interaction and support each of these service components gives to the other produces an outcome which justifies the resource implications of expanding both.”

Although some respondents described model examples of integration of mental health into primary health care,
many discussed the past failures of attempted integration with primary health-care systems\(^{5}\) (see also panel 3 on Nigeria). They identified three key barriers. First, primary health-care systems in low-income and middle-income countries tend to be overburdened with multiple tasks and patient loads, and primary health-care workers do not always have the necessary time to provide proper care for people with mental disorders. Second, primary health-care workers do not receive sufficient supervision and support by specialised services for the effects of training to be sustainable. This observation is in line with the 1978 Declaration of Alma-Ata,\(^4\) which promoted a primary-care model “sustained by integrated, functional and mutually supportive referral systems” as an integral part of the country’s health system. This influential declaration does not seem to have been closely read by many of the mental health leaders, who have tried to develop mental health as a free-standing activity in primary health-care settings, since the most common strategy for organisation of mental health in primary health care has been short-term training of workers without meaningful follow-up or supervision and without cultivation of district-level specialised services to act as backup. Training primary health-care staff in mental health care without considering their workloads and supervision can cause help-seekers to be exposed to inappropriate treatment. Respondents working in Afghanistan and the occupied Palestinian territory expressed concern that the ease of prescribing medicines in primary health-care settings can lead to over-prescription when workers neither have the skills nor time to differentiate between normal distress and disorder, and cannot offer or organise psychosocial supports. Paradoxically, a third barrier raised by respondents was that in many low-income and middle-income countries essential psychotropic medicines are not continuously available through primary health care, which can hinder appropriate care for people whose disorders can be effectively treated with medication.

**Human resources for mental health**

One well-established barrier to scaling-up of mental health services is the inadequate number of people who are trained to provide care.\(^{5,6}\) Respondents pointed out that in many countries poor working conditions and low status of the profession mean that few people enter the mental health professions. At the same time, higher salaries in private practice and overseas mean that scarce psychiatrists are encouraged to leave governmental employment. Moreover, mental health professionals—whether they are psychiatrists, nurses, or social workers—have few incentives to live in rural areas where most people in low-income and middle-income countries tend to live.

Respondents suggested that more flexibility and creativity was needed to diversify the workforce, as far as possible by building on existing formal and non-formal resources (panel 4). Some argued that family members are the prime resource for care, and several advocated support for, and even formalisation of, their caregiver role. Yet, professional institutions might resist such flexible solutions. For example, P Delgado (Ministry of Health, Brazil), discussed the decision to train general medical doctors to take on the role of psychiatrists in small municipal districts. He wrote:

> “This public health decision found great resistance in the medical establishment. As a result, it has been subject to continuous negotiations.”

Respondents described the mental health training received by general health workers during their formal

Panel 3: *Nigeria: An ill-fated attempt to integrate mental health into primary care*

The high rate of mental health problems and the associated disability and burden in general health-care settings and in the community in Nigeria have been well-documented.\(^{4,6}\) Recommendations by the WHO in the 1970s\(^{4,6}\) provided the necessary impetus for policymakers in Nigeria to consider integration of mental health into primary care. A programme to do so was formulated as part of the National Mental Health Programme and Action Plan, which was formally published as a policy document in 1991.\(^6\) By promulgation of this programme, mental health has become the ninth component of the nation’s primary-care service. The programme envisages that mental health services be scaled-up so that essential treatment, including psychotropic medications, is available to those in need in the community. Services are to be delivered by trained primary health workers, with coordinated supervision provided by specialist mental health workers.

15 years later, the programme has not had the desired effect on provision of mental health service to Nigerians. The service reaches only a minority of those in need; estimates suggest that fewer than 20% of people with mental health problems receive any services.\(^6\) Of those who do receive service, hardly any get adequate treatment, even though research shows that evidence-based interventions can be delivered at affordable costs in the country.\(^4\) The programme’s laudable goal—to reduce stigma of mental disorders in the community and improve the knowledge and attitude of primary-care workers about mental health—has not been a success.\(^4,6\) Most primary-care settings still do not have the basic psychotropic medications that were included in the essential drug list.

Many reasons can be advanced for the failure of the programme. First, primary-care workers remain poorly trained and supervised in mental health issues. Crucially, the Nigerian programme does not articulate a structured and clear link between primary-care workers and specialist mental health professionals. No effort was made to first develop secondary mental health services to sustain mental health training or support for primary-care services. Inadequate support for the primary health workers trained in basic mental health care could have resulted in their isolation and poor morale. Second, mental health services, in general and primary or community mental health service, remain especially poorly funded. Also, 91% of the 2005 mental health budget is directed at mental hospitals. This lopsided interest is further exemplified by the recent development of new stand-alone mental hospitals in the country, with the result that more than 80% of psychiatric beds are now located in mental hospitals. Third, political will to improve mental health services might not have yet increased. For example, no serious effort has been made to appoint designated senior officials to oversee mental health issues at the Ministries of Health across the nation. This official neglect receives little civil-society attention. Non-governmental organisation is active in Nigeria with advocacy for mental health reforms or the rights of mentally ill persons as its goal. The experience in Nigeria shows that although the goal of integrating mental health services into the primary-care system sounds attractive, implementation might fall below expectations because of inadequate planning and implementation of policy objectives.
Respondents were concerned that many national public mental health leadership methods to train and supervise others. and will require specialists to be trained in adult-learning services in low-income and middle-income countries, of specialists is essential to reforming mental health mental health problems. Accordingly, redefining the role of specialists was not redefined, and that they should only take clinical responsibility for people who present with complex mental health problems. Accordingly, redefining the role of specialists is essential to reforming mental health services in low-income and middle-income countries, and will require specialists to be trained in adult-learning methods to train and supervise others.

Public mental health leadership

Respondents were concerned that many national mental health leaders have insufficient public-health skills, and that this might hinder rapid progress of service development. In the words of one respondent:

“Leadership cannot be expected from clinicians turned-by-default-into-administrators/planners. Their views, experience, and training are not compatible with population-oriented mental health action.”

I Levav (formerly WHO Regional Office for the Americas)

Mental health leaders in low-income and middle-income countries have responsibility for complex tasks such as development of policy, strengthening of services, and advise on population-level interventions to prevent mental health problems. In addition to general management and leadership skills, these tasks require a population-wide vision. Many respondents noted the absence of mental health leaders with experience or training in public health in many low-income and middle-income countries. Often, senior psychiatrists who are promoted to become national mental health leaders focus on clinical management of individuals, rather than on population-oriented actions. Although psychiatrists might resist the promotion of non-psychiatrists as leaders, several respondents recommended appointment of general public-health leaders in mental health leadership positions.

Respondents identified various reasons that leaders tend not to have public mental health skills. In many

Panel 4: Expert opinions on mobilisation of non-formal resources

“In my view, the first principle to guide every development of services would be to build on what already exists. It looks as if up to now, extended families and neighbours are doing a lot. In that sense, support to families and communities should be the first move. Any plan should be implemented with the view of examining, in the first place, how the local communities can contribute to this plan. This means that the actions should not be so much oriented towards more services but towards more ‘resources’, including informal resources and formal services.”

C Mercier (University of Montreal, Canada)

“Family members in community mental health care must be recognised as KEY resource to community mental health system. Therefore psychosocial education can be promoted to impart some skills/knowledge on how to manage the burden of mental illness and increase their effectiveness as care providers.”

C S Katontoka (Mental Health Users Network of Zambia)

“I am a firm believer in expanding the category of mental health care givers. I am enthusiastic about the training of hairdressers, barbers and priests to recognise the simple mental health disturbances and refer them for further therapy if necessary.”

G Alleyne (Pan American Health Organisation)

Panel 5: Barriers to effective mental health training

“Training has to take place where people go for services. Often, medical students and psychiatric residents are trained in mental hospitals. The same goes for nursing students...The content of their teaching is often non-relevant for the programs and services that are needed in low and middle income countries.”

I Levav (formerly WHO Regional Office for the Americas)

“Theoretical training without continuous on the job supervision is a very poor investment. In and out short courses, even with excellent trainers and on vital topics tend to be a waste of time without some form of follow up.”

L Jones (International Medical Corps)

“As for the training process I am becoming less and less enthusiastic for the workshops, especially with ‘training of trainers’ notions. Often a lot of money is spent on workshops and not much output comes and there is seldom any follow-up.”

I Patkai (Christoffel-Blindenmission)

“Training is perceived as a quick fix...Our experience here is that the best form of training is in the form of ongoing hands-on supervision, problem solving and managing the sometimes huge structural constraints that can turn the training into practice.”

R Giacaman (Institute of Community and Public Health, occupied Palestinian territory)
countries, general public-health training and health-services delivery have never addressed mental health, which has been left to psychiatrists. Internationally, few universities offer courses in public mental health to train future mental health administrators, planners, and leaders. Only a few international training and exchange opportunities exist to strengthen public mental health skills for leaders. Moreover, mental health leaders often have many clinical and hospital-management duties, in addition to private practices (to boost their meagre governmental salaries); as a result, they do not have the time or incentives to develop their public mental health knowledge through self-study or courses.

**Lessons learned**

Many lessons can be drawn from our survey about barriers to mental health service reform. First, many of the barriers to progress in development of mental health services can be overcome by generation of sufficient political will to improve availability of and access to humane mental health care. The words “politics” and “political” were repeated 145 times in the answers of the 57 respondents in our survey, without being prompted by use in the survey questions. Political will, in this context, refers to the inclination, shaped by convictions or incentives, for policymakers to take action and to make or block change. Political will is likely to be directly affected by national and international factors, such as lobbying by professionals, consumers’ groups, and other advocacy groups; expressions of public opinion; and donors’ political priorities. Factors that affect political will can be divided into three categories: the national political environment, domestic advocacy, and transnational influence. Agents at all three levels create incentives and norms that influence the behaviour of policymakers. As we have seen, mental disorders are usually low on the public-health priority agenda of national and international agencies and donors. At the national level, political will in government ministries responsible for health and social welfare is necessary to counter the resistance of various groups with vested interests—whether trade unions, managers of government departments, or professional associations—who might object to reforms. Strong political support is needed to realise modest innovations, such as acceptance of unconventional solutions for diversification of the workforce; creation of mental health units in ministries of health; appointments (where necessary) of public-health experts in mental health leadership positions; collaboration with ministries of social welfare; engagement of all relevant stakeholders to ensure community-based housing and livelihood supports for people with severe mental disorders; and implementation of powerful legislation and policies that protect people with mental disorders from human-rights violations.

Second, advocacy for people with mental disorders needs to be substantially improved and expanded. Advocacy for mental health services will be more likely to succeed if such advocacy is informed by much needed research on the factors that shape political will for improvement of mental health services among different types of policymakers. Moreover, advocacy has not been sufficiently clear, informative, consensus-based, or focused. This observation has implications for national-level mental health planning. Indeed, few countries have consensus-based national mental health plans that have been written in consultation with key stakeholders, including non-governmental organisations, representatives of clients and consumers, and sectors other than health. Yet, in our experience, and according to our surveyed experts, such plans are vital—not just because sound planning is invaluable to successful development, but also because consensus-based plans are forceful vehicles for advocacy. By functioning as a coherent proposal for services, a well-developed national plan for mental health, that has been developed in a participatory way by the government, and with the participation of all key stakeholders, can lead to progress. For example, national-level consensus on mental health services in Albania, Sri Lanka, and the occupied Palestine territory has ensured support from both within these countries and from international donors. Such plans need to be developed over a short timespan in a
participatory manner to communicate to political decisionmakers and funding sources that mental health stakeholders can agree and act swiftly, and to create the necessary momentum for implementation.

Third, development of secondary care-level community mental health services should be prioritised. Although this review does not cover the detailed technical aspects of developing mental health services,37 survey respondents offered observations. They argued that mental health care delivered via primary health care and non-formal community resources require supervision and specialist back-up support, and that downsizing mental hospitals requires availability of a range of services and supports in the community. From these observations we infer that specialist community mental health services should be developed first when creating a mental health system in a district or province, to support responsible downsizing of mental hospitals and to sustain mental health investment in primary health-care clinics, which is essential for proper population coverage. Nonetheless, investment in primary care or existing tertiary care (eg, improvement of conditions in mental hospitals) is vital, and opportunities to invest in such care should be taken. Yet, such investments will probably go furthest if they are preceded by, or are at least in tandem with, development of community mental health services.

The fourth and final lesson of our review is an old lesson: people responsible for service development need to be much more effective in the way they use formal and informal resources that are already available in the community. The need for deinstitutionalisation and decentralisation of resources was covered in detail in the 2001 World Health Report.3 The suggestion by respondents in our survey that specialist staff should be used mainly as supervisors, rather than as clinicians, was also raised by the report of the Institute of Medicine.3 Moreover, our review highlighted substantial unused opportunities to engage non-formal human resources. The scarcity of formally trained mental health professionals in many low-income and middle-income countries suggests that more action is needed to ensure that non-professional community members take part in mental health programming. Survey respondents repeatedly advocated not only training and supervision for general health workers but also involvement of people with mental disorders, their family members, and other non-formal resources in the community. This viewpoint is consistent with the use of participatory action methods, which are common practice in community development.51,4 These approaches have been increasingly applied to develop community mental health care for people with severe mental disorders in a range of low-income and middle-income countries.52–57 Moreover, networks of people with mental disorders—organised into movements such as the Pan African Network of Users and Survivors of Psychiatry—promise to play a substantial part in increasing the availability of humane care in low-income and middle-income countries. Non-formal community resources will need to be recognised and mobilised to ensure access to care for the millions of people who need it. Accordingly, researchers who implement the research agenda described in the call for action in this Series58 will need to design innovative research that involves use of non-formal community resources.

Our findings are largely consistent with existing mental health policy recommendations (table 1). The major difficulty has not usually been policy but its implementation. Most notably we highlight misinterpretation of the Alma-Ata Health for All declaration46 to mean that development of mental health in primary health care can be a free-standing activity. Our review sends a clear message to all stakeholders involved in implementation of the call for action:56 scaling-up of evidence-based mental health interventions will depend on strengthening mental health components of many levels of the health system, together with renewed attention to politics, leadership, planning, advocacy, and participation.

Contributors
BS initiated and provided overall supervision of this review, including the formulation of its framework and research question. BS and MVo designed the expert survey. BS, MVo, and CU drafted the questionnaire. RB and DS were responsible for qualitative data analyses. RB drafted the background paper with inputs from all authors. JM and MVo drafted the Sri Lanka case study, and OG drafted the Nigeria case study. AC compared the three high-level reports. All authors have seen and approved the final version. The final article was written by MVo with input from all authors. The views expressed in this review are those of the authors, and do not necessarily represent the decisions, policies, or views of the institutions which they serve.

Survey respondents
Atalay Alem (Addis Ababa University, Addis Ababa, Ethiopia); George Alleyne (Pan American Health Organization (PAHO), Washington, USA); Florence Bainienga (Makerere University, Kampala; formerly World Bank, Washington, DC and Ministry of Health, Kampala, Uganda); Gary Belkin (Millenium Villages Project, New York and New York University, New York, USA); Jafar Bolhari (Institute of Psychiatry, Tehran, Iran); Thom Bornemann (Cartier Center, Atlanta, USA and formerly WHO, Geneva, Switzerland); John Bowis (European Parliament, Brussels, Belgium); Bhargavi Dhar (Bapu Trust, Pune, India); Mike Davies (Christoffel-Blindenmission, Manila, Philippines); Paramesheva Deva (Mara University of Technology, Shah Alam, Malaysia and formerly WHO, Manila, Philippines); Joop de Jong (City of Amsterdam, Amsterdam, and formerly Transcultural Psychosocial Organisation (TPO) and HealthNet-TPO, Amsterdam, the Netherlands); Pedro G Delgado (Ministry of Health, Brasilia, Brazil); André Delorme (WHO Collaborating Centre for Research and Training in Mental Health, Montreal, Canada); Domingos Savio do Nascimento (formerly, Ministry of Health, Brasilia, Brazil); Mervyn Freeman (Human Sciences Research Council, Pretoria, South Africa; and formerly Ministry of Health, Pretoria, South Africa); Rita Giacaman (Institute of Community and Public Health, West Bank, occupied Palestinian territory); DS Goel (formerly Government of India, Delhi, India); Karen Hetherington (WHO Collaborating Centre for Research and Training in Mental Health, Montreal, Canada); Gaston Harmois (WHO Collaborating Centre for Research and Training in Mental Health, Montreal, Canada); Rachel Jenkins (Institute of Psychiatry, London, UK; formerly UK Ministry of Health, London, UK); Lynne Jones (International Medical Corps, Los Angeles, USA); C. Sylvester Katontoka (Mental Health Users Network of Zambia, Lusaka, Zambia); Pikko Lahti (formerly Finnish Association for Mental Health, Helsinki, Finland); Izhak Levav...
Conflict of interest statement
We declare that we have no conflict of interest.

Acknowledgments
We thank Alastair Ager (Columbia University, New York, USA) and Mitchell Weiss (University of Basel, Switzerland); and Harvey Whiteford (University of Queensland, Australia); Naotaka Shinfuku (Seinan Gakuin University, Fukuoka, Japan); formerly WHO, Geneva and World Psychiatry Organisation, Geneva, Switzerland; Emran M Razaghi (Ministry of Health, Teheran, Iran); A Rahimi Movaghar (Ministry of Health, Teheran, Iran); Srinivasa Murthy (WHO Iraq, Amman, Jordan; formerly WHO, Cairo, Egypt and WHO, Geneva, Switzerland; formerly National Institute of Mental Health and Neurosciences, Bangalore, India); Ravi Narayan (People’s Health Movement and Society for Community Health Awareness, Research and Action, Bangalore, India); Istvan Patkai (Christoffel-Blindenmission, Manila, Philippines); Pau Perez-Sales (Community Action Group and Hospital La Paz, Madrid, Spain); Michel Perreault (WHO Collaborating Centre for Research and Training in Mental Health, Montreal); Michael Phillips (Beijing Suicide Research and Prevention Center, Beijing, China); Dainius Puras (Vilnius University, Lithuania); and Harvey Whiteford (WHO, Geneva, Switzerland), for reviewing drafts of the manuscript. BS Mitchell Weiss (University of Basel, Switzerland); and Harvey Whiteford (University of Queensland, Brisbane, Australia; formerly World Bank, Washington, USA).

References
9 National Steering Committee for Mental Health. The policy and the operational plan for mental health services development in Albania, Tirana, Republic of Albania: Ministry of Health, 2003.
28 Sartorius N, Kaelber CT, Cooper JE, et al. Progress toward achieving a common language in psychiatry. Results from the field trial of the clinical guidelines accompanying the WHO classification of mental and behavioral disorders in ICD-10. Arch Gen Psychiatry 1993; 50: 115–124.


Energy and Health 4

Energy, energy efficiency, and the built environment

Paul Wilkinson, Kirk R Smith, Sean Beevers, Cathryn Tonne, Tadj Oreszczyn

Since the last decades of the 19th century, technological advances have brought substantial improvements in the efficiency with which energy can be exploited to service human needs. That trend has been accompanied by an equally notable increase in energy consumption, which strongly correlates with socioeconomic development. Nonetheless, feasible gains in the efficiency and technology of energy use in towns and cities and in homes have the potential to contribute to the mitigation of greenhouse-gas emissions, and to improve health, for example, through protection against temperature-related morbidity and mortality, and the alleviation of fuel poverty. A shift towards renewable energy production would also put increasing focus on cleaner energy carriers, especially electricity, but possibly also hydrogen, which would have benefits to urban air quality. In low-income countries, a vital priority remains the dissemination of affordable technology to alleviate the burdens of indoor air pollution and other health effects in individuals obliged to rely on biomass fuels for cooking and heating, as well as the improvement in access to electricity, which would have many benefits to health and wellbeing.

The built environment includes the buildings in which people live and work, and the spaces and infrastructure in cities, towns, and villages. It is where most human activity takes place, where most energy services are used, and where many of the advantages and disadvantages of energy use arise.

The world is becoming increasingly urbanised. In 1950, only 30% of the world’s population lived in urban areas; currently the proportion is almost 50%.1 Net population growth of the next few decades will nearly all accrue in the urban centres of developing countries.2 With urban and industrial development comes growing demands for energy and rising expectations of material goods.

This article analyses the connections between the built environment, energy, and human health. The global context is the need to ensure the adequate, equitable, and secure access to clean and safe energy for all individuals, while minimising greenhouse-gas emissions.

Energy efficiency—an important goal for health

Efficient use of energy is seemingly a very attractive means to reduce energy-related effects on the environment and health. To achieve the same services with less energy use should, in theory, reduce burdens on infrastructure, decrease occupational risks, lower costs, cut emissions of local pollutants and greenhouse gases, and lessen harmful exposures. Efficiency improvement also seems to have enormous potential: currently only 20–30% of the chemical energy of the fuel burned is typically transformed to useful work or heating (figure 1).2

Although behavioural factors have a part in such use, greater energy efficiency—ie, higher ratio of useful energy output to input energy—essentially means more efficient technology. In society, as efficiency rises, the direct

Key messages

- Global trends in urbanisation and industrial development will probably be a continuing major driver of increasing fossil-fuel use over coming decades
- Technological advances have a contribution to reversing raised greenhouse-gas emissions, but evidence of past trends shows even notable improvements in technological efficiency tend to be accompanied by increased, not reduced, use of fossil fuels—underlining need for instruments to promote decreased energy use or to decarbonise use
- Nonetheless, evidence indicates health benefits from improved energy efficiency—eg, in home environments and protection against temperature-related morbidity and mortality
- Shift towards renewable energy production will put increasing focus on cleaner energy carriers—electricity and probably hydrogen—which would have particular benefits for health in urban environments
- One particularly difficult challenge is to tackle lack of access to clean energy and dependence of many people in low-income settings on inefficient and inadequately ventilated burning of biomass for household energy needs

Key indicators

- Number or proportion of homes in low-income countries reliant on inefficient burning of biomass or coal for household energy needs
- Concentration of outdoor air pollutants, especially smaller particles (PM_{10} and PM_{2·5}), in urban centres
- Rate of mortality and morbidity related to combustion-derived air pollution, indoors and outdoors
- CO_{2} emissions per dwelling
- Energy needed per dwelling to maintain essential fuel needs, specifically adequate heating and cooling
energy-related health effects of energy use tend to fall because of increasingly clean, often centralised combustion of fuels, and cleaner end-use technology—often accompanied, as wealth increases, by more effective systems and legislation for control of health-damaging emissions. This pattern is the wealth-related risk transition described in the first article of this Series. Emissions of carbon dioxide (CO₂), the dominant anthropogenic greenhouse gas, are also, in theory, reduced. But until now, and perhaps still into the future, an apparently immutable law of socioeconomic development has been that use of energy services grows at faster pace than improvement in efficiency. The net result is that richer societies, with generally cleaner energy use, also contribute most per head to overall energy use and consequently to CO₂ and other greenhouse-gas emissions (figure 2). Until the evidence about its role in climate change, CO₂ was regarded an innocuous by-product of fossil-fuel combustion.

This association between wealth and energy signals a fundamental challenge for tackling climate change, and is an important reason why efficiency alone will not sufficiently reduce greenhouse-gas emissions without additional attention to the character of energy used. Without some form of direct control, increase in energy efficiency for an energy service often leads to substantially increased demand for the service, because of the lower price it creates per unit service. Vaclav Smil provides several examples. The efficiency of street lighting in the UK increased 20-times between the 1920s and the end of the 20th century, from about 10 lumens per watt (W) for incandescent bulbs to about 200 lumens per W for low-pressure sodium lamps. Yet during the same period, the average intensity of street lighting also rose steeply, with the net result that lighting-related energy consumption per km of road increased 25-fold. In the

Figure 1: Schematic of UK energy flow
1 petajoule (PJ)=10¹⁵ joules.

Figure 2: Scatter plot of (A) energy consumption and (B) energy consumption per gross national income (GNI) vs per head GNI
Graphs constructed from online data sources. Symbol sizes are proportional to country populations. Kg oil eq=energy equivalent to that produced by combustion of kg of oil.
transport sector, the internal combustion engine went through similarly advanced technological development during the 20th century, improving its mass-to-power ratio from 30 g/W in 1900 to about 1 g/W in 2000. Yet this change, along with other largely technological achievements, has allowed a spectacular rise in car ownership and in the yearly distances travelled by road. The changes in aircraft engines (from propeller, to turbojet, to turbofan designs) has been even more striking, with the most modern high-bypass turbofans (0·1 g of weight per W of power) making intercontinental travel almost routine—and at a rate of energy consumption per passenger-mile approaching that of some road vehicles. During the past 30 years, the theoretical energy efficiency of the UK housing stock has increased by 30%, although the net energy use has also increased by 30%. Of course, the effects of the different factors involved are difficult to separate, including price, efficiency, rising wealth, technological improvement, and investment in infrastructure.

These trends in individual pieces of technology add up to macroeconomic patterns (figure 2) that are visible in investment in infrastructure. A second limitation on what energy efficiency can achieve, at least in the short term, is one of practicality. Urban layout, building structure, and the devices used in buildings typically have long lifespans of decades or longer (table 1) and require substantial capital investment to replace.

Consequently, energy properties of the built environment include much inertia, and long lead-times are needed to achieve substantial change, apart from when there are opportunities for retro-fitting. But any attempt to adapt old buildings is often expensive and less effective than designing efficiency into new infrastructure. Moreover, new energy efficiency technology is also not always compatible with existing infrastructure: a wall cavity cannot be filled with insulation if no cavity exists. For many adaptations, the cost-effectiveness might seem unattractive without full account of environmental and health effects, and many households, companies, and institutions may have little incentive to make the necessary capital outlay.

Nevertheless, because of poor information and other barriers, major gains can be made for all societies in enhanced energy efficiency. Furthermore, many efficiency measures are actually cost-negative—ie, they save money. Thus, even though some of the potential energy savings will not be realised because of increased activity, energy efficiency comprises the major “low hanging fruit” in nearly all energy studies. Efficiency can be achieved in terms of urban structure and form, in building form and construction, and in energy-using appliances within buildings. The degree of planning, timescale for change, and capital investments needed vary substantially across these categories. Another important question for health in the built environment is the nature of the energy carriers (the fuels) used to deliver energy services. Cleaner forms of power generation, based on renewable or nuclear technology, could also favour the use of cleaner energy carriers, such as electricity and hydrogen, as the main modes of energy delivery that can reduce human exposures to health-damaging pollutants. The benefits of energy efficiency and of modal shifts in energy carriers will be considered in the final section of this article.

Before turning to specific aspects of the built environment, we should note that health systems themselves are substantial users of energy. For example, in 2001, the UK National Health Service (NHS) estate consumed an estimated 12 650 gigawatts per h of energy—about 0·8% of the total energy consumed in England and Wales. This figure is almost doubled if other health-service buildings, including administration, are included. Additionally, NHS staff, visitors, and patients travelled some 25 billion passenger-km (about 3·5% of the national total), and if energy expenditure by the pharmaceutical industry is also taken into account, the energy consumption by the health sector could be between 3% and 5% of the national total. In terms of per head per year, this level of energy consumption is not far
below the typical total energy consumption of a person living in Bangladesh. The health sector itself therefore has an important role in leading efforts to find and deal with energy issues.

**Urban structure**

Urban design and infrastructure has bearing on various aspects of energy use and health effects. First, it is an important determinant of energy use in buildings and of choices in transport (as described in the third article of this Series). Compact urban areas that avoid large distances between buildings and with few physical barriers are among the most important factors that could make the urban environment more conducive to physical activity, including walking and cycling.\(^{10-12}\) Conversely, low-density urban areas tend to lead to poor access to public transport; high car use; and large heating, cooling, and lighting loads per individual. Lower urban density largely accounts for the much greater energy use per head in US cities than in European cities, for example. Separate but related debates have been made about the extent to which factors (such as socioeconomic mix) are important for social wellbeing. Thus, urban design and land-use choices are, in theory, determinants of energy demand, but even more important is that local environments could also affect health. As shown in article three of this Series, the main health connections are self-evident, and relate to effects on physical activity and weight management\(^{11}\) (with their many physical and psychosocial benefits), as well as effects on injury risks, air pollution, and social cohesion, as seen in article three of this Series. However, specific epidemiological evidence about environmental interventions is comparatively limited, and is an area of much needed further research.\(^{14-15}\)

High urban density also makes possible efficiency options of transferring by-product heat between power plants and buildings, and of having district heating systems. Such solutions could substantially improve the efficiency with which the energy from fuel is captured for useful work and heat. However, many local combined heat and power sources, particularly in areas of high population density, could have unwelcome effects on air quality compared with centralised generation and distribution of electricity, for example. The potential effects on health of such choices in energy delivery remain largely unquantified and are the focus of current research.

Urban density also affects two important human exposures that are under increasing attention in view of climate change and our apparently unbreakable dependence on motor vehicles—namely exposure to heat and outdoor air pollution. Outdoor temperatures within cities often exceed those of the surrounding countryside by several °C—a phenomenon referred to as the urban heat island effect.\(^7\) The reasons relate to the high heat capacity of various elements of the urban environment, the reduced thermal radiation with sheltering by tall buildings, and lack of evapotranspiration because of the small number of trees and other plants;\(^7\) heat generated by buildings, transportation, and other aspects of human activity can also add to ambient temperature warming. The magnitude of the temperature excess is variable, and depends on such factors as meteorological conditions and time of day. Its potential importance lies in the fact that, under climate change, the frequency, intensity, and duration of heat waves is expected to increase substantially,\(^{19-20}\) with potentially important adverse effect on health,\(^{21-23}\) as was shown, for example, by the heatwaves in Paris, France, in 2003\(^ {21}\) and Chicago, USA, in 1995.\(^ {22-23}\) Current urban environments could compound the risks because of the heat island effect, and also because of the way some buildings capture heat.

The available evidence does not yet allow precise quantification of the effect of the heat island effect on mortality during heatwaves. However, there is evidence that air conditioning protects against the risk of heat death,\(^ {21-26}\) and in consequence increased attention has been given to improved access to air-conditioned rooms as a health protection measure for heatwaves. Unfortunately, the energy demands of air conditioning are typically high, so its widespread use only adds to the problem of climate change. The alternative is to adapt urban spaces and buildings to use simpler, passive means of temperature control. Such options include: measures to increase shading from the sun (for example by planting trees);\(^ {26}\) provision for controllable ventilation during the day and high levels of ventilation at night; use of heavier-weight building materials; and improvement of insulation.\(^ {28}\)

The evidence for adverse effects of urban air pollution clearly shows that particle pollution in particular is

---

**Figure 3:** Seasonal average variation in mortality in relation to energy efficiency of English homes

Figure adapted from reference 37. Energy-inefficient homes are in the lowest quartile of standardised heating costs and energy-efficient homes are in the highest quartile.
responsible for a large global burden of mortality and morbidity.39 Transport of air masses means that air pollution is not a uniquely urban problem, but it is predominantly urban because of the density of traffic and stationary sources in cities. Street canyons and other buildings in cities can also affect dispersal of pollutants and, thus, local pollutant concentrations. But in the context of climate change, interest is increasing in potential interactions between the weather and local air quality. Most notable is the possible effect on concentrations of summer ozone (which has well recognised adverse health effects)30–33 because of the importance of temperature and sunlight to the air chemistry that leads to its formation. Although the effects of climate change on ozone are complicated, where the concentrations of precursors are high, ozone levels are likely to increase.44 Ozone could have been responsible for an appreciable proportion of the deaths occurring during the 2003 European heatwave,35,36 and the interaction of heat and ozone is an issue of increasing research interest, particularly to identify options for reducing the adverse health effects in the context of climate change.

**Buildings: energy efficiency in the home**

**High-income countries**

Because of the long lifespan of housing, in most countries, most of the existing stock (which typically accounts for more than a quarter of carbon emissions) predates modern energy and thermal comfort standards. In consequence, not only is it inefficient in energy use, but the cost and difficulty of space heating seem to contribute to ill health and even mortality risk, although research evidence remains notably sparse.45 For example, a study of mortality patterns in England found that people who live in older, poorly heated homes of low-energy efficiency seem to be at increased risk of winter-related and cold-related death from cardiovascular disease. Even as a simple comparison, figure 344 shows the much larger seasonal fluctuation in mortality in people living in energy-inefficient homes than in energy-efficient homes. The apparent conclusion, although still not formally tested through randomised controlled trials, is that more energy-efficient stock would reduce this mortality burden. In the UK, the yearly winter excess of deaths is typically between 20 000 and 50 000 deaths, mostly from cardiovascular causes.38,39 Although exposures to the cold during outdoor excursions could be important in determining adverse health risks,40,41 some research evidence and reasonable theoretical grounds suggest that the indoor environment is important as well. Thus, at least a theoretical rationale can be used to pursue energy efficiency on health as well as environmental grounds. Figure 4 shows the main connections between household energy efficiency and health.

A scheme to implement energy-efficiency improvements for low-income households in England has been running over recent years, with the aim of tackling fuel poverty through packages of insulation (eg, loft or cavity-wall insulation, draught proofing) and heating system upgrades. Those improvements, which in theory reduce standardised CO2 emissions and energy costs, have been shown to increase winter indoor temperatures (thus with probable benefits in terms of cardiorespiratory morbidity and mortality)46 and to reduce normalised relative humidity, condensation, and visible mould growth.47 Additionally, psychosocial benefits have been reported, consequent to improved thermal comfort, expanded use of space, increased privacy, and improved social interaction.48 Notably, no evidence from this study of interventions in low-income households in England has
yet shown that energy-efficiency improvements lower fuel consumption.52 Evidence of a similar range of benefits has been obtained from one of very few randomised trials of energy-efficiency interventions. This New Zealand study46 showed that insulation of existing houses led to positive changes to the indoor environment, improved self-rated health, self-reported wheezing, days off school and work, and reduced visits to general practitioners.

Improved energy efficiency could also affect health in other, largely unquantified, ways—through cost savings (potentially important for fuel-poor households—ie, those that need to spend >10% of household income on fuel), reduced emissions of air pollutants to the local environment, and (potentially) contribution to mitigation of climate change (figure 4). But another important benefit could arise from the replacement or refurbishment of old, inefficient, combustion appliances (boilers, burners, cooking stoves). Carbon monoxide (CO) is generated by incomplete combustion, and in high concentrations it can be rapidly fatal. There is suspicion that the occurrence of CO poisoning is under-diagnosed, and some debate—but no firm evidence—that chronic exposure to low-level CO can adversely affect cognitive function.60–63 Although few surveys of indoor CO concentrations have been done, some surveys have found that peaks in indoor CO concentrations (up to or above 100 ppm [parts per million]) occur with non-negligible frequency. Whether measurable adverse effects exist from regular exposure to such levels, and how often appliances deteriorate further to produce acutely dangerous CO concentrations remains unknown.

Adverse effects could also result from improved energy efficiency if adequate measures are not taken to guard against reduced air exchange and if additional thermal insulation exacerbates rather than reduces summertime overheating.39 The balance between adequate ventilation for health and reduced ventilation to minimise heat loss is probably one of the greatest challenges in the design and refurbishment of buildings to use low energy.33,132 Figure 553,54 shows how the internal relative humidity and energy consumption changes with ventilation rate in a typical UK dwelling. Relative humidity is the main determinant of mould growth and the prevalence of house dust mites, both of which produce allergens. In northern European countries, a minimum air exchange rate of 0·5 air changes per h is recommended to protect against adverse effects.

For new buildings, the opportunities are much greater to incorporate energy efficiency measures from the outset. However, over the short term, new houses add to, or replace, only a small part of the total housing stock, and so can make only a modest contribution to efficiency gains. Moreover, with current constructions being predominantly of energy-intensive concrete, brick, steel, and glass, the construction of new buildings adds to carbon emissions. For example, production of a new house typically results in carbon emissions equivalent to 5 years’ energy use. A large demolition and new building programme would, in the short term, contribute to climate change rather than mitigate it. However, most buildings undergo several phases of refurbishment during their lifecycle (windows, for example, tend to be replaced every 20 years or so), and for existing properties, the key is to take the opportunity of refurbishment to improve energy efficiency.35

Although energy efficiency is one of the key strategies to help meet energy needs, another strategy is the use of renewable energy generators integrated into the fabric of buildings. Renewable energy technologies such as solar thermal water heaters and photovoltaic solar panels can be located on the façade of buildings and generate solar electricity and heat. In many settings, the façade of a building can (in theory) generate nearly as much energy as the building requires, but the mismatch in time between generation and demand necessitates either expensive storage or a sophisticated trading of energy via a grid system. Furthermore, the capital cost of renewable technologies is high and the running cost low. This high capital cost is most often a challenge for vulnerable individuals, and a bigger divide between fuel-poor and fuel-rich homes (both between and within nations) can be envisaged in the future, unless policies are put in place to prevent this outcome. In the future, moving from a traditional centralised energy supply to a mixed system of local generation could, if not carefully managed, lead to new health and safety issues as home owners become responsible for energy-generating technology.

### Low-income countries

In the first article of this Series, we referred to the 2 billion people without access to electricity and to the health burdens of indoor air pollution from household use of solid fuels in developing countries. These problems stem from poverty, but also pose some difficult technical

---

**Table 2: Health outcomes of indoor air pollution**

<table>
<thead>
<tr>
<th>Health outcome</th>
<th>Strength of evidence</th>
<th>Population group</th>
<th>Relative risk (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute infections of lower respiratory tract</td>
<td>Strong</td>
<td>Children aged 0–4 years</td>
<td>2.3 (1.9–2.7)</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>Strong</td>
<td>Women aged ≥30 years</td>
<td>3.2 (2.3–4.8)</td>
</tr>
<tr>
<td>Lung cancer (coal)</td>
<td>Moderate I</td>
<td>Men aged ≥30 years</td>
<td>1.8 (1.0–3.2)</td>
</tr>
<tr>
<td>Lung cancer (biomass)</td>
<td>Moderate II</td>
<td>Women aged ≥30 years</td>
<td>1.9 (1.1–3.5)</td>
</tr>
<tr>
<td>Asthma</td>
<td>Moderate I</td>
<td>Children aged 5–14 years</td>
<td>1.6 (1.0–2.5)</td>
</tr>
<tr>
<td>Cataracts</td>
<td>Moderate II</td>
<td>Adults ≥15 years</td>
<td>1.2 (1.0–1.5)</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>Moderate II</td>
<td>Adults ≥15 years</td>
<td>1.3 (1.0–1.7)</td>
</tr>
</tbody>
</table>

Data adapted from reference 57. Strong=many studies of solid-fuel use in developing countries, supported by evidence from studies of active and passive smoking, urban air pollution, and biochemical or laboratory studies. Moderate=at least three studies of solid-fuel use in developing countries, supported by evidence from studies on active smoking and on animals (moderate I=strong evidence for specific age/sex groups; moderate II=limited evidence).
problems of achieving reliable energy services at very low cost. They suffer from the same sort of social, economic, and political barriers that have long frustrated aspirations to ensure basic health needs are met for all people.

Lack of access to electricity and clean fuel lies behind many aspects of poor health and poverty in the developing world. This situation is partly and indirectly due to energy’s central role in supporting basic education and health infrastructure; but it also derives from direct effects of energy use at the household level and the quality of indoor air. Lack of access to clean household fuels and electricity has direct links to several Millennium Development Goals (MDGs). Household use of solid fuel, particularly biomass for cooking and heating, has perhaps the most direct link to the MDGs through its effects on the health of children and women. In unvented homes, the concentrations of health-damaging pollutants to which households are exposed typically reach levels many times higher than those found in urban outdoor environments where health effects are well established. The largest exposures occur to women, who are normally responsible for food preparation and cooking inside the home, and infants and young children who are usually with their mothers near the cooking area. Clear epidemiological evidence links these pollutant exposures to acute infections of the lower respiratory tract in children, which is the chief cause of child mortality in the developing world; to chronic obstructive pulmonary disease, especially in women; and (for coal, at least) to risk of lung cancer (table 2).

Growing evidence indicates that these exposures are important risk factors for lung and other cancers (from biomass as well as coal burning), cataracts and other eye diseases, low birthweight and other adverse pregnancy outcomes, and tuberculosis. Increased risks for asthma and cardiovascular diseases are also suspected.

Although the epidemiological evidence base is growing rapidly and is consistent with animal and toxicological information, passive and active tobacco smoking studies, and outdoor air pollution epidemiology, household studies of solid-fuel air pollution have had important limitations in proving causality and quantifying the benefits of real interventions. These challenges include the residual confounding potential in the use of observational designs to study diseases that have many causes linked to poverty, which is also closely linked to use of low-quality fuels. Additionally, because of the problems in taking measurements in these settings, exposure-response associations have been difficult to determine. Thus, although links with several disease endpoints are becoming well established, what the effect of particular improvements might be is less clear. Therefore in a world in which public-health resources are extremely scarce for the populations that need them most, it has been difficult to argue that precious funds be diverted from the many other urgent health needs—such as vaccines, antibiotics, and food supplements to pregnant women—to improved fuels and stoves.

To address the need to provide high-quality evidence, the first randomised trial in air pollution to our knowledge was done in Guatemala. Much more detailed outcome assessment was done than in previous studies, which probably confused many upper respiratory infections (which have little public-health effect) with lower respiratory infections (which are difficult to diagnose in field settings). Furthermore, much more detailed exposure assessment was done than in any previous study. It showed decreases in blood pressure in women within 1–2 months after introduction of an improved chimney woodstove in a population using open wood fires for cooking. The reduction in both systolic and diastolic blood pressure, which was seen in both cross-sectional and longitudinal analyses, substantially exceeded what is usually found in salt reduction studies. Furthermore, among infants in the intervention group, serious, physician-diagnosed, bacterial pneumonia, which is thought to have a higher case-fatality rate than viral pneumonia, was reduced by about 40%. This difference is less than that found in the observational studies, but indicates (without confounding) what can be achieved with a real intervention. No effect was found for infants with pneumonia who were positive for respiratory syncytial virus, which had been a study hypothesis on the basis of previous studies.

This Guatemalan efficacy study, however, is only one in one region of the world. History has shown that, to have a substantial effect on policy, additional efficacy studies and large-scale effectiveness studies will be needed.

Although household stoves are the oldest of human combustion devices and have seen many innovations over thousands of years, the technology need to meet current expectations for protecting health is, perhaps surprisingly, not yet well developed in developing countries. In recent decades, several improved cook-stove programmes have been implemented by countries concerned mainly with improving fuel efficiency to protect local natural environments and to enhance energy services for the poor from existing biomass supplies. Even the largest and most successful of these, the Chinese programme, which was responsible for dissemination of 180 million stoves from the early 1980s to the late 1990s, did not have health as a major objective, although it reduced exposures to some extent through increased efficiency and use of chimneys. No large-scale programmes and only a handful of small ones have yet addressed health directly by designing, testing, and monitoring their efforts in the context of exposure reduction or health improvement.

One problem with existing improved stove technologies is illustrated by the Guatemalan trial, which is consistent with studies in several regions including China. Even a well operating chimney stove only moves the smoke 1–2 m and does not actually reduce smoke emissions. Thus, pollution levels in and around the rest of the house...
do not change much and actual personal exposures do not reduce nearly as much as kitchen levels because people do not spend all day in the kitchen. The implication of this drawback for designing large-scale chimney-stove interventions needs to be explored, but the long-term message is clear: affordable, reliable stoves are needed that do not generate pollution in the first place.

Countries such as South Korea that transited from poor income to middle income in the past century before the rise in petroleum prices, simply switched household fuel use during this period from biomass to kerosene and liquefied petroleum gas, products of the oil fuel cycle. By comparison with biomass (and coal), these fuels are very clean and efficient for household use. Unfortunately, however, current oil prices and associated uncertainties in the worldwide petroleum market make such a transition extremely difficult for most poor individuals in developing countries today. With even more large price rises in the past years, the price gap between biomass and these household fossil fuels has been widened further, probably driving some groups back to biomass, and greatly straining the budgets not only of households but also of the many governments, including those of India and Indonesia, that highly subsidise such fuels.

This widening price gap, however, also provides an opportunity to develop biomass-based stoves and fuels that have improved combustion and thus do not produce pollution at all. Several technical approaches seem attractive in this regard, most of which involve means to assure good secondary combustion. Such technologies probably cannot be developed at costs comparable to those of the cheap stoves found in the poorest households at present, which are often nearly costless, but would seem able to operate in the gap between these and the now quite costly fossil alternatives. Over time, as the market operates, economies of scale develop, technology improves, and the evidence needed to promote societal assistance grows, perhaps even the poorest groups could be served.

Although most people at risk of exposure to indoor air pollution live in rural areas of the world’s poorest countries, this risk is increasingly becoming a problem for poor urban dwellers, a trend that will probably increase with the urban transition. Additionally, the effect on health of household fuel use go beyond indoor air pollution and affect the household economy, women’s time and activities, gender roles and relationships, safety and hygiene, and the local and global environment. For example, half the worldwide wood harvest has been estimated to be used as fuel. Furthermore, in some settings, poor families spend more than 20% of their disposable household income on biomass fuel (compared with 9% in the UK for household expenditure on housing fuel and power), or devote more than 25% of total household labour to wood collection. These additional benefits to improvement in the household fuel cycle relate to other MDGs, as seen in the first article of this Series.

Transition to clean energy
The contrast in the energy needs and priorities of rich and poor countries highlights a central tension that has been referred to earlier in this Series: in health terms, the poorest populations would gain from improved access to electricity and other modern energy sources, yet improved access to energy also means increased consumption and potentially increased emissions of greenhouse gases.

The solution lies not only in international agreements about equitable CO₂ emissions targets, but also in technology transfer—rich countries helping poorer populations to adopt clean energy technology (clean in terms of health-damaging pollutants and CO₂ emissions), thereby in part modifying the conventional pattern of environmental risk transition associated with economic development.

Clean technology is usually more expensive than conventional technology, and if not widely affordable in high-income countries, it is far less affordable (and generally a much lower priority) in less developed countries. Thus, to counter potential economic barriers, technology transfer should also carry transfer of resources, and on a large scale. Because of the often scarce international aid available, such resources should be deployed in ways that meet local public-health and development priorities, yet evidence indicating the most cost-effective deployment is unclear. Nevertheless, almost all measures aimed at reducing greenhouse-gas emissions and accelerating the transition to newer energy technology will probably have beneficial effects on outdoor air quality, particularly in urban settings. This aim is a major public-health argument for greenhouse-gas mitigation policies, and the potential for benefit is theoretically greatest in some of the cities of low-income to middle-income countries that are undergoing rapid economic development.

Co-benefits to outdoor air quality of greenhouse-gas reductions and clean energy technology
Estimation of the effect of greenhouse-gas mitigation policies on air quality is an uncertain process. However, several theoretical calculations have been attempted, on the basis that good evidence now exists about the associations between outdoor air pollution and health, and reasonable models of the contribution of emission sources to air pollution concentrations. One such study was done by Cifuentes and colleagues, who developed scenarios for Mexico City, Santiago de Chile, Sao Paulo, and New York, using air pollution health effect factors appropriate to every city. They found that the adoption of readily available technology could lead to appreciable reductions in premature deaths, chronic respiratory disease, and person-days of work loss or other restricted activity—illustrating, in semiquantitative terms, at least,
the principle that immediate co-benefits to health could accrue from efforts aimed at the control of greenhouse-gas emissions. However, such studies can only be indicative. The effects of control policies are infinitely complex to characterise in reality, and innumerable assumptions need to be made about health effects of air pollution and their reversibility. Nevertheless, it is quite reasonable to expect substantial benefits of this kind, even if uncertain in magnitude. Standard scoping methods for such analyses are also now being developed.3

Cleaner energy carriers

Similar arguments about health effects arise in relation to a switch towards cleaner energy carriers. While oil, coal, and gas supplies decline as primary energy sources, current forms of energy delivery will be gradually substituted by other, often cleaner, energy carriers, including electricity, and potentially biofuels and hydrogen. The switch to electricity and hydrogen in particular would have a substantial effect on the local emission of air pollutants in urban environments.

Electricity is a high-quality and very versatile energy carrier, which is almost certain to have an increasing role in the delivery of energy in future. In 1900, electricity accounted for only around 2% of energy consumption, whereas it is around 30% currently,7 which largely stems from its many advantages. For individuals with access to mains supplies, electricity is instantly available, effectively free of emissions at the point of consumption (apart from low-frequency electromagnetic fields), and it can drive a wide range of electric and electronic devices which are integral to modern living. Although any possible health effects of low-frequency electric and magnetic fields are unclear, no conclusive evidence so far has indicated substantial health risks: such risks are certainly low by comparison with those associated with emissions from the combustion of carbonaceous fuels. Death and serious injury from electrocution in the home remain uncommon:72 figures from the Office for National Statistics show that in England and Wales, for example, 20–50 deaths per year were recorded from accidental electrocution in 1994–2003. Of course, combustion-related emissions, even from distant power stations, contribute to background levels of air pollution, so the full benefit for health of a modal shift towards electricity will depend also on the extent to which it is generated by clean technology.

The chief disadvantages, however, are that electricity can be comparatively inefficient to generate and transmit, and cannot be directly stored. Fluctuations in demand therefore need a (networked) supply that can cope with variations from base-load to peak-load within very short time-periods. However, such a coping mechanism is theoretically achievable even with largely renewable electricity generation by use of efficient demand management systems, time-of-day pricing, and other load-levelling techniques, used in concert with energy storage, including pumped water, thermal inertia, and other technology. However, most such storage holds energy at much lower density than fossil fuels (table 3).7

The impetus to develop hydrogen as an energy carrier is based on several desirable properties. By mass, its combustion liberates much more energy than conventional hydrocarbon fuels (although it is not as dense an energy store as fossil fuels, even when highly compressed, table 3); transport of energy through pipelines could be more efficient than through electric power lines (>1500 km of hydrogen pipeline in Europe already exists); it can be ignited with air in a normal internal combustion engine; it can be turned into electric current in a fuel cell; and, with some technical constraints, it can be stored.

Perhaps its most important benefit, however, is that its combustion with air generates water vapour with very little pollution other than some mononitrogen oxides (NO), and no CO₂. There are no substantial pollutant emissions from hydrogen fuel cells. Although water vapour is an important greenhouse gas, the burning of hydrogen as a fuel at ground level will not affect water vapour concentrations in the atmosphere in ways that could materially influence the Earth’s radiative balance. Hydrogen is not a primary fuel, however. At present its production is dominated by processes that depend on conventional energy sources. Reformation of natural gas

<table>
<thead>
<tr>
<th>Weight (kJ/kg)</th>
<th>Volume (MJ/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conventional fuels</strong></td>
<td></td>
</tr>
<tr>
<td>Crude oil</td>
<td>42 000</td>
</tr>
<tr>
<td>Coal</td>
<td>32 000</td>
</tr>
<tr>
<td>Wood</td>
<td>12 500</td>
</tr>
<tr>
<td><strong>Synthetic fuels</strong></td>
<td></td>
</tr>
<tr>
<td>Hydrogen, gas</td>
<td>120 000</td>
</tr>
<tr>
<td>Hydrogen, liquid</td>
<td>120 000</td>
</tr>
<tr>
<td>Hydrogen, metal hydride</td>
<td>2000–9000</td>
</tr>
<tr>
<td>Methanol</td>
<td>21 000</td>
</tr>
<tr>
<td>Ethanol</td>
<td>28 000</td>
</tr>
<tr>
<td><strong>Thermal energy (low-quality)</strong></td>
<td></td>
</tr>
<tr>
<td>Water (100–40°C)</td>
<td>250</td>
</tr>
<tr>
<td>Rocks (100–40°C)</td>
<td>40–50</td>
</tr>
<tr>
<td>Iron (100–40°C)</td>
<td>Roughly 30</td>
</tr>
<tr>
<td><strong>Thermal energy (high-quality)</strong></td>
<td></td>
</tr>
<tr>
<td>Rocks (eg, 400–200°C)</td>
<td>Roughly 160</td>
</tr>
<tr>
<td>Iron (eg, 100–40°C)</td>
<td>Roughly 100</td>
</tr>
<tr>
<td><strong>Mechanical energy</strong></td>
<td></td>
</tr>
<tr>
<td>Pumped hydro, 100-m head</td>
<td>1</td>
</tr>
<tr>
<td>Compressed air</td>
<td>–</td>
</tr>
<tr>
<td>Flywheels (steel)</td>
<td>30–120</td>
</tr>
<tr>
<td><strong>Electrochemical energy</strong></td>
<td></td>
</tr>
<tr>
<td>Lead-acid</td>
<td>40–140</td>
</tr>
<tr>
<td>Nickel cadmium</td>
<td>Roughly 350</td>
</tr>
<tr>
<td>Lithium ion</td>
<td>700</td>
</tr>
</tbody>
</table>

*Table 3: Energy density of various forms of energy storage*
(heating with steam to about 1000°C), the most common form of production, yields about 11 tonnes of CO₂ emissions per tonne of hydrogen. Hydrogen is seldom used as an energy fuel, but is an important chemical product used mainly in the production of ammonia for nitrogen fertilisers and in the conversion of heavier crude oils to lighter fuels. Its worldwide production, currently about 50 million tonnes a year, is rapidly growing. Whether it has a substantial role as a fuel mainly depends on two issues: whether it can be economically produced without substantial greenhouse-gas emissions; and whether the technology and systems can be developed for its safe storage and use in everyday life.

Hydrogen could be produced from electrolysis of water in off-peak periods by use of electricity from nuclear power or other low-carbon generation. Norway already produces hydrogen for the fertiliser industry by electrolysis using hydroelectric power. In future, substantial production could occur by direct thermochemical processes, by use of the heat of various designs of high-temperature nuclear reactors. The economics are not yet favourable, but the equation could alter as oil prices rise and closer attention is paid to the environmental and health costs of burning fossil fuels.

The second issue has important technological hurdles. Hydrogen’s low boiling point (−253°C) and density (0·09 g/L under atmospheric pressure and ambient temperature) mean that it has to be stored under high pressure, at very low temperatures, or by adsorption onto alloys (eg, titanium and iron or magnesium and nickel). Hydrides concentrate hydrogen almost as effectively as storing it in liquid form, but over time the efficacy of this process wanes. The buoyancy of hydrogen means that it is rapidly dispersed from a leak, but it is flammable at a low concentration in air and its ignition energy is an order of magnitude lower than for petrol or methane (0·02 mJ vs 0·24 mJ).

With appropriate technology, however, the risk of a combustion accident is probably no greater than for petrol or natural gas. Hydrogen could be delivered to buildings via a reticulated network of pipes, and its use has been successfully shown in buses and cars, and even in trial aircraft. Technological barriers and cost mean that its use will initially be restricted, probably to selected vehicle fleets (eg, bus fleets served by garages with the necessary technology adaptations) and small-scale stationary plants. The case for hydrogen use as a fuel for energy in buildings is particularly unclear, because an effective electricity grid can deliver controlled power directly without need to generate hydrogen as an intermediate carrier. Nonetheless, hydrogen as a fuel is familiar from the days of “town gas” (which was largely a mixture of hydrogen and CO) produced from coal before the arrival of natural gas, and it can be added to the existing (natural) gas network in concentrations up to about 20% without need for special adaptation. Despite some technical hurdles, the potential contribution of a hydrogen economy to a cleaner, healthier environment seems attractive, even if the economics are currently unfavourable. Mark Jacobson’s assessment of the effect of converting all US road vehicles to using hydrogen fuel cells suggests that 3700–6400 lives per year could be saved.

**London case study**

The webappendix provides a case study illustration of the possible effect on air pollution and health of policies aimed at greenhouse-gas emissions controls and possible switch in energy carriers in London. The calculations are
entirely hypothetical, based on very broad indicative scenarios, and compare theoretical futures with the status quo. The scenarios include changes to the fuels (energy carriers) for transport, changes in buildings-related energy use, and assumptions about patterns of vehicle use. The hypothetical experiments assume immediate and complete implementation of broad policy objectives, and are thus somewhat artificial because, among other assumptions, no account is taken of the underlying trends in (for example) technology, building regulations, and vehicle numbers, which mean that the future energy-related emissions in London would be very different in the future even without these scenario changes. And such substantial shifts as these scenarios suggest would in reality take time to achieve. Nonetheless, the air pollution modelling used is reasonably well developed; and the use of life-table analysis based on published exposure-response associations is an established approach to estimate the health effects.79

The results (webappendix) should be interpreted not as precise estimates but as signposts of the air-pollution-related health gains in the short term. They indicate modest but worthwhile short-term co-benefits of bold (but theoretically achievable) policies aimed at reducing greenhouse gases, irrespective of the (unquantified) long-term benefits through limiting climate change. Although the legitimacy and detail of these scenarios can be challenged, they provide two important observations.

First, action is necessary on many fronts—in buildings, transport, human behaviours, power generation—if the net result is to achieve the necessary reduction in CO2 emissions. Even with the fairly bold objectives set out in these scenarios, which imply quite substantial change, the combined effect on overall CO2 emissions is still only part way toward the required medium-term reduction in greenhouse gases for a city in a high-income country.

The second observation is the interdependence of cities and regions in terms of air pollution. In the scenarios, we assumed changes in emissions in London only, and not in the surrounding region. But as figure 6 shows, regional air quality has an important bearing on air quality in London.

For particle pollution (PM_{10}) in particular, a high proportion of the concentrations arise from long-range transport of pollutants from outside London, much of it transported from the near continent, including France and the low countries—although such secondary pollutants could be less relevant for health than primary particles generated locally. The local urban-area sources of particles in London contribute a modest addition to PM_{10} levels, and vehicle emissions along the road network account for the local spikes. Thus, action in London alone is necessarily limited in the extent to which it can reduce overall particle pollution: more substantial reductions will require collective action in neighbouring countries and beyond, including changes to cleaner modes of power generation and industrial energy sources. Concentrations of oxides of nitrogen (webappendix), which are largely traffic-related, more closely related to local sources.

**Conclusions**

As with so many aspects of the energy debate, the factors that can and do have bearing on future policies for the built environment are innumerable. However, the mechanisms to include proper assessment of the health costs and benefits of those complex choices so far have not been developed and are often not sought. Scientific evidence is always imperfect, but for many questions, such as the effects of indoor air pollution, the evidence of health links is already strong and partly quantified; for others, such as the effects of energy-efficient homes, the evidence is meagre; yet even in this area, such available evidence amounts to a persuasive case that health and environmental goals are generally served by broadly the same policy course. What is needed to help embed this understanding into policy action are large studies that can provide a sound basis for assessment of cost-effectiveness, taking account of multiple direct (immediate-term) and indirect health links. These studies will be complex, but the questions are important. The energy policy implemented for urban environments, and for international development, must surely be one of the current priorities for public health.

**Conflict of interest statement**

We declare that we have no conflict of interest.

**References**


76 Jacobson MZ, Colella WG, Golden DM. Cleaning the air and
75 Hore-Lacy I. Nuclear energy in the 21st century. London and
74 Kunitomi K, Yan X, Nishihara T, Sakaba N, Mouri T. JAEA’s VHTR
73 Sorensen B. Renewable energy, 3rd edn. Burlington, MA: Elsevier
72 Ormandy D, Battersby S, Landon M, Moore R, Wilkinson P.
70 Cifuentes L, Borja Aburto VH, Gouveia N, Thurston G, Davis DL.
69 Cifuentes L, Borja Aburto VH, Gouveia N, Thurston G, Davis DL.
68 Smith KR. Biofuels, air pollution, and health: a global review. New
66 Smith KR. You don’t get what you expect, you get what you inspect.
64 McCracken JM, Smith KR, Mittleman M, Diaz A, Schwartz J.
63 Bruce N, Weber M, Arana B, et al. Pneumonia case-finding in the
60 Ciferri MC, Carbone L, Carbone A, et al. The impact of indoor air
57 Fuller G, Carslaw D, Lodge H. An empirical approach for the
56 Carslaw D, Beevers S, Fuller G. An empirical approach for the
55 Miller KA, Siscovick DS, Sheppard L, et al. Long-term exposure to
54 Miller KA, Siscovick DS, Sheppard L, et al. Long-term exposure to
In October, 2006, a 37-year-old man self-presented to our emergency department, with blurred vision and a persistent dull occipital headache, both of which had been present for 4 weeks. He had no other symptoms; there was no history of head injury or loss of consciousness; his past medical history was unremarkable, and he was taking no regular medications. The patient was fully alert, had no fever, and his blood pressure was normal. Systemic examination, including a neurological examination, revealed nothing abnormal—except for fundoscopy, which was thought to show retinal haemorrhages.

An ophthalmological opinion was requested. The patient described the “blurring” of his vision as consisting of episodes in which objects looked “wavy”. On examination, visual acuity was preserved at 6/5 in both eyes, and there was no relative afferent pupillary defect. The visual fields had generalised constriction, and greatly enlarged blind spots. Repeat fundoscopy showed swollen optic discs with nerve fibre layer haemorrhages (“flame haemorrhages”), consistent with papilloedema (figure)—signifying raised intracranial pressure (ICP). In the absence of any features of other syndromes that cause raised ICP, we suspected idiopathic intracranial hypertension (IIH). We sought a more detailed history. The patient revealed that, after a domestic crisis, he had consumed about 60 pints of beer in 4 days. His symptoms had begun immediately after the binge. He had also experienced severe headache and vomiting for a day after the binge, but had attributed these to a bad hangover. The apparent onset of raised ICP, at a time of dehydrating alcohol consumption, indicated the possibility of cerebral venous sinus thrombosis (CVST). We did urgent CT of the brain, with and without venography. Plain CT showed no abnormality; but venography showed extensive thrombosis, extending from the sagittal sinus to the jugular bulb. A lumbar puncture confirmed raised ICP; blood tests showed a raised concentration of lupus anticoagulant, indicating probable antiphospholipid syndrome. All other blood tests gave normal results, except for a slightly increased partial thromboplastin time ratio of 1·3 (upper limit of normal range 1·2). We offered the patient long-term anticoagulation, which he accepted. When last seen, in July, 2007, the patient had no headache or blurred vision, although slight papilloedema remained; his visual fields had mild generalised constriction, with normal blind spots.

The estimated annual incidence of CVST is 3–4 cases per million; CVST mainly occurs in children and young adults. About 75% of adult patients are women. CVST causes venous infarction and intracranial hypertension. Venous infarction, and the resulting localised oedema, can cause seizures, impaired consciousness, and focal neurological impairment. Intracranial hypertension can cause headache, vomiting, and loss of vision—which can be caused by reduced visual acuity, or constriction of the visual fields. In 20–40% of cases, CVST has no features of venous infarction, and is particularly likely to be mistaken for IIH—which, like CVST, is particularly common in young women. Plain CT has a false-negative rate of 25–30% for CVST; however, venography increases the sensitivity of CT to 95%. As shown by this case, however, a detailed history is also crucial. The management of CVST differs radically from that of IIH, with CVST requiring urgent anticoagulation.

References