THE LANCET

Volume 370, Issue 9586, Pages 457-538 (11 August 2007-17 August 2007)

1. Medical students as champions for social justice
   Page 457
   The Lancet

2. Peace-keeping efforts in Darfur
   Page 458
   The Lancet

3. Looking for a volunteer to help fight dengue fever
   Page 458
   The Lancet

4. Ductal carcinoma in situ and breast MRI
   Pages 459-460
   Carla Boetes and Ritse M Mann

5. Stroke prevention in elderly patients with atrial fibrillation
   Pages 460-461
   David Garcia and Elaine Hylek

6. Tetrahydrobiopterin for patients with phenylketonuria
   Pages 462-463
   Angel L Pey and Aurora Martinez

7. Protecting the rights of those in conflict
   Pages 463-464
   Gilbert Burnham and Courtland Robinson

8. When rumours derail a mass deworming exercise
   Pages 465-466
   Alexander Dodoo, Sam Adjei, Mary Couper, Bruce Hugman and Ralph Edwards
9. Health care for refused asylum seekers in the UK
   Pages 466-467
   Peter Lawrence Hall

10. Clinical update: sporadic primary hyperparathyroidism
    Pages 468-470
    Antonio Sitges-Serra and Anders Bergenfelz

11. Scientists tackle water contamination in Bangladesh
    Pages 471-472
    Samuel Loewenberg

12. In Venezuela, two public-health systems grow apart
    Pages 473-474
    Daniel Cancel

13. Ethnicity and access to health care
    Pages 475-476
    Pascale Allotey

14. Theodore de Mayerne
    Page 476
    Christopher Lawrence

15. Joseph Graziano: tackling arsenic poisoning in Bangladesh
    Page 477
    Samuel Loewenberg

16. Harvey R Colten
    Page 478
    Stephen Pincock

17. Health workers and vaccination coverage in developing countries
    Page 479
    Carsten Krüger

18. Health workers and vaccination coverage in developing countries
    Pages 479-480
    Guoqing Hu

19. Health workers and vaccination coverage in developing countries
    Page 480
    Onyebuchi A Arah
20. Health workers and vaccination coverage in developing countries
Page 480
Nishith K Singh and Vikas Singh

21. Health workers and vaccination coverage in developing countries – Authors’
reply
Page 481
Sudhir Anand and Till Bärnighausen

22. Long-term risks of increased use of intravenous iron
Pages 481-482
Jerome L Sullivan

23. Long-term risks of increased use of intravenous iron
Page 482
Behdad Afzali and David JA Goldsmith

24. Long-term risks of increased use of intravenous iron – Authors' reply
Pages 482-483
Michael Auerbach, Harold Ballard and John Glaspy

25. Risk-minimisation strategies for peanut allergy
Page 483
Robyn E O'Hehir and Jo A Douglass

26. Measurement of health and disability
Pages 483-484
Jose M Valderas

27. Measurement of health and disability – Author's reply
Page 484
Daniel Mont

28. Healing the wounds of Tiananmen: Beijing Olympics and beyond
Page 484
Jian Zhou

29. MRI for diagnosis of pure ductal carcinoma in situ: a prospective observational
study
Pages 485-492
Christiane K Kuhl, Simone Schrading, Heribert B Bieling, Eva Wardelmann,
Claudia C Leutner, Roy Koenig, Walther Kuhn and Hans H Schild
30. Warfarin versus aspirin for stroke prevention in an elderly community population with atrial fibrillation (the Birmingham Atrial Fibrillation Treatment of the Aged Study, BAFTA): a randomised controlled trial
   Pages 493-503
   Jonathan Mant, FD Richard Hobbs, Kate Fletcher, Andrea Roalfe, David Fitzmaurice, Gregory YH Lip and Ellen Murray

31. Efficacy of sapropterin dihydrochloride (tetrahydrobiopterin, 6R-BH4) for reduction of phenylalanine concentration in patients with phenylketonuria: a phase III randomised placebo-controlled study
   Pages 504-510

32. Nutritional iron deficiency
   Pages 511-520
   Michael B Zimmermann and Richard F Hurrell

33. Do human rights matter to health?
   Pages 521-527
   Jerome Amir Singh, Michelle Govender and Edward J Mills

34. Photoprotection
   Pages 528-537
   Stephan Lautenschlager, Hans Christian Wulf and Mark R Pittelkow

35. Hearing dysphasic voices
   Page 538
   Daniela Hubl, Martinus Hauf, Claudia van Swam, René Müri, Thomas Dierks and Werner Strik
Medical students as champions for social justice

Last week we launched TheLancetStudent.com—a forum to encourage medical students to become more involved in global health—at one of the largest gatherings of international medical students to date. Over 900 medical students from 92 countries gathered in Canterbury, UK, for the 56th General Assembly of the International Federation of Medical Students Association (IFMSA). The theme of the conference, access to essential medicines, demonstrates the growing enthusiasm of medical students to learn more about global health issues—not just through the orthodox disease-focused model but also through debate and discussion on the wider political, economic, and social factors intricate to global health.

In the UK for example, following the lead at University College London in offering an intercalating Batchelor of Science degree in International Health, Bristol and Leeds also have similar degrees and Birmingham is due to start one this October. And the Global Health Education Consortium is now active in over 70 health profession schools and training programmes in the USA, Canada, Central America, and the Caribbean.

However, the general feeling at the IFMSA conference was that there is not enough teaching on global health in the medical curriculum to equip the doctors of the future to best help patients around the world. Many delegates agreed that in order to be more relevant, the rigidity of the current medical curriculum needs to be replaced by one better suited to help create committed, motivated, and engaged doctors who are best adapted to the future uncertainties of global medicine. Some also thought that teaching in global health should include more multi-disciplinary approaches where input from medical professionals, sociologists, economists, and anthropologists is given equal weight and attention.

Such improvements in global-health teaching raise some controversial issues, for individual medical students, medical educators, teaching institutions, and medical organisations, such as the IFMSA. For example, how can medical students learn more about global-health issues, including the social, economic, and political determinants of health, without becoming active participants in society, being prepared to become more involved in challenging and addressing the huge inequalities and injustices at the core of global health?

Many delegates at the conference agreed that it is the responsibility of the medical profession to not only speak out about such injustices, but also to tackle and address them. However, campaigning for social justice often requires political activism and it is therefore unsurprising that many individual doctors and medical students, and medical organisations and institutions, shy away from such activity. The IFMSA, for example, stresses that it is a non-political organisation.

IFMSA delegates at last week’s conference heard how many of the factors which prevent the world’s poorest people from accessing medicines involve wider social, economic, and political issues. The challenge then for medical students as individuals, and for medical student organisations such as the IFMSA, is how far they are prepared to go to address such issues.

Some medical students are already highly active in campaigning to bring about change. For example, within the context of access to essential medicines, the Universities Allied for Essential Medicines works with student groups across the USA, Europe, and Canada to determine how universities can help ensure that medicines, are made more accessible in poor countries. It also campaigns to increase the amount of research done on neglected diseases which predominantly affect people who are too poor to constitute a market attractive to private-sector investment.

In addition, many doctors and medical students signed, and actively promoted, the petition of over 450 000 signatures organised by Médecins Sans Frontières Access to Essential Medicines campaign to call on the drug company Novartis to drop its case against the Indian patent office—which it lost last week—for not granting a patent for the drug imatinib mesilate.

It is therefore a most welcome and positive step forward that, as a result of the conference, the IFMSA proposes to devise a Canterbury Declaration on access to essential medicines—a consensus statement from the IFMSA which will support and encourage action in three areas: intellectual property, human resources, and health-care systems. The willingness to devise such a declaration by the world’s largest body of medical students hopefully signals the desire of the next generation of doctors to not only be the barometer of social justice, but also to be its champion. The Lancet
Peace-keeping efforts in Darfur

In a speech at the UN last week, UK Prime Minister, Gordon Brown, described the war in Darfur as “the greatest humanitarian disaster the world faces today”. Since the conflict began 4 years ago, 200 000 people are believed to have died and 2·5 million people have fled their homes to escape the violence. Now, there is renewed hope as the UN Security Council with the alleged support of the Sudanese Government has approved the largest peacekeeping mission in the world—the deployment of a 26 000-strong hybrid UN-African force to bring security to the region. But can this UN mission succeed when past missions in Darfur have failed?

It is unclear who will supply the needed troops. The stipulation that the troops have to be mainly African will be a challenge; the existing 7000-African Union peace force is already overstretched. A firm strategy for the new deployment and the kind of peace these troops will be looking to monitor and implement are vague. Furthermore, deployment of 26 000 troops brings all sorts of logistical problems, including access to water and allocation of land. Additionally, training troops on gender issues to protect women and children from the continued violence needs to be ensured.

As attacks on aid organisations increase, many displaced civilians are simply not getting assistance. On top of the psychological trauma, the lack of access to clean water and the poor sanitation have driven up the number of cases of and deaths from water-borne diseases, and there has been a sharp increase in malnutrition. This mission must provide protection for all civilians, and access and security for humanitarian operations. It must have the authority to make both government and opposition groups accountable for acts of violence against civilian populations, and re-enforce the implementation of pledges made under the Darfur Peace Agreement. The mission can only be deemed a success when the millions of displaced people feel they are able to go back to their homes in an environment of safety, security, and dignity.

The peace mission in theory is a good first step but to bring lasting peace to the region, undoubtedly, a political solution must be urgently sought. ■ The Lancet

Looking for a volunteer to help fight dengue fever

“April is the cruellest month”, the poet TS Eliot famously opined. But for millions in the Asia-Pacific region, August and September will be crueler. The rainy season of the tropics and subtropics brings not only the monsoons but also dengue fever. Dengue is endemic in more than 100 countries and its vector, the Aedes aegypti mosquito, which thrives on warm-weather, heavy-rain, and crowded-cities, annually infects more than 50 million people around the world. This year has already seen even larger outbreaks in several countries, including Singapore, Cambodia, Malaysia, the Philippines, and Vietnam, with many deaths (182 in Cambodia alone) reported. The most severe form of the disease, the haemorrhagic type is a leading cause of deaths in children in endemic countries.

Vector control, by insecticides, cleaning household water-storage containers, and draining pools of stagnant water, is the main component of dengue prevention. Other efforts are also required—public-health activities, better hospital care, and education, an always critical but usually formidable task. Earlier recognition of the signs of the disease, especially in children, is important. A permanent decrease in disease incidence may only be achievable with a vaccine. Fortunately, candidate vaccines exist; unfortunately, safety and efficacy data will not be available for another few years.

As with other diseases, funding is a problem, and so is health-care infrastructure. And cooperation and coordination among governmental and non-governmental efforts are in short supply. The many players in the dengue-prevention field need to be brought, at least for the present crisis, under one umbrella. Health ministers in the Asia-Pacific region, with input from organisations, such as the International Committee of the Red Cross and WHO, should quickly appoint a dengue czar. That person needs to understand public health, of course, but even more pressing at the moment is a person who can organise, coordinate, influence, and persuade. We can think of a few obvious candidates. Jimmy Carter, are you available? ■ The Lancet
Ductal carcinoma in situ and breast MRI

In today’s *Lancet*, Christiane Kuhl and colleagues present the results of a large prospective assessment of the comparative sensitivity of mammography and MRI for detection of pure ductal carcinoma in situ (DCIS).1 It is widely believed that mammography is more sensitive in detecting DCIS than is MRI. However, Kuhl found that the sensitivity of MRI for DCIS is much higher than that of mammography, especially for high-grade lesions, which are thought to be more prone to progress to invasive carcinomas. Almost half of all DCIS lesions are mammographically occult, and high-grade lesions without necrosis are even less likely to be detected. Although these results were unexpected, the pathophysiology of breast cancer provides ample justification for the findings.

Before the spread of screening mammography, about 2% of all detected breast tumours were DCIS, yet autopsy studies have shown that almost 9% of women have undetected DCIS.2 Since the start of screening mammography, the incidence of DCIS has increased nearly ten-fold, and about 20% of all tumours detected at screening are now pure DCIS. On the basis of circumstantial evidence, almost all invasive carcinomas are believed to begin as DCIS lesions. However, the time course of transition from in-situ to invasive carcinoma is unknown, and whether all DCIS will ultimately evolve to invasive disease is unclear.3,4 Nevertheless, treatment of DCIS by complete resection, or, when breast-conserving therapy is used, radiotherapy, is deemed appropriate for all DCIS lesions.3

On mammography, DCIS usually manifests as micro-calculifications, which are caused by necrosis and subsequent calcification of debris. These calcifications are usually very small, but need to be bigger than 100 μm for mammographic detection. Only 27% of mammographically detectable DCIS lesions present with soft-tissue changes on mammography.5 Recently, digital mammography was shown to outperform analogue mammography for the detection of breast cancer. However, in a large study by Pisano and colleagues,7 only 60% of in-situ lesions could be detected by digital mammography. Despite the increase in DCIS detection by mammography, many DCIS do not contain observable calcifications, and will therefore be mammographically occult.

The assumption has been that DCIS cannot be detected by MRI, because MRI does not visualise calcium and cannot be done at a sufficient resolution. Yet contrast-enhanced breast MRI visualises neovascularisation. Normally, the contrast agent is confined to the intravascular space, except in places where the vessel wall is corrupted (eg, by neoplasm). In DCIS, neovascularisation occurs; however, the vessels formed are more mature than vessels in invasive carcinomas. Therefore the typical wash-out patterns, indicative of malignancy, are often absent. Instead, more subtle asymmetric enhancement patterns can be seen. In high-grade DCIS, microvessel density is higher and consequently enhancement is stronger, which explains why these lesions are more easily identified on MRI than are low-grade DCIS.8

MRI has a higher sensitivity for DCIS than mammography,9 and about half of contralateral carcinomas detected on MRI are DCIS.10,11 Kuhl and colleagues’ results should therefore be expected theoretically.

That only 20% of tumours detected through screening are pure DCIS is disappointing, when one keeps in mind that most breast tumours probably evolve from DCIS. The observation that MRI detects many DCIS lesions that go unnoticed on mammography implies that some invasive carcinomas can be prevented by timely intervention on the basis of MRI findings. As such, MRI has the potential to increase survival when used to detect breast cancer. However, currently MRI is considered an adjunct to mammography, and all series are biased by the fact that there needs to be an indication before MRI

Maximum intensity projection of subtracted contrast-enhanced breast MRI
In right breast, non-mass-like enhancement corresponding to area of grade-2 DCIS is visible.
is done. For instance, in Kuhl and colleagues’ study, one in eight women had findings that demanded biopsy. These results would not be expected in the general population. Despite the high recall rates, 52% of biopsy specimens in Kuhl’s study showed breast cancer, a figure that is also accepted in screening mammography. In MRI screening of high-risk patients, MRI also doubled the recall rate but the rate of detected lesions per biopsy did not change—one in three biopsies was positive for cancer.

These findings can only lead to the conclusion that MRI outperforms mammography in tumour detection and diagnosis. MRI should thus no longer be regarded as an adjunct to mammography but as a distinct method to detect breast cancer in its earliest stage. A large multicentre breast-screening trial with MRI in the general population is essential.

*Carla Boetes, Ritse M Mann
Radboud University Nijmegen Medical Centre, Department of Radiology, 6500 HB Nijmegen, Netherlands
c.boetes@rad.umcn.nl

We declare that we have no conflict of interest.

Stroke prevention in elderly patients with atrial fibrillation

Warfarin is highly effective in the prevention of stroke in atrial fibrillation, with a 64% risk reduction compared with 22% for aspirin. However, there are concerns about whether this benefit can be extrapolated to older age groups because few patients older than 75 years were enrolled in early clinical trials. Older patients are at the highest risk of stroke, but old age has been identified as an independent risk factor for warfarin-associated haemorrhage.

In today’s *Lancet*, Jonathan Mant and colleagues report the results of the Birmingham Atrial Fibrillation Treatment of the Aged (BAFTA) study, in which 973 patients, age 75 years and older and with atrial fibrillation, were randomised to warfarin (target international normalised ratio 2−3) or aspirin 75 mg. The primary endpoint was fatal or disabling strokes (ischaemic and haemorrhagic) or clinically significant arterial embolism. Warfarin was superior to aspirin in the prevention of stroke (1.8% vs 3.8% per year) and was no more hazardous than aspirin in terms of major haemorrhage (1.9% vs 2.0% per year).

To what extent can the BAFTA results be applied to a real-world population of elderly patients with atrial fibrillation? To answer this question, we must consider the possible influence that selection bias may have had in BAFTA. Once enrolled in BAFTA, warfarin-eligible patients had a 50% chance of receiving aspirin, which is less effective than warfarin at preventing stroke. Because it would be unethical to deny warfarin to patients for whom it is clearly indicated, participation in BAFTA was restricted to patients for whom there was...
clinical uncertainty about which of the two treatments should be used. From a physician’s perspective, this eligibility requirement probably encouraged the recruitment of patients at a lower risk of stroke. For example, recently revised guidelines recommend either aspirin or warfarin (depending on the individual physician’s and patient’s preference) for patients with a CHADS2 score of 1 (CHADS2 scores patients with atrial fibrillation to determine their risk of stroke. For a patient older than 75 years, this would mean he or she had no other stroke risk factors, such as hypertension, diabetes, history of stroke, or congestive heart failure). ¹

Therefore, compared with other study populations, participants in BAFTA had a lower prevalence of risk factors for stroke (table). It is noteworthy that just over a fifth of the 4639 patients identified were enrolled, and 1570 patients were excluded because warfarin was the only appropriate treatment for these patients. The low prevalence of risk factors for stroke might, in addition to good control of blood pressure and the large proportion of patients already taking a vitamin K antagonist at study entry, explain the lower than anticipated rate of thrombotic events in BAFTA. Nevertheless, the fact that BAFTA showed that warfarin is more effective than aspirin, even in this relatively low-risk group of patients, adds to other evidence² that, in patients with atrial fibrillation, anticoagulation protects patients against stroke more effectively than antiplatelet therapy.

We agree with Mant and colleagues that the lack of difference in major haemorrhage between the two groups is surprising. Additionally, the rates of haemorrhage in BAFTA were significantly lower than rates in another study of elderly patients treated with warfarin.³ How might we explain these observations? First, just over four-fifths of the patients in BAFTA were taking warfarin or aspirin before enrolment, which means that BAFTA selected a group of individuals who had already survived exposure to antithrombotic therapy. Second, we presume that combination therapy with an antiplatelet and a vitamin K antagonist (e.g., clopidogrel and warfarin), a known independent risk factor for warfarin-associated haemorrhage, was not permitted. Third, the number of potential patients initially invited for an electroencephalogram is not clear. Possibly, the individuals who responded to the invitations to participate represent an overall healthier population (i.e., less likely to be recently discharged from hospital). Knowledge of the prevalence of risk factors for haemorrhage (including history of haemorrhage, renal disease, and anaemia) would allow a more informed interpretation of the bleeding rates.

Despite these considerations, BAFTA adds important new information for the care of elderly patients with atrial fibrillation. Mant and colleagues enrolled an unprecedented number of patients in an age group that has been largely under-represented in randomised trials. BAFTA firmly establishes the superior efficacy of warfarin as a stroke-prevention strategy in elderly patients with atrial fibrillation. However, in the future, our greatest challenge will be to identify those patients (elderly or not) who are truly at the highest risk of major bleeding, particularly intracranial haemorrhage. For everyone else, no matter the age group, the benefits of well-managed warfarin substantially outweigh its risks.

*David Garcia, Elaine Hylek
Department of Internal Medicine, University of New Mexico, Albuquerque, NM 87131, USA (DG); and Department of Internal Medicine, Boston University, Boston, MA, USA (EH)
davegarcia@salud.unm.edu

DG and EH have received research funding from Bristol-Myers Squibb and acted as consultants for Bristol-Myers Squibb. EH has also received research funding from AstraZeneca.
Phenylketonuria is a metabolic disease of human beings that is caused by mutations in the phenylalanine hydroxylase (PAH) gene. Catabolism by PAH is reduced and hyperphenylalaninaemia ensues. Phenylketonuria is inherited in a Mendelian recessive manner, and the severity of the metabolic phenotypes is ranked by plasma concentration of phenylalanine before treatment and phenylalanine daily tolerance.

In the past 40 years, patients with phenylketonuria had only the choice of a synthetic low-phenylalanine diet to prevent major disease manifestation (eg, mental retardation). In today’s *Lancet*, Harvey Levy and co-workers assess the treatment of phenylketonuria by use of an alternative approach based on sapropterin—the 6R-epimer of the natural cofactor of PAH, tetrahydrobiopterin (BH4). The investigators concluded that sapropterin is a realistic adjunct in the treatment of some patients with phenylketonuria, combined with a low-phenylalanine diet; for some patients sapropterin treatment might even be an alternative to the diet.

In 1999, Kure and co-workers showed promising results for short-term treatment of phenylketonuria with BH4. Subsequent studies have shown effective supplementation with BH4 for short-term and long-term treatment of patients with phenylketonuria who present with various metabolic phenotypes. The phase III study by Levy and co-workers confirms and extends previous findings. Furthermore, the study highlights the potentially important effect of this treatment for patients with phenylketonuria: about half the patients had substantial correction of phenylalanine plasma concentration. Importantly, Levy shows that long-term sapropterin treatment is efficacious and safe.

PAH catalyses the rate-limiting step of phenylalanine catabolism in the liver, with BH4 as a cofactor and dioxygen as an additional substrate. Researchers first proposed that BH4-responsive patients carry at least one mutant allele, have substantial residual activity of PAH, and that some patients with BH4-responsive mutations might have reduced affinity for the cofactor—as seen in other cofactor-responsive metabolic diseases. Some genotypes are associated with BH4-responsive phenylketonuria. However, the supportive role of BH4 in PAH activity is complex because it acts as a cofactor (ie, necessary for catalysis) and as an inhibitor (ie, keeping PAH in a low active, highly stable state ready for activation on an increase in plasma concentration of phenylalanine; figure).

Moreover, lack of information for the molecular basis of the action of BH4 supplementation in patients with phenylketonuria prompted research to investigate the molecular mechanisms for BH4-responsiveness. These studies lend support to BH4 having a role in the protection of wildtype and mutant PAH against degradation in vitro and in vivo, and towards enzyme inactivation. By contrast, only a few mutations affect the binding affinity for the cofactor.

Additional mechanistic explanations for responsiveness focus on the substoichiometric relation between PAH and BH4 concentrations under physiological conditions in the liver, which could increase PAH activity on cofactor administration by increased saturation of the enzyme by the cofactor (especially for mutations that lead to low affinity for BH4). Structural and biophysical studies further suggest that the 1,2-dihydroxypropyl side-chain of the 6R-epimer of BH4 is essential for enzyme regulation and the large stabilisation induced...
by the natural cofactor compared with other functional tetrahydropterins.9

Clinical trials such as that done by Levy and co-workers enable consolidation of the feasibility of long-term and safe sapropterin-based treatment for patients with phenylketonuria. Such treatment would ease the social burden of classic phenylketonuria treatment that is based on dietary restrictions.

Angel L Pey, *Aurora Martinez
Department of Biomedicine, University of Bergen, N-5009 Bergen, Norway
aurora.martinez@biomed.uib.no

We declare that we have no conflict of interest.


Protecting the rights of those in conflict

The world’s 40 most unstable states contain 38% of the world’s population and have a per-capita gross domestic product of US$750, just 10.3% of the global average.1 Within these states, much of the world’s violence and human rights abuses occur. Beset by conflict, poverty, structural violence, and rights abuses, fragile states fail to meet the basic needs of their citizens. The humanitarian community is now almost always present in fragile states. They often maintain activities throughout the protracted phases of crisis: during conflict, in the descent into violence, and in the unstable peace and the chronic uncertainty that characterise the fragile state’s post-conflict status.

In the context of armed conflict, disaster, and generalised violence, it is now widely accepted that humanitarian organisations must not only bring altruism and technical competence, but also provide assistance in a rights-based approach.2 The UN Special Rapporteur on the Right to Health, Paul Hunt, states: “The right to health can be understood as a right to an effective and integrated health system, encompassing health care and the underlying determinants of health, which is responsive to national and local priorities, and accessible to all.”3 These rights are reflected in the standards and duties the humanitarian community has set for itself through a code of conduct, minimum service standards, and a charter which guide the humanitarian response through the volatile twists and turns that a fragile state environment presents.4 To these responsibilities should be added the duty to observe and record the circumstances that affect their beneficiaries.

Before a conflict spreads and deepens, humanitarian organisations can work to stabilise fragile states through efforts that include capacity building for accountable governance, strengthening of education (particularly for girls), creating microfinance schemes, and delivering health services, especially to vulnerable and marginalised peoples. These actions help build civil society, increase access to livelihoods, and improve human rights awareness, which can help stabilise the state and lessen the risks of collapse into violence.

During violent conflict, the Universal Declaration of Human Rights, the Geneva Conventions, and other core instruments of human rights and humanitarian law

www.thelancet.com Vol 370 August 11, 2007 463

See Editorial page 458
See Series page 521
set out behaviour in conflict, but are routinely violated with impunity, even by their signatories. Humanitarian organisations may be the only source of support and protection to displaced and other affected populations in circumstances of violence. Amidst the chaos of war, both relief work and the humanitarian presence itself can help stabilise communities and prevent further displacement from vital services. While humanitarian organisations are front-line responders in conflict, they are also witnesses to the fate of populations beset by conflict, and often are their sole advocates. In this role, they can document abuses of human rights, and affirm that aid is delivered on the basis of need alone, without discrimination on excludable grounds (eg, race, religion, nationality). Yet aid organisations themselves are at potential risk of unwittingly acting as instruments of local government or donor government policy.

The postconflict period is often only a narrow window during which fragile governments must establish their legitimacy or slide back into collapse. Postconflict states are beset by inequities, discrimination, and lack of access to services, which must all be addressed if stability is to return. Wars leave populations seriously traumatised, but an unjust peace that does not seek to redress the loss of basic human rights and human security is likely to come unstuck, leading to further violence. Without constructive efforts that visibly improve rights and livelihoods, the emotional scars of war can undermine confidence in a frail government. Visible efforts to improve health services help to provide some of that confidence. Evidence for improvement in a population’s health status can be a measure of the improved right of access to care. Conflict, indeed violent conflict, will certainly continue through the 21st century. The challenge is to improve respect for human rights, reduce the risks of conflict, and protect the health of populations unable to escape violence. Documentation can improve our ability to protect rights in conflict in the same way that evidence has improved the practice of medicine. Examining the patterns of gender-based violence in Liberia helped improve the protection of women in that and other conflicts. Population-based data provided important evidence on the scale and nature of human rights abuses in Kosovo. The careful documentation of injuries from antipersonnel mines was a key factor in building support for the 1997 Treaty of Ottawa, which banned the production, development, and stockpiling of such mines. The collection and use of evidence to better protect populations now needs to become a standard activity of humanitarian organisations and others concerned with the safety of populations in the midst of violence. Information on rights in a population should be added to the measures of health status to help us to more fully understand the needs of populations in the various phases of conflict.

*Gilbert Burnham, Courtland Robinson
Center for Refugee and Disaster Response, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
gburnham@jhsph.edu

We declare that we have no conflict of interest.

When rumours derail a mass deworming exercise

The role of misleading or scaremongering news media reports in causing drug crises is well known. Pharmacovigilance systems can play an important educational and preventive role through safety monitoring of products and effective communication about safety issues with health professionals and the general public. That failure to use an existing pharmacovigilance system in implementation planning and weaknesses in communications about medication could lead to mass hysteria and civil unrest might thus seem implausible. But this reaction happened in several regions of Ghana on Feb 12, 2007.

Ghana had its first national deworming exercise, jointly organised by the Ghana Health Service, the Ghana Education Service, and UNICEF, on Feb 12–16, 2007. Trained teachers supervised the administration of a single 500 mg mebendazole tablet to nearly 4·5 million children, aged 3–15 years in 28 043 public schools. The drug had been imported by UNICEF; it was manufactured in 2005 and the expiry year was 2010. Within hours of the start of the programme, there were reports on local radio stations about deaths and serious side-effects affecting several children in three administrative regions. These reports led to considerable public disorder. In some instances, teachers were attacked and schools were shut.

The Ministry of Health immediately commissioned the independent pharmacovigilance centre at the University of Ghana Medical School to investigate. The investigators, in collaboration with the WHO Programme for International Drug Monitoring, found no deaths, three admissions to hospital for suspected Plasmodium falciparum malaria, and scattered reports of mild stomach aches, nausea, and cramps (known adverse events of mebendazole). The investigators also noted localised mass hysteria, including parents rushing their children to hospital or giving them palm oil in the belief that it would induce emesis; attacks on teachers by irate parents and carers, and attendance at hospital by over 350 children, all of whom were reassured and discharged. A possible source of the rumours was the emergency activities associated with the death of a child, killed by a falling wall, 2 h after the start of the deworming exercise. The report by the investigating team was presented by the Deputy Minister for Health at a press conference. The news that the Ghana Health Service had asked an independent team to investigate helped restore confidence and calm nerves. Other assurances came from UNICEF, which attested to the safety of the drug and widespread coverage of these findings in the print and electronic news media seemed to assure parents and calm the situation.

Incidents such as this one point to the need for active pharmacovigilance, excellent communication, and crisis-management planning to accompany public-health programmes that involve mass administration of a drug, especially in countries where resources are limited, rumourmongering rife, and populations volatile. The possible effects of this crisis and its management on further public-health interventions in Ghana are a worry. Interestingly, anxieties about side-effects of deworming drugs (praziquantel and albendazole) and the lack of proper communication and effective collaboration between education and health workers had been previously noted as barriers to success in a school-based deworming exercise in Ghana. Similar findings have been made in Turkey, where allegations that mebendazole, used in a school-based deworming exercise, caused sexual sterility and was an investigational drug being used for experimental aims by foreign researchers led to a 50% fall in drug administration.

Panel: Lessons from reactions to the Ghana deworming programme

- Comprehensive information about safety profiles and side-effects is needed
- Parents and the general public should be briefed about the purpose of public-health programmes that involve mass drug administration, especially in children or pregnant women
- Children and parents (in this case) should be informed about possible side-effects and how to react should they occur
- All stakeholders need effective briefing, before the start, about any programme that involves mass treatment. Such briefing should include journalists, especially in areas where ethics and good practice in reporting are weakly established. Health workers also need briefing, because in this case they were not directly involved (teachers were)
- People and their culture, and problems that could emerge, need understanding
- Independent experts with knowledge of pharmacovigilance must be on hand to quickly investigate unexpected effects
- Standing emergency committee is needed to manage sudden crises and public relations, and for rapid deployment of experts on the ground or by telephone if necessary
- General education and communication about drugs and public-health programmes is needed so that citizens can understand what is happening
- In planning and communication, provision is needed to deal with effects of poor education and knowledge, and how this might lead to irrationality and paranoia in large populations
There are several lessons from the recent Ghana experience (panel). The political and sociological aspects of such events are also important. On the internet forum\(^3\) associated with the first press report about the incident, there was massive outpouring of anger, resentment, paranoia, and vitriol. The situation suddenly became the focus for a wide range of volatile and dangerous political and social prejudices and hatred.

It is worrying that a responsible activity such as the deworming programme can potentially cause huge loss of confidence in the public-health system and give rise to civil disorder. However, the Ministry of Health’s rapid involvement of the country’s independent pharmacovigilance centre allowed the facts to be established\(^10\) and order to be restored. The Government was seen to react urgently and effectively.

\*Alexander Dodoo, Sam Adjei, Mary Couper, Bruce Hugman, Ralph Edwards
Centre for Tropical Clinical Pharmacology & Therapeutics, University of Ghana Medical School, Korle Bu Teaching Hospital, Accra, Ghana (AD); Ghana Health Service, Accra, Ghana (SA); PSM/QSM, World Health Organization, Geneva, Switzerland (MC); and Uppsala Monitoring Centre, Uppsala, Sweden (BH, RE)
alexooo@yahoo.com

---

**Health care for refused asylum seekers in the UK**

For all its faults, the UK National Health Service (NHS) provides a standard of health care that most people in the world can only covet. Although few in the UK would claim that privileged access to good-quality health care is a divine birthright, there is an expectation that services will be protected from unreasonable demands by foreigners. In 2004, the UK Government, without evidence of health tourism\(^1\) but anxious for its electoral popularity, enacted legislation that violated international law and the human rights of the nearly half a million impoverished refused asylum seekers, by charging them for hospital care.\(^2\) The responsibility lies with a Cabinet that, despite—or perhaps because of—consisting of an unusually large number of lawyers (including the then Prime Minister, Tony Blair), has been criticised for disregarding international and national law.\(^3\)

The importance of health-specific human rights to health is traditionally underplayed by governments through policy, health professionals by default, and the general public through ignorance. Health-specific human rights form part of what are known as economic, social, and cultural rights (ESCR), which is one of the two major human-rights categories (the other being civil and political rights). There has been much debate about which category is the most important. Some argue that the so-called negative rights of political and civil liberties, such as the Human Rights Act, form the true core of human rights. Others champion the fundamental right to basic necessities, including employment, education, and the highest attainable standard of health care. Either way, once ratified, all human-rights treaties are equally binding on governments.

The beginnings of the stratagem date back to 1997, when the UN Committee that monitors the UK’s observation of its ESCR obligations reminded the Government that it is legally required to incorporate the treaty within domestic law.\(^4\) The UN Committee was...
compelled to reiterate the admonition more explicitly in 2002. 5 Thus, when access to free hospital-care was withdrawn 2 years later, refused asylum seekers were illegally denied not only their right to the highest attainable standard of health, but also their right to judicial remedy. Theoretically, access to free primary-health care is retained because family doctors have a discretion to enrol anyone, but NHS guidelines illegally stigmatise applications from refused asylum seekers as ineligible and charging is encouraged. 6 Furthermore, every application takes place against a backdrop of repeated threats from the Government to withdraw the discretionary status.

Where should doctors stand on the issue of health rights? A legitimate motivation for aspiring to medical practice, which could be consonant with principled objections to health rights, is difficult to conceive. The UK’s General Medical Council 7 ranks the duty to “Protect and promote the health of the patients and the public” second only to “Make the care of your patient your first concern”—a ringing endorsement of health rights. The executive summary of the new NHS policy document Human Rights in Healthcare 8 begins uncompromisingly with “Neglecting people’s human rights is bad for their health”. In 1993, the genocide scholar Raul Hilberg introduced the triangular notion of victim, perpetrator, and bystander as a prerequisite for genocide—an observer’s indifference provides encouragement. 9 Although genocide is heinous, every preventable death is equally tragic. Tenably, any reluctance from a profession which claims the pre-eminent role in health care—to recognise, validate, and engage with health rights—is analogous and amounts to bystanding.

By contrast with the Government’s pronouncements, the report on asylum seekers from Parliament’s Joint Committee on Human Rights serves as a model of clarity and probity. 1 “All asylum seekers including those whose claims have been refused and the Home Office intends to remove from the UK are still ‘within the jurisdiction’ and therefore beneficiaries of the rights set out in the panoply of international human rights treaties that the UK has adopted. They have simply asserted a fundamental right in seeking asylum. Regardless of their reasons for coming to the UK, asylum seekers must be treated with humanity before and after their applications have been decided. All are owed the human rights obligations successive Governments have assumed.” If, as Tony Blair declared, 10 the great question of our age is what values should govern the future of the world, one answer covers more bases than any other—respect for international laws on human rights.

Peter Lawrence Hall
Doctors for Human Rights, Abbots Langley, Herts WD5 0BE, UK
peterhall@doctorsforhumanrights.org
I played a role in the development of General Comment 14 of the International Covenant on Economic, Social, and Cultural Rights.

Clinical update: sporadic primary hyperparathyroidism

Primary hyperparathyroidism is a biochemical syndrome caused by increased secretion of parathyroid hormone (PTH) from one or more of the parathyroid glands. About 5% of cases are an autosomal dominant inherited disease, mostly caused by multiple adenomas and presenting as an isolated condition (familial primary hyperparathyroidism) or in the setting of more complex syndromes such as multiple endocrine neoplasia types 1/2 or primary-hyperparathyroidism/jaw-tumour syndrome. Most cases are sporadic and caused by a single adenoma (85–95%) or multiglandular disease (5–10%). Less than 1% of cases are caused by a parathyroid carcinoma. External radiation and chronic lithium therapy are the cause of a few sporadic parathyroid adenomas. The origin of most benign parathyroid tumours is unknown. Alterations in the vitamin D receptor in parathyroid cells are currently being investigated as a potential cause of cell proliferation.

Persistently increased serum PTH is associated with activation of PTH and calcitriol receptors in peripheral target organs, the most important of which are the kidney, small intestine, and bone. In the kidney, PTH activates 1-α-hydroxylase which promotes the synthesis of calcitriol and, in turn, leads to increased calcium absorption in the small intestine and to bone resorption. The biochemical hallmark of primary hyperparathyroidism is hypercalcaemia (serum calcium >2·55 mmol/L [10·2 mg/dL, 1 mg/dL=0·25 mmol/L]), which is often associated with hypophosphataemia and hypercalciuria of the absorptive type. Diminished bone-mineral density (osteopenia or osteoporosis) and increased markers of bone turnover are also a common finding. Parathyroid adenoma weight, serum PTH concentrations, and the severity of hypercalcaemia are related, although less conspicuously so for accompanying symptoms. If extreme deficiency of vitamin D occurs, serum calcium may be normal or only moderately increased, despite very high PTH concentrations and severe bone disease.

Primary hyperparathyroidism is a common endocrine disease with an annual incidence of about 20 cases per 100 000. Women account for three-quarters of cases that come to surgery, and a screening investigation suggested a prevalence of primary hyperparathyroidism of up to 3·4% in menopausal women. Chronic hypercalcaemia causes renal lithiasis, nephrocalcinosis, calcification of vascular and heart valve-cusps, acute or chronic calcifying pancreatitis, and chondrocalcinosis, all of which may occur in primary hyperparathyroidism. In 5% of cases, the syndrome is diagnosed after acute admissions for severe hypercalcaemia (>3·50 mmol/L) with dehydration, renal failure, neurological deterioration and even coma (parathyrotoxic crisis). Although data on the risk of bone fracture are conflicting, recent cohort studies suggest that primary hyperparathyroidism is associated with increased risk of vertebral crushing and femur, rib, or pelvic fractures, which peaks 5 years before parathyroidectomy. Furthermore, non-traditional symptoms of primary hyperparathyroidism, such as fatigue or irritability, are more common than in the general population. The prevalence of arterial hypertension doubles that of the general population and continues to increase even after parathyroidectomy. The reasons for the association between primary hyperparathyroidism and arterial hypertension remain obscure. Both renal impairment and altered intracellular calcium metabolism might be implicated.

The proportion of asymptomatic primary hyperparathyroidism is continuously increasing as a result of routine calcium measurement. In certain units, asymptomatic patients account for more than half the cases. Some investigators, however, feel that truly asymptomatic primary hyperparathyroidism is not so common if history taking is accurate enough and if target organs and mental status are properly assessed. For example, renal ultrasound can detect asymptomatic renal lithiasis and bone densitometry can show unexpectedly low bone-mineral density.

Primary hyperparathyroidism is diagnosed either because presenting symptoms lead to assay of serum calcium, or because hypercalcaemia is detected in patients undergoing a blood test for an apparently unrelated reason. In settings in which calcium is not included in routine blood testing, assay should be requested in patients with the conditions listed in the panel. Primary hyperparathyroidism is the only cause of hypercalcaemia associated with high concentrations of serum PTH (>55 pg/mL). Hypercalcaemias of tumoral, dietary, pharmacological, or endocrine non-parathyroid
Comment

origin inhibit parathyroid PTH secretion. Tumoral hypercalcaemias (pseudohyperparathyroidism) can be appropriately diagnosed by measuring the PTH-related peptide. Familial hypocalciuric hypercalcaemia is a hereditary disorder, usually asymptomatic, and with mutations of the parathyroid calcium-sensing receptor gene. It should be considered in the differential diagnosis of hypercalcaemia, particularly when the calcium urinary excretion is low (<2·5 mmol or 100 mg per 24 h) or hypercalcaemia persists after an unsuccessful cervical exploration.

Increased serum PTH is being found in normocalcaemic patients undergoing metabolic bone assessment. These patients probably represent the earliest manifestation of primary hyperparathyroidism after secondary causes of increases in serum PTH have been ruled out.5

Medical treatment is reserved for patients with severe hypercalcaemia or with a parathyrotoxic crisis, who benefit from lowering of calcium levels before parathyroidectomy. These patients need aggressive fluid replacement, correction of electrolyte imbalance, and forced diuresis. For severe hypercalcaemia, refractory to these initial measures, intravenous bisphosphonates should be administered. Asymptomatic, mild, and moderate primary hyperparathyroidism should not be treated medically. If, for whatever reason, surgery cannot be done, yearly surveillance is advisable, with good hydration, mobility, and a regular calcium dietary supply.6

Parathyroidectomy is the only curative treatment for the disease. For patients with asymptomatic primary hyperparathyroidism, agreed indications for surgical treatment include: calciuria (>10 mmol per 24 h), 30% reduction in creatinine clearance, osteoporosis (T score less than –2·5 at any site), age below 50 years, serum calcium 1 mg above reference values, and medical surveillance impossible or undesirable.6 If no surgical contraindications exist, however, surgery should be discussed with and offered to almost all patients with asymptomatic primary hyperparathyroidism because the disease will progress in 30% of cases, long-term medical follow-up is expensive, with compliance problems, and there is an increased risk of bone fractures and cardiovascular morbidity and mortality. On the other hand, surgery cures over 95% of the patients with less than 2% permanent complications and virtually no mortality.

The surgical treatment of primary hyperparathyroidism has undergone a radical change. The increasing use and accuracy of parathyroid scintigraphy and ultrasonography allows preoperative localisation of solitary adenomas in 75–90% of the patients. Better localisation techniques have generated a move towards less invasive surgery. Currently, focused parathyroidectomy (minimal incision, adenoma excision, and no identification of the normal glands) can be done in 60–75% of all patients with sporadic primary hyperparathyroidism. Focused or selective parathyroidectomy has been shown in randomised trials to lower the risk of postoperative hypocalcaemia and to reduce operation time compared with bilateral neck exploration, with excellent success rates.7–9 For patients with inconclusive localisation studies or suspected multi-glandular disease, bilateral parathyroid exploration is the procedure of choice.

Most commonly, focused parathyroidectomy is done under general anaesthesia, through a small lateral or central neck incision. Patients are discharged on the same day or after an overnight stay. Variations of the procedure include endoscopic or video-assisted approaches, local anaesthesia with early (<4 h) discharge, and radioguided surgery. All of these procedures are being prospectively assessed.

To improve the results of adenomectomy, quick intraoperative measurement of serum PTH has been recommended,10 and is currently the focus of much investigation. Because the half-life of PTH is 3–5 min, serum PTH can be expected to drop soon after focused parathyroidectomy if no other diseased gland is present. The most used and tested criteria for cure calls for a drop

Panel: Disorders in which serum calcium should be routinely checked

- Renal lithiasis (calcium)
- Osteoporosis
- Chondrocalcinosis
- Arterial hypertension
- Cardiovascular disease
- Peptic ulcer
- Pancreatitis of all types
- Functional abdominal symptoms
- Acute or chronic mental disturbances
- Postmenopausal women
- Hypophyseal and pancreatic endocrine tumours (multiple endocrine neoplasia 1)
- Medullary carcinoma of the thyroid and phaeochromocytoma (multiple endocrine neoplasia type 2a)

Comment

of serum PTH values, 10 min after adenoma resection, of at least 50% below the highest previous value (either at induction of anaesthesia or pre-excision). Two recent reports\textsuperscript{11,13} indicate that patients with a single adenoma localised by concordant imaging (ultrasonography and scintigraphy) might be operated on successfully without intraoperative assay of PTH. Gawande and colleagues reported 322 consecutive focused parathyroidectomies in patients with concordant imaging studies.\textsuperscript{11} In only three cases did the quick intraoperative assay of serum PTH identify additional diseased glands. In three of the remaining patients, primary hyperparathyroidism persisted despite an appropriate drop in serum PTH.

The variables associated with postoperative hypocalcaemia include: preoperative serum calcium above 3.0 mmol/L, multiple gland excision, increased alkaline phosphatase, and reoperative and bilateral surgery. Patients at risk of developing postoperative hypocalcaemia should be given oral calcium (1–3 g or calcium ion daily) with or without calcitriol (0.5–1 μg daily). Patients at risk of developing postoperative hypocalcaemia include: preoperative serum calcium above 3.0 mmol/L, multiple gland excision, increased alkaline phosphatase, and reoperative and bilateral surgery. Patients at risk of developing postoperative hypocalcaemia should be given oral calcium (1–3 g or calcium ion daily) with or without calcitriol (0.5–1 μg daily).

Parathyroidectomy improves neuromuscular symptoms and quality of life, decreases bone turnover,\textsuperscript{13} increases bone mineral density, and halts renal lithiasis,\textsuperscript{14} but does not reverse long-standing arterial hypertension, chondrocalcinosis, soft-tissue calcifications, or renal failure. In postmenopausal women with primary hyperparathyroidism, parathyroidectomy increases bone mineral density at all sites although the size of the effect is variable (particularly at the lumbar spine) and difficult to predict for an individual. Improvement is greater in young patients and in those patients with preoperative osteoporosis.\textsuperscript{15} After parathyroidectomy, a reduction in fracture risk is already detectable at 1 year and persists at 10 years.\textsuperscript{13,16} VanderWalde and colleagues reported that the 10-year fracture-free survival after a diagnosis of primary hyperparathyroidism was 73% in patients treated surgically compared with 59% in those just observed.\textsuperscript{16} Fracture risk-reduction was greatest at the hip level.

An improvement is also seen in fine variables of cardiovascular function investigated with echocardiography during exercise, such as reduction of maximum blood pressure, ST-segment depression, and ventricular extrasystolic beats.\textsuperscript{17} A higher prevalence of left ventricular hypertrophy has also been seen in primary hyperparathyroidism compared with controls, even when adjusted for the presence of arterial hypertension. Parathyroidectomy resulted in a measurable decrease of left ventricular wall-thickness.\textsuperscript{18} In agreement with these findings, risk of myocardial infarction increases up to 10 years before surgery and declines to a normal level more than 1 year after parathyroidectomy.\textsuperscript{19}

\*Antonio Sitges-Serra, Anders Bergenfelz
Department of Surgery, Hospital del Mar, 08003 Barcelona, Spain (AS-S); and Lund University Hospital, Lund, Sweden (AB)

asitges@imas.imim.es

We declare that we have no conflict of interest. AS-S has a 4-year grant for his research from AGUR (Agència de Gestió d’Ajuts Universitaris i de Recerca, #2005SGR 00961).


More than 70 million people in Bangladesh are estimated to be exposed to toxic levels of arsenic from their drinking water in what WHO has called the “largest mass poisoning of a population in history”.

The sad irony is that the problem is the unintended consequence of a campaign in the 1970s and 1980s by international development organisations, including UNICEF, to get villagers to stop drinking dirty surface water. More than 10 million wells were installed in Bangladesh. Unfortunately, nobody thought to test the wells for the presence of arsenic, which has since been found in a large percentage of the wells at amounts sometimes five, and even ten times over, what is deemed safe by WHO. The arsenic is thought to be naturally occurring, possibly drawn down from the Himalayas over thousands of years, and then released by the sudden and massive withdrawal of water from the aquifers.

Estimates vary widely about how many people are affected. A 1999 survey found that some 27 million people were drinking water tainted with 50 parts per billion (ppb) arsenic. Another 50 million were exposed to over 10 ppb—the WHO safety limit that many now say is out-of-date. Some estimates based on arsenic tests in Taiwan have put the safe amount at 1·7 ppb.

Although the world’s international institutions have paid scant attention to this continuing catastrophe, it has drawn the interest of public-health experts and scientists from among the world’s elite research institutions. Joseph H Graziano vividly remembers the first time he went to Bangladesh to assess the crisis with a small team from Columbia University.

“We saw with our own eyes for the first time some of the epidemiology associated with arsenic exposure. We all came away saying that this was a life altering experience”, said Graziano, the associate dean for research and a professor of environmental health sciences at Columbia University’s Mailman School of Public Health.

The symptoms of long-term exposure to arsenic begin with the blackening of the hands and feet, progressing to nodular growths, and later to open sores and gangrene. Eventually, it can lead to cardiovascular and reproductive damage and to virulent cancers of the bladder, skin, lungs, and liver. In children, the exposure is also thought to lead to learning disabilities and other neurological effects. Researchers at Columbia University say that the arsenic poisoning could double Bangladesh’s cancer mortality rate within two decades.

For such a massive disaster, the response by international aid agencies has been small, especially since researchers estimate that substantial mitigation could be achieved for less than US$100 million.

Arsenic poisoning from drinking water occurs to a much lesser extent in 60 countries around the world, from Taiwan to the USA. Worldwide, over 150 million people are estimated to be drinking the poisonous water. Bangladesh is one of the world’s poorest countries. The cash-strapped government does not have the capacity to deal with this problem, and recent political upheavals have not helped the situation.

But along with this exceptional disaster has come an unusual opportunity for collaboration between researchers from myriad branches of science, including experts in epidemiology, environmental health, medicine, economics, statistics, sociology, hydrology, and geochemistry. “This is a really good example of true multidisciplinary work”, said David C Christiani, a professor at Harvard’s Schools of Public Health and Medicine who is developing a research project with researchers from the Massachusetts Institute of Technology.
Many Bangladeshis drink and cook rice with well water that is contaminated with arsenic (MIT). “Everybody comes with their different perspective on these things. This cross-talk gives you a holistic perspective”, says Christiani.

At Columbia, one of Graziano’s principal collaborators is Alexander van Geen, a senior research scientist at the university’s Lamont-Doherty Earth Observatory, who specialises in geochemistry. He says that his team of earth scientists worked closely with the public-health researchers and the two sides realised that they could not only study the problem, but also that it was fixable most immediately by creating new, safe wells for their research participants and actively testing and labelling the contaminated wells.

Other researchers are looking at how to mitigate the effects of arsenic ingestion. Once people have been exposed for several years, conservatively estimated at a decade, the DNA damage almost guarantees that cancer is imminent, said Habibul Ahsan, a cancer researcher at the University of Chicago. Ahsan’s research is now focused on how to undo the genetic damage caused by arsenic. Supplements of micronutrients such as selenium and vitamin E are showing positive effects.

Scientists have been working on the arsenic problem for less than a decade, but they have already developed innovative solutions like inexpensive filters, methods to cheaply test wells for contamination, and more elaborate programmes to map out the extent of the arsenic contamination, so that safe water can be drawn from either existing wells or new ones that tap into deeper, uncontaminated aquifers.

Graziano and Christiani’s teams have assessed arsenic removal strategies, and have found that seemingly simple solutions may have unintended side-effects. Scientists in both Bangladesh and the USA, for instance, have come up with cheap and easy ways to make filters that can safely remove arsenic, using such commonly available ingredients as rusty nails and coal ash. But there are drawbacks to these approaches; the filters eventually get clogged up and need to be disposed of, which could put the arsenic back into the ground.

Although in practice many of the solutions have technical and effectiveness hurdles, everybody working on the problem agrees it is solvable. What is lacking is political will to act. And, even more importantly, the hundred million or so dollars in funding needed to put the fixes into practice—a relatively small amount to save millions of lives.

Meanwhile, the search for a solution, and even to understand the source and means of distribution of the arsenic, continues. Peter Ravenscroft, who is working on a book about the subject at Cambridge University for the Royal Geographic Society, cites research from west Bengal that Bangladeshis are actually exposed to. Bangladeshis, for example, may get water from many different sources when they go out to work, and the hard physical labour they do means they may consume nearly twice as much water as the average person.

Christiani has been sharing his findings with Charles Harvey, a professor of hydrogeology at MIT. The two are putting together a collaborative project that would link their seemingly disparate research activities.

In two separate locales, Harvey and his team from MIT are surveying wells to learn about factors such as the prevalence of arsenic in wells, how the rates vary with the season, and the effects of irrigation in spreading arsenic. Meanwhile, Christiani’s team is working with researchers from a local non-governmental organisation—the Dhaka Community Hospital Trust—to map out the extent of arsenic contamination, so that safe water can be drawn from either existing wells or new ones that tap into deeper, uncontaminated aquifers.

These types of complications have further emphasised the need for epidemiological studies. “We know a lot about arsenic as a classic poison but there is a lot we still do not know about how it operates in people”, said Christiani.

Christiani’s work has focused on using biological markers of cumulative exposure, like toenails and urine, to get a more accurate measure of how much arsenic Bangladeshis are actually exposed to. Bangladeshis men, for example, may get water from many different sources when they go out to work, and the hard physical labour they do means they may consume nearly twice as much water as the average person.

Christiani has been sharing his findings with Charles Harvey, a professor of hydrogeology at MIT. The two are putting together a collaborative project that would link their seemingly disparate research activities.

In two separate locales, Harvey and his team from MIT are surveying wells to learn about factors such as the prevalence of arsenic in wells, how the rates vary with the season, and the effects of irrigation in spreading arsenic. Meanwhile, Christiani’s team is working with researchers from a local non-governmental organisation—the Dhaka Community Hospital Trust—to measure the effects of arsenic on pregnant women and small children.

Along with the scientific advances that could lead to mitigation, the research could also have another effect—a political one. “Having the academic research lends the issue credibility and helps put pressure on the government and the international community to set stricter standards on arsenic and to start coming up with innovative technologies to provide safe water to people”, said Christiani.

Samuel Loewenberg
In the southwest corner of Caracas, the 54-year-old Coche public hospital has only one radiograph machine working; the second was recently disabled after rats chewed through the electrical wires. Just 5 km away in the neighbourhood of El Valle, brand new health clinics have been opened as part of President Hugo Chavez’s social mission, with state of the art equipment funded by the government. Both treat patients free of charge, but the differences are stark.

This contrast illustrates perfectly the reality of the current state of the country’s public-health system. Instead of re-equipping and improving existing hospitals, the government has poured millions of dollars into creating a parallel social health-care system, which still is not growing fast enough to meet people’s needs.

In the 1999 constitution of Venezuela, article 83 guarantees the right of all citizens to access free health care and receive adequate treatment. This policy forms part of the socialist agenda by Chavez, the controversial former paratrooper turned revolutionary leader. His focus on the poor was largely seen as a giant step forward in settling part of the social debt accumulated after years of abandonment and inefficiency in the public-health system. 8 years on from that bold initiative, citizens are still receiving free care. But now, as a result of the fragmented Venezuelan health-care system, poor management, constant changes in administration, and periodic shortages of supplies, the quality is not what many had anticipated.

Maria Elena Rodriguez, a health research coordinator with the Caracas-based human rights watch group Provea, says of the government’s health reforms “I think a lot of measures in the past 5 years have been admirable but they have been followed by plenty of management and implementation failures as well as a lack of transparency in spending”.

The major advances under Chavez’s government in health revolve around the social development programmes known as missions. These programmes are funded principally by the state oil company Petroleos de Venezuela, which, by riding the wave of an oil boom, has fuelled the government’s political projects and kept the economy afloat.

The mission Barrio Adentro (inside the neighbourhood), the banner social programme of the Chavez government in 2003, has brought primary health care directly to the poorest Venezuelans. The programme has been tremendously popular with citizens who are unable to pay private health insurance and has boosted the president’s ratings. The patients are attended to in their own neighbourhoods by Cuban doctors, who are on loan in exchange for oil shipments. The mission began with just 54 Cuban doctors in 2003 and has expanded throughout the country. According to figures from the health ministry, some 26,819 doctors are now practising in Venezuela.

The original Barrio Adentro I programme was initially characterised by two-storey, octagonal brick clinics located in poor urban areas which serve both as offices and residences for doctors. That model has been giving way to well equipped clinics capable of providing advanced health care to Venezuelans. The second phase of the programme, Barrio Adentro II, was inaugurated in 2005 with the opening of 30 diagnostic centres, 30 rehabilitation centres, and several high technology centres. The high-tech centres provide nuclear magnetic resonance tests, three dimensional ultrasound, mammography, video endoscopy, and electrocardiography, among other services.

The original goal to build 1235 new clinics during 2006 was not achieved, although 175 diagnostic, 183 rehabilitation, and six high-tech centres were opened last year according to the health ministry.

With this programme, the public-health system in Venezuela has diversified greatly. It has reached
The printed journal includes an image merely for illustration

Panos Pictures

Chavez’s new health-care clinics have been popular with the Venezuelan public

low-income communities with the intention of relieving the old, overcrowded hospitals. Critics, however, say Barrio Adentro has grown at the expense of the former health-care system and has yet to lessen the number of patients in traditional institutions.

The attention the government gives to Barrio Adentro has also been an issue with Venezuelan public-health workers, who complain of low wages, shortages of basic supplies, and poor working conditions. Many see the social projects as getting preferential treatment while hospitals are left to fend for themselves.

Olga Machado de Castillo, a member of the board of directors and secretary of labour relations in the Venezuelan Medical Federation, a staunch opposition group to Chavez, believes the neglected public hospitals have worsened under the current government. The presence of Cuban doctors in the country, she says, forms part of an active discrimination campaign against Venezuelan health professionals. She calls the arrival of Cuban doctors “an invasion”.

“One of the problems in the country is the presence of supposed Cuban doctors. We have determined through studies that only one in every ten of these doctors is really qualified to be practising in medicine while the others are simple technicians”, said Castillo.

The Federation refuses to acknowledge any advancements or improvements in health under Chavez; it also claims preventable diseases are on the rise, and diseases that had been eradicated in the past have returned once again. The government denies these allegations, saying that reports of tuberculosis and dengue fever outbreaks are media-provoked slander. They also defend Barrio Adentro saying no Venezuelan doctors would do the work of the Cuban medics in impoverished and violent city slums.

But the equipment, personnel, and propaganda that has been used in Chavez’s social missions has come at a steep financial cost, and has left the staff of the country’s long-established public hospitals asking about their share of the money.

Elisabeth Chacón has been working at the Coche Hospital for 26 years. Trained as a paediatrician, she is now the hospital’s director, and is thus very aware of the woes that have beset the hospital. The deteriorating façade and out-of-service lifts convey maintenance problems in the hospital, though the deficit in doctors, nurses, and therapeutic beds highlight a more profound crisis. “The money we receive for 1 year of medicine does not cover our costs for 3 months, and the rest of the year we have to ask for credits and request help from the health ministry”, she said.

The Health Secretariat is in charge of funding, and answers to the needs of 14 hospitals and 87 outpatient clinics in the metropolitan area. Their annual budget equals $248 million but 90% is spent on paying the salaries of doctors, nurses, and other labourers. This leaves little cash for buying supplies, maintaining the infrastructure and ambulances, and buying new equipment.

The government’s response to the deteriorated state of the public hospitals has been the creation of the Barrio Adentro III programme. This initiative has been intended to supply the hospitals with the vital equipment and supplies they have lacked for so long, although so far hospitals say they have seen little progress.

Luisana Melo, the newly appointed director of the Health Secretariat, admits that the public health-care system is in “crisis”, but she remains optimistic about the possibility of changes under her mandate. In her opinion, the principal cause for the degradation of the country’s health-care system is the years of neglect it faced at the hands of the country’s previous “neoliberal” administrations. The policies enacted during these years, as Melo sees it, tried to “weaken the public-health system in order to strengthen the private system”. Melo mentioned the need to audit the number of workers and the deficits in hospitals to better distribute the funds.

Venezuela’s hospitals are still popular with the public. Elena Rodríguez says that despite all the government propaganda, many Venezuelans tend to turn to the decade-old hospitals first, even with new clinics around the corner. “As part of the culture or habit people go straight to hospital when they have a minor ache or pain, which contributes to the hospitals being overrun”, she says.

Melo says this habit is because of a pattern she calls “ruleteo” (roulette table), where patients are repeatedly told to go to the next hospital or clinic to be attended to. “Instead of going to Barrio Adentro for a consultation, the people go directly to the hospitals. This is to avoid being sent all over town in case the diagnostic centres are unable to treat their ailment”, she said.

A proposal by the Lord Mayor’s Office to organise the different public and private health care elements under one control centre could help to fix the roulette problem, according to Melo. Or it could be another good idea that is not followed through to the end, as critics suggest.

Daniel Cancel
Book

Ethnicity and access to health care

My young daughter needed to go to the toilet during a shopping trip in London during the last school holidays. While I stood next to the hand basins waiting for her, a woman came out of one of the cubicles and as she washed her hands, looked up and admonished me in no uncertain terms for the poor job I had done in keeping the toilets clean. Speechless, I watched her storm out. I am not an indolent cleaner; but I am Black, of African descent, and an Australian national. My daughter, having overheard, asked why I had not explained that I was a professor. But why do I have to? And more importantly, how do I explain to my 11-year-old that being part African, cleaner; but I am Black, of African descent, and an Australian national. My daughter, having overheard, asked why I had not explained that I was a professor. But why do I have to? And more importantly, how do I explain to my 11-year-old that being part African, status and ability?

The reality is there is no escaping the social and biological imperative to recognise similarities and differences in other human beings. It proffers clear advantages; indeed, is often protective. However, it is in the interpretation and values assigned to the differences that problems arise. The large body of evidence on health inequalities and the social determinants of health has unravelled the historical effects of discrimination against groups on the basis of particular markers of difference, including race with a more recent focus on ethnicity. These effects are so insidious that addressing them often seems overwhelming.

However, inaction is simply not an option. Health-care professionals have moral and ethical obligations to address the needs of vulnerable populations irrespective of the drivers of vulnerability. A widening human rights discourse on the right to the highest attainable standard of health, in addition to legislative support from international treaties on non-discrimination, provides a further impetus to action. The ongoing challenge is what that action might be. Raj Bhopal’s important and comprehensive Ethnicity, Race, and Health in Multicultural Societies challenges us to achieve better health for ethnic minority populations. The standard and current approach in public health is to first ensure that there is a robust body of evidence to establish causation and to monitor interventions and policy initiatives. Progress in this area has been driven by the increasing voluntary and involuntary global movement of people, highlighting the trend towards multiethnic societies. However, as Bhopal makes clear, the use of race and ethnicity as variables involves conceptual and analytic tensions in how we interpret those health effects that are attributable to genetic predispositions, culturally patterned behaviours of the group being studied (such as high-risk behaviours, health beliefs, and health-seeking behaviour), or to the behaviour of others towards the group being studied.

These tensions also highlight the range of approaches and challenges in responding to ethnic disparities in health. A few years ago I co-authored a series of cultural competency guides for health professionals commissioned by a state health department in Australia. The idea was to provide snapshot community profiles of disadvantaged communities—a range of minority ethnic as well as refugee and asylum-seeking groups—to enhance the delivery and quality of care to individuals from these communities. While acknowledging the complexities of diversity and the risk of stereotyping, the resources aimed to “celebrate cultural differences and sensitise health-care providers to the needs of individuals from different culturally and linguistically diverse backgrounds.” I do not hold this up as one of my better bodies of work; however, I was astounded at the negative reaction from senior South African public-health colleagues a few years later. The notion that quality services could be provided on the basis of focusing on difference seemed to run counter to their efforts of achieving equality for their disadvantaged populations. Bhopal also provides a critical analysis of comparative priority setting in ethnicity and health that results in a focus on differences in disease profiles and interventions to the detriment of addressing conditions that are similar across the whole population.

The fundamental issue is one of equity versus equality. Given the evidence on ethnic disparities, and within a context of finite resources, should the response be to channel resources into eliminating those vulnerabilities in the high-risk, disadvantaged groups? This necessarily focuses on difference, with strategies that target those differences. On the other hand, a system that emphasises equality and universal improvement of quality and access has the potential, with a wider net, to draw in and improve health outcomes for all; an approach supported by scholars such as Geoffrey Rose. It is also the “colour blind” approach favoured by my South African colleagues and invites health practitioners to treat every presenting patient to the best of their ability, without giving primacy to a provider-imposed identity. This approach does
hold some personal appeal; given we all have multiple identities, having a choice of which identity we assume in any particular context would be refreshing. At the doctor’s surgery, I am happy to be a patient; in London with my daughter, I am happy to be a mum and a shopper and not happy to be mistakenly identified as a Black cleaner. But this approach also denies the reality of the complex and dynamic factors that cause and sustain ethnic and, indeed, other health disparities, and ignores the pragmatics of limited health resources available to treat individuals and populations.

So what is the right approach to redressing the very real ethnic disparities in health? In the integration of social sciences with epidemiological data, Ethnicity, Race, and Health in Multicultural Societies provides critical and thought-provoking insights into public-health research and clinical practice with multiethnic populations. Bhopal challenges the reader to reflect on their own ethnicity and identity and how these might interact to affect their health and access to care. There are clearly no universal right approaches and he concludes that the response of health systems needs to be shaped by local circumstances. The sentiment is echoed by Sir Liam Donaldson in the foreword, highlighting the need for flexibility as a critical component of an equitable health service.

Irrespective of the differences in local context, however, there are some common and arguably universal features in ethnicity and health. The focus on health care is a necessary response to disparities in disease specific morbidity and mortality, inequalities in access, and other barriers to health care. But a focus on care is not sufficient to redress the broader societal causes that produce and sustain disadvantage. We accept that societies are diverse and the inevitability of changing demographics and growing multiethnicity. The broader response to ethnic disparities, therefore, is a collective social and political responsibility. We need to engage politically and work towards a shared goal; the type of inclusive society that provides the structures that enable all, irrespective of race, ethnicity, gender, and other markers of exclusion, to enjoy “the highest attainable standard of health conducive to living a life in dignity”. We need to hold our governments and ourselves accountable in achieving this. Like many others, I am Black and a woman—those are facts that will not change. What we as a society choose to make of that and the consequences of that choice are in our control.

Pascale Allotey
pascale.allotey@brunel.ac.uk

In brief

Book Theodore de Mayerne
Hugh Trevor-Roper, who died in 2003, was one of the most distinguished scholars of his generation. He shot to fame in 1947 with his enthralling The Last Days of Hitler and built his reputation as a brilliant essayist and student of early modern history. An error of judgment, however, took some of the lustre off his reputation when he authenticated the so-called Hitler Diaries, which proved to be fakes. The satirical magazine Private Eye nicknamed him Hugh Very-Ropey.

The son of a doctor, Trevor-Roper had a longstanding interest in the history of medicine. His broader interests are apparent in this rich, wide-ranging study, edited from his papers after his death. The subject is Theodore de Mayerne, a physician and polymath who was at the centre of medical politics and courtly diplomacy and intrigue during the reigns of James I and Charles I.

Mayerne was born in 1573 into a Huguenot family. His medical training at Montpellier equipped him to be an urbane, learned physician, but the French Catholic medical faculty despised the upstart Huguenots and their chemical remedies so, in 1611, he moved to England where he gained renown as a clinician as Physician-in-Ordinary to the Stuart court.

Mayerne attended the noblest in the land and continued to experiment and prescribe chemical drugs that were by no means welcome to the fellows of the Royal College of Physicians, but his tact and clinical reputation kept him safe. His record of the constitution and the life-long afflictions and final illness of the sickly James I is one of the jewels of medical history. After James’s death in 1625, Mayerne remained a mighty figure but Charles I distrusted his continental protestant friends and kept him in England. Here he compiled an official report on measures to prevent the spread of plague. During the civil war, he balanced himself between parliamentarians and royalists and he even negotiated Oliver Cromwell’s rule comparatively unscathed. He died in 1655.

Mayerne represents a type of doctor who has vanished and the clue is in the book’s title, Europe’s Physician. He was part of the European medical clerisy: elite physicians who could converse in Latin, Greek, and French. The idea of, say, a British medical profession was a long way off. Mind you, when, around 1600, the Paris Medical Faculty caught in Mayerne a whiff of German chemistry coupled with Protestantism they smelled the future.

Christopher Lawrence
ucgalaw@ucl.ac.uk
Profile
Joseph Graziano: tackling arsenic poisoning in Bangladesh

In his office in Harlem overlooking the Hudson River, Joseph Graziano’s hands carve up the air as he tries to demonstrate the flow of pollutants between aquifers beneath the surface soil of Bangladesh. He gestures to a picture on his wall of a purple landscape that represents a 25 km² area, explaining how the groupings of blue, green, and red markings show varied concentrations of the arsenic that poisons the region’s underground water. He pauses to reflect: “This is my life.”

For Graziano, a pharmacologist by training and now the dean of research at Columbia University’s Mailman School of Public Health, New York, USA, this foray into hydrogeology might seem far afield. But Graziano is no dilettante. For the past 8 years, he has been meeting regularly with experts in hydrogeology, sociology, economics, and many other fields in a concerted effort to tackle one of the worst public-health disasters in history: the arsenic-contaminated water in Bangladesh that affects more than 70 million people.

“It is a very unusual collection of disciplines, and we have learned from each other”, says Graziano. In monthly seminars, 15 scientists discuss their findings and give each other informal lessons in biomedical science, earth sciences, and economics. This process means that the scientists must not only cooperate, but also have a good understanding of each other’s expertise. “You live with these guys long enough, you really learn their language. It has been one of the most wonderful experiences of my life, to work with such a multidisciplinary group”, says Graziano.

Alexander van Geen, a senior research scientist at the university’s Lamont-Doherty Earth Observatory, says that collaborating on the Bangladesh effort has had a profound effect on him. “This has been the most exciting project I have worked on in my career. Without Joe it would not have happened.” The idea to bring different disciplines together to tackle this issue was Graziano’s, says Habibul Ahsan, a cancer specialist at the University of Chicago. It was Ahsan, a former Columbia University researcher who was born in Bangladesh, who first brought the problem to Graziano’s attention in 1999. “He has a big-picture coordinating ability. He can bring people from diverse disciplines to the same table and get things done”, says Ahsan.

When he spoke to The Lancet, Graziano had just returned from Dhaka, where he led a symposium on the arsenic disaster. The meeting brought together more than 200 experts from Bangladesh and around the world, as well as senior officials from UNICEF and WHO. Graziano has lost count, but he estimates this is the 20th time he has been to Bangladesh since he first got involved in 1999. At that time, he went with a small team to investigate the problem at first hand. He says it was a life-changing experience. “Now, we all wonder what we used to do before.”

Graziano’s previous focus had been on the health consequences of exposure to metals. Between 1983 and 1998, he led a team to investigate environmental lead exposure in children in Kosovo. His pharmacological research laboratory developed the oral drug now used to treat lead poisoning. It was the study of toxicology that brought Graziano into public health, and it has been a uniting thread through his many endeavours. In addition to his work abroad, he helped start various community-based environmental health research projects in West Harlem, where the Mailman School is located.

One of Graziano’s best traits, say his colleagues, is his ability not only to think about creative ways to tackle a problem, but his openness to approaches that are different from the ones he was trained in. Graziano’s openness is crucial, says van Geen: “I find him very responsive. He has a vision of what the larger issues are. But he is willing to reconsider his position. In a group like ours, if one of the leaders was not open to new ideas, the group would fall apart.”

Van Geen’s geological team, working with Graziano’s public-health team, found that 90% of the people in the area surveyed lived within 100 m of a safe well. After an extensive information campaign, in which the unsafe wells were labelled, they found that two-thirds of people using contaminated wells had switched to safe sources of water. “The challenge now is to try to influence policies in Bangladesh, but very cautiously”, says Graziano. “We don’t want to be a bunch of white guys going in and saying, ‘we know what to do’—the way the international agencies that promoted the original well-digging programme did.”

Graziano’s team worked with 60 local public-health researchers in Bangladesh to study the long-term effects of arsenic poisoning. Graziano recalls that when his team from Columbia was putting together their project, the ethics review committee told them: “you can’t just make this a natural history of arsenic poisoning”. They had a moral obligation to help mitigate the poisoning. So far, the team have tested 6000 wells, which are used by more than 70 000 people. 60% of those wells proved to have arsenic concentrations above 50 μg/L. After the wells were tested, the populations using the affected wells were given health education programmes to encourage them to switch to safe sources of water. “It is an exhausting project, it requires going to Bangladesh frequently, it’s not a light workload. And somehow he manages to be available, make the right decision when necessary”, says van Geen. “I worry a little bit about what is going to happen when he retires.”

Samuel Loewenberg
c/o The Lancet, London, UK
Harvey R Colten

Paediatric immunologist who described the complement cascade and mentored many colleagues. He was born on Jan 11, 1939, in Houston, TX, USA, and died on May 24, 2007, in New York, NY, USA, of complications from colon cancer, aged 68 years.

During the course of his career, Harvey Colten made many important contributions to paediatric medicine. His particular interest lay in the immune system, specifically the complement cascade, and his publications in that field attest to his brilliance as a scientist-physician. But perhaps more than his individual work, Colten’s contribution has come from the way he nurtured the careers of the brightest young doctors and scientists who crossed his path.

In 1986, for example, Colten decided to leave Boston’s Children’s Hospital, after 16 years, to take up the chair of paediatrics at Washington University, in St Louis, MO, USA. Losing Colten himself was bad enough, remembers his friend David Nathan, president emeritus of the Dana-Farber Cancer Institute, Boston, MA, but what made it worse was that many of the department’s best members went with him. “It was like a great plague had swept through the department”, he says. “He was like the Thief of Baghdad, but in the end it was good for paediatrics because he was extremely supportive of them.” Lawrence Nogee, associate professor of neonatology at Johns Hopkins School of Medicine, Baltimore, MD, who worked in Colten’s group in the late 1980s and early 1990s, agrees: “He trained some real leaders in the field. He left a very important legacy in terms of training and moulding the next generation.”

Colten received his medical degree from Western Reserve University, now Case Western Reserve, in Cleveland, OH, in 1963. He spent 5 years at the National Institutes of Health and joined Harvard University in Cambridge, MA, in 1970. From 1986 to 1997 he headed the paediatrics department at Washington University before moving to Northwestern for 2 years and then Columbia University New York in 2002. Colten was always interested in translating scientific study into patients’ care, says Nathan. Some of his important contributions included recognising early that the common pathway of much lung disease was inflammation, and describing the synthesis and role of elements of the complement cascade. “He was one of the people who brought the techniques of molecular biology to the study of complement deficiencies”, said F Sessions Cole, a professor of paediatrics, cell biology, and physiology at Washington. “He also helped each of us [who worked with him] to see patients not only as people with emotional and medical needs, but also as questions in nature.”

Colten was an aficionado of good coffee, and was fond of asking his collaborators to meet in his extraordinarily well organised office over espresso. “One could determine very quickly whether one was going to have a good meeting with him or not by whether he offered you espresso”, Cole remembers with a laugh. That’s not to say he was undiscriminating. He pushed himself and his co-workers hard, and judged his colleagues on the basis of their ability and contribution. “He believed in egalitarianism based on intellect”, Cole says. “He had an incisive mind”, adds Daphne deMello, a paediatric pathologist who worked with him in the early 1990s. “He got a grip of any problem you presented him and had answers and possible routes for further investigation straight away.”

2 years ago, when he was already seriously ill, Colten travelled to attend the funeral of his former colleague Stanley Korsmeyer, remembers Nathan: “I was surprised to see him there, considering his condition, but he came to salute Stanley’s excellence and skill. He respected excellence.” Colten is survived by his wife of 48 years, Susan; two daughters, Jennifer and Lora; his son, Charles; and six grandchildren.

Stephen Pincock
stephenpincock@gmail.com
Health workers and vaccination coverage in developing countries

Sudhir Anand and Till Bärnighausen (April 14, p 1277) are to be congratulated on their econometric analysis of health-worker density and vaccination coverage. Nevertheless, I am concerned about their conclusion that increased health-worker density, especially that of nurses and midwives, will produce higher vaccination coverage.

I have done a reanalysis of the same data using the numbers of nurses and midwives per 100 000 population as an independent variable and vaccination coverage as a dependent variable (figure). My analysis shows a general increase in vaccination coverage from lower to higher nurse-midwife densities, but also large variations in vaccination coverage between different countries with similar nurse-midwife densities. At a nurse-midwife density of less than 30 per 100 000, no sufficient vaccination coverage seems to be achievable. At a density of 30–100 per 100 000, several countries achieve vaccination coverage close to or greater than 80%. Not all of these countries are small, and female educational levels are low in some of them (eg, Egypt, Ghana, Tanzania). A nurse-midwife density of greater than 100 per 100 000 correlates with good vaccination coverage in most, but not all, instances (eg, Nigeria).

What are the conclusions then? Public-health experts should examine more closely why some countries are so successful in achieving high vaccination coverage despite low numbers of health workers, and why others fail to achieve this goal. In some, political and health-system support might be very strong. In others, people’s attitude towards vaccination might be an important obstacle. Hence there must be more research on the political, economic, educational, health-system, and societal factors that contribute to vaccination coverage. Health workers are extremely important, but one cannot just assume that with increasing health-worker density developing countries will achieve better vaccination coverage.

I declare that I have no conflict of interest.

Carsten Krüger
theca.carsten.krueger@web.de
Department of Paediatrics, St Franziskus Hospital, Robert Koch Strasse 55, D-59227 Ahlen, Germany


One of the conclusions of the article by Sudhir Anand and Till Bärnighausen,1 that a higher density of health workers (nurses) increases the availability of vaccination services over time and space, is very difficult for most public-health decisionmakers and practitioners to understand, especially considering that doctor density was found to be insignificant. Possibly, the following two reasons could help us understand this unusual conclusion.

First, the distribution of health workers at different levels probably affected the results. Only health workers who take charge of vaccination should have been taken into account in the model. The ratio of such health workers to the total number of health workers would inevitably have an effect on the results.

Second, the average national data used by Anand and Bärnighausen in this study could possibly have affected the results in terms of the great gap that exists between urban and rural areas.2–5 The result would be different if the model was built separately based on the data from urban and rural areas. Of course, it is very difficult to collect these data. But the possible bias should be considered when giving policy suggestions.

We declare that we have no conflict of interest.

Guoqing Hu
ghu@jhph.edu
Department of International Health, Johns Hopkins Bloomberg School of Public Health, 624 North Broadway, Baltimore, MD 21205, USA; and Department of Epidemiology and Health Statistics, School of Public Health, Central South University, Changsha, Hunan, China

Figure: Vaccination coverage in relation to nurse-midwife density per 100 000 population

MCV=measles-containing vaccine. DPT3=full course of diphtheria, pertussis, and tetanus vaccine. Polio3=full course of polio vaccine.

Submissions should be made via our electronic submission system at http://ees.elsevier.com/thelancet/
I welcome the interesting analysis by Sudhir Anand and Till Barnighausen, who found that higher health-worker densities contribute significantly to vaccination coverage in developing countries. Here, I offer some additional thoughts. First, that doctor density mattered little could be partly explained by the substitutive contributions of supervised medical trainees and staff from foreign non-governmental organisations during vaccinations. I have experienced this in Nigeria.

Second, the effect of land area on vaccination coverage is dependent on the distribution of health workers. A country with a relatively small land area and average health-worker density could have low vaccination coverage if the health workers were poorly distributed, working in places with less vaccination needs. Third, national income is known to be positively associated with female literacy rates and workforce supply. Therefore, the effects of national income on vaccination coverage would seem insignificant in models already adjusted for female literacy and health-worker densities. Fourth, the strong positive association between female literacy and vaccination coverage suggests that education by fostering awareness, acceptance, and action remains an important means to improve vaccination coverage. The recent polio vaccination drawbacks seen in parts of Nigeria known for their low female literacy levels are a sad reminder.

Finally, I agree with Claudio Lanata's Comment: Health workers are essential for achieving the Millennium Development Goals. However, I disagree with the assertion that we need to "move away from training"; we need to re-engineer health-workforce training with emphasis on retention and sustainability. It is unfortunate that the developed countries that promote these vaccinations are also partly responsible for draining away developing countries' health workforce.

I declare that I have no conflict of interest.

Osyenbuchi A Arah
o.a.arah@amc.uva.nl

Department of Social Medicine, Academic Medical Center, University of Amsterdam, PO Box 22700, Amsterdam 1100 DE, Netherlands


The association shown between health-worker density and vaccine coverage in the analysis by Sudhir Anand and Till Barnighausen is logically sound. Additionally, other locally prevalent factors should be stressed.

In India, the report of the Universal Immunization Program Review of six poorly performing states concluded that the infrastructure to reach every child was at place but that the problem had been the delivery of services. Auxiliary nurse midwives, who provide basic immunisation services, are overburdened with duties such as keeping files and other operational logistics. Surprisingly, in one study, increased availability of supportive manpower, the community health workers who are the main community outreach people, had a less than expected effect on coverage. The review thus recommends training of community health workers, supervision, monitoring, social mobilisation, and appointment of state immunisation task forces to guide and improve intersectoral coordination, as key measures to increase the current coverage. Lower vaccine coverage (non-polio vaccines) in India was also associated with other variables: low maternal literacy (as shown in Anand and Barnighausen's analysis), girl child, and children born in lower caste families.

Second, parallel mass campaigns (eg, polio) can adversely affect routine immunisation coverage, possibly through human resource consumption. Proper planning with a focus on human resources in terms of targetted and efficient delivery of services (eg, targeting less educated mothers through trained community workers) might thus be an important priority.

We declare that we have no conflict of interest.

*Nishith K Singh, Vikas Singh nishith_singh2004@yahoo.co.in

Southern University Illinois School of Medicine, 701 N First Street, Springfield, IL 62794, USA (NKS); and Department of Emergency Medicine, Vardhman Mahaveer Medical College and Safdarjang Hospital, New Delhi, India (VS)

Authors’ reply

Carsten Krüger shows a scatter plot of vaccination coverage against combined nurse and midwife density and observes that, for a given healthworker density, different countries achieve different levels of vaccination coverage. It is precisely the deficiencies of this type of univariate analysis that motivated our article.

In our article, we did multivariate regressions of vaccination coverage and included independent variables other than health-worker density: national income per person, female adult literacy, and land area. We found that health-worker density is a significant determinant of vaccination coverage after controlling for these variables. In the Discussion section, we suggested that an analysis of outliers from the fitted regression plane might help us to understand yet other factors that could affect vaccination coverage; as examples, we mentioned political factors (eg, conflict) and health-system performance factors.

Notwithstanding Guoqing Hu’s first point, we find it very difficult to understand how a higher density of health workers (nurses) can decrease the availability of vaccination services over time and space. Since self-administered vaccination technologies are not used routinely for the vaccinations that we investigated, it is difficult to see how a higher health-worker density would decrease the availability of vaccination services. As we stated in our article, “the level of income has no independent effect after controlling for other determinants.” This implies an important policy conclusion: economic growth is not a necessary condition for increasing childhood vaccination coverage in a country.

We declare that we have no conflict of interest.

*Sudhir Anand, Till Bärnighausen
sudhir.anand@economics.ox.ac.uk

St Catherine’s College, Oxford OX1 3UJ, UK (SA), and Africa Centre for Health and Population Studies, University of KwaZulu-Natal, Mtubatuba, South Africa (TB)


Long-term risks of increased use of intravenous iron

Michael Auerbach and colleagues (May 5, p 1502)1 maintain that, apart from high-molecular-weight iron dextran, “there is no substantially increased risk with the administration of intravenous iron” for anaemia. Their argument is based on two conclusions: (1) that risk of short-term adverse reactions is minimal, and (2) that intravenous iron “improves erythropoietic response”. Lack of immediate adverse reactions and intensification of the erythropoietic response are not adequate justification for increased use of intravenous iron. Auerbach and colleagues’ recommendation ignores possible delayed risks of achieving a maximum erythropoietic response in patients with chronic disease.2 More importantly, it also ignores potential long-term adverse effects of storage iron load on survival.3-4 Perspectives on iron-related risk come from a randomised trial of iron reduction in peripheral arterial disease.5 A decrease in mean serum ferritin from 122.5 to 79.7 μg/L was induced and maintained by phlebotomy for an average of 3.13 years. Iron depletion was avoided. Although there was no overall cardiovascular benefit of iron reduction, there was a significant 54% decrease in total mortality in the youngest quartile at entry (ages 43–61 years) in association with iron reduction. Administration of intravenous iron to maximise erythropoietic response in chronic anaemias often produces far greater stored iron loads than those seen in these patients with peripheral arterial disease before iron reduction therapy.
Correspondence

Much additional research on long-term adverse consequences of maintaining iron in storage is urgently needed; however, lack of definitive data does not mean that there is no risk.\(^2\) The burden of proof for long-term safety rests with advocates of increased use of parenteral iron.

I declare that I have no conflict of interest.

Jerome L Sullivan
jlsullivan3@gmail.com

University of Central Florida, Burnett College of Biomedical Sciences, Orlando, FL 32826, USA

2 Strippoli GF, Tognoni G, Navaneethan SD, Nicolucci A, Craig JC. Haemoglobin targets: were we wrong, time to move on. Lancet 2007; 369: 346–50.

The enjoyable, informative Clinical Update by Michael Auerbach and colleagues\(^1\) highlights the efficacy of intravenous iron products for total or “top-up” repletion in the settings of anaemia of chronic kidney disease (CKD) and anaemia of malignancy. However, it concentrates solely on the immediate period after administration and neglects the potential long-term adverse effects of parenteral iron.

Patients with CKD have an exaggerated risk of cardiovascular death which can partly derive from anaemia, but correction to a normal haemoglobin concentration does not ameliorate, and can exacerbate, this risk.\(^2,3\) Giving boluses of iron salts can aggravate oxidative stress and thereby potentiate existing, or cause de novo, cardiovascular disease.\(^4\) Infection is already a major cause of mortality and morbidity in dialysis (and cancer) patients; certain microorganisms use iron as a metabolic cofactor and can thus become either more frequent, or more virulent, as causes of infection as a result of iron use.\(^3\)

Our available markers to monitor total body iron (stored and usable) are poor;\(^4\) much the more so in CKD (in which ferritin, also an acute-phase reactant, is even less reliable as a marker of iron stores). The “safe” (ie, anaphylaxis-free) short-term repeated administration of intravenous iron to patients on dialysis might in fact come at the expense of increased risk of medium-term to long-term complications. This is a controversial area in nephrology, and we need longer-term trials to examine the overall net benefit to patients of using iron in this way in the context of CKD-related anaemia.

We declare that we have no conflict of interest.

Behdad Afzali, *David J A Goldsmith
david.goldsmith@gstt.nhs.uk

Department of Nephrology and Transplantation, Guy’s Hospital, London SE1 9RT, UK


Authors’ reply

We appreciate Jerome Sullivan’s concerns. However, he incorrectly attributes the harm in recent trials\(^3,5\) to higher haemoglobin concentrations per se rather than to concomitant higher epoetin doses. Singh and colleagues\(^3\) found that the nearly 50% increase in the relative risk of congestive heart failure and death was not related to the rate of rise of haemoglobin. Further, negative outcomes could not have been attributed to iron administration because only 4–2% of the patients in that trial received intravenous iron. What we do know is that, in studies to date, intravenous iron decreased exposure to erythropoiesis stimulatory activity (ESA) for similar haemoglobin increments. As an adjuvant to ESA therapy, the total dose of intravenous iron is not clinically relevant outside of chronic kidney disease.

Clinical trials show no relation between negative outcomes and serum ferritins up to 1200 μg/L, total iron dose, or percentage transferrin saturation.\(^5\) In a study of all-cause mortality in haemodialysis patients,\(^3\) there was no relation to iron variables or intravenous iron dose, whereas hypoalbuminaemia was predictive.

It is worth quoting the conclusions of the randomised trial of iron reduction in peripheral arterial disease (Sullivan’s reference 5): “reduction of body iron stores in patients with symptomatic [peripheral arterial disease] did not significantly decrease all-cause mortality or death, nonfatal myocardial infarction and stroke.” The authors further concluded, “because of lower than expected accrual, the study was underpowered overall and particularly underpowered to definitively assess outcomes in younger patients and smokers.”

In response to Behdad Afzali and David Goldsmith, efficacy trials of intravenous iron designed to show haemoglobin increases independent of ESAs have not seen increased infectious or cardiovascular events. More importantly, in clinical practice, ESAs are decreased as haemoglobin approaches a safe target. This finding is consistent with the study in 58 000 dialysis patients where those given intravenous iron were significantly less likely to die than those not.\(^4\)

Lastly, we concur with the need for long-term trials of iron in anaemia management that recognise iron deficiency and harm from excess ESA exposure as significant confounders. But if we follow Sullivan’s “burden of proof” postulate we will eschew a group of drugs known to be safe while increasing use of a class of drugs that is potentially harmful.
Risk-minimisation strategies for peanut allergy

Recent high-profile Coroner’s cases in Australia have ignited public awareness of the catastrophic consequences of the increasing prevalence of peanut allergy.1,2

The community has leapt to avoid legal liability by establishing “nut-free zones” in schools and kindergartens, often not acknowledging the impossibility of guaranteeing such a venue. The medical community often zealously attempts to ensure patients’ safety by cocooning patients at home together with their EpiPens. Meanwhile the medical research world strives to identify safe treatments.

As allergists with more than 40 years’ experience between us in the care of adolescent and adult patients, we believe that it is timely to remind medical practitioners and the public that additional strategies exist to improve the safety of peanut-allergic patients while allowing them to lead full and active lives. We teach patients in our clinic to avoid peanuts and also guide them to follow risk-minimisation strategies, especially when dining out:

1. They ensure that asthma control is optimum (asthma is the main risk factor for death due to anaphylaxis).3

2. They phone ahead to notify friends or restaurants of their particular allergies (to reduce risks from inadvertent exposure).

3. They carry a written anaphylaxis emergency plan and kit, including adrenaline, a non-sedating anti-histimine, and oral corticosteroids to minimise rebound while seeking appropriate emergency-room follow-up.4

4. Finally, and perhaps most usefully, patients are trained to pause before eating food that they have not prepared themselves, and cautiously touch-test a trace of food on their external lip before putting it into their mouths. Tell-tale warnings of a chilli-like reaction, or tingling, burning, or swelling, alerts them not to proceed.

We stress that patients are not told to take risks with food. They should only eat if they are confident that the food is nut-free. Having made a decision to proceed, touch testing provides a useful brake and has often protected our patients from further inadvertent exposure.

We believe that excellent asthma control and cautious touch-testing should be keystones of worldwide education strategies for preventing deaths from nut anaphylaxis alongside adrenaline autoinjectors. This strategy would decrease the huge impact that these allergies have on the lives of patients, their families and carers, and the community. Meanwhile the search for a safe and effective desensitising vaccine for nut allergy remains a priority.5

We declare that we have no conflict of interest.

*Robyn E O’Hehir, Jo A Douglass r.ohehir@alfred.org.au

Department of Allergy, Immunology and Respiratory Medicine, Alfred Hospital and Monash University, Commercial Road, Melbourne, Victoria 2004, Australia


Measurement of health and disability

In a Viewpoint on the measurement of health and disability (May 12, p 1658),1 Daniel Mont argues that disability-adjusted life-years (DALYS) do not accord with WHO’s International Classification of Functioning (ICF). When substantiating his point, Mont makes two claims that deserve some comment.

First, in his comparison between the ICF classification and DALYS, an essential and frequently overlooked distinction is missing, namely the one between the ICF conceptual framework (based on conceptual developments that extend and update previous theoretical models) and the tool embodying that framework, the classification itself. The criticism of DALYS under the ICF framework is fair, but it should not be based on an inappropriate comparison. Whereas the ICF system serves the standardised classification of people on the basis of a detailed amount of individualised information, DALYS are essentially a population health measure, and therefore nobody should be surprised if they do not seem to be sensitive to individual factors.

Second, the general statement that “with the ICF approach, the
medical diagnosis does not matter”, should be put into proper context. Although it is true that no explicit link has been established between codes of the International Classification of Diseases1 and of the ICF, one of the ICF components specifically deals with the integrity of anatomical structures. Furthermore, disease-specific ICF core sets are actually being developed for a range of disorders,2,3 and a recently proposed definition of disability explicitly acknowledged the involvement of a health condition under this framework.4 Joint use of both WHO classifications systems is needed for enhanced information on health and disability.

I declare that I have no conflict of interest.

Jose M Valderas
jose.valderas@manchester.ac.uk
National Primary Care Research and Development Centre, University of Manchester, Williamson Building, Oxford Road, Manchester M13 9PL, UK


Author’s reply

As Jose Valderas states, the ICF—as a classification system—is different from a population health measure such as the DALYS. Still, the fact that the basic conceptualisations of disability in the two tools are very much at odds with one another is problematic.

I agree that health conditions are generally involved in disability. However, a given medical diagnosis can be associated with a wide variety of functional limitations in terms of both their type and extent. This point is not accounted for by DALYS. It must be acknowledged, as the ICF core sets will no doubt show, that a cluster of limitations of body function, activity, and participation are associated with particular diagnoses. However, we must also keep in mind that a particular limitation is also often associated with a wide range of diagnoses.

Many health interventions will naturally be aimed at the prevention of particular diseases, as they should. But habilitation and rehabilitation interventions must be aimed at addressing the various barriers that interact with functional limitations—whatever their cause—to create disability.

I declare that I have no conflict of interest.

Daniel Mont
dmont@worldbank.org
Disability and Development Team, World Bank, 1818 H St NW, Washington, DC 20433, USA

Healing the wounds of Tiananmen: Beijing Olympics and beyond

Thanks to their government’s efforts to erase history, many Chinese people, especially the younger generation, barely have any idea of what happened on the night of June 3–4, 1989, when the People’s Liberation Army entered the city of Beijing to end a student-led democracy demonstration. Live ammunition was fired against unarmed civilians, and tanks rolled over the bodies of the victims, causing hundreds of deaths and thousands of injuries. In the aftermath of the massacre, many student leaders and their supporters were arrested, imprisoned, and some subjected to torture.

China’s rapid economic growth and increasing influence in international politics make it difficult for leaders of Western countries to criticise Beijing’s human rights practices. For these reasons, concerned parties such as The Lancet1 and the Independent Federation of Chinese Students and Scholars in the US feel more obliged than ever to keep a watchful eye on China. For those who lost their loved ones and who were injured on June 4, memories of the violence have not faded away and they continue to live in injustice. We hence firmly stand by our position that any praise of China’s achievements in human rights is built on a shaky ground unless Beijing takes serious measures towards putting an end to the enduring Tiananmen tragedy.

As Beijing makes its final dash towards Olympic glory in 2008, we wish to appeal to the Chinese government to use the opportunity not only to showcase the country’s economic miracles, but also to demonstrate its commitment to human rights. We urge President Hu Jintao and Premier Wen Jiaobao to take steps to honour the following requests made previously by victims’ families and other concerned groups:

(1) Appoint an independent committee and conduct a fair investigation into the tragedy, including a full account of the dead and the injured; make the findings of the investigation available to the public.

(2) Offer compensation to the families of the dead and the injured.

(3) Immediately release those who are still in jail because of their participation in the 1989 protests or because of their support for the movement during subsequent years.

(4) Allow the domestic media to freely report on the 1989 student movement and the tragedy that followed.

The Olympic torch will be lit 1 year from now, and China is working hard to ensure that, by that time, the world will see the image of a modern and civilised nation. That image, however, will be tarnished if Beijing fails to resolve this long-running human rights issue.

I declare that I have no conflict of interest.

Jian Zhou
ifcss@ifcss-us.org
Independent Federation of Chinese Students and Scholars in the US, c/o Wei Jingsheng Foundation, PO Box 35449, Washington, DC 20003, USA

MRI for diagnosis of pure ductal carcinoma in situ: a prospective observational study

Christiane K Kuhl, Simone Schrading, Heribert B Bieling, Eva Wardelmann, Claudia C Leutner, Roy Koenig, Walther Kuhn, Hans H Schild

Summary

Background Diagnosing breast cancer in its intraductal stage might be helpful to prevent the development of invasive cancer. Our aim was to investigate the sensitivity with which ductal carcinoma in situ (DCIS) is diagnosed by mammography and by breast MRI.

Methods During a 5-year period, 7319 women who were referred to an academic national breast centre received MRI in addition to mammography for diagnostic assessment and screening. Mammograms and breast MRI studies were assessed independently by different radiologists. We investigated the sensitivity of each method of detection and compared the biological profiles of mammography-diagnosed DCIS versus DCIS detected by MRI alone. We also compared the risk profiles of women with mammography-detected DCIS with those of MRI-detected DCIS.

Findings 193 women received a final surgical pathology diagnosis of pure DCIS. Of those, 167 had undergone both imaging tests preoperatively. 93 (56%) of these cases were diagnosed by mammography and 153 (92%) by MRI (p<0·0001). Of the 89 high-grade DCIS, 43 (48%) were missed by mammography, but diagnosed by MRI alone; all 43 cases missed by mammography were detected by MRI. By contrast, MRI detected 87 (98%) of these lesions; the two cases missed by MRI were detected by mammography. Age, menopausal status, personal or family history of breast cancer or of benign breast disease, and breast density of women with MRI-only diagnosed DCIS did not differ significantly from those of women with mammography-diagnosed DCIS.

Interpretation MRI could help improve the ability to diagnose DCIS, especially DCIS with high nuclear grade.

Introduction Although commonly held to be a direct precursor of invasive breast cancer, ductal carcinoma in situ (DCIS) is a heterogeneous disease.1 High-grade DCIS seems to progress to invasive breast cancer more often and more rapidly than does low-grade DCIS.2–7 Molecular markers and genetic signatures indicate that high-grade lesions and low-grade lesions develop through distinct pathways, rather than by progressive dedifferentiation.8–10 Whereas high-grade DCIS is likely to progress to high-grade invasive cancer, low-grade DCIS can be more indolent or might progress to only certain types of cancer, usually well differentiated. There is agreement in that DCIS should be treated (at least) by local excision to avoid recurrence or progression to invasive breast cancer.11–13 Diagnosis of DCIS, especially high-grade DCIS, is therefore considered desirable in principle.

Before the advent of mammographic screening, DCIS used to be a rare diagnosis, making up only 2% of cancers treated in 1980.14 Today, about 20% of breast cancers are diagnosed in the pre-invasive stage.15–18 Mammography is the mainstay for diagnosing DCIS, whereas other breast imaging techniques—eg, sonography, MRI, scintimammography, and positron emission tomography—have been previously shown to be unreliable.19 Specifically, a number of studies investigated the diagnostic yield of breast MRI in patients who, on the basis of the demonstration of mammographic microcalcifications, had been diagnosed with DCIS. The results of these studies were concordant in that MRI did not visualise all intraductal cancers that were mammographically apparent and thus MRI was deemed to be less sensitive than mammography for the diagnosis of pure intraductal cancer.20–23 Accordingly, the use of breast MRI for diagnosing DCIS is mostly discouraged.24–26 In recent years, however, it has become evident that the diagnosis of intraductal cancer is feasible with MRI, although it requires diagnostic criteria that are different from those that are used to diagnose invasive cancer.25–27 Since these diagnostic criteria were published, our group and others have found that MRI does allow the prospective diagnosis of DCIS, and even allows the diagnosis of DCIS that could go undetected by mammography.26–32 Whether breast MRI can be used to diagnose DCIS prospectively if its use is not restricted to patients with mammographic abnormalities (especially microcalcifications) is unknown. Our aim was to compare the respective sensitivities of mammography and breast MRI, and to compare the biological profiles of DCIS detected by mammography with those detected by breast MRI.

Methods

Patients Data were collected at the breast centre at the University of Bonn Hospital and Medical School, an academic tertiary care institution. The breast centre consists of the Departments of Gynaecological Oncology, Radiology, Pathology, and Radiation Oncology. The centre houses a breast cancer screening site and, as such, offers...
Articles

diagnostic assessment for its own screening participants and for patients with questionable or suspicious clinical or imaging findings identified elsewhere. The centre also offers intensified surveillance for women at increased risk for breast cancer. This group consists of women with a family history of breast cancer and a calculated lifetime risk of 20% or more, as based on geneticist’s assessment, and women in follow-up after breast conserving treatment. Intensified screening is also offered to concerned women at average risk who, after careful explanation of the implications of an average risk and of possible disadvantages of screening, opt to undergo additional screening tests. Moreover, it is standard of care to offer breast MRI as part of the work-up of women with non-negative screening mammogram and for screening the contralateral breast in women with biopsy-proven unilateral breast cancer. Screening, surveillance, diagnostic work-up, and treatment of patients with breast cancer are done in close accordance with European guidelines; all therapeutic decisions are established in multidisciplinary consensus.

The study protocol was reviewed and approved by the authors’ institutional review board. All study participants provided written or oral informed consent. During a 5-year period between Jan 2, 2002, and Dec 31, 2006, breast MRI was done in women who had (1) a non-normal screening mammogram, (2) normal conventional imaging studies, but clinical symptoms of breast cancer, (3) normal conventional imaging studies, but at an increased risk for (primary or recurrent) breast cancer, or (4) normal conventional imaging and an average risk, but were concerned about breast cancer and wished to undergo MRI as an additional screening test.

The diagnoses of the respective imaging methods were prospectively collected and recorded. The breast centre’s pathology database was used to identify all women who received the final surgical diagnosis of DCIS during the study period—ie, independent of their detectability on imaging studies. Only women with pure DCIS (ie, without associated invasive breast cancer or microinvasion) were considered in this study.

Procedures

Imaging studies (mammography and breast MRI) were both completed and interpreted before any biopsy procedure (core, vacuum, or surgical) to allow an unbiased interpretation.

Two-view mammography was done in accordance with national quality assurance guidelines. Between January, 2002, and December, 2004, film-screen mammography was used (Siemens Mam모mat, Siemens Medical Solutions, Erlangen, Germany); thereafter, full-field digital mammography was applied (Lorad Selenia, Hologic Women’s Health). Additional views were obtained at the discretion of the attending breast radiologist. In all patients with questionable or suspicious calcifications, additional spot compression and magnification views were obtained.

Breast MRI was done according to a standardised published protocol. In short, a 1.5T system (Intera and Intera Achieva, Philips Medical Systems, Best, Netherlands) equipped with a dedicated bilateral multi-element breast surface coil (four-channel Breast Array Coil, In Vivo and Philips Medical Systems, Best, Netherlands) was used to image both breasts in the axial orientation. Gentle fixation of the breasts in the slice encoding direction (cranio-caudal direction) was applied to immobilise the breast and to reduce the number of sections needed to cover the fibroglandular tissue. The imaging protocol consists of a T2-weighted axial turbo spin echo pulse sequence (repetition time 3000 ms, echo time 110 ms) without fat suppression, followed by the dynamic contrast enhanced series. This series is a two-dimensional gradient echo pulse sequence (repetition time 290 ms, echo time 4-6 ms, flip angle 90°) that consisted of 31 sections, 3 mm thick, obtained before and four times immediately after bolus injection of 0.1 mmol/kg bodyweight gadopentetate dimeglumine (Magnevist, BayerSchering Healthcare, Berlin, Germany) and 20 mL saline solution by a power injector set to an infusion rate of 3 mL/s. Both breasts were covered with an in-plane spatial resolution of 0.6x0.6 mm and a temporal resolution of 110 s per dynamic acquisition. No fat suppression was used. Precontrast images of the dynamic series were subtracted from the postcontrast images to selectively highlight enhancing structures. No parallel imaging was applied. The entire MRI protocol took 14 min to complete.

Mammography and MRI scans were read and scored independently by different radiologists trained in breast imaging. Four breast radiologists were involved, each of whom had 10–15 years of experience in interpreting both mammograms and breast MRI studies.

Table 1: Reasons for referral to breast MRI

<table>
<thead>
<tr>
<th>Reason for MRI</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal mammogram*</td>
<td>93 (56%)</td>
</tr>
<tr>
<td>Normal mammogram</td>
<td>74 (44%)</td>
</tr>
<tr>
<td>MRI for screening concerned women at average risk</td>
<td>29 (17%)</td>
</tr>
<tr>
<td>MRI for follow up after breast cancer†</td>
<td>18 (11%)</td>
</tr>
<tr>
<td>Incidental finding at MRI‡</td>
<td>14 (8%)</td>
</tr>
<tr>
<td>MRI for screening women at increased familial risk</td>
<td>8 (5%)</td>
</tr>
<tr>
<td>Patients with normal mammogram but clinical symptoms</td>
<td>5 (3%)</td>
</tr>
<tr>
<td>Total</td>
<td>167 (100%)</td>
</tr>
</tbody>
</table>

*MRI studies were read blinded to the mammogram. †Does not include patients with a recent diagnosis of cancer who underwent MRI for staging to rule out residual disease, but does include patients with a recent diagnosis of mammographically unilateral breast cancer who presented for screening of the opposite breast. ‡Incidental findings: patients who were referred for work up of a mammographic abnormality unrelated to the DCIS.
As described previously all MRI scans were first read without mammographic films to avoid biased readings.\(^a\) The diagnoses of these first readings were used for further analysis in this study. A synoptic reading of mammograms and MRI scans was done only in a second step to develop clinical management recommendations.

All mammograms were independently double read in accordance with EU guidelines. The final diagnoses were coded and recorded. Coding was done according to the American College of Radiology Breast Imaging and Data System (BI-RADS), which ranges from 1 (negative) to 5 (almost certainly malignant). A BI-RADS score of 0 (assessment incomplete) was not accepted as final diagnosis. Breast density was scored according to the four-point ACR scale, which ranges from ACR1 (almost entirely fat) to ACR4 (extremely dense).

MRI findings were described and diagnoses coded in accordance with the magnetic resonance BI-RADS lexicon;\(^b\) the final assessment categories were the same as those used for mammography. Interpretation of MRI scans was based on morphology and enhancement kinetics of contrast enhancing areas. If early asymmetric (unilateral) non-mass-like enhancement with segmental or ductal distribution and granular internal enhancement was noted, this was deemed to be a BI-RADS score of 5. If enhancement was very subtle until the late post-contrast phase, or if the distribution did not unambiguously follow the milk ducts, or if there was bilateral (but not symmetric) enhancement, this was deemed to be a BI-RADS score of 4.

If at least one of the imaging methods was positive, the patient underwent biopsy. Biopsy was also obtained if neither of the imaging studies was positive, but clinical findings (bloody nipple discharge, palpable lumps, nipple retraction, Paget’s disease of the nipple) suggested a breast abnormality. Histological proof of lesions that were primarily identified by breast MRI was obtained by MRI-guided wire placement with excisional biopsy, or by MRI-guided vacuum core biopsy (Vacora, Bard Medical, [Karlsruhe, Germany] or ATEC Suros Surgical Systems [Indianapolis, IN, USA]).

Vacuum core biopsies and excisional biopsies were fixed overnight in 4% formalin. Core biopsies were embedded in paraffin and were cut in to at least six (4 µm thick) serial sections per block. DCIS was categorised as being low, intermediate (non-high grade with necroses), or high grade. Where the differential diagnosis was between ductal hyperplasia and low-grade DCIS, additional immunohistochemical stainings (cytokeratin 5/6 and p63) were done as described elsewhere\(^c\) to determine the presence or absence of myoepithelia. Excisional biopsies were stained with four different colours to allow a three-dimensional orientation and correlation with the radiological findings. All tissues had been marked with at least two sutures by the surgeon to facilitate spatial orientation of the specimen.

All specimens were completely embedded. DCIS size was assessed by counting how many slides included intraductal proliferations and by measuring the maximum diameter of the lesion on the mounted sections. All margins were measured and reported with relation to the different stains on the surface of the specimen. Oestrogen and progesterone receptor status was assessed immunohistochemically and was reported according to the Remmele score.\(^6\) The Van-Nuys prognostic index (VNPI) was determined on the basis of DCIS size, margin width, and nuclear grading.\(^1\)

---

**Table 2: Sensitivity of mammography and MRI for DCIS detection**

<table>
<thead>
<tr>
<th>Women with mammography-positive DCIS (n=93)</th>
<th>Women with mammography-negative DCIS (n=74)</th>
<th>p value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td>0.7</td>
</tr>
<tr>
<td>Median (range) 55 (31-84)</td>
<td>Median (range) 55 (32-76)</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) 54.3 (11.0)</td>
<td>Mean (SD) 53.7 (10.4)</td>
<td></td>
</tr>
<tr>
<td>Menopausal status</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Pre-menopausal 33 (35%)</td>
<td>Pre-menopausal 26 (35%)</td>
<td></td>
</tr>
<tr>
<td>Post-menopausal 60 (65%)</td>
<td>Post-menopausal 48 (65%)</td>
<td></td>
</tr>
<tr>
<td>Breast density</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>ACR1 21 (23%)</td>
<td>ACR1 11 (15%)</td>
<td></td>
</tr>
<tr>
<td>ACR2 31 (33%)</td>
<td>ACR2 32 (43%)</td>
<td></td>
</tr>
<tr>
<td>ACR3 31 (33%)</td>
<td>ACR3 22 (30%)</td>
<td></td>
</tr>
<tr>
<td>ACR4 10 (11%)</td>
<td>ACR4 9 (12%)</td>
<td></td>
</tr>
<tr>
<td>Mean ACR density 2.3</td>
<td>Mean ACR density 2.3</td>
<td></td>
</tr>
<tr>
<td>Family history</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>None 70 (75%)</td>
<td>None 50 (68%)</td>
<td></td>
</tr>
<tr>
<td>One case 11 (12%)</td>
<td>One case 7 (10%)</td>
<td></td>
</tr>
<tr>
<td>Moderate risk 8 (9%)</td>
<td>Moderate risk 11 (15%)</td>
<td></td>
</tr>
<tr>
<td>High risk 4 (4%)</td>
<td>High risk 6 (8%)</td>
<td></td>
</tr>
<tr>
<td>Personal history of breast cancer</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>No 68 (73%)</td>
<td>No 53 (72%)</td>
<td></td>
</tr>
<tr>
<td>Yes 25 (27%)</td>
<td>Yes 21 (28%)</td>
<td></td>
</tr>
<tr>
<td>History of benign breast biopsy</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>No 77 (83%)</td>
<td>No 57 (77%)</td>
<td></td>
</tr>
<tr>
<td>Yes 16 (17%)</td>
<td>Yes 17 (23%)</td>
<td></td>
</tr>
<tr>
<td>p value* Calculated by the Wilcoxon two sample test and the two sample Student t test.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3: Demographic and risk profiles of women with mammography-positive DCIS versus those with mammography-negative DCIS**

---

\(^a\) Calculated by the McNemar’s test. **Non-high-grade without necroses. †Non-high-grade with necroses.**
Statistical analysis

The imaging diagnoses were dichotomised as follows: diagnoses coded as BI-RADS 1, 2, and 3 were deemed to be negative; BI-RADS categories 4 and 5 were deemed to be positive.

The demographic and risk profiles of the women with mammography-positive DCIS were compared with those of the women with mammography-negative (MRI-positive) DCIS. Furthermore, the mammographic detectability of DCIS was correlated with the respective breast density (ACR scores). The biological profiles (size, nuclear grading, presence or absence of necroses, VNPI, oestrogen and progesterone receptor status) of mammography-positive versus mammography-negative DCIS were compared. The Wilcoxon matched pairs signed ranks and the Student t test for paired samples were used as appropriate to check for statistical significance, where p<0.05 was considered significant.

The positive predictive value of mammography and of MRI was determined as follows. The number of women who received an imaging diagnosis of (possible) cancer—ie, a diagnosis category BI-RADS 4 or 5—was recorded. The respective final histological diagnosis was used as the standard of reference to determine the number of true positive diagnoses (ie, an imaging diagnosis of BI-RADS 4 or 5 with a final pathology diagnosis of invasive or intraductal cancer) and the number of false positive diagnoses (ie, an imaging diagnosis of BI-RADS 4 or 5 with a final pathology diagnosis of non-cancer). In women with malignant lesions, the final surgical pathology result was used. The positive predictive value was calculated as TP/(TP+FP), where TP is the number of true-positive diagnoses and FP is the number of false-positive diagnoses.

Statistical analyses were done with SPSS version 12.0.

Role of the funding source

The study sponsors approved the proposals regarding study design and data analysis, but had no role in data collection, analysis, interpretation, or drafting or reviewing the manuscript. The corresponding author had full access to all the data in the study, and had final responsibility for the decisions to submit for publication.

Results

Of the 7319 women who had undergone both mammography and MRI during the study period, 1208 (15%) women received a positive imaging diagnosis (BI-RADS 4 or 5). 574 of these women received a benign pathological result; 469 were finally diagnosed with invasive breast cancer, and 165 women had an imaging diagnosis of BI-RADS 4 or 5 and received the final pathological diagnosis of pure DCIS. Two further cases of DCIS were diagnosed on the basis of clinical symptoms only.

During the study period, 26 cases of DCIS were diagnosed in women who underwent mammography, but not MRI before biopsy. Reasons for not being given breast MRI included contraindications to MRI (n=4); patients’ preference (7), and other reasons (5). In ten women, MRI was done only after mammographic vacuum core biopsy or only after surgery. None of the patients had received an MRI without mammography. These women were thus excluded from our study. The final study cohort consisted of the 167 patients who had undergone both mammography and MRI before biopsy, and who received the final surgical pathology diagnosis of pure DCIS.

The mean age of the 167 women with DCIS was 54.1 (SD 10.7) years (range 31–84 years). 97 (58%) had an average risk for breast cancer and had been referred for regular screening; 40% (%) were in follow-up after breast cancer, 14 (8%) underwent screening for familial breast cancer, and 12 (7%) had clinical symptoms (nipple discharge in six, palpable lump in three, nipple retraction in two, and Paget’s disease in one). Indications for breast MRI are shown in table 1.

Table 2 shows the sensitivity of mammography compared with that of MRI for detection of DCIS. MRI...
detected significantly more cases of any grade of DCIS than did mammography (p<0.0001). Mammography was falsely negative in 74 (44%) cases, whereas MRI was falsely negative in 14 (8%) cases (p<0.0001). In 81 (49%) of the cases, the DCIS was seen with both imaging methods. In two (1%) patients, both with low-grade DCIS, both methods were negative, and the diagnosis was established because of clinical findings (Paget’s disease of the nipple or nipple discharge). In 12 (7%) patients, only mammography was positive. In 72 (43%) patients, only MRI was positive.

Patients with mammography-positive DCIS did not differ from patients with mammography-negative DCIS in terms of their age, menopausal status, personal history of breast cancer in the same or the opposite breast, history of benign breast disease, or familial risk (table 3). The distribution of mammographic breast density patterns was also much the same in the two groups (table 3).

Across all three grading classes (low, intermediate, and high grade), MRI detected more cases of DCIS than did mammography (figure 1), although this was significant only for intermediate and high-grade DCIS (table 2). For low-grade DCIS, each method depicted a sizeable number of low-grade DCIS that went undetected by the other imaging method (figure 1). Two of the high-grade DCIS were detected by mammography only; 43 were detected by MRI only.

The sensitivity of mammography decreased with higher nuclear grade, whereas that of MRI increased (table 2). The sensitivity of mammography for detecting high-grade DCIS was significantly higher in high-grade DCIS with necroses than in high-grade DCIS without necroses (p=0.012; figure 2); the sensitivity of MRI did not depend on presence or absence of necroses (table 2).

The biological profiles of the DCIS, broken down by the method of detection, are shown in table 4. Of the 12 intraductal cancers that were only depicted by mammography, 10 (83%) were non-high-grade. Of the 72 cases of DCIS that were only diagnosed by MRI, 43 (60%) were high grade. The mean size of DCIS that were diagnosed by mammography did not differ significantly from the size of those that were detected by MRI alone (p=0.56; table 4). The mean size of the 12 DCIS which were detected only by mammography was significantly smaller than those that were detected by mammography and MRI or by MRI alone (p=0.56; table 4).

During the entire 5-year study period, 574 women had a positive imaging diagnosis (BI-RADS 4 or 5), but received a benign pathological diagnosis. Accordingly, the corresponding imaging diagnoses are to be considered false-positive. 634 women received the final pathological diagnosis of cancer (DCIS or invasive), and the corresponding imaging diagnoses can be considered true-positive. This translates into an overall positive predictive value of 47%. Stratified by imaging method, the data are as follows: with mammography, 421 true-positive diag-

noses were established and 344 false positive (benign at histology), giving a positive predictive value of 55%. With MRI, 599 true-positive diagnoses were established and 413 false positive, giving a positive predictive value of 59%.

Discussion

Our study suggests that the sensitivity of film screen or digital mammography for diagnosing DCIS is limited. Of the 167 intraductal cancers that had been diagnosed during the study period, 72 (43%) were mammographically occult, but were diagnosed by MRI alone.

What are the diagnostic and prognostic implications of these mammographically occult lesions? Mammography tends to identify breast cancers with comparatively benign biological profiles, a fact referred to as length time bias. Some of the cases of DCIS diagnosed by mammography are biologically benign and would never threaten a woman’s life, a situation known as overdiagnosis bias. Indeed, in our cohort, almost a third of all mammography-detected DCIS were low grade, as were almost all the intraductal cancers which were diagnosed by mammography alone (table 4). Although more low-grade DCIS will be diagnosed if MRI is used in
addition to mammography and thus more women could be overdiagnosed with a possibly prognostically irrelevant disease, one should note that 60% of the cases of DCIS diagnosed by MRI alone were high grade. Notwithstanding the many uncertainties regarding the natural history of DCIS, there is broad agreement that the diagnosis of high-grade DCIS is prognostically relevant and should help prevent the development of invasive cancer. Therefore, there is reason to assume that MRI helps anticipate the diagnosis of lesions that, if left undetected, would progress to (high grade) invasive cancer. Therefore, there is reason to assume that MRI helps anticipate the diagnosis of lesions that, if left undetected, would progress to (high grade) invasive breast cancer.

On pathophysiological grounds, that breast MRI is more likely to detect high-grade DCIS than it is to detect low-grade DCIS is plausible. The diagnosis of breast diseases in MRI is based on tissue contrast material enhancement. Contrast enhancement in breast MRI depends on a locally increased microvessel density or capillary permeability. Several studies have shown a significantly higher vessel density in high-grade DCIS than in those with non-high-grade DCIS, with or without necroses, and the enhancement pattern in breast MRI has been shown to correlate with the biological profile of DCIS.

Is it plausible that the mammographic sensitivity for DCIS should indeed be so low? There is evidence to suggest that most invasive cancers evolve through the intraductal stage. If this is the case, and if mammography was very sensitive for DCIS, then one should expect that, in women who undergo systematic yearly mammographic screening, most breast cancers would be diagnosed in the intraductal stage. However, only around 20% of screen-detected cancers are DCIS, implying that either a substantial number of cases of DCIS are mammographically occult and missed by mammographic screening, or that the time interval of the intraductal stage is too short to be captured. In our cohort, the cases of DCIS missed by mammography were of the same size as those that were diagnosed by mammography; the smallest were the ones that were detected by mammography only. This finding suggests that MRI detects a different subset of DCIS than does mammography—those cases that do not exhibit calcifications and, therefore, will remain mammographically occult. Thus, MRI does not simply detect cases of DCIS at an earlier stage than mammography—ie, cases that, if left alone, would later be detectable by mammography as well. Missing DCIS by mammographic screening seems to be a matter of mammographic detectability per se (ie, whether or not a DCIS develops calcifications or not) rather than a matter of timing (ie, screening intervals).

Who were the patients with DCIS that was diagnosed by MRI alone? The cohort of women with mammography-detected DCIS had the same age, menopausal status, and breast density, and were as likely to have a history of breast cancer or of benign breast biopsy as patients whose DCIS was not detected by mammography. Thus, we could not identify factors that would predict the likelihood with which the mammographic diagnosis of DCIS fails. This finding suggests that DCIS could be missed by mammography—not only in women with dense breast tissue or in women who are at increased risk for breast cancer, but also in women with non-dense breasts and in women at average risk.

By contrast with our findings, previous studies on the diagnostic yield of breast MRI in the setting of mammographic calcifications found a consistently lower sensitivity of MRI compared with mammography. However, these studies had substantial selection bias since they included only women with mammographic abnormalities, with the published sensitivity rates of MRI for mammography-detected DCIS. One should note that we did not do follow-up imaging studies for the 6111 women with normal clinical breast examination, normal mammogram, and normal MRI. Some of these women could conceivably have had DCIS that remained undetected clinically and by imaging studies. However,

<table>
<thead>
<tr>
<th></th>
<th>Only mammography positive (MRI false negative)</th>
<th>Mammography positive (with or without positive MRI)</th>
<th>Only MRI positive (mammography false negative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>12 (100%)</td>
<td>93 (100%)</td>
<td>72 (100%)</td>
</tr>
<tr>
<td>Grading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low grade</td>
<td>7 (58%)</td>
<td>27 (29%)</td>
<td>15 (21%)</td>
</tr>
<tr>
<td>All non-high grade</td>
<td>10 (83%)</td>
<td>47 (51%)</td>
<td>29 (40%)</td>
</tr>
<tr>
<td>High grade</td>
<td>2 (17%)</td>
<td>46 (50%)</td>
<td>43 (60%)</td>
</tr>
<tr>
<td>Necroses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>7 (58%)</td>
<td>54 (58%)</td>
<td>29 (40%)</td>
</tr>
<tr>
<td>Absent</td>
<td>5 (42%)</td>
<td>39 (42%)</td>
<td>43 (60%)</td>
</tr>
<tr>
<td>Oestrogen-receptor status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not available</td>
<td>2 (17%)</td>
<td>10 (11%)</td>
<td>7 (10%)</td>
</tr>
<tr>
<td>Positive</td>
<td>10 (83%)</td>
<td>66 (71%)</td>
<td>48 (67%)</td>
</tr>
<tr>
<td>Negative</td>
<td>0 (0%)</td>
<td>17 (18%)</td>
<td>17 (24%)</td>
</tr>
<tr>
<td>Progesterone-receptor status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not available</td>
<td>2 (17%)</td>
<td>10 (11%)</td>
<td>7 (10%)</td>
</tr>
<tr>
<td>Positive</td>
<td>9 (75%)</td>
<td>65 (70%)</td>
<td>44 (61%)</td>
</tr>
<tr>
<td>Negative</td>
<td>1 (8%)</td>
<td>18 (19%)</td>
<td>21 (29%)</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>4-50</td>
<td>4-80</td>
<td>4-70</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>15.2 (13.4)</td>
<td>28.4 (17.4)</td>
<td>26.4 (16.1)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>9.5 (7.3-18.5)</td>
<td>23.0 (15.0-40.0)</td>
<td>23.5 (14.0-35.0)</td>
</tr>
<tr>
<td>VNP1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>3-8</td>
<td>3-9</td>
<td>3-9</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>5.2 (1.8)</td>
<td>6.4 (1.7)</td>
<td>5.9 (1.4)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>5.5 (3.0-6.3)</td>
<td>6.5 (0-8.0)</td>
<td>6.5 (0-7.0)</td>
</tr>
</tbody>
</table>

* Determined by scoring DCIS size, nuclear grade, presence or absence of necroses, and margin width, with VNP1 values ranging between 3 and 9.
such triple negative cases of DCIS would then affect the diagnostic accuracy of MRI and mammography alike. Finally, although we did our best to ensure an independent, unbiased reading of the MRI scans, it is conceivable that the fact the readers were aware that women frequently undergo MRI for further assessment of mammographic abnormalities could have biased the interpretation of MRI studies.

Can we draw any conclusions with regard the appropriate use of MRI for diagnosis of DCIS? Recommendations regarding the use of MRI for screening for DCIS are not appropriate to date. Our results are not representative of the regular mass screening setting, and have to be interpreted with care. We report on observations from one breast centre that offers a high level of expertise for interpreting both mammography and breast MRI scans, a fact that is reflected by the high positive predictive values for both methods during the 5-year study period. Since breast MRI is currently used only rarely in clinical practice, for the time being, few radiologists can offer a level of expertise for MRI that comes close to that required for diagnostic mammography—eg, by the US Mammography Quality Standard Act. Additionally, there are currently no standards that would define appropriate technical requirements for breast MRI. Therefore, our results are unlikely to be reproducible in a community breast imaging service at present. However, the number of breast centres that offer MRI is increasing. The ACRIN 6667 multi-institutional trial for screening the contralateral breast included many different institutions that operated on different settings (academic, hospital, private practice). The results of this trial indicated that the performance of breast MRI was fairly stable across the different levels of practice. With the publication of the American Cancer Society Guidelines, according to which MRI is proposed as an adjunct to mammography for screening women with a lifetime risk of 20% or more, more radiologists should be able to do and read more breast MRI scans and, by doing so, gain and maintain a sufficient level of expertise. Moreover, the American College of Radiology and the European Society of Breast Imaging are about to issue an accreditation programme and technical standards for breast MRI. There is thus reason to assume that the field of breast imaging is moving towards a broader use of MRI in the different levels of clinical practice.

A systematic, multi-institutional screening trial will be necessary to further investigate the clinical role of MRI for diagnosing DCIS, and randomised trials are required to investigate its effect (or lack thereof) on recurrence rates or mortality. Our data justify further investigations on the role of MRI for diagnosing DCIS, even if, in view of the low absolute incidence of intraductal cancer, doing such a prospective trial would clearly be associated with a substantial financial investment.

Do our findings have any implications on our knowledge of DCIS as a whole? If further studies confirm that only half of all cases of high-grade DCIS are diagnosed by mammography, these data would suggest that our knowledge on the natural behaviour of high-grade DCIS is biased towards the mammography-visible subgroup. This finding seems important, since histological criteria of DCIS—eg, presence or absence of necroses—are increasingly used to predict the outcome of patients with DCIS, and, thus, to guide management decisions. Current concepts with regard to the prognostic importance of necroses could conceivably be different if the non-calciﬁng, mammography occult, MRI-only DCIS are also taken into consideration.

Contributors
All authors participated in the data collection and reporting stage of this manuscript, and have seen and approved the final version of the manuscript.

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

Acknowledgments
This study was supported by Förderverein für Radiologie an der Universität Bonn (grant 01-12), the Ricomagus Moeschke Donation, and the BMBF (grant 01 EZ 0012).

References


Warfarin versus aspirin for stroke prevention in an elderly community population with atrial fibrillation (the Birmingham Atrial Fibrillation Treatment of the Aged Study, BAFTA): a randomised controlled trial

Jonathan Mant, FD Richard Hobbs, Kate Fletcher, Andrea Roalfe, David Fitzmaurice, Gregory YH Lip, Ellen Murray, on behalf of the BAFTA investigators and the Midland Research Practices Network (MidReC)

Summary

Background Anticoagulants are more effective than antiplatelet agents at reducing stroke risk in patients with atrial fibrillation, but whether this benefit outweighs the increased risk of bleeding in elderly patients is unknown. We assessed whether warfarin reduced risk of major stroke, arterial embolism, or other intracranial haemorrhage compared with aspirin in elderly patients.

Methods 973 patients aged 75 years or over (mean age 81·5 years, SD 4·2) with atrial fibrillation were recruited from primary care and randomly assigned to warfarin (target international normalised ratio 2–3) or aspirin (75 mg per day). Follow-up was for a mean of 2·7 years (SD 1·2). The primary endpoint was fatal or disabling stroke (ischaemic or haemorrhagic), intracranial haemorrhage, or clinically significant arterial embolism. Analysis was by intention to treat. This study is registered as an International Standard Randomised Controlled Trial, number ISRCTN89345269.

Findings There were 24 primary events (21 strokes, two other intracranial haemorrhages, and one systemic embolus) in people assigned to warfarin and 48 primary events (44 strokes, one other intracranial haemorrhage, and three systemic emboli) in people assigned to aspirin (yearly risk 1·8% vs 3·8%, relative risk 0·48, 95% CI 0·28–0·80, p=0·003; absolute yearly risk reduction 2%, 95% CI 0·7–3·2). Yearly risk of extracranial haemorrhage was 1·4% (warfarin) versus 1·6% (aspirin) (relative risk 0·87, 0·43–1·73; absolute risk reduction 0·2%, –0·7 to 1·2).

Interpretation These data support the use of anticoagulation therapy for people aged over 75 who have atrial fibrillation, unless there are contraindications or the patient decides that the benefits are not worth the inconvenience.

Introduction

12% of people aged over 75 years have atrial fibrillation,1 and 56% of people with this arrhythmia are over the age of 75.2 Atrial fibrillation is a major risk factor for stroke, leading to a fivefold increase in risk.3 Because risk of stroke increases with age,4 stroke prevention in elderly people with atrial fibrillation is a key aspect of management for this group.

Anticoagulation therapy with warfarin is highly effective at reducing stroke risk, but is associated with monitoring costs and a higher risk of haemorrhage compared with other treatments.5 Antiplaetelet agents such as aspirin provide a more convenient but less effective alternative.6 A meta-analysis of individual-patient data from trials showed that anticoagulants are significantly more effective than aspirin at preventing stroke, but that this benefit is at the cost of higher risk of major bleeding.7 Concerns have been expressed over the applicability of the aforementioned evidence to elderly patients with atrial fibrillation, particularly in primary care settings.8–10 Older patients were significantly under-represented in the trials: the mean age of participants in trials that compared anticoagulation therapy with no treatment was 69 years,11 and the mean age of participants in trials that compared anticoagulants with antiplatelet agents was 72 years.12 In the large Atrial fibrillation Clopidogrel Trial with Irbesartan for prevention of Vascular Events (ACTIVE W),13 in which treatment with clopidogrel plus aspirin was compared with oral anticoagulation therapy, the mean age of participants was 70 years. This younger age of participants in trials as compared with routine clinical practice is a potential drawback because several studies have shown that risk of serious haemorrhage in patients on anticoagulation increases with age.14–17 For example, in the Stroke Prevention in Atrial Fibrillation (SPAF) II trial,18 the annual risk of stroke with haemorrhagic or ischaemic residual deficit was slightly higher in the subgroup of patients aged over 75 years assigned to warfarin as opposed to aspirin (4·6% vs 4·3%). In a meta-analysis of data from individual patients aged 75 years or over who were included in trials of aspirin versus anticoagulants, a 2·2% lower risk of ischaemic stroke in those on warfarin was potentially offset by a 1·7% greater risk of a major bleed.19 However, the CIs around risk of stroke and risk of bleed were wide, because few patients aged 75 years or over were included within the trials. In addition, the frequency of anticoagulant-associated intracerebral haemorrhage has risen substantially through the 1990s, particularly in the elderly, raising further concerns about the possible overuse of anticoagulation.20
Evidence from trials of warfarin versus aspirin in primary care populations, which might be at lower risk of stroke than hospital-based populations, is mixed. Although the Second Copenhagen Atrial Fibrillation, Aspirin, and Anticoagulation (AFASAK) study reported lower risk of thromboembolism in patients on warfarin than in those on aspirin, Hellemons and colleagues and Gullov and colleagues in their trials did not show a difference between effects of the treatments; however, these studies were underpowered.

Uncertainty over the optimum treatment of elderly people with atrial fibrillation is evident in current guidelines. Guidelines produced jointly by the American College of Cardiology, the American Heart Association, and the European Society of Cardiology for the management of atrial fibrillation recommend use of anticoagulants for patients who have two or more risk factors for stroke (of which age 75 years or over is one), but these guidelines also suggest that patients aged 75 years or older at high risk of bleeding can be treated with a lower international normalised ratio (INR) target than was used in the aforementioned trials. Guidelines in England and Wales on the one hand recommend that patients aged 75 years or over with an additional risk factor should be given anticoagulants, but on the other hand recommend that the ages of these patients as a risk factor should be taken into account. This confusion in guidelines reflects practice, and currently less than half of elderly patients receive warfarin.

In view of these uncertainties and concerns over thromboprophylaxis for elderly people who have atrial fibrillation, the Birmingham Atrial Fibrillation Treatment of the Aged (BAFTA) study compared the efficacy of warfarin with that of aspirin for the prevention of stroke in a primary care population of patients aged 75 years or over who have atrial fibrillation.

Methods

Study design and participants

BAFTA was a prospective randomised open-label trial with blind assessment of endpoints. The primary aim was to compare the frequency of fatal and non-fatal disabling stroke (ischaemic or haemorrhagic), intracranial haemorrhage, and other clinically significant arterial embolism in patients who had been randomly assigned to warfarin versus aspirin. The methods used for the BAFTA study are reported in detail elsewhere. Secondary aims were to compare the frequency of major haemorrhage, other vascular events, and all-cause mortality in patients assigned to warfarin versus aspirin.

Patients were recruited from 260 general practices in England and Wales between April, 2001, and November, 2004. Patients were eligible for inclusion in the study if they were aged 75 years or over and had atrial fibrillation or atrial flutter demonstrated by a study electrocardiogram (ECG) or by an ECG done within the previous two years. Potential patients were identified through computer searches of primary care records for diagnoses of atrial fibrillation and through opportunistic pulse measurements; patients identified in either way were invited to attend the practice for an ECG, and evidence of atrial fibrillation was verified by a consultant cardiologist. Patients were excluded if they had any of the following: rheumatic heart disease; a major non-traumatic haemorrhage within the previous 5 years; intracranial haemorrhage; endoscopically proven peptic ulcer disease in the previous year; oesophageal varices; allergic hypersensitivity to either of the study drugs; a terminal illness, as judged by their primary care physician; surgery within the past 3 months; or blood pressure greater than 180/110 mm Hg. Patients were also excluded if their primary care physician judged, on the basis of risk factors for stroke and haemorrhage, that the patient either should or should not be on warfarin. Thus, random allocation of participants to treatments was ethical, because inclusion was restricted to patients for whom there was clinical uncertainty as to which of the two treatments should be used. Patients with confirmed atrial fibrillation whom the primary care physician judged to be potentially eligible were invited to attend a randomisation clinic for allocation of treatment.

Consent was a two-stage process. An information sheet that explained the aim of the trial and the potential risks and benefits of warfarin versus aspirin in this age group was sent to patients with their invitation to attend a randomisation clinic. At the clinic, the primary care physician went through the information sheet with the patient, and obtained their written informed consent to take part. Randomisation was stratified into six groups on the basis of sex and age (75–79, 80–84, and 85+ years). Within each stratum, randomly permuted
blocks of eight were generated to produce allocation tables that were held by the Birmingham Cancer Trials Unit, which primary care physicians telephoned for the treatment allocation when they had an eligible patient. The study was approved by the West Midlands Multi-centre Research Ethics Committee (MREC/99/7/57).

Procedures
Patients assigned aspirin were prescribed 75 mg daily. Patients assigned warfarin were treated with a target INR of 2·5, with an acceptable range of 2–3, in line with standard UK policy. The frequency or method of INR testing was not altered by the study protocol, which was intended to test real-life control of INR. There are three components to the INR testing: where the blood is taken; and who is responsible for adjusting the warfarin dose. Different general practices undertake none, some, or all of these activities; for those that do none, the patients attend a hospital-based anticoagulation clinic. Frequency of INR testing ranged from once per week or less, if control needed to be established, to every 12 weeks, if the INR were stable. If a patient already being treated with warfarin was randomly assigned to aspirin, then the warfarin therapy was stopped, and vice versa.

The primary outcome was first occurrence of fatal or non-fatal disabling stroke (ischaemic or haemorrhagic), other intracranial haemorrhage, or clinically significant arterial embolism. A stroke was defined by clinical symptoms that developed rapidly, or by signs of focal or widespread loss of cerebral function with symptoms that lasted for more than 24 h or led to death, and had no apparent cause other than one of vascular origin. A disabling stroke was defined by a modified Rankin score of 2–5 at 1 month or longer after stroke, or deterioration in the score if the baseline Rankin score was greater than or equal to 2. A modified Rankin score of 2 is defined as a slight disability, in which the individual cannot undertake all previous activities, but can look after their own affairs without assistance. A stroke that led to a hospital admission of 30 days or more was also classified as disabling. A stroke was classified as fatal if, in the opinion of the endpoint committee, the stroke initiated a sequence of events that led to death. If a death certificate specified stroke but the endpoint committee judged that there was insufficient corroborative clinical information, then the death was not classified as being due to stroke. Classification as intracranial haemorrhage required verification through brain imaging. An arterial embolism was defined as clinically significant if the diagnosis had been confirmed by vascular imaging, scintigraphy, surgery, or autopsy. Pulmonary embolism was not included. Clinical details on possible primary events (ie, clinical notes, discharge summaries, post-mortem reports, results of brain imaging, and death certificate, as applicable) were sent to two independent neurologists who were blind to treatment allocation. They determined whether a primary endpoint had occurred and, if this were a stroke, whether it was ischaemic or haemorrhagic.

Secondary outcomes were major extracranial haemorrhage (defined as a fatal haemorrhage, or one that resulted in the need for transfusion or surgery), other admissions to hospital for haemorrhage, hospital admission or death as a result of a non-stroke vascular event, and all-cause mortality. Clinical details on possible secondary events were sent to an independent geriatrician blind to treatment allocation, to determine whether a secondary event had occurred.

Primary care physicians reviewed patients every 6 months after treatment allocation, and copies of patients’ primary care records were reviewed by researchers between these visits. Use of other risk-modifying treatments (eg, agents to lower blood pressure or blood lipid concentrations) and whether the patient was still taking the trial drugs were monitored. Patients were sent yearly postal questionnaires, and all were flagged with the UK National Health Service Central Register to ensure that BAFTA researchers would be notified of all deaths. For patients whose warfarin treatment was managed in hospital clinics, INR records were obtained from the relevant hospital. Follow-up ceased in September, 2006, at the planned termination of the research funding.

<table>
<thead>
<tr>
<th>n (%)</th>
<th>Did not meet inclusion criteria</th>
<th>224 (6.1%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rheumatic heart disease</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>History of major haemorrhage</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>History of intracranial haemorrhage</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Active peptic ulcer</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Oesophageal varices</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Allergy to study drugs</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Blood pressure &gt;180/110 mm Hg</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Recent surgery</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>No longer had atrial fibrillation</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Terminal illness</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Wrong age</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n (%)</th>
<th>One treatment judged to be beneficial</th>
<th>1995 (54.4%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Needs warfarin</td>
<td>1570</td>
</tr>
<tr>
<td></td>
<td>Should not be on warfarin</td>
<td>417</td>
</tr>
<tr>
<td></td>
<td>Specialist referral</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n (%)</th>
<th>Patient decision</th>
<th>905 (24.7%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wanted to be on warfarin</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>Did not want to be on warfarin</td>
<td>338</td>
</tr>
<tr>
<td></td>
<td>Did not want to be on aspirin</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Did not want to take part in research</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td>Did not attend</td>
<td>27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n (%)</th>
<th>Died or moved away</th>
<th>110 (3.0%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No reason given</td>
<td>422 (11.8%)</td>
</tr>
</tbody>
</table>

| n (%) | Total                             | 3666        |

Table 1: Reasons why patients identified with atrial fibrillation did not take part in the study
We estimated that a sample size of 1240 patients followed up for an average of 3 years would lead to detection of a 33% difference in the relative risk (RR) of the primary endpoint with 90% power at 5% significance. This assumed a 9% yearly risk of stroke in patients using aspirin, a loss-to-follow-up rate of 1% per year, and a 14% yearly rate of primary endpoints or death. The study was powered on the basis of its primary endpoint alone, not on the basis of the prespecified subgroup analyses or secondary outcomes. Because the frequency of major haemorrhage is lower than the risk of stroke in the age-group used in this study, and any differential effect on other vascular events and all-cause mortality is likely to be small, we recognised that the study would be underpowered to detect differences in secondary outcomes.

The primary analysis was an intention-to-treat comparison of warfarin versus aspirin for prevention of the primary endpoint and of the secondary outcome measures, including major haemorrhage. We calculated RR of yearly event rates and their corresponding 95% CI using the Poisson exact method with SAS version 9.1 and Stata version 9.2. For all analyses, patients were censored at the time of the first event relevant to that analysis—thus, although a patient might have had multiple events, only the first event was counted in each calculation. Hazard ratios were also computed with the log-rank method. These ratios gave the same results as the RR, so are not reported. Subgroup analyses were prespecified for age, method of identification of atrial fibrillation (screening versus case-note review), previous stroke or transient ischaemic attack (TIA), use of warfarin before study entry, and baseline risk of stroke (congestive heart failure, hypertension, age of 75 years or over, diabetes mellitus, and previous stroke or TIA [CHADS2] score 1–2 or 3–5). Kaplan-Meier curves were constructed for the primary event, and to show adherence with study drugs. An on-treatment analysis was prespecified to explore the risks of major haemorrhage (intracranial haemorrhage, including haemorrhagic stroke, or extracranial haemorrhage that was fatal or led to a need for transfusion or surgery). For this analysis, risks were computed according to treatment received. Thus, a patient who switched treatments would contribute exposure time to both warfarin and to aspirin. We did not include time that a patient did not take warfarin or an antiplatelet agent, or any events that occurred during this time. We calculated the RR of a major haemorrhage and its 95% CI using the Poisson exact method. INR control was monitored by the percentage of time spent within the therapeutic range. During the study, interim data were monitored by an independent data and safety monitoring committee; this committee would have halted the trial if a difference of three or more SDs was recorded for a major outcome (terms of reference are available from the corresponding author on request). This study is registered as an
International Standard Randomised Controlled Trial, number ISRCTN89345269.

Role of the funding source
The Medical Research Council had an observer on the trial steering committee, but had no direct role in study design, in data collection, analysis or interpretation, in writing the report, or in the decision to submit for publication.

Results
Figure 1 shows the trial profile. 3231 (70%) of the 4639 patients confirmed to have atrial fibrillation were identified because their primary care records featured atrial fibrillation, and the other 1408 (30%) were identified because they had an irregular pulse in opportunistic screening. Of the whole cohort, 973 (21%) people entered the study. These patients came from 234 of the 260 participating practices; the number of patients recruited per practice ranged from 1 (53 practices) to 21 (1 practice). Of the 181 practices that randomly assigned more than one patient to treatment, 152 allocated at least one patient to each arm of the trial. Table 1 shows the baseline patient characteristics. The patients taking warfarin before the study were of similar status to those not on warfarin (mean age 81·1 vs 81·7 years), but had more risk factors for stroke (133 [35%] vs 142 [24%] had a CHADS2 score of 3–6).

Figure 2 shows adherence to study therapy. Of 488 patients allocated to warfarin, 326 (67%) remained on this treatment throughout their study time. Of those who stopped taking warfarin or did not start taking it, most (127, 78%) were put on an antiplatelet agent (aspirin in 124 cases, clopidogrel in three). Patients on warfarin had INR values in the therapeutic range (2–3) 67% of the time, and were below range 19% and above range 14% of the time. The median INR was 2·3 (IQR 2·0–2·8), and the mean INR was 2·4 (SD 0·84). Of the 234 practices that entered at least one patient for randomisation, 190 allocated at least one patient to warfarin therapy. For 41 (22%) of these practices, the blood taking, INR analysis, and warfarin dosing was done entirely at the hospital; for 58 (30%), the INR analysis and dosing were done at the hospital but the blood was taken at the practice. 47 (25%) practices that allocated at least one patient to warfarin were responsible for all three stages of INR testing, and 36 (19%) were responsible for blood taking and dose adjustment, but the INR was analysed at a hospital laboratory. Eight practices (4%) did not report how they monitored the INR. Of patients allocated to aspirin, 368 (76%) remained on this drug the whole time they were in the study. However, 82 (70%) of those who stopped taking aspirin (or did not start that treatment) were switched to or stayed on warfarin.

Table 3: Nature of primary events

<table>
<thead>
<tr>
<th>Type of stroke*</th>
<th>Warfarin (n=488)</th>
<th>Aspirin (n=485)</th>
<th>Warfarin vs aspirin</th>
</tr>
</thead>
<tbody>
<tr>
<td>n Risk per year</td>
<td>n Risk per year</td>
<td>RR (95% CI) p</td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>21 1·6%</td>
<td>44 3·4%</td>
<td>0·46 (0·26–0·79) 0·003</td>
</tr>
<tr>
<td>By severity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatal</td>
<td>13 1·0%</td>
<td>21 1·6%</td>
<td>0·59 (0·27–1·24) 0·14</td>
</tr>
<tr>
<td>Disabling non-fatal</td>
<td>8 0·6%</td>
<td>23 1·8%</td>
<td>0·33 (0·13–0·77) 0·005</td>
</tr>
<tr>
<td>Type of stroke*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischaemic</td>
<td>10 0·8%</td>
<td>32 2·5%</td>
<td>0·30 (0·13–0·63) 0·0004</td>
</tr>
<tr>
<td>Haemorrhagic</td>
<td>6 0·5%</td>
<td>5 0·4%</td>
<td>1·15 (0·29–4·77) 0·83</td>
</tr>
<tr>
<td>Unknown</td>
<td>5 0·4%</td>
<td>7 0·5%</td>
<td>0·69 (0·17–2·51) 0·53</td>
</tr>
<tr>
<td>Other intracranial haemorrhage†</td>
<td>2 0·2%</td>
<td>1 0·1%</td>
<td>1·92 (0·10–113·3) 0·65</td>
</tr>
<tr>
<td>Systemic embolism†</td>
<td>1 0·1%</td>
<td>3 0·2%</td>
<td>0·32 (0·01–3·99) 0·36</td>
</tr>
<tr>
<td>Total number of events</td>
<td>24 1·8%</td>
<td>48 3·8%</td>
<td>0·48 (0·28–0·80) 0·0027</td>
</tr>
</tbody>
</table>

RR=relative risk. Type of stroke was determined by the endpoint committee on the basis of brain imaging or post-mortem findings. If neither of these was available, the stroke was classified as unknown. The three other intracranial haemorrhages were subdural; two of these were fatal (one in each treatment group). Two of the systemic emboli were fatal (one in each treatment group).

Figure 3: Kaplan-Meier plot of time to primary event

Articles
or aspirin, and the mean diastolic blood pressure was 75 (11) mm Hg for patients on warfarin and 76 (11) mm Hg for patients on aspirin.

There were fewer primary events in patients assigned to warfarin than in those assigned to aspirin (24, 1·8% per year vs 48, 3·8% per year, RR 0·48, 95% CI 0·28–0·80) (table 3, figure 3). We used these results to calculate the number needed to treat for one year to prevent one primary event as 50.

Risk of stroke rose with age, was particularly high in people who had already had a stroke or TIA, was higher in people already known to have atrial fibrillation than in those identified by opportunistic screening, and was higher in people on warfarin before study entry than in those new to this treatment (figure 4, table 4). There was no evidence of an interaction between the effectiveness of warfarin and any of these patient subgroups. In particular, warfarin was as effective in people aged 85 years or over as it was in younger people.

We also compared the effect of warfarin with that of aspirin on risk of major haemorrhage in these subgroups, but CIs are wide because there were only 50 events (table 4). There were no significant interactions. Yearly risk of haemorrhage did not differ between people assigned to warfarin who were on anticoagulant therapy at study entry and those who were not. Risks of bleeding rose by similar amounts with age in both groups.

Table 5 shows secondary outcomes. Warfarin was no better than aspirin at prevention of non-stroke vascular events, and overall mortality rates were the same in both groups. The composite outcome of major vascular events (ie, stroke, myocardial infarction, pulmonary embolus, or vascular death) was slightly lower in people on warfarin than in people on aspirin. The risk of admission to hospital with any non-stroke vascular event did not differ between the treatment groups.

There was no evidence of increased risk of a major haemorrhage (ie, an intracranial haemorrhage, or one that was fatal or resulted in the need for transfusion or surgery) in patients on warfarin compared with those on aspirin (table 5). The net benefit therefore favoured oral anticoagulation. The on-treatment analysis for major haemorrhage similarly identified no difference between warfarin and aspirin (RR 0·88, 95% CI 0·46–1·63).

Discussion

We have shown that warfarin is more effective than aspirin in prevention of stroke in people with atrial fibrillation who are aged 75 or over. With respect to our primary aim, we showed that the frequency of major stroke, arterial embolism, and intracranial haemorrhage was significantly lower in patients on warfarin than in those on aspirin. With respect to our secondary aims, we recorded no evidence that anticoagulants were more hazardous than aspirin therapy in this age-group, although the study had limited power to detect those differences. Nevertheless, the rate of major haemorrhage on warfarin was gratifyingly low in this primary care setting. We also recorded no difference in all-cause mortality or in other vascular events, but the CIs were not sufficiently narrow to enable us to be certain that warfarin had no effect.

We showed a significant result for the primary endpoint despite the lower-than-predicted event rates and study size. This finding indicates the benefit of warfarin over aspirin for stroke prevention. Figure 5 compares our results with an individual-patient data meta-analysis8 of warfarin versus aspirin from subgroups of patients within the same age-group as used in BAFTA. The effect on stroke prevention was of a similar magnitude in BAFTA as in the meta-analysis. The yearly risk of a major bleed in people on aspirin was similar in the two data sets: 1·5% in the meta-analysis and 2·0% in BAFTA. The key difference between the two sets of results is that the meta-analysis reported a doubling of risk of major haemorrhage in people on oral anticoagulants compared with those on aspirin, whereas we showed no such difference, even though we used a low dose (75 mg) of aspirin. If we had used a larger dose of aspirin, the frequency of bleeding in this group might have been even higher.8

Figure 4: Relative risk of all primary events by subgroup
Bars show the 95% CIs calculated by Poisson exact method. The broken line shows relative risk of primary events in the whole trial population. TIA=transient ischaemic attack.
### Table 4: Risk of primary event and major haemorrhage by treatment allocation for different patient subgroups

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Warfarin</th>
<th>Aspirin</th>
<th>Warfarin vs aspirin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary event Major haemorrhage</strong>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>10/267</td>
<td>27/264</td>
<td>0.35 (0.15–0.75)</td>
</tr>
<tr>
<td>Female</td>
<td>14/221</td>
<td>21/221</td>
<td>0.65 (0.30–1.33)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75–79</td>
<td>11/197</td>
<td>15/200</td>
<td>0.71 (0.29–1.65)</td>
</tr>
<tr>
<td>80–84</td>
<td>6/196</td>
<td>19/190</td>
<td>0.30 (0.10–0.77)</td>
</tr>
<tr>
<td>85+†</td>
<td>7/95</td>
<td>14/95</td>
<td>0.50 (0.17–1.31)</td>
</tr>
<tr>
<td><strong>Method of identification§</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice register</td>
<td>15/342</td>
<td>38/341</td>
<td>0.38 (0.20–0.71)</td>
</tr>
<tr>
<td>Screening</td>
<td>9/146</td>
<td>10/144</td>
<td>0.85 (0.31–2.33)</td>
</tr>
<tr>
<td><strong>On warfarin before entry</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>6/194</td>
<td>21/187</td>
<td>0.26 (0.09–0.68)</td>
</tr>
<tr>
<td>No</td>
<td>18/294</td>
<td>27/298</td>
<td>0.65 (0.34–1.23)</td>
</tr>
<tr>
<td><strong>History of stroke or TIA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5/64</td>
<td>12/60</td>
<td>0.39 (0.11–1.19)</td>
</tr>
<tr>
<td>No</td>
<td>19/424</td>
<td>36/425</td>
<td>0.65 (0.34–1.23)</td>
</tr>
<tr>
<td><strong>CHADS2 score¶</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–2</td>
<td>15/349</td>
<td>31/349</td>
<td>0.47 (0.23–0.89)</td>
</tr>
<tr>
<td>3–6</td>
<td>9/139</td>
<td>17/136</td>
<td>0.50 (0.20–1.20)</td>
</tr>
</tbody>
</table>

*Major haemorrhage was defined as extracranial bleeds that were fatal or required transfusion or surgery, or intracranial haemorrhage (including haemorrhagic stroke). †p values are for interactions between treatments and subgroups (ie, whether the treatment has different effects in different groups). ‡Interaction reference category. §Patients identified through practice registers were already known to have atrial fibrillation; patients identified through screening were mostly newly identified as having atrial fibrillation. ¶See footnote to table 2 for explanation of CHADS2 score.

### Table 5: Risk of secondary and composite outcomes by treatment allocation for different patient subgroups

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Warfarin</th>
<th>Aspirin</th>
<th>Warfarin vs aspirin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Death</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All causes</td>
<td>107</td>
<td>108</td>
<td>0.95 (0.72–1.26)</td>
</tr>
<tr>
<td>Fatal primary endpoint</td>
<td>15</td>
<td>23</td>
<td>0.63 (0.31–1.26)</td>
</tr>
<tr>
<td>Other vascular death*</td>
<td>41</td>
<td>34</td>
<td>1.16 (0.72–1.88)</td>
</tr>
<tr>
<td>Non-vascular death*</td>
<td>51</td>
<td>51</td>
<td>0.96 (0.64–1.45)</td>
</tr>
<tr>
<td><strong>Secondary vascular outcomes (fatal and non-fatal)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All strokes</td>
<td>33</td>
<td>61</td>
<td>0.52 (0.33–0.80)</td>
</tr>
<tr>
<td>All strokes plus TIA</td>
<td>40</td>
<td>70</td>
<td>0.55 (0.36–0.82)</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>15</td>
<td>15</td>
<td>0.96 (0.44–2.11)</td>
</tr>
<tr>
<td>Heart failure</td>
<td>38</td>
<td>23</td>
<td>1.19 (0.92–2.79)</td>
</tr>
<tr>
<td>Other vascular events†</td>
<td>34</td>
<td>45</td>
<td>0.71 (0.44–1.13)</td>
</tr>
<tr>
<td>All non-stroke vascular events</td>
<td>78</td>
<td>76</td>
<td>0.97 (0.70–1.35)</td>
</tr>
<tr>
<td><strong>Haemorrhage (fatal and non-fatal)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major extracranial haemorrhage</td>
<td>18</td>
<td>20</td>
<td>0.87 (0.43–2.17)</td>
</tr>
<tr>
<td>Other hospital admission for haemorrhage</td>
<td>24</td>
<td>19</td>
<td>1.22 (0.64–2.36)</td>
</tr>
<tr>
<td>All major haemorrhages (including intracranial and haemorrhagic stroke)</td>
<td>25</td>
<td>25</td>
<td>0.96 (0.53–1.75)</td>
</tr>
<tr>
<td><strong>Composite outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major vascular events (stroke, myocardial infarction, pulmonary emboli, I vascular death)</td>
<td>76</td>
<td>100</td>
<td>0.73 (0.53–0.99)</td>
</tr>
<tr>
<td>Primary events plus major haemorrhage</td>
<td>39</td>
<td>64</td>
<td>0.59 (0.39–0.89)</td>
</tr>
</tbody>
</table>

Analyses are censored at first event, so the composite outcomes are not the sum of the individual categories of event. *Includes deaths that occurred after non-fatal primary endpoints, including four deaths from stroke (as ‘other vascular death’). †Other events leading to hospital admission or death, such as angina, deep vein thrombosis, acute bowel ischaemia, pulmonary embolism, acute arrhythmia, and elective vascular surgery. ‡There were five pulmonary emboli, one in the warfarin group and four in the aspirin group.
The similarity in risk of major haemorrhage between patients on warfarin and those on aspirin in this study is surprising. There are several possible explanations. First, the CIs for our estimate of risk of major haemorrhage were wide, so we cannot exclude the possibility that warfarin does increase risk compared with aspirin. However, if warfarin does increase the risk, it does so by less than the twofold increase estimated in the meta-analysis of individual-patient data. Second, some of the anticoagulation trials that were used for the meta-analysis had higher target INR ranges than BAFTA: the upper end of the range was 4·5 in SPAF II,4·2 in AFASAK,20 and 4·0 in the European Atrial Fibrillation Trial,27 compared with the BAFTA target of 2·5 (acceptable range 2–3). Risk of intracranial bleeding is much higher in patients who have an INR greater than 4.8 Notably, the beneficial effects on stroke prevention in BAFTA were similar to those in the meta-analysis, in which the mean INR target was higher.8 A further possible explanation for the low occurrence of bleeding in people on anticoagulation therapy in BAFTA is that 40% of patients were on warfarin before entry to the study, since the hazards of warfarin seem to be greater in people who are new to the treatment. In ACTIVE W,13 the yearly risk of a major haemorrhage was 2·0% in people randomly assigned to warfarin who were already on the drug, compared with 2·9% in people who were started on warfarin as part of the trial. (In the studies used for the meta-analysis, patients were all new to anticoagulation.)9 However, we did not record significant differences in haemorrhage risk between people who were naive to warfarin and those who had previously been on the treatment (2·1% vs 1·6%). Importantly, the degree of INR control in our study resembles the control in typical primary care practice in the UK, and thus is likely to show the true hazard of warfarin in non-trial settings. Indeed, a cohort study of 27 Scottish practices reported that INR control was in the target range (2–3) 68% of the time—very similar to our control.8 The degree of INR control achieved in this study was similar to that in other trials that have aimed for a target INR of 2–3. Patients on warfarin were within the target INR range for 64% of the time in ACTIVE W, and for 61% of the time in SPAF III.11,22

The low safety of aspirin in this age-group was also seen in the Warfarin Versus Aspirin in Stroke Prevention in Octogenarians with Atrial Fibrillation trial. This trial randomly assigned 75 patients aged 80–90 years to warfarin or aspirin, and reported a higher rate of side-effects and intolerance to aspirin (albeit at the higher dose of 300 mg per day) than to warfarin.19 Furthermore, people with atrial fibrillation in the ACTIVE W trial who were randomly assigned to anticoagulation therapy with a target INR of 2–3 or to aspirin plus clopidogrel had a similar yearly risk of haemorrhage (2·2% on oral anticoagulation vs 2·4% on the antiplatelet combination).13

Another possible explanation for the absence of difference in bleeding risk between warfarin and aspirin is the crossover between treatment. A third of people randomly assigned to warfarin did not start the treatment or started it but later stopped, and 17% of patients randomly assigned to aspirin were taking warfarin by the end of the study. These crossovers are likely to dilute differences in effect between the agents. Thus, although we have probably underestimated the benefit of warfarin over aspirin in prevention of ischaemic stroke and thromboembolism, the crossovers might also have led to underestimation of bleeding risk, if people perceived to be at high risk of haemorrhage were taken off warfarin. However, the effect is likely to be small, because no differences were reported in either the intention-to-treat or the on-treatment analyses.

The slightly greater number of hospital admissions for heart failure in the warfarin group is probably a chance finding. The difference was not significant, and admission for heart failure was one of multiple secondary outcomes that were tested. Other vascular and all-cause mortality did not differ between the warfarin and aspirin groups. This finding might be because of low power, but is consistent with the meta-analysis of the oral anticoagulant versus aspirin trials,9 which reported hazard ratios of 0·95 for vascular death and 0·93 for all-cause death at all ages. Only 20% of deaths in BAFTA were attributable to fatal primary endpoints or subsequent fatal strokes, so the efficacy of warfarin on stroke risk

---

![Figure 5: Indirect comparison of results from BAFTA for ischaemic stroke and major haemorrhage with results from six other randomised trials of aspirin versus anticoagulation](image-url)

BAFTA results are compared with those from a meta-analysis by Van Walraven and colleagues8 that used data from individual patients aged ≥75 years from six previous studies.
reduction is diluted by the 80% of deaths not related to stroke. The apparent absence of difference between warfarin and aspirin with respect to other vascular events might be due to the similar effectiveness of aspirin and warfarin in prevention of these events.40

We recorded no significant differences in risk of primary event or major haemorrhage in the different patient subgroups that we looked at. However, the study was not powered to detect interactions, so these analyses might be thought of as only exploratory, particularly for the risk of major haemorrhage, for which there were few events.

Warfarin was associated with a lower risk of primary endpoint than aspirin for all age-groups, and there was no evidence that the risks of warfarin relative to aspirin increased with age. Indeed, in people aged 85 years or over, the risk of a major haemorrhage was 27% lower on warfarin than it was on aspirin (this difference was not significant). Nevertheless, the yearly risks of major haemorrhage rose with age for both treatments, being on average over 3% in people aged 85 years or over. We did not allocate a group to no treatment in this study, so we do not know what the risk of major haemorrhage in this age-group would have been without treatment.

An advantage of a study based on primary care, such as BAFTA, is that the nature of the interventions are as close as possible to the real clinical situation. Furthermore, the results are related to the effectiveness of the treatments in the general population, by contrast with studies that recruit from hospital settings, where patients might be atypical. Nevertheless, only 21% of the study population that we identified as having atrial fibrillation took part in the study; we estimate that this is about 10% of the total study population of people over the age of 75 years who had atrial fibrillation, because only about half of people we identified as potentially having atrial fibrillation responded to an initial invitation to attend the practice for an ECG. This recruitment rate is similar to that in other atrial fibrillation trials that attempted to identify the total study population. For example, the SPAF trial included 7% of patients identified in atrial fibrillation. The primary-care-based studies AFASAK1 and AFASAK2 recruited respectively 40% and 24% of patients who attended an ECG laboratory and had an ECG showing atrial fibrillation. In the primary-care-based Primary Prevention of Arterial Thromboembolism in Nonrheumatic Atrial Fibrillation study, 21% of identified patients were randomly assigned to a trial that involved treatment with anticoagulants.23

Do these low proportions of the total population affect the applicability of the findings? In BAFTA, the most common reason provided for exclusion of patients was a clinical preference for anticoagulation. This finding suggests that the effect of the patient selection was to underestimate the potential benefit of anticoagulation, because these patients are likely to have been those in whom anticoagulation was perceived to be safer, and in whom the risk of stroke was likely to be higher. Indeed, the risk of stroke in patients randomly assigned to aspirin was higher in BAFTA in people on warfarin at entry to the study than in people who were not on warfarin. Few patients were excluded because of perceived high risk of haemorrhage: only 20% of the patients excluded because the treatments were not judged to have equal potential benefit were excluded because warfarin was felt to be inappropriate. Interestingly, patients were more likely to state a preference not to be on warfarin than one to be on it, but these preferences might change as a result of the data made available by this trial.

In this study, strokes needed to be of the most clinically severe types (fatal or disabling) to qualify as a primary endpoint. Therefore, the recorded benefits are in terms of prevention of serious clinical events. We probably did not capture all the minor strokes and TIsAs that occurred. This is a possible explanation for the higher proportion of fatal strokes reported in this study than in ACTIVE W, although our proportion of fatal strokes was similar to that in the Stroke Prevention using an Oral Thrombin Inhibitor in Atrial Fibrillation III trial.44

For the BAFTA study, we used 75 mg aspirin, the dose most commonly used in the UK for stroke prevention in patients with atrial fibrillation. There is strong evidence that this dose is at least as effective as higher doses for stroke prevention for indications other than atrial fibrillation, and the risk reduction for stroke achieved by aspirin in atrial fibrillation is of the same order of magnitude as is achieved in other high-risk groups.5,45 There is no evidence that different aspirin doses have different effects in the atrial fibrillation trials, so a higher dose of aspirin in this study would probably not have been any more effective at preventing stroke, whereas as it would have been likely to increase haemorrhage rates.

The primary event rate in BAFTA was lower than we had anticipated, especially in patients with additional risk factors for stroke. For example, people on aspirin with a CHADS2 score of 3–6 had a yearly risk of a primary event of only 3–3%, compared with an anticipated average risk nearer to 9%. Lower than anticipated event rates were also reported in ACTIVE W. These lower rates are probably caused by a decline in the frequency of stroke secondary to unrelated changes in management of other risk factors, such as blood pressure and blood cholesterol. The low rates also mean that the benefit of warfarin over aspirin in terms of absolute risk reduction is low—2% per year. Nevertheless, more primary events were recorded in people with a history of stroke or TIA, confirming the importance of this as a risk factor for stroke in atrial fibrillation, as emphasised in guidelines for England and Wales.7

We might have underestimated the benefits of anticoagulant therapy compared with aspirin in the elderly population as a whole, because many patients at high risk of stroke will have been excluded from the study.
and because of the treatment crossovers. Studies have reported that only 46% of people with atrial fibrillation over the age of 75 years with a history of stroke or TIA were on warfarin in primary care,46 and that less than half of people aged over 75 years with atrial fibrillation were discharged from hospital on warfarin.47 Our results show that warfarin could safely be used much more widely in this age-group. In summary, these data lend support to the use of anticoagulation for all people aged over 75 years with atrial fibrillation and thrombosis—a range in the benefit is not worth the inconvenience of the treatment.46 The target INR should be 2–3—a range in which there is clear evidence of benefit and no evidence from this study of harm compared with aspirin. Aspirin itself should not be regarded as a contraindication to anticoagulation therapy.

Contributors
JM, FDRH, DF, and GYHL participated in conception of the study, and JM, FDRH, DF, EM, and GYHL designed the study. KF contributed to amendments to study design. JM, FDRH, and DF were responsible for data acquisition. JM, DF, and AR were responsible for data analysis, and JM, FDRH, DF, and GYHL were responsible for data interpretation. JM wrote the first draft of the paper, and all authors contributed to subsequent drafts. All authors have seen and approved the final version.

BAFTA committee members and investigators

Conflict of interest statement
GYHL has received funding for research, educational symposia, consultancy, and lecturing from manufacturers of drugs used to treat atrial fibrillation and thrombosis (AstraZeneca, Sanofi-Aventis, Bayer, Astellas, and Daiichi-Sankyo). GYHL was clinical advisor to the Guideline Development Group that wrote the UK National Institute for Health and Clinical Excellence guidelines on management of atrial fibrillation, and is on the writing committee for the American College of Chest Physicians guidelines on antithrombotic therapy for atrial fibrillation. All other authors declare that they have no conflict of interest.

Acknowledgements
BAFTA was funded by the Medical Research Council and supported by MidReC and the Primary Care Research Trust. We thank other primary care research networks for their support: Hertfordshire Research Network (HertNet); Institute for Health Research, University of Bedfordshire; Cymru Alliance of Primary Care Oriented Research Networks (CAPRICORN); Northern Primary Care Research Network (NoReN); Somerset and North and East Devon Primary Care Research Network (SaNDNeR); London South Thames Research Network (STaRNet); and Trent Focus. We also thank the Stroke Research Network for support, and thank Robert Hart and Peter Sandercroft for their contributions to the final draft of the paper.

References


Morgan SV. Between the devil and the deep blue sea—balancing the risks and potential benefits of warfarin for older people with atrial fibrillation. Age Ageing 2004; 33: 544–47.


Mant J, Fitzmaurice D, Murray E, Hobbs F. Study does not have power to show that aspirin is as good as anticoagulation. BMJ 2000; 320: 1309–14.


Efficacy of sapropterin dihydrochloride (tetrahydrobiopterin, 6R-BH4) for reduction of phenylalanine concentration in patients with phenylketonuria: a phase III randomised placebo-controlled study

Harvey L Levy, Andrzej Milanowski, Anupam Chakrapani, Maureen Cleary, Philip Lee, Friedrich K Trefz, Chester B Whitley, François Feillet, Annette S Feigenbaum, Judith D Bebchuk, Heidi Christ-Schmidt, Alex Dorenbaum, for the Sapropterin Research Group*

Summary

Background Early and strict dietary management of phenylketonuria is the only option to prevent mental retardation. We aimed to test the efficacy of sapropterin, a synthetic form of tetrahydrobiopterin (BH4), for reduction of blood phenylalanine concentration.

Methods We enrolled 89 patients with phenylketonuria in a Phase III, multicentre, randomised, double-blind, placebo-controlled trial. We randomly assigned 42 patients to receive oral doses of sapropterin (10 mg/kg) and 47 patients to receive placebo, once daily for 6 weeks. The primary endpoint was mean change from baseline in concentration of phenylalanine in blood after 6 weeks. Analysis was on an intention-to-treat basis. The study is registered with ClinicalTrials.gov, number NCT00104247.

Findings 88 of 89 enrolled patients received at least one dose of study drug, and 87 attended the week 6 visit. Mean age was 20 (SD 9.7) years. At baseline, mean concentration of phenylalanine in blood was 843 (300) µmol/L in patients assigned to receive sapropterin, and 888 (323) µmol/L in controls. After 6 weeks of treatment, patients given sapropterin had a decrease in mean blood phenylalanine of 236 (257) µmol/L, compared with a 3 (240) µmol/L increase in the placebo group (p=0.0001). After 6 weeks, 18/41 (44%) patients (95% CI 28–60) in the sapropterin group and 4/47 (9%) controls (95% CI 2–20) had a reduction in blood phenylalanine concentration of 30% or greater from baseline. Blood phenylalanine concentrations fell by about 200 µmol/L after 1 week in the sapropterin group and this reduction persisted for the remaining 5 weeks of the study (p<0.0001). After 6 weeks, 18/41 (44%) patients (95% CI 28–60) in the sapropterin group and 4/47 (9%) controls (95% CI 2–20) had a reduction in blood phenylalanine concentration of 30% or greater from baseline. Blood phenylalanine concentrations fell by about 200 µmol/L after 1 week in the sapropterin group and this reduction persisted for the remaining 5 weeks of the study (p<0.0001). 11/47 (23%) patients in the sapropterin group and 8/41 (20%) in the placebo group experienced adverse events that might have been drug-related (p=0.80). Upper respiratory tract infections were the most common disorder.

Interpretation In some patients with phenylketonuria who are responsive to BH4, sapropterin treatment to reduce blood phenylalanine could be used as an adjunct to a restrictive low-phenylalanine diet, and might even replace the diet in some instances.

Introduction One of the most compelling and repeated stories in medical genetics is the effective prevention of mental retardation from phenylketonuria by newborn screening and early dietary treatment to reduce the concentration of phenylalanine in blood.\(^1\) Mutations in the gene that encode a resulting increase in blood phenylalanine.\(^4\) Non-compliance after successful treatment in early childhood can result in below-average IQ scores,\(^5\) behavioural problems,\(^3\) and severe emotional dysfunction, including attention deficit disorders,\(^6\) depression, anxiety, and agoraphobia.\(^7,8\)

One novel approach to treatment is to enhance the activity of residual PAH by treatment with pharmacological amounts of its cofactor, tetrahydrobiopterin (BH4), or its biologically active synthetic form 6R-BH4 (sapropterin dihydrochloride, known as sapropterin). Since 1999, studies have shown that a subset of patients with phenylketonuria respond to treatment with BH4\(^14\) and that this effect is caused by increased oxidation of phenylalanine.\(^15\) This responsiveness to BH4 is probably associated with mutations in the PAH gene that encode a variant form of the enzyme with some residual activity; however, the genotype–phenotype correlation is weak.\(^9\) In some patients with phenylketonuria, 6R-BH4 treatment has been shown to increase the tolerance for phenyl-
alanine sufficiently to allow a less restrictive diet\textsuperscript{49} or even, in some cases, discontinuation of the diet.\textsuperscript{50,51}

We aimed to assess the efficacy of sapropterin compared with placebo, for reduction of phenylalanine in blood in patients with phenylketonuria. Our secondary objective was to assess the safety of sapropterin compared with placebo.

**Methods**

**Participants**

We enrolled patients between March, 2005, and February, 2006, at 16 centres in North America and 14 centres in six countries in Europe. The protocol (PKU-003) was developed in collaboration with a Phenylketonuria Advisory Committee and approved at each centre by its institutional review board, research ethics board, or ethics committee. Each patient, or a parent or guardian for those aged younger than 18 years, gave written informed consent for participation in the study.

Figure 1 shows that we screened 89 patients with phenylketonuria who had relaxed or abandoned a strict low-phenylalanine diet and who had participated in a previous phase-I screening study (Protocol PKU-001).\textsuperscript{21} Eligibility criteria included responsiveness in PKU-001, which was defined as a reduction of 30% or more in blood phenylalanine concentration after 8 days of treatment with sapropterin at a dose of 10 mg/kg per day. We started to screen patients 6 weeks before randomisation. Eligibility criteria for our study were blood phenylalanine of 600 µmol/L or greater (or ≥450 µmol/L after a protocol amendment) at the screening visit, age of 8 years or older, and willingness and ability to comply with study procedures and to adhere to their current diet. Women with childbearing potential had to have a negative urine pregnancy test to be eligible, and sexually active men and women had to adopt acceptable birth control measures to prevent pregnancy.

**Procedures**

We took two baseline measurements of blood phenylalanine concentration at 1 and 2 weeks before randomisation. At week 0, we randomised all eligible patients in a 1:1 ratio. We used an interactive voice-response telephone system to maintain blinding. Patients were stratified by study centre and by blood phenylalanine at screening visit (<600 µmol/L and ≥600 µmol/L). Patients in the treatment group were allocated to receive 10 mg/kg sapropterin, and controls were given a placebo (which was similar in taste and appearance), orally once daily for 6 weeks.

We obtained randomisation lists from Statistics Collaborative (Washington, DC, USA); they were generated by a computer program that was verified for accuracy with strict quality-control procedures. Each randomisation list started with a block of two, followed by blocks of four, which ensured that the first two assignments in each stratum were to both the sapropterin and placebo groups, and that subsequent assignments in each stratum consisted of two sapropterin and two placebo assignments in random order.

Block size was not divulged to the study sponsor or to the study investigators until the study was completed. Investigators, patients, and sponsors were kept unaware of the treatment allocation until the database was locked. Sapropterin and placebo tablets were dissolved in 120–240 mL of water, apple juice, or orange juice. Patients were instructed to continue their usual diet without modification. The concentration of blood phenylalanine was measured 2–5 h after breakfast at the screening assessment, each of the two baseline assessments, and treatment weeks 0, 1, 2, 4, and 6. We also measured the phenylalanine concentration in blood at the final visit of any patient who withdrew from the study early. Blood had been collected for assessment of PAH genotype during the screening study (PKU-001).

A central laboratory measured the concentration of phenylalanine in blood with the aminoacid analyser method. Centres and patients were forbidden to obtain such concentrations independently. Neither investigators nor patients were aware of postbaseline blood phenylalanine concentrations during the study. We also recorded medical histories, monitored adverse events and vital signs, and did physical examinations, thyroid function tests (thyroxine [T4] and thyroid-stimulating hormone [TSH]), and clinical laboratory tests (including chemistry, haematology, and urinalysis).

The primary endpoint, to measure efficacy of sapropterin compared with placebo, was the change in blood phenylalanine concentration from baseline to week 6. Secondary endpoints were changes in phenylalanine concentrations in blood at each of the 6 weeks of treatment, and the proportion of patients who had blood

![Figure 1: Trial profile](image-url)
phenylalanine of less than 600 µmol/L at week 6. To assess safety, we also compared the frequencies of all adverse events and serious adverse events in the sapropterin and placebo groups. Adverse events were classified according to the Medical Dictionary for Regulatory Activities.22

Statistical analysis
We analysed all randomised patients who received at least one dose of study drug. Based on the results of a previous pilot study,19 we assumed a difference between treatment groups in mean change at week 6 of 150 (SD 85) µmol/L, and a two-sided type I error rate of 0·05. Therefore, we calculated that a sample size of 80 patients (40 in each group) would provide over 95% power to detect a significant difference between treatment groups at week 6.

To assess change in blood phenylalanine concentration from baseline to week 6, we defined baseline blood phenylalanine as the average of two baseline assessments (1 and 2 weeks before randomisation) and the week 0 assessment. If the week 6 assessment of blood phenylalanine was missing, we used a last observation carried forward method to impute the data. We compared the mean change in blood phenylalanine concentration at week 6 with an analysis of covariance model, using change in blood phenylalanine from baseline to week 6 as the outcome variable, and treatment group as a categorical covariate.

We compared the mean change in blood phenylalanine concentration at week 6 by analysis of covariance, with change in blood phenylalanine from baseline to week 6 as the outcome variable, treatment group as a categorical covariate, and baseline blood phenylalanine concentration as a continuous covariate. We used Fisher’s exact test to compare the proportion of patients whose blood phenylalanine concentration was less than 600 µmol/L at week 6. We also did a post-hoc analysis of the proportion of patients with a blood phenylalanine of less than 360 µmol/L at week 6. We analysed data with SAS statistical software (version 9.1.3). The study is registered with ClinicalTrials.gov, number NCT00104247.

Role of the funding source
The study protocol was drafted and developed by the study sponsor, BioMarin Pharmaceuticals, in collaboration with Merck Serono. The study sponsor collected and analysed the data. All authors had full access to all data and statistical analyses in this study and participated in writing the manuscript. The corresponding author had final responsibility for the decision to submit for publication.

Results
42 patients were randomly assigned to receive sapropterin, and 47 to placebo (figure 1). Sex was the only baseline demographic for which the treatment groups differed (table 1). Treatment compliance was high; 72/88 (82%) patients took all doses of the study drug correctly throughout the 6 weeks of the study. Numbers of missed doses or dosing errors were similar in the two treatment groups. Dietary compliance was also high; only 7/41 (17%) patients in the sapropterin group and 12/47 (26%) controls reported that their diet contained an increased or decreased intake of phenylalanine. None of these deviations from protocol were regarded as serious.

Concentrations of blood phenylalanine for each patient during the 6 weeks of the study are shown in figure 2. Figure 3 shows that mean blood phenylalanine concentration at baseline was 842·7 (SD 299·6) µmol/L for the Sapropterin treatment group (n=41) and 842·7 (SD 300·6) µmol/L for the Placebo group (n=47). The total mean blood phenylalanine concentration was 842·7 (SD 299·6) µmol/L.

<table>
<thead>
<tr>
<th>Baseline characteristics</th>
<th>Sapropterin treatment group (n=41)</th>
<th>Placebo group (n=47)</th>
<th>Total (n=88)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>27 (66%)</td>
<td>24 (51%)</td>
<td>51 (58%)</td>
</tr>
<tr>
<td>Female</td>
<td>14 (34%)</td>
<td>23 (49%)</td>
<td>37 (42%)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age range (IQR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥8 and ≤12 years</td>
<td>6 (15)</td>
<td>11 (23%)</td>
<td>17 (19%)</td>
</tr>
<tr>
<td>&gt;12 years</td>
<td>35 (85)</td>
<td>36 (77%)</td>
<td>71 (81%)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>39 (95%)</td>
<td>47 (100%)</td>
<td>86 (98%)</td>
</tr>
<tr>
<td>Non-white</td>
<td>2 (5%)</td>
<td>0</td>
<td>2 (2%)</td>
</tr>
<tr>
<td><strong>Blood phenylalanine at screening</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylalanine &lt;600 µmol/L</td>
<td>7 (17%)</td>
<td>9 (19%)</td>
<td>16 (18%)</td>
</tr>
<tr>
<td>Phenylalanine ≥600 µmol/L</td>
<td>34 (83%)</td>
<td>38 (81%)</td>
<td>72 (82%)</td>
</tr>
</tbody>
</table>

Data are n (%) or mean (SD). For categorical variables, percentages include all patients with available data. For continuous variables, all patients with available data are included.

Table 1: Baseline characteristics

Figure 2: Blood phenylalanine concentrations at baseline and week 6
Black line shows no change from baseline; dotted line shows 30% reduction.

Figure 3: Blood phenylalanine concentration at week 6 from baseline.
sapropterin group and 888.3 (323.1) µmol/L for controls. At 6 weeks, 18/41 (44%) patients given sapropterin (95% CI 28–60) and 4/47 (9%) controls (95% CI 2–20) had a blood phenylalanine reduction of 30% or more from baseline (p=0·0002). Blood phenylalanine concentration was reduced by 50% or more in 13/41 (32%) patients who received sapropterin (95% CI 18–48) and 1/47 (2%) controls (95% CI 0–11).

Figure 4 shows that mean blood phenylalanine in controls fluctuated only slightly from baseline over the 6 weeks of the study. By contrast, in the sapropterin group, the mean fell from 842.7 (299.6) µmol/L at baseline to 619.9 (354.7) µmol/L at week 1 and remained at this lower concentration for the duration of the study. At week 6, the mean change in blood phenylalanine from baseline for the sapropterin group was −235.9 (257.0) µmol/L, compared with 2.9 (239.5) µmol/L for controls (p<0.0001). By week 6, the mean blood phenylalanine concentration was 606.9 (377.0) µmol/L in the sapropterin group. The estimated difference between treatment groups in mean change in concentration of blood phenylalanine from baseline at week 6 was −245 (52.5) µmol/L (p=0·0002).

Table 2 shows analyses of the efficacy of sapropterin at week 6. 7/41 (17%) patients assigned to sapropterin had blood phenylalanine of less than 600 µmol/L at screening; this proportion had increased to 22/41 (54%, 95% CI 38–69) after 6 weeks (p=0·004). In controls, this proportion was 9/47 (19%) at screening and 11/47 (23%, 95% CI 11–36) at week 6. 15/34 (44%) of patients assigned
to sapropterin and 4/38 (11%) of controls had blood phenylalanine of 600 μmol/L or more at screening, and of less than 600 μmol/L at week 6 (p=0.003). 13/41 (32%) in the sapropterin group and 1/47 (2%) controls had blood phenylalanine concentrations of less than 360 μmol/L at week 6 (p<0.001).

11/47 (23%) patients in the sapropterin group and 8/41 (20%) patients in the control group experienced adverse events that might have been drug-related (p=0.80). No patient withdrew from the study because of adverse events. Most adverse events were deemed to be unrelated to the study drugs. The most commonly reported adverse event was upper respiratory tract infection. Nervous system disorders (eg, headache) were more frequent in the placebo group than in the sapropterin group (table 3). No serious adverse events were recorded in either group, and no patient died during the study.

No patients in the sapropterin group and two controls had clinically significant changes in liver enzymes (alanine aminotransferase in one and aspartate transaminase in the other). One patient in the sapropterin group had a clinically significant low T4 at week 0 (before sapropterin exposure) and again at week 6. This patient had normal TSH concentration at week 0 and high TSH after 6 weeks.

Full PAH genotypes, obtained for the PKU-001 trial, were available for 17 of the 19 patients with a reduction of 30% or more in blood phenylalanine in response to sapropterin. In one patient, only a single mutation was identified and in the other a genotype was not available. Responsiveness to BH4 is thought to require at least one PAH mutation that allows for residual enzyme activity (non-null mutation). 16/17 fully genotyped patients had at least one non-null mutation, as did the patient with only one identified mutation. However, responsiveness was not consistently linked with specific mutations; six mutations were associated with both responsiveness and non-responsiveness. Moreover, the two PAH mutations identified in a patient who had a 63% reduction in blood phenylalanine after 6 weeks of sapropterin treatment were IVS10-3C→T and G272X, neither of which would seem to allow for residual PAH activity.

**Discussion**

This 6-week study showed that sapropterin treatment can reduce the concentration of phenylalanine in blood in some patients with phenylketonuria. Mean blood phenylalanine fell in the sapropterin group at week 1 and remained lower than in patients in the placebo group for the duration of the study. Sapropterin treatment over a 6-week period was safe. Our results suggest that sapropterin treatment might be used as an adjunct to the restrictive low-phenylalanine diet in patients with phenylketonuria, and could even replace the diet in some instances. These findings accord with those of earlier pilot studies that also described substantial reductions in blood phenylalanine in patients with phenylketonuria who were given oral doses of 5–20 mg/kg per day of 6R-BH4, and were able to normalise their diet.

The proportion of patients assigned to receive sapropterin who had decreases in blood phenylalanine concentrations after 6 weeks of at least 30% was larger than that in controls (44% vs 9%). However, almost the same number of patients in each group, 12 given sapropterin and 10 given placebo, had an 11–29% decrease in blood phenylalanine concentration at week 6. Moreover, 7 (17%) patients given sapropterin and 21 (45%) controls had increased phenylalanine in blood. Some variations in blood phenylalanine could possibly have been caused by chance, or by undisclosed changes in dietary phenylalanine. Nevertheless, the much higher proportion of patients who had a blood phenylalanine decrease of at least 30% in the sapropterin group, compared with controls, suggests that the cause was a sapropterin response rather than a dietary difference. All patients with phenylketonuria who were enrolled in our study had prestudy reductions of at least 30% in blood phenylalanine while taking sapropterin. However, not all were necessarily truly responsive to BH4. Therefore, patients in our study for whom sapropterin treatment did not reduce blood phenylalanine might not have been BH4-responsive.
14/17 mutations we identified have been previously linked to a BH4-responsive phenotype;27 two mutations, G218V and Q226H, have not been reported in responsive patients. BH4-responsiveness is thought to occur only when at least one PAH mutation allows for residual enzyme activity. However, one patient in our study, who had a 63% decrease in concentration of blood phenylalanine while taking sapropterin, had two PAH mutations, IVS10-3C→T and G272X, both of which would seem to be null. Because two intronic mutations have been linked with BH4-responsiveness,27 the intronic mutation IVS10-3C→T could possibly be associated with residual PAH activity, and thus responsiveness. If so, the explanation could lie in alternative splicing, which would result in multiple transcripts, at least one of which might produce enzyme activity.30

Despite these genotype–phenotype associations, BH4-responsiveness cannot reliably be predicted from PAH genotype,14,29 and must therefore be determined on the basis of blood phenylalanine in response to a loading dose of BH4. The most commonly used assessment method consists of administration of two 20 mg/kg doses of 6R-BH4, 24 h apart, followed by blood phenylalanine measurements at 0, 4, 8, 12, 24, and 48 h after the first dose.30 A reduction in blood phenylalanine concentration of 30% or more at any of these timepoints is usually regarded as indicative of responsiveness, although a lesser decrease could also be therapeutically beneficial.

Several short-term dietary-intervention crossover studies in patients with early treated, well controlled phenylketonuria, and with IQs in the normal range, have shown that executive function and higher cognitive function deteriorate when the concentration of phenylalanine in blood exceeds 600 µmol/L and improve when phenylalanine is reduced to 600 µmol/L or below, even above 10 years of age.15–17 In our study, more than half the patients in the sapropterin group had blood phenylalanine concentrations of 600 µmol/L or greater by week 6. Smith and colleagues showed a 4–10 point reduction in IQ, dependent on age, at 4–10 years of age.16 Together, these findings emphasise the importance of maintaining low blood phenylalanine concentrations at all times, and support the use of sapropterin to lower blood phenylalanine in BH4-responsive patients with phenylketonuria. Furthermore, lifelong, reliable control of blood phenylalanine could also promote better overall central nervous system development and function, and alleviate the burden of a limited dietary regime.

Contributors
HLL drafted the manuscript. All authors have read and approved the manuscript and contributed to the study design, analysis, or interpretation of data, and the drafting and revision of the manuscript.

Study organisation
Sapropterin Research Group: University of British Columbia, Children’s and Women’s Hospital, Vancouver, BC, Canada (L Clarke); Hospital for Sick Children, Toronto, Canada (A Feigenbaum); Hôpital Jeanne de Flandres, Lille (D Dobbelare); Hôpital d’Enfants, Vandoeuvre-lès-Nancy, France (F Feillet); Hôpital Necker—Enfants Malades, Paris, France (V Absalir); Otto-Huemer-Centrum für Kinder-und-Jugendmedizin, Berlin, Germany (J Hennemann); Klinik für Kinder und Jugendmedizin, Reutlingen, Germany (F Trefz); The Children’s University Hospital, Dublin, Ireland (E Trayac); Azienda Ospedaliero San Paolo Clinica, Milan, Italy (M Giovannini); Azienda Ospedaliera Universitaria, Catania, Italy (C Meli); Instytutu Matki i Dziecka, Warsaw, Poland (A Milanowski); Birmingham Children’s Hospital, Birmingham, UK (A Chakravani); Great Ormond Street Hospital, London, UK (M Cleary); National Hospital for Neurology and Neurosurgery, London, UK (P Lee); Royal Manchester Children’s Hospital, Manchester, UK (A Morris); Children’s Memorial Hospital, Chicago, IL, USA (B Burton); University of California, Los Angeles, CA, USA (E Crombez); Washington University School of Medicine, St Louis, MO, USA (D Grange); Oregon Health and Sciences University, Portland, OR, USA (C Harding); Kaiser Permanente Medical Centre, Oakland, CA, USA (J Baker); Kaiser Point West Medical Offices, Sacramento, CA, USA (M Lipson); Kaiser Santa Teresa, San Jose, CA, USA (J Bergoffen); Children’s Hospital and Harvard Medical School, Boston, MA, USA (R Levy); University of Utah, Salt Lake City, UT, USA (N Longo); Children’s Hospital, Los Angeles, CA, USA (L Randolph); Yale University School of Medicine, New Haven, CT, USA (M Seashore); University of Texas Southwestern Medical Centre, Dallas, TX, USA (I Waber); Mount Sinai School of Medicine, New York, NY, USA (M Wasserstein); University of Wisconsin, Madison, WI, USA (J Wolff).

Conflict of interest statement
HLL serves on the PKU Advisory Board and is a consultant to BioMarin Pharmaceuticals. JDB and HC-S have received funding from BioMarin Pharmaceuticals for statistical analysis. AD is an employee of BioMarin Pharmaceuticals.

Acknowledgments
This study was sponsored by BioMarin Pharmaceuticals, the Children’s Hospital Boston General Clinical Research Centre, the University of Minnesota General Clinical Research Centre, and grants M01-RR00400 and M01-RR02172 from the National Centre for Research Resources, National Institutes of Health. We thank Emil Kakkis, Stuart Swiedler, and Sandra Shipberg of BioMarin Pharmaceuticals and Inna Furman of the National Institutes of Health Consensus Development Conference for statistical analysis. We also thank the patients in the trial, the study investigators and nurses, and especially the coordinators without whose tireless work this study could not have been accomplished.

References
Nutritional iron deficiency

Michael B Zimmermann, Richard F Hurrell

Iron deficiency is one of the leading risk factors for disability and death worldwide, affecting an estimated 2 billion people. Nutritional iron deficiency arises when physiological requirements cannot be met by iron absorption from diet. Dietary iron bioavailability is low in populations consuming monotonous plant-based diets. The high prevalence of iron deficiency in the developing world has substantial health and economic costs, including poor pregnancy outcome, impaired school performance, and decreased productivity. Recent studies have reported how the body regulates iron absorption and metabolism in response to changing iron status by upregulation or downregulation of key intestinal and hepatic proteins. Targeted iron supplementation, iron fortification of foods, or both, can control iron deficiency in populations. Although technical challenges limit the amount of bioavailable iron compounds that can be used in food fortification, studies show that iron fortification can be an effective strategy against nutritional iron deficiency. Specific laboratory measures of iron status should be used to assess the need for fortification and to monitor these interventions. Selective plant breeding and genetic engineering are promising new approaches to improve dietary iron nutritional quality.

Epidemiology

Estimates of occurrence of iron deficiency in industrialised countries are usually derived from nationally representative samples with specific indicators of iron status.1 By contrast, estimates from developing countries are often based only on haemoglobin measurements from restricted regions or target populations, and should be interpreted with caution. Prevalence estimates of iron deficiency anaemia (ie, iron deficiency and low haemoglobin) based on haemoglobin alone are over-estimations because they fail to account for other causes of anaemia, such as nutritional deficiencies (eg, vitamin A deficiency), infectious disorders (particularly malaria, HIV disease, and tuberculosis), haemoglobinopathies, and ethnic differences in normal haemoglobin distributions.2,3 For example, in Côte d’Ivoire, iron deficiency was detected with specific indicators of iron status in about 50% of anaemic women and children.4 Even in industrialised countries, haemoglobin alone, which is used to detect iron deficiency anaemia, has poor sensitivity and specificity.5 Anaemia is regarded as a public health problem when the frequency of low haemoglobin values is more than 5% in the population.6 WHO estimates that 39% of children younger than 5 years, 48% of children between 5 and 14 years, 42% of all women, and 52% of pregnant women in developing countries are anaemic,7 with half having iron deficiency anaemia.8 According to WHO, the frequency of iron deficiency in developing countries is about 2.5 times that of anaemia.9 Iron deficiency is also common in women and young children in industrialised countries. In the UK, 21% of female teenagers between 11 and 18 years, and 18% of women between 16 and 64 years are iron deficient.10 In the USA, 9–11% of non-pregnant women aged between 16 and 49 years are iron deficient, and 2–5% have iron deficiency anaemia, with more than twofold higher frequency in poorer, less educated, and minority populations.11 In pregnant women of low-income areas in the USA, the frequency of iron deficiency anaemia in the first, second, and third trimesters is 2%, 8%, and 27%, respectively.12 In France, iron deficiency and iron deficiency anaemia affect 29% and 4% of children younger than 2 years;13 in the USA, 2% of children between 1 and 2 years have iron deficiency anaemia.1

Physiology

Human beings are unable to excrete iron actively, so its concentration in the body must be regulated at the site of iron absorption in the proximal small intestine (figure). Diets contain both haem and non-haem (inorganic) iron; each form has specific transporters. A putative intestinal haem iron transporter (HCP1) has been identified, which is upregulated by hypoxia and iron deficiency, and might also transport folate.14,15 Transport of non-haem iron from the intestinal lumen into the enterocytes is mediated by the divalent metal ion transporter 1 (DMT1).16 DMT1 transports only ferrous iron, but most dietary iron that enters the duodenum is in the ferric form. Therefore, ferric iron must be first reduced to ferrous iron, possibly by the brush border ferric reductase, duodenal cytochrome b (DCYTb),17 or by other reducing agents, such as ascorbic acid. Once inside the enterocyte, iron that is not directly transferred to the circulation is stored as ferritin and ultimately is lost when the cell is sloughed at the villus tip. Efflux of iron across the basolateral membrane into the blood is mediated by the transport protein ferroportin 1, and the iron oxidase, hephaestin. Ferroportin 1 also mediates export of iron from other

Search strategy and selection criteria

We searched PubMed, Current Contents Connect, and ISI Web of Science for articles in English, French, German, and Spanish. We searched for “iron”, “iron deficiency”, “anaemia”, “nutrition”, “haemoglobin”, “bioavailability”, “supplementation”, “fortification”, “plant breeding”, and “genetic engineering”. We mainly selected publications from the past 5 years, but did not exclude highly regarded earlier publications.
cells, including macrophages. Hepcidin is a regulatory hormone secreted by the liver that inhibits both the absorption and release of iron from macrophages and other cell types. Hepcidin seems to bind to ferroportin 1 at the basolateral membrane of the enterocyte, causing its internalisation and degradation. The internalisation and degradation processes decrease iron transfer into the blood, and additional iron is lost in sloughed enterocytes. In iron deficiency, hepcidin release from the liver is decreased, thereby increasing iron absorption to the maximum. In the erythroid iron cycle, senescent red cells are broken down mainly by macrophages in the spleen, and the extracted iron is returned to the circulation where it binds to transferrin. Transferrin binds to specific transferrin receptors (TfRs) on erythroid precursors in the bone marrow, and the cycle is completed when new erythrocytes enter the circulation in the following 7–10 days. Iron deficiency increases iron

Figure: Regulation of intestinal iron uptake

Haem iron is taken up by the haem iron transporter (HCP), undergoes endocytosis, and Fe²⁺ (ferrous iron) is liberated within the endosome or lysosome. Non-haem iron includes Fe³⁺ (ferric iron) salts. Fe³⁺ is reduced to Fe²⁺ by ascorbic acid in the lumen or by membrane ferrireductases that include duodenal cytochrome B (DCYTB). At the apical membrane, the acid microclimate provides an H⁺ electrochemical gradient that drives Fe²⁺ transport into the enterocyte via the divalent metal-ion transporter (DMT1). At the basolateral membrane, iron transport to transferrin in the circulation is mediated by ferroportin 1, in association with hephaestin. Hepcidin is produced by the liver, binds to ferroportin 1, causing its internalisation and degradation and decreasing iron transfer into the blood.

Dietary iron bioavailability is low in populations consuming monotonous plant-based diets with little meat. In meat, 30–70% of iron is haem iron, of which 15–35% is absorbed. However, in plant-based diets in developing countries most dietary iron is non-haem iron, and its absorption is often less than 10%. The absorption of non-haem iron is increased by meat and ascorbic acid, but inhibited by phytates, polyphenols, and calcium. Because iron is present in many foods, and its intake is directly related to energy intake, the risk of deficiency is highest when iron requirements are greater than energy needs. This
situations happen in infants and young children, adolescents, and in menstruating and pregnant women. During infancy, rapid growth exhausts iron stores accumulated during gestation and often results in deficiency, if iron-fortified formula or weaning foods are not supplied. Excessive early consumption of cows’ milk can also contribute to early-childhood iron deficiency. In a study of infants aged 6 months, frequency of iron deficiency anaemia was lowest in infants fed iron-fortified formula (about 1%) but occurred in 15% of breastfed infants, and 20% of infants fed cows’ milk or non-fortified formula. In the USA, the introduction of iron-fortified weaning foods in the 1970s was associated with a reduction in the frequency of iron deficiency anaemia in infants and preschool children. Many developing countries, plant-based weaning foods are rarely fortified with iron, and the frequency of anaemia exceeds 50% in children younger than 4 years. In school age children, iron status typically improves as growth slows and diets become more varied. The frequency of iron deficiency begins to rise again, mainly in female individuals, during adolescence, when menstrual iron losses are superimposed with needs for rapid growth. Because a 1 mL loss of blood translates into a 0.5 mg loss of iron, heavy menstrual blood loss (>80 mL per month in about 10% of women) sharply increases the risk for iron deficiency. Other risk factors for iron deficiency in young women are high parity, use of an intrauterine device, and vegetarian diets. During pregnancy, iron requirements increase three-fold because of expansion of maternal red-cell mass and growth of the fetal–placental unit. The net iron requirement during pregnancy is about 1 g (equal to that contained in about 4 units of blood), most of which is needed in the last 2 trimesters. During lactation, because only about 0.25 mg of iron per day is excreted into breastmilk and most women are amenorrhoeic, iron requirement is low—only half of that of non-pregnant, non-lactating women.

Increased blood loss from gastrointestinal parasites aggravates dietary deficiencies in many developing countries. Infections with Trichuris trichiura (whipworm) and Necator americanus (hookworm) cause intestinal blood loss and are important causes of iron deficiency anaemia. Reversed estimates indicate that hookworms afflict more than 700 million people in tropical and subtropical regions. In endemic areas, hookworm infection is estimated to account for 35% of iron deficiency anaemia and 73% of its severe form; and deworming decreases the occurrence of anaemia. In a trial in Nepal, women who were given albendazole in the second trimester of pregnancy had a lower rate of severe anaemia during the third trimester, gave birth to infants of greater weight, and mortality of infants at 6 months decreased. Iron deficiency anaemia can also be caused by impaired iron absorption. Gastric acid is needed to maintain ferric iron forms in solution, and achlorhydria might be a substantial cause of iron deficiency, mainly in elderly people, in whom atrophic gastritis is common. Other common causes of lowered iron absorption and iron deficiency are mucosal atrophy in coeliac disease and, possibly, Helicobacter pylori infection, although a study of iron absorption showed no effect of H pylori.

**Adverse effects**

The high frequency of iron deficiency anaemia in the developing world has substantial health and economic costs. In an analysis of ten developing countries, the median value of physical productivity losses per year due to iron deficiency was about US$0.32 per head, or 0.57% of the gross domestic product. In the WHO African subregion, it is estimated that if iron fortification reached 50% of the population, it would avert 570 000 disability adjusted life years (DALYs) every year. During the first two trimesters of pregnancy, iron deficiency anaemia increases the risk for preterm labour, low birthweight, infant mortality, and predicts iron deficiency in infants after 4 months of age. Estimates are that anaemia accounts for 3.7% and 12.8% of maternal deaths during pregnancy and childbirth in Africa and Asia, respectively. Data for the adverse effects of iron deficiency on cognitive and motor development in children are equivocal because environmental factors limit their interpretation. Several studies reported adverse effects of iron deficiency anaemia on infant development that might be only partly reversible. Other studies suggest that no convincing evidence exists that iron deficiency anaemia affects mental or motor development in children younger than 2 years, but that iron deficiency adversely affects cognition in school children. Anaemic school-children have decreased motor activity, social inattention, and decreased school performance. Whether adverse effects of iron deficiency on neuromotor development are due to anaemia or absence of iron in the developing brain is unclear. Iron deficiency anaemia increases susceptibility to infections, mainly of the upper respiratory tract, which happen more often and have a longer duration in anaemic than in healthy children. A recent study showed no positive effect of iron supplementation on physical growth during childhood. The response to iodine prophylaxis is reduced in goitrous children with deficiencies of both iodine and iron, probably because of impairment of the haem-dependent enzyme, thyroid peroxidase. Iron supplementation can increase low serum retinol concentrations in iron-deficient children. Iron deficiency might increase the risk for chronic lead poisoning in children exposed to environmental lead. In adults, physical activity is reduced, and manual labourers in developing countries are more productive if they are given iron and treated for hookworm and other infections. Iron deficiency, even in the absence of anaemia, might cause fatigue and reduce work performance.
Laboratory diagnosis

Table 2 shows useful indicators for diagnosis of iron deficiency anaemia in population studies. The major diagnostic challenge is to differentiate between iron deficiency anaemia in otherwise healthy individuals and anaemia of chronic disease. Inflammatory disorders increase circulating hepcidin concentrations, and hepcidin blocks iron release from enterocytes and the reticuloendothelial system, resulting in iron-deficient erythropoiesis. If chronic, inflammation can produce anaemia of chronic disease. The distinction between anaemia of chronic disease and iron deficiency anaemia is difficult because increased serum ferritin concentration in anaemia does not exclude iron deficiency anaemia in the presence of inflammation. A widely used marker of inflammation is the C-reactive protein (CRP), but the extent of increase of CRP concentration that invalidates the use of serum ferritin to diagnose iron deficiency is unclear; CRP values higher than 10–30 mg/L have been used. Moreover, during the acute-phase response, the increase of CRP concentration is typically of shorter duration than the increase of serum ferritin. Alternative markers such as α1-acid glycoprotein (AGP) might be useful because AGP tends to increase later during infection than CRP, and remains high for several weeks. A distinct advantage of the soluble transferrin receptor (sTfR) is that it might differentiate iron deficiency anaemia from anaemia of chronic disease. Thus, in surveys in developing countries with a high frequency of infection, in addition to serum ferritin and haemoglobin measurements, laboratory assessment should include sTfR, zinc protoporphyrin (ZPP), and CRP, AGP, or both, although the sensitivity and specificity of sTfR and ZPP are low in these settings. In an anaemic individual with high CRP, AGP, or both, high sTfR and ZPP concentrations are likely to mean concurrent iron deficiency, despite high serum ferritin.

<table>
<thead>
<tr>
<th>Selected cutoff values to define iron deficiency</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Haemoglobin (g/L)</strong></td>
<td></td>
</tr>
<tr>
<td>6 months–5 years &lt;110</td>
<td>When used alone, it has low specificity and sensitivity</td>
</tr>
<tr>
<td>6 years–11 years &lt;115</td>
<td></td>
</tr>
<tr>
<td>Pregnant women &lt;120</td>
<td></td>
</tr>
<tr>
<td>Non-pregnant women &lt;110</td>
<td></td>
</tr>
<tr>
<td>Mean corpuscular volume (MCV) (μm)</td>
<td></td>
</tr>
<tr>
<td>Children older than 11 years and adults &lt;82</td>
<td>A reliable, but late indicator of iron deficiency</td>
</tr>
<tr>
<td>Reticulocyte haemoglobin content (CHr) (pg)</td>
<td>Low values can also be due to thalassaemia</td>
</tr>
<tr>
<td>In adults ≥28.0</td>
<td>Wider use is limited because it can only be measured on a few models of analyser</td>
</tr>
<tr>
<td>Erythrocyte zinc protoporphyrin (ZPP) (μmol/mol haem)</td>
<td></td>
</tr>
<tr>
<td>5 years or younger &gt;70</td>
<td>It can be measured directly on a drop of blood with a portable haematofluorometer</td>
</tr>
<tr>
<td>Children older than 5 years &gt;80</td>
<td>A useful screening test in field surveys, particularly in children, in whom uncomplicated iron deficiency is the primary cause of anaemia</td>
</tr>
<tr>
<td>Children older than 5 years on washed red cells &gt;60</td>
<td>Red cells should be washed before measurement because circulating factors, including serum bilirubin, can spuriously increase values</td>
</tr>
<tr>
<td>Transferrin saturation &lt;16%</td>
<td>Lead poisoning can increase values, particularly in urban and industrial settings</td>
</tr>
<tr>
<td>Serum ferritin (SF) (μg/L)</td>
<td></td>
</tr>
<tr>
<td>5 years or younger &lt;12</td>
<td>It is probably the most useful laboratory measure of iron status; a low value of SF is diagnostic of iron deficiency anaemia in a patient with anaemia</td>
</tr>
<tr>
<td>Children older than 5 years &lt;15</td>
<td>In healthy individuals, SF is directly proportional to iron stores: 1 μg/L SF corresponds to 8–10 mg body iron or 120 μg storage iron per kg bodyweight</td>
</tr>
<tr>
<td>In all age groups in the presence of infection &lt;30</td>
<td>As an acute-phase protein, SF increases independent of iron status by acute or chronic inflammation; it is also unreliable in patients with malignancy, hyperthyroidism, liver disease, or heavy alcohol intake</td>
</tr>
<tr>
<td>Serum transferrin receptor (sTfR)</td>
<td></td>
</tr>
<tr>
<td>Cutoff varies with assay, and with patient age and ethnic origin</td>
<td>Main determinants are the erythroid mass in the bone marrow and iron status; thus, sTfR is increased by enhanced erythropoiesis and iron deficiency; sTfR is not substantially affected by the acute-phase response, but it might be affected by malaria, age, and ethnicity</td>
</tr>
<tr>
<td>Its application limited by high cost of commercial assays and lack of an international standard</td>
<td></td>
</tr>
<tr>
<td>sTfR-to-SF ratio</td>
<td></td>
</tr>
<tr>
<td>This ratio is a quantitative estimate of total body iron; the logarithm of this ratio is directly proportional to the amount of stored iron in iron-replete patients and the tissue iron deficit in iron deficiency</td>
<td></td>
</tr>
<tr>
<td>In elderly people, this ratio might be more sensitive than other laboratory tests for iron deficiency</td>
<td></td>
</tr>
<tr>
<td>This ratio cannot be used in individuals with inflammation because SF might be high independent of iron stores</td>
<td></td>
</tr>
<tr>
<td>This ratio is assay specific; although it is only validated for adults, this ratio has been used in children</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Indicators of iron deficiency anaemia
Iron deficieny anaemia. In industrialised countries, childbearing age improved iron nutrition and reduced folic acid supplementation every week to women of studies supported by WHO in southeast Asia, iron and trimester. Iron supplementation during pregnancy is not aff ect prevalence of anaemia during the third in the USA, iron supplementation increased birthweight, iron-replete, non-anaemic low-income pregnant women that it improves maternal or fetal outcomes. However, in widespread advocated even though so far little evidence exists universal iron supplementation of pregnant women is advisable in developing countries, where women often anti-folate antimalarial drugs, and thereby contributed folic acid might have reduced the eff ectiveness of supplements were thought to be the cause, provision of oral iron supplements can be supplied every few days; this regimen might increase fractional iron absorption. In studies supported by WHO in southeast Asia, iron and folic acid supplementation every week to women of childbearing age improved iron nutrition and reduced iron deficieny anaemia. In industrialised countries, universal iron supplementation of pregnant women is widely advocated even though so far little evidence exists that improves maternal or fetal outcomes. However, in two controlled trials of prenatal iron supplementation in iron-replete, non-anaemic low-income pregnant women in the USA, iron supplementation increased birthweight, reduced incidence of preterm delivery, or both, but did not affect prevalence of anaemia during the third trimester. Iron supplementation during pregnancy is advisable in developing countries, where women often enter pregnancy with low iron stores.1 Untargeted iron supplementation in children in tropical countries, mainly in areas of high transmission of malaria, is associated with increased risk of serious infections. In a region of endemic malaria in east Africa, untargeted supplementation with iron (12.5 mg per day) and folic acid in preschool children increased risk of severe illness and death. Although iron supplements were thought to be the cause, provision of folic acid might have reduced the effectiveness of anti-folate antimalarial drugs, and thereby contributed to morbidity. A similar study in Nepal, which is a non-malarial area, showed no effects of iron and folic acid on infection-related morbidity. A recent WHO report stated that iron and folic acid supplementation should be targeted to children who are anaemic and at risk of iron deficiency, and concurrent protection from malaria and other infectious diseases should be provided.

Supplementation
Iron supplementation can be targeted to high-risk groups (eg, pregnant women), and can be cost effective, but the logistics of distribution and absence of compliance are major limitations. For oral supplementation, ferrous iron salts (ferrous sulphate and ferrous gluconate) are preferred because of their low cost and high bioavailability. Standard therapy for iron deficiency anaemia in adults is a 300-mg tablet of ferrous sulphate (60 mg of iron) three or four times per day. Although absorption is enhanced when given on an empty stomach, nausea and epigastric pain might develop. If these side-effects arise, lower doses between meals should be attempted, or iron should be provided with meals, although food reduces absorption of medicinal iron by about two-thirds. Alternatively, oral iron supplements can be supplied every few days; this regimen might increase fractional iron absorption. In studies supported by WHO in southeast Asia, iron and folic acid supplementation every week to women of childbearing age improved iron nutrition and reduced iron deficieny anaemia. In industrialised countries, universal iron supplementation of pregnant women is widely advocated even though so far little evidence exists that improves maternal or fetal outcomes. However, in two controlled trials of prenatal iron supplementation in iron-replete, non-anaemic low-income pregnant women in the USA, iron supplementation increased birthweight, reduced incidence of preterm delivery, or both, but did not affect prevalence of anaemia during the third trimester. Iron supplementation during pregnancy is advisable in developing countries, where women often enter pregnancy with low iron stores.

Strategies
Three main strategies for correcting iron deficiency in populations exist, alone or in combination: education combined with dietary modification or diversifi cation, or both, to improve iron intake and bioavailability; iron supplementation (provision of iron, usually in higher doses, without food); and iron fortifi cation of foods. A new approach is biofortifi cation via plant breeding or genetic engineering. Although dietary modifi cation and diversifi cation is the most sustainable approach, change of dietary practices and preferences is difficult, and foods that provide highly bioavailable iron (such as meat) are expensive.

Fortification
Iron fortifi cation is probably the most practical, sustainable, and cost-effective long-term solution to control iron deficiency at the national level. Overall cost-effectiveness for iron fortifi cation is estimated to be $66–70 per DALY averted. Fortifi cation of foods with iron is more diffi cult than it is with other nutrients, such as iodine in salt and vitamin A in cooking oil. The most bioavailable iron compounds are soluble in water or diluted acid, but often react with other food components to cause off-fl avours, and colour changes, fat oxidation, or both. Thus, less soluble forms of iron, although less well absorbed, are often chosen for fortifi cation to avoid unwanted sensory changes. Fortifi cation with low iron doses is more similar to the physiological environment than is supplementation and might be the safest intervention. Iron fortifi cation of milk or cereals does not increase infection-related morbidity in children younger than 18 months. In an analysis of four studies of infants receiving iron-fortifi ed foods, the regimen did not cause visible adverse eff ects and signifi cantly protected against the development of respiratory tract infections (incidence rate ratio 0.92, 95% CI 0.86–0.98; p=0.02).

Industrialised countries
Although little direct evidence exists, the reduction in occurrence of iron deficiency in young children in industrialised countries has been attributed to iron fortifi cation of infant formulas and weaning foods. Iron-fortifi ed foods distributed through the Special Suplemental Nutrition Program for Women, Infants, and Children (WIC) have probably contributed to the fall of iron deficiency in underprivileged preschool children in the USA. At present, the low frequency of iron deficiency anaemia in adolescent and young women in the USA might be at least partly due to consumption of iron-fortifi ed wheat fi our, although other factors, including open-market fortifi cation of food products, and use of vitamin and mineral supplements, have also had a role. More-specifi c evidence is provided by retrospective studies from Sweden that reported decrease of iron intake and increase of iron deficiency in young women since iron fortifi cation of wheat fi our was discontinued in 1994. By contrast, fi ndings from Denmark, where iron fortifi cation of wheat fi our was discontinued in 1987, suggest no change in the frequency of iron deficiency in adults older than 40 years but the data might have been confounded by the effects of increasing bodyweight, alcohol consumption, or both, contributing to increased values or serum ferritin.
Iron fortification efforts have been accelerated by the fortification of soy sauce in China, fish sauce in Vietnam, fortification programmes in 14 countries, including iron fortification programmes in developing countries. GAIN has awarded $38 million in grants to food fortification programmes, funded mainly by the Bill & Melinda Gates Foundation, development agencies, and the private sector through the Global Alliance for Improved Nutrition (GAIN), an alliance of United Nation agencies, national governments, development agencies, and the private sector. GAIN has awarded $38 million in grants to food fortification programmes in 14 countries, including iron fortification of staple cereal flours and is the only soluble iron compound that does not precipitate peptides in fish and soy sauces. Use of micronised ferric pyrophosphate, a white-coloured iron compound with good bioavailability, has allowed successful fortification of colour-sensitive food vehicles, such as low-salt in Africa and rice in India. A micronised, dispersible ferric pyrophosphate and ferrous bisglycinate, an aminoacid chelate, are iron fortificants particularly useful for liquid products. Infants and young children in developing countries are at high risk of iron deficiency and might not be reached by universal fortification programmes. Chile has shown convincing evidence of the benefit of targeted fortification in women, yet several factors contribute to the failure of mass fortification. Poor programme control and enforcement, failure to detect effectiveness, and failure to determine the effectiveness of iron fortification programmes in developing countries are generally recommended for women of childbearing age and postmenopausal women. To date, no clear indication of efficacy of iron fortification in developing countries where the risk of developing iron deficiency is high for all groups other than adult men and postmenopausal women. Up to now, no clear indication of efficacy of iron fortification in developing countries exists, because of several factors (panel 1). However, recent studies have shown convincingly that iron fortification can be effective. The iron compound and type of fortification should be chosen on the basis of the fortification vehicle, iron requirements of the target population, and iron bioavailability of the local diet (panel 2). Efficacy should be monitored with measurements of serum ferritin and, when possible, serum transferrin receptor, in addition to haemoglobin. Failure to define iron status with specific indicators clearly impairs iron metabolism and erythropoiesis (eg, malaria) and type of fortification should be chosen on the basis of the fortification vehicle, iron requirements of the target population, and iron bioavailability of the local diet (panel 2). Efficacy should be monitored with measurements of serum ferritin and, when possible, serum transferrin receptor, in addition to haemoglobin. Findings from an efficacy trial in Thailand suggest that two forms of elemental iron, electrolytic iron and hydrogen-reduced iron, might be useful for fortification, but their bioavailability is only 50–79% that of ferrous sulphate. Two other forms of reduced iron, carbon-monoxide-reduced and atomised iron, are poorly absorbed and unlikely to be useful for food fortification. A trial in Sri Lanka failed to show a reduction in anaemia occurrence after 2 years of fortification of low-extraction wheat flour with either electrolytic or reduced iron, but fortification was probably too low. Wheat flour fortification with ferrous sulphate in Chile at 30 mg/kg has probably contributed to a strong decrease in iron deficiency. Fortification of maize flour in South Africa with ferrous fumarate has shown effectiveness in lowering anaemia, and improving iron status and motor development of infants in poor settings. Clear guidelines on wheat flour fortification have recently been published. Sodium iron ethylenediaminetetraacetic acid (NaFeEDTA) has shown effectiveness as a fortificant in sugar in Guatemala, curry powder in South Africa, soy sauce in China, fish sauce in Vietnam, and maize flour in Kenya. NaFeEDTA is absorbed 2–3 times more than ferrous sulphate from diets high in phytic acid, but is approved as a food additive only at 0.2 mg iron per kg bodyweight, which limits its usefulness as a fortificant for infants and children. NaFeEDTA does not promote fat oxidation in stored cereals and is the only soluble iron compound that does not precipitate peptides in fish and soy sauces. Use of micronised ground ferric pyrophosphate, a white-coloured iron compound with good bioavailability, has allowed successful fortification of colour-sensitive food vehicles, such as low-salt in Africa and rice in India. A micronised, dispersible ferric pyrophosphate and ferrous bisglycinate, an aminoacid chelate, are iron fortificants particularly useful for liquid products. Infants and young children in developing countries are at high risk of iron deficiency and might not be reached by universal fortification programmes. Chile has shown convincing evidence of the benefit of targeted fortification.
of powdered milk with ferrous sulphate and ascorbic acid, with frequency of anaemia decreasing from 27% to 9%. By contrast, distribution of a milk-based iron-fortified weaning food in Mexico for 1 year did not improve iron status, possibly because of the poor bioavailability of the reduced iron used as a fortificant. Complementary food supplements that are added to the infant’s food immediately before consumption have been developed. Three types of supplements have been tested: powders (sprinkles), crushable tablets, and fat-based spreads. Iron status was improved in Ghanaian infants with home fortification with powder containing encapsulated ferrous fumarate.

Biofortification

The variation in the iron content of cultivars of wheat, bean, cassava, maize, rice, and yam suggests that selective breeding might increase iron content of staple foods. However, although differences in iron content exist in wheat (25–56 mg/kg) and rice (7–23 mg/kg), most of the iron is removed during the milling process. Thus, to increase iron concentration in milled wheat up to 40 mg/kg, which is the fortification level commonly used in wheat flour, might be difficult. This problem was evident when the effectiveness of a rice cultivar high in iron was tested in a feeding trial in Filipino women consuming either the high-iron rice (3·21 mg/kg) or a local variety (0·57 mg/kg) for 9 months. Possibly because the high-iron rice added only an extra 1·5 mg of iron a day to the diet, no clear benefit of iron status was seen. Iron absorption from other cereals and legumes (many of which have high native iron content) is low because of their high contents of phytate and polyphenols. Donangelo and colleagues compared iron bioavailability from two varieties of red beans: an iron-rich genotype (containing 65% extra iron) and a low-density genotype. Only a small amount of iron was absorbed from both cultivars, probably because of their high phytate and polyphenol content. Decrease of the content of these inhibitors in high-iron cultivars might be needed to have a positive effect on human nutrition. Genotypes of maize, barley, and rice have been identified that are low-phytic-acid mutants, with phytic acid phosphorus content decreased by up to two-thirds compared with wild type. Although such reductions might improve iron absorption from diets containing small amounts of meat and ascorbic acid, phytic acid content might be needed to be lowered by more than 90% to increase iron absorption from the monotonous cereal-based diets seen in many developing countries.

Because of these limitations, genetic engineering might prove to be the most effective way to have a useful amount of absorbable iron in plant foods. iron content in rice can be increased two-to-three fold by introduction of the ferritin gene from soy bean or phaseolus vulgaris. Iron uptake from soils might be increased by introduction of a ferric reductase gene into plant root systems. To lower the phytic acid content of rice, Lucca and colleagues introduced a phytase from Aspergillus fumigatus that was developed to withstand food processing. Although phytase activity increased seven-fold, it proved to be unstable and was destroyed when rice was cooked. Overall, these studies suggest that iron content can be increased in staple foods by plant breeding, genetic engineering, or both.

Conclusions

Nutritional iron deficiency is still common in young women and children in developing countries where monotonous, plant-based diets provide low amounts of bioavailable iron. The high prevalence of iron deficiency in the developing world has substantial health and economic costs. However, more data are needed on the functional consequences of iron deficiency; for example, the effect of iron status on immune function and cognition in infants and children needs to be clarified. Continuing rapid advances in understanding the molecular mechanisms of iron absorption and metabolism might enable development of new strategies to combat iron deficiency. Although technical challenges limit the amount of bioavailable iron that can be added to many foods, evidence from controlled trials has shown that iron fortification can effectively control iron deficiency. Whether iron fortification can be successful in tropical areas without concurrent control of malaria and hookworm infections remains to be seen. Specific laboratory measures of iron status—eg, serum ferritin, sTfR, and zinc protoporphyrin—should be used to assess the need for fortification and for monitoring. Because of findings showing the risks of untargeted iron supplementation in young children, development of new strategies are urgently needed to provide additional dietary iron to susceptible infants and young children in developing countries who might not be reached by universal fortification programmes. New methods to enhance native iron content of plant-based staple foods are also needed. Selective plant breeding and genetic engineering are promising new approaches to improve dietary iron bioavailability; however, a major challenge is to show that they can increase iron content to nutritionally useful levels and that the additional iron is bioavailable.

Conflict of interest statement

We declare that we have no conflict of interest.

References

Seminar

33 Hurrell RF. How to ensure adequate iron absorption from iron-fortified food. Nutr Rev 2002; 60: 57–58.


130 Briend A, Solomons NW. The evolving applications of spreads as a FOODlet for improving the diets of infants and young children. Food Nutr Bull 2003; 24: S34–S38.


144 Lonnerdal B. Genetically modified plants for improved trace element nutrition. J Nutr 2003; 133 (suppl 1): 1490S–93S.


Health and Human Rights 2

Do human rights matter to health?

Jerome Amr Singh, Michelle Govender, Edward J Mills

Legal instruments and litigation as a way to enforce the rights to life and to health is a relatively new strategy that is increasingly common. We show how legal measures have been used to attain health and human rights with case examples from India and South Africa that resulted in large public-health benefits.

The adoption of the Universal Declaration of Human Rights by the UN in 1948 gave hope to the most vulnerable and oppressed people worldwide. Although not legally binding, this declaration was designed to inspire a culture of respect for human rights throughout the world and signalled a commitment to human development. Although the right to health could arguably be included in article 25 of the Universal Declaration of Human Rights, it was explicitly included in article 12 (table) of the legally binding International Covenant on Economic, Social, and Cultural rights in December, 1966. This covenant was optimistically expected to herald improved health-care delivery in participating countries once it came into force in January, 1976. Health-related demands from activists were therefore no longer empty requests but a legitimate and enforceable right.

Some countries have since established the right to health or analogous rights in their national constitutions; others have interpreted this right in non-binding policies. However, despite their ratification of the covenant, many countries have failed to honour its stipulations. Unsurprisingly, this disregard has resulted in questions about the value and effect of human rights on health. Yet, although many nations fail to honour their legal or moral obligations to health rights, several countries, including two notable developing countries, India and South Africa, show how respect for, and promotion of human rights can lead to improved health outcomes.

Although we will later briefly outline how human rights have brought about health reforms in civil-law jurisdictions such as Argentina and Ecuador, in this paper we focus on India and South Africa, both common-law jurisdictions. India and South Africa are the best known examples of how human rights matter to health. India is the most renowned example of a country in which the courts have directed health reforms, even in the absence of a codified right to health, and South Africa is the best known example of a country in which an explicit codified right to health has prompted health reforms despite the country’s failure to ratify the covenant on economic, social, and cultural rights.

Despite the differences between these two countries—South Africa is more racially diverse than India, India’s population is substantially larger than that of South Africa, and South Africa has a substantially higher gross domestic product per head than does India—when they became democracies, they both inherited fragmented and disparate health systems needing urgent reform (a factor common to many developing countries). Both countries have had effective public-interest litigation that led to health reforms, and are increasingly assuming leadership roles for worldwide health.

As discussed previously, health and human rights is both a social and a medical movement in addition to a legal practice. For people with health-related training, we aim to provide some insight into how appropriate documentation of health problems coupled with epidemiological and medical research can prove human rights inadequacies, in a legal setting. Although medical workers generally do not need knowledge of human rights and law, their ethical duties sometimes need them to assume a role as advocate on behalf of patients and communities, and we feel that knowledge of this subject matter will better equip them for such a role.

Common law jurisdictions

India

Despite India’s status as the world’s largest democracy and one of the fastest growing economies, its successive governments have been slow to fulfil the right to health. Despite this country’s high burden of disease, it spent only 4%-8% of its gross domestic product on health care in 2003. The nature of India’s constitution’ might have

<table>
<thead>
<tr>
<th>International instrument</th>
<th>Wording</th>
<th>Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1948 Universal Declaration of Human Rights, article 25</td>
<td>“Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control.”</td>
<td>Legally non-binding on signatory states (but arguably morally binding)</td>
</tr>
<tr>
<td>1966 International Covenant on Cultural, Economic, and Social Rights, article 12</td>
<td>“1. The States Parties to the present Covenant recognize the right of everyone to the enjoyment of the highest attainable standard of physical and mental health. 2. The steps to be taken by the States Parties to the present Covenant to achieve the full realization of this right shall include those necessary for: (a) The provision for the reduction of the stillbirth-rate and of infant mortality and for the healthy development of the child; (b) The improvement of all aspects of environmental and industrial hygiene; (c) The prevention, treatment and control of epidemic, endemic, occupational and other diseases; (d) The creation of conditions which would assure to all medical service and medical attention in the event of sickness.”</td>
<td>Legally binding for ratifying states and theoretically enforceable in domestic courts</td>
</tr>
</tbody>
</table>

Table: Important international human rights instruments of law that are relevant to health
The printed journal includes an image merely for illustration

Child labourer in pencil factory, India

corresponded to government indifference to public-health problems; although its non-enforceable chapter IV contains directive principles for policy that are intended as guidelines for minimum standards of living and for the ways in which health-related rights should be fulfilled, it does not include an explicit right to health.

Article 42 of the directive principles, which pertains to just and humane conditions of work and maternity relief, states: “The State shall make provision for securing just and humane conditions of work and for maternity relief.” Moreover, article 47 says: “The State shall regard the raising of the level of nutrition and the standard of living of its people and the improvement of public health as among its primary duties and, in particular, the State shall endeavour to bring about prohibition of the consumption, except for medicinal purposes, of intoxicating drinks and of drugs which are injurious to health.” The non-binding nature of these principles arguably allowed for complacency and lethargy by Indian policymakers and consequently encouraged systemic violations of health-related rights in that country. However, in the past two decades, robust activism and public-interest litigation, coupled with creative judicial interpretation have successfully challenged indifferent Indian health authorities, and brought relief to many people.

Public-interest litigation in India has centered on the fundamental right to life, article 21 of the Indian Constitution, which guarantees an individual’s right to life and is enforceable in court. Although this article does not contain an explicit and enforceable right to health, in a landmark case in 1980 regarding deaths due to accumulation of soot in the lungs of young workers in state-run pencil factories, the court ordered the government to ensure installation of safety measures in the factories. A year later the Indian Supreme Court further signalled its intention to make the right to health enforceable by inference from the right to life. It ruled that the right to life includes:

“the right to live with human dignity and all that goes along with it, namely, the bare necessities of life such as adequate nutrition, clothing and shelter and facilities for reading, writing, and expressing oneself in diverse forms, freely moving about and mixing and conningling with fellow human beings. Every act which offends against or impairs human dignity would constitute deprivation pro tanto of this right to live and it would have to be in accordance with reasonable, fair and just procedure established by law which stands the test of other fundamental rights. The magnitude and components of this right would depend upon the extent of economic development of the country, but it must, in any view of the matter, include the bare necessities of life and also the right to carry on such functions and activities as constitute the bare minimum expression of the human self.”

In a subsequent case in 1984, dealing with deplorable working conditions of bonded labourers, the court went further and issued detailed directives for how the state should fulfil its constitutional obligations to the labourers by interpreting the right to life widely to include the right to live with dignity. In the 1995 case of Consumer Education and Research Centre versus Union of India, the court considered the plight of workers in the asbestos industry and ruled that “social security, just and humane conditions of work and leisure to workmen are as a part of [the worker’s] meaningful right to life…. Noting that years of exposure to asbestos could result in asbestosis, the court judged that the fundamental right to health and medical care should continue after retirement. Importantly, the court noted that directions could be issued to the state or to private employers to protect the environment, prevent pollution in the workplace to safeguard the health of workers, or preserve free and unpolluted water for the safety and health of all people.

In the case of Kirloskar Brothers Limited versus Employee’s State Insurance Corporation, the Supreme Court ruled that the right to health is the fundamental right of workers and is equally binding against the state and the private sector. These judgments show the fulfilment of article 12(2)(b) improvement of industrial hygiene, and article 12(2)(c), control of occupational diseases, of the international covenant on economic, social, and cultural rights, and cumulatively have resulted in improved health and working conditions for many Indian people.

In Parmahand Katara versus Union of India and others, the Supreme Court held that physicians are obliged to preserve life irrespective of a patient’s innocence or guilt in an alleged crime. The adequacy of health services has also been brought before the Indian judiciary. In Paschim Baga Khet Mazoor Samiti versus State of West Bengal, the court was asked to establish whether non-availability of services in government health
centres amounted to a violation of article 21. In this case the claimant was refused treatment at eight state-run medical institutions in succession because of no availability of beds or insufficient technical capacity and eventually got treatment in the private sector, at great personal expense. In awarding compensation to the claimant, the court ruled that the right to emergency medical care was a core component of the right to health, which in turn was an integral part of the right to life. As a result of this case, state hospitals in India and the medical workers employed therein are now obliged to provide timely medical treatment to people in need. Similarly, in Mahendra Pratap Singh versus State of Orissa, the court ruled that the Indian government’s failure to open a health-care centre in a village amounted to a violation of the right to life, and by extension, the right to health.

By including the right to health in the right to life the Indian judiciary has also, among other orders upheld the state’s obligation to maintain health services; given the High Court duty to monitor the conditions of mentally ill and insane women and children in prisons; control pollution hazards; ban hazardous drugs; ban inhuman conditions in care homes; protect the health rights of mentally ill patients subjected to inhumane conditions; and prohibit passive smoking in public places. The Indian courts have also brought access to clean drinking water, and stopped animals straying onto public roads, thereby increasing road safety. Most recently, the courts have turned their attention to HIV/AIDS, specifically, people experiencing discrimination as a result of their HIV/AIDS status, and those at risk of HIV infection as a result of non-disclosure of their partner’s HIV status by health workers (the court ruled that the interests of third parties overrode privacy concerns of the patient). These two judgments could be important in the years to come. According to UNAIDS estimates, India has the world’s third highest number of HIV-infected people, after South Africa and Nigeria, thus such landmark rulings will undoubtedly play a crucial part in shaping the government’s response to this growing health crisis.

India’s remarkable range of judicial cases show that although India does not have an explicit and binding right to health in its constitution, its judiciary is using creative reasoning to force the government to fulfil this right. Cumulatively, this approach has brought relief from suffering to many people and shows that a rights-based approach can have great effects on human health in a functional democracy. Hopefully, the Indian government will assume greater or sole responsibility for tackling the country’s neglected diseases such as tuberculosis, lymphatic filariasis, leishmaniasis, and leprosy; a task it disproportionately delegates to non-governmental organisations at present. A human rights-based approach to health in South Africa has similarly affected many people in slightly more than a decade of democracy, although many health-related challenges remain.

South Africa

Although millions of South African people still experience discrimination and deprivation because of the legacy of apartheid, today South Africa is classified by the World Bank as an upper middle-income country, and it spends about 8.5% of its gross domestic product on health care, a higher proportion than many other countries. Despite this high expenditure, substantial health disparities still persist between the country’s different racial groups. Most poor South Africans are black people who rely on state health facilities that are unable to give a quality service. Only people who can afford the expensive services of the private health sector can access high-quality medical care. Health-care reforms are taking place slowly, as a result of obligations imposed on the state by the bill of rights in the South African constitution. The country’s widely celebrated bill of rights was first included in the interim constitution, and was later included in the country’s slightly revised final constitution. It contains several provisions pertaining to health care.

Section 27 of South Africa’s constitution addresses the right of access to health care, food, water, and social security. The legislature grouped these rights together and recognised, like the Indian judiciary, that such rights are linked and all contribute to the overall wellbeing of an individual. However, contrary to India, all rights in the South African constitution are enforceable and binding to the state. As a result, since 1994, the country introduced many new policies including: free primary health-care services for all citizens; free health care for children under the age of 6 years, pregnant women, and disabled people; a detailed national drug policy with amendments to legislation to reduce drug prices; legislation to enable access to private health care; a patient’s rights charter; free housing for homeless people (although a large backlog still persists); and free basic water (adequate supply of clean water sufficient for a basic standard of personal and domestic hygiene). The number of people accessing social security grants has increased from 2 million in 1994 to more than 10 million in 2005. The bill of rights has not only inspired substantial reforms in social security and health policies, but it has also given South Africans a way to accomplish such reforms. These reforms have had, or could have a great effect on the lives of many people, particularly those living with HIV, the country’s biggest health threat. Selected landmark cases illustrate the effect of litigation on health.

After groundbreaking constitutional court rulings on the right to emergency medical treatment and the right to housing, a civil-society AIDS advocacy group the Treatment Action Campaign (TAC), brought an action against the South African government to compel it to provide nevirapine (an antiretroviral drug), to pregnant
HIV-positive women by arguing the child’s right to health. This matter culminated in the 2002 constitutional court case, Minister of Health and others versus TAC and others.\(^a\) The government had restricted the provision of nevirapine to only eighteen test sites, claiming it was doing research to assess the feasibility of using this drug. The TAC demanded that this programme be instituted nationwide so that all women and children could benefit from treatment.

The court ruled that the government’s policy was unreasonable in two respects: first that access was restricted to research and training sites, because such restrictions failed to address the needs of mothers and their newborn babies who could not access these sites; and second, that it did not allow nevirapine to be given elsewhere in the public health system even though the system had capacity to deliver the drug, and its use was medically indicated. The court also ruled that although the government was unable to provide all medical services to the general public, health policy should be reformed to meet constitutional obligations (the right to health). Recognising that this ruling did not mean that everyone could immediately claim access to such treatment, the court judged that access to treatment to prevent mother-to-child transmission should be the ultimate aim. It ruled that every effort had to be made to reach this goal as soon as reasonably possible.

This ruling is consistent with the court’s previous ruling on the right to housing,\(^b\) in which it judged that any government programme that excludes a substantial segment of society cannot be reasonable. By addressing the health needs of women and children with HIV, who are mainly black, the court helped the most historically marginalised segment of South African society in a realistic and reasonable manner. Although further pressure was needed for government to act on the judgment, nevirapine is now freely available to thousands of pregnant mothers at government treatment facilities throughout the country. Activists note, though, that the government does not adequately monitor the success of the programme, and that disparities exist between provinces in their access to the intervention and in the standard of care provided.\(^c\)

Access to health care is meaningless if most South African people cannot afford essential medication, thus the South African legislature made changes to the medicines and related substances control amendment in 1997\(^d\) and 2002\(^e\) that allow for the supply of affordable medicines in specific circumstances, and for only specific licensed individuals to dispense medicines. Although these measures are laudable because they will probably make essential medication affordable to poor people (fulfilling the right to health for these individuals), they sparked outrage among doctors, pharmacists, and stakeholders in the pharmaceutical industry. In a high profile case in 1998\(^f\) that was watched closely by the international community, the pharmaceutical industry challenged the government on grounds of intellectual property, regarding provisions of the Medicines and Related Substances Control Act authorising parallel imports and generic substitution of brand name drugs.\(^g\) After much international pressure from lobby groups, and facing the realistic possibility of losing the case because of the right to health, the pharmaceutical companies withdrew their case against the government within a few days of the start of the trial. This case has since been hailed internationally as a milestone victory against the pharmaceutical industry.

The right to health has also been used to provoke health reform in the private sector without using the court system. In September, 2002, the AIDS Law Project, a human-rights advocacy organisation with a focus on HIV/AIDS in South Africa, lodged a complaint on behalf of several organisations with the Competition Commission of South Africa protesting against the high prices of drugs for the treatment of HIV/AIDS. Essentially, the complaint alleged that two large drug companies, GlaxoSmithKline and Boehringer Ingelheim were charging excessive prices for their antiretroviral drugs, even if the costs of research and development, higher profits, licensing fees, and incentives were taken into account. According to the complainants, these high prices resulted in the drugs being inaccessible to the general public and therefore led to the “premature, predictable and avoidable deaths”\(^h\) of thousands of people with HIV/AIDS. They argued that the right to life should be placed before profiteering.

In October, 2003, the Competition Commission, a statutory body, ruled that it had obtained enough evidence to support a referral to the Competition Tribunal for adjudication on the basis of prohibited excessive pricing, and of failure to license generic manufacturers in return for payment of a reasonable royalty.\(^i\) As a result, both GlaxoSmithKline and Boehringer Ingelheim agreed to a settlement in December, 2003, and the complaint was withdrawn. The companies agreed to the following: to grant more licences to generic companies to manufacture or import specific generic antiretroviral drugs; licensees could sell their products to both the public and private sectors, with sales being subject to a maximum royalty rate of 5%; and licensees that manufacture antiretroviral drugs in South Africa would be entitled to export their products to all countries in sub-Saharan Africa. This settlement resulted in a substantial reduction in the price of antiretroviral drugs in South Africa, and by extension, in other African countries that import drugs from South Africa. Moreover, since this landmark settlement, antiretroviral drug prices have continued to decrease, thus antiretroviral drug treatment is much more accessible to the general public. Although additional concessions need to be made to further reduce the cost of antiretroviral drugs, the settlement reached with GlaxoSmithKline and Boehringer Ingelheim was sup-
ported by the right to access to health and has had a great effect on such negotiations, and will continue to do so.

As in India, the courts in South Africa have also had to address the rights of prisoners to access health care. In 1996 the Cape High Court ordered that prisons in Cape province: treat the status of all prisoners with HIV/AIDS as confidential; protect prisoners from stigma associated with their HIV status or sexual orientation; provide prisoners with condoms; make treatment available to prisoners with HIV/AIDS; and only test prisoners for HIV with their informed consent.50 However, despite this judgment and others that have ruled in favour of prisoner health rights, access to antiretroviral treatment has not been made available uniformly in prisons around the country, leading to further court action. In 2005, the AIDS Law Project became aware that HIV-positive prisoners at Westville Prison in Durban, South Africa, were unable to access lifesaving antiretroviral treatment despite qualifying for treatment in terms of national guidelines. After failed attempts to remedy the situation with government, the AIDS Law Project brought urgent legal action on behalf of the TAC and prisoners against the government to compel it to provide antiretroviral treatment to HIV-positive prisoners.48

The AIDS Law Project argued that the South African constitution gives prisoners the right to receive adequate medical treatment at the expense of the state. Moreover, the country’s Operational plan for comprehensive HIV and AIDS care, management and treatment for South Africa,49 published by the Department of Health in 2004, makes clear reference to the right of prisoners to access antiretroviral treatment at accredited public-health facilities. In June, 2006, the Durban High Court ruled in favour of the prisoners and ordered the Department of Correctional Services to begin immediate treatment of the prisoners.49 The court gave the government two weeks to present its plan for provision of treatment to prisoners at Durban Prison although two further court orders were needed before the government took action. After repeated defeats in court and sustained pressure from both local and international civil society, in October, 2006, the South African government announced a radical change in its approach to the country’s worsening HIV/AIDS crisis.51 South Africa’s Deputy President, Phumzile Mlambo-Ngcuka, was appointed as Chair of the South African National AIDS Council. Unlike the Minister for Health, who repeatedly clashed with the TAC and openly championed untested traditional remedies for HIV/AIDS, the Deputy President promptly met with the TAC and openly endorsed antiretroviral treatment. Moreover, in a thinly-veiled rebuke of both South Africa’s President and Health Minister, the Deputy Health Minister, Nozizwe Madlalala-Routledge acknowledged the weakness of government leadership on HIV/AIDS despite its clear fiscal commitments to that end.52 This substantial policy change seriously the weight of its human rights duties and obligations and proactively addressing the health needs of people living with HIV. In addition to the cases discussed here, the courts in South Africa have used the constitutional right to health and related rights to make landmark rulings pertaining to reproductive freedom,53 discrimination in access to health insurance because of sexual orientation,54 and the right to water.55

Civil law jurisdictions

Although many of the most famous cases about health and human rights were undertaken in the common law countries of India and South Africa, the enforcement of human rights has also instigated health reforms in civil law jurisdictions. Hogerzeii and colleagues55 showed empirically that access to essential medicines in the fulfilment of the right to health had been enforced by courts in several low-income and medium-income countries, mostly in Latin America.

Latin America

Although Argentina, like India, does not have the right to health written in its constitution, the case in 1998 of Mariela Viceconte versus Ministry of Health6 is one of the world’s leading cases for how human rights can be used to challenge a state’s inadequate response to serious pandemic disease. In this case the plaintiffs, Mariela Viceconte and others, argued that at least 3.5 million people living in a region affected by Argentine haemorrhagic fever did not have adequate access to preventive medical services necessary to guard against the infection. The government argued that it did not have the means to do a massive immunisation campaign because of inadequate supplies of vaccine. The plaintiff argued that in instances in which a state is facing a substantial health crisis threatening many lives, its legal obligation to act in terms of international law (including the international covenant on economic, social, and cultural rights, which Argentina had ratified) was especially strong. She requested that the court order the Argentine government to take protective measures against the disease, to produce the appropriate vaccine, and to rehabilitate environments that were reservoirs for the disease. The court cited international law, including the right to health outlined in the international covenant on economic, social, and cultural rights and the incorporation of these treaties into domestic law, in its ruling that the government was legally obliged to intervene, and made the Ministers of Health and Economy personally liable for production of the vaccine within a specified time schedule.6 Although further litigation and action by the claimants was necessary to enforce the ruling, the case is regarded as important because:

“it reaffirmed the judicial process as a method for enabling ordinary citizens to challenge state agencies regarding the merit of health policies, saw the direct
application by a domestic court of international standards on the right to health...and affirmed the role of the state as guarantor of the right to health in the event that the private sector is unable or (more likely) unwilling to provide the necessary services.”

As a result of this case the Argentine government developed a social plan to deliver basic medicines to those in need within 5 years of the ruling. Another noteworthy case about the right to health in Latin America is the 2004 case, Mendoza and Others versus Ministry of Public Health in Ecuador. This country’s constitutional court referred to the right to health in international instruments that Ecuador had ratified and which had been added to the constitution, when it ruled that the Ministry of Health had failed to honour its obligation to right by suspending an HIV treatment programme. Although respect for the rule of law and judicial independence has recently been compromised in Argentina60 (the Argentine Constitution stipulates that the Council of the Judiciary should have a balance of legislators, judges, lawyers, and academics, which was thought to be essential to shelter judges from sudden shifts in politics, however in October, 2006, Argentina’s Congress passed a law reducing the council from 20 to 13 members and increased the proportion of politicians so they outnumber the experts on the Council, thereby opening the judicial appointment process to political manipulation) and Ecuador61 (on Dec 8, 2005, Ecuador’s President dismissed 27 of the Supreme Court’s 31 judges and replaced them with his political allies; on April 15, 2006, the President fired the entire Supreme Court and declared a state of emergency) the cases described here show that the right to health can be equally effective in provoking health reform in both civil law and common law jurisdictions.

Conclusion

The repeated successes in India and South Africa of human rights prompting health reforms can be attributed largely to at least four factors: intense and sustained pressure by strong and competent civil society organisations in those countries; fairly independent, competent, and progressive judicial authorities; governments having respect for the rule of law; and use of medical evidence to support legal arguments. Unfortunately, many countries do not enjoy this situation (now including Argentina and Ecuador). In such instances donors, international lobby groups, and democratic governments should support civil society organisations dedicated to health or democratic reforms or both that operate in oppressive environments, and urge such governments to initiate health reforms. They should pressure governments into promotion, protection, and fulfilment of the human rights of their citizens by citation of international law. Admittedly, this tactic is easier said than done. However, these measures were effective in South Africa in the apartheid era, and could enable the right to health in other oppressed nations. Although much work still needs to be done in India and South Africa to improve conditions for sick and destitute people, the lesson that these countries have taught and continue to teach the rest of the world is that human rights do matter and can positively affect the health of all people.

These countries represent only a handful of successes, and to believe that all litigations on health have been or will be successful would be incorrect. However, to expect that other countries can achieve similar results is reasonable. Hopefully, the examples described here will inspire stakeholders in countries that still do not respect human rights to pressure their respective governments into eventual respect for these rights and thus improve the health of their citizens.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgments

JAS receives financial support from the Centre for the AIDS Programme of Research in South Africa (CAPRISA), through the US National Institute of Allergy and Infectious Diseases, Department of Health and Human Services grant number U19AI13794. JAS also receives financial support from the University of Toronto Joint Centre for Bioethics and Department of Public Health Sciences through a grant from the Bill and Melinda Gates Foundation Grand Challenges in Global Health Initiative.

References

9. Francis Corale Mullin vs The Administrator, Union Territory of Delhi, 2 SCR 516, 1981.
12. Kirloskar Brothers Ltd. vs Employee’s State Insurance Corporation, 2 SCC 682, 1996.
Photoprotection

Stephan Lautenschlager, Hans Christian Wulf, Mark R Pittelkow

Photocarcinogenesis (weak)
Immunosuppression

Sun exposure is the main cause of photocarcinogenesis, photoageing, and photosensitivity; thus, photoprotection is an important issue. In a skin cancer prevention strategy, behavioural measures—e.g., wearing sun protective clothes and a hat and reducing sun exposure to a minimum—should be preferred to sunscreens. Often this solution is deemed to be unacceptable in our global, outdoor society, and sunscreens could become the predominant mode of sun protection for various societal reasons (e.g., healthiness of a tan, relaxation in the sun). The application of a liberal quantity of sunscreen has been shown to be by far the most important factor for effectiveness of the sunscreen, followed by the uniformity of application and the specific absorption spectrum of the agent used. The sunscreen market—crowded by numerous products—shows various differences worldwide. Nevertheless, sunscreens should not be abused in an attempt to increase time in the sun to a maximum. Controversies about safety of sunscreens and clinical recommendations are discussed.

Unprotected exposure to ultraviolet radiation is a major causal factor in the development of skin cancer. Non-melanoma skin cancers are initiated for the most part by chronic sunlight exposure and can be readily produced by experimental exposure to ultraviolet radiation in animal models. Ultraviolet B (UVB) is the major active terrestrial waveband region that causes direct photochemical damage to DNA, from which gene mutations arise. Unlike UVB, ultraviolet A (UVA) could have more indirect effects on DNA via the generation of reactive oxygen species. The most lethal of the skin cancers, cutaneous malignant melanoma, is more commonly associated with sporadic burning exposure to sunlight—especially early in life—but the wavelengths responsible have not been clearly identified. There are several indications that UVA might have an important role in the pathogenesis of melanoma. However, this involvement has recently been questioned, since only UVB induced melanoma in a transgenic mouse model. There is also mounting evidence that the clinical heterogeneity of malignant melanoma and the variable susceptibility of individuals to its development after exposure to ultraviolet radiation could be explained by specific genetic mutations and polymorphisms, respectively.

Ultraviolet radiation also causes skin ageing and photo-dermatoses; responses in human skin vary according to wavelength (table 1). However, the action spectrum for ultraviolet-induced tanning and erythema are almost identical. Indirect evidence suggests that UVA has a greater role in long-term sun damage (figure 1) than it does in acute effects such as sunburn or vitamin D synthesis, which are overwhelmingly attributable to UVB.

To reduce the deleterious effects of ultraviolet radiation to a minimum, public education on photoprotective measures should be promoted continually. In decreasing order of efficacy and lifestyle disruption, we discuss the following measures: complete avoidance of sun exposure, seeking shade at times when disease-inducing wavelengths are relatively intense, wearing clothing protective against ultraviolet radiation penetration, and the use of topical sunscreens to specifically prevent or reduce ultraviolet-induced cellular damage to a minimum.

Environmental photoprotection

Ozone (O₃) is a photoabsorbing molecule present mainly in the stratosphere between 10 and 50 km above the surface of the earth. It absorbs high quantities of shortwave UVB and all ultraviolet C (UVC) radiation, but only a little UVA. The ozone concentration varies naturally according to temperature, weather, latitude, and altitude. The atmosphere is thinner at higher altitudes, resulting in an increase of the intensity of ultraviolet radiation by 4% for every 300 metres of elevation. The average increase in UVB intensity per degree of latitude is about 3%. The time of day also affects the intensity of ultraviolet radiation. At the solar zenith, the sun’s rays pass through less of the
atmosphere than they do at any other time point; thus, the atmosphere absorbs substantially less ultraviolet radiation. In Denmark, an open prospective observational study showed that 50% of the total daily solar ultraviolet dose reaches the earth between 1200 and 1500 h. During the summer, the sun is higher in the sky, and less ultraviolet radiation is absorbed during its passage through the atmosphere.

Substances that deplete ozone (eg, chlorofluorocarbons) have a substantial effect on terrestrial ultraviolet exposure. Although ozone levels vary seasonally, stratospheric ozone levels have decreased annually since the 1970s, especially in the southern hemisphere, and have only recently begun to stabilise. Sunburn and photosensitivity could have been estimated that a 1% decrease in ozone levels is followed by a 1–2% increase in melanoma mortality. Although ozone levels vary seasonally, stratospheric ozone has a substantial effect on terrestrial ultraviolet exposure.

Although there is a recognised way of measuring protection from UVB, there is not a standardised method to measure the efficacy of UVA blocking. The Australian/New Zealand Standard, which has been in use since 1983, is based on in-vitro testing. This standard specifies that a “broad spectrum” claim can be made if the product fulfils one of the following criteria: either an 8-µm layer of the product does not transmit more than 10% of radiation between 320 and 360 nm or an 20-µm layer of the product does not transmit more than 1% of radiation between 320 and 360 nm. In addition to its use in Australia and New Zealand, this standard has been adopted in many European countries. Commonly used in-vivo methods are immediate pigment darkening, persistent pigment darkening, and the protection factor test method. Persistent pigment darkening is more reliable than immediate pigment darkening because pigmentation remains stable for 2–24 hours.

Photoprotective clothing
During the past two decades, it has been recognised that textiles are a reliable means of photoprotection. Australia has had a leading role in establishing skin cancer education programmes that have urged the use of clothing in conjunction with hats and sunscreens for ultraviolet protection. Nevertheless, several studies have shown that, by contrast with popular opinion, some textiles provide only limited ultraviolet protection. Wright and colleagues reported that the protection afforded by a light-coloured cotton shirt was equivalent to a sun-
excellent protection against the hazards of solar radiation, especially garments specifically manufactured to be ultraviolet protective. There are conflicting reports as to whether clothing can prevent melanocytic nevi, known to be a strong predictor of risk of subsequent cutaneous malignant melanoma. No effect was seen in adolescent twin study from the UK, whereas a recent analysis in Germany showed a protective effect of clothing in children. These results suggest that everyday clothing is protective, whereas shirts worn on the beach might not offer the same protection as everyday clothing or that the wearing of such items is unreliably reported.

**Sunscreens**

Several studies have shown a reduction in the number of actinic keratoses and squamous cell carcinomas, but not basal cell carcinomas, in careful and regular sunscreen users. Also, mathematical models based on a 3-year, randomised controlled trial of sunscreen use in children showed a 30–40% reduction in new nevi for users of SPF 30 sunscreen compared with control individuals who used the vehicle only. Sunscreen use has recently been shown to attenuate new nevus development on intermittently sun-exposed body sites for white school children, especially in children with freckles. The density of melanocytic nevi is known to be a strong predictor of the risk of subsequent cutaneous malignant melanoma. Furthermore, prevention of acute effects of exposure to ultraviolet radiation—eg, sunburn and formation of sunburn cells, cutaneous DNA damage, and immunosuppression—has been shown, as well as prevention of chronic effects including photoageing. However, a recent systematic review failed to show a beneficial effect in preventing malignant melanomas by sunscreen. Even at the beach, only a few people regularly use sunscreens systematically.

Active ingredients in sunscreens differ considerably worldwide (table 2). Even the maximum allowed concentrations of active agents show variation among national regulatory agencies. The list of permitted ultraviolet filters in cosmetic products in the EU has been updated regularly in the past two decades. Currently, there are 27 different sunscreens listed. 28 different sunscreens are approved in Australia and four additional agents are currently under review. By contrast, only 16 agents are listed in the latest Food and Drug Administration (FDA) monograph in the USA. Since 1978, the FDA has allowed only the addition of avobenzone, zinc oxide, and—in 2006—ecamsule to the list. Regulatory agencies in Europe and elsewhere treat sunscreens as cosmetics, where the regulatory approval process is faster. In the USA, sunscreen active ingredients are treated like drugs.

Historical aspects of the development of sunscreens have been reviewed recently. The most successful of the early 20th century sunscreens was certainly Ambre Solaire, containing benzyl salicylate, prepared by Eugene Schueller in 1935. Shortly afterwards Delial, containing benzylimidazole sulfonic acid, was introduced.

Today, topical sunscreens are divided into two broad categories: organic (formerly designated chemical) and inorganic (formerly designated physical) agents.

**Inorganic agents**

Inorganic agents (titanium dioxide and zinc oxide) reflect and scatter ultraviolet and visible radiation from a film of inert metal particles, which forms an opaque barrier. Depending on the particle size, there is less variability in the photoprotective properties, there is less variability in the photoprotective effect of inorganic agents when compared with organic sunscreens. Opaque inorganic sunscreens might give some protection against visible light-induced photosensitivity diseases and there are currently no documented reports of sensitisation reactions. However, inorganic sunscreens are often cosmetically unacceptable because of their opaque quality and occlusiveness. The higher refractive index of titanium dioxide compared with zinc oxide (2.6 vs 1.9) explains its whiter appearance and therefore the lower cosmetic acceptability. Recently, modern pharmaceutical approaches such as micronisation and encapsulation have allowed the development of high-quality inorganic sunscreens. Decreasing particle size to 10–50 nm (micronised form) results in less scattering of visible light, leading to a more cosmically acceptable product.

---

**Panel 2: Factors that influence ultraviolet protection of clothing**

**Factors that increase protection**
- Tightly woven fibres
- Thicker fabrics
- Denim, wool, synthetic materials (eg, polyester)
- Lax materials
- Dry materials
- Shrink after washing
- Treatment with a broad-spectrum ultraviolet absorber (eg, Tinosorb)
- Dark-colour fabrics
- Unbleached fabrics

**Factors that offer lower levels of protection**
- Loosely woven fibres
- Thinner fabrics
- Cotton, linen, acetate, rayon
- Stretched textiles
- Wet materials
- Hydration
- Water washing alone
- Light colour fabric
- Bleached fabrics
### UVB filters

<table>
<thead>
<tr>
<th>Synonyms, abbreviations, and trade names</th>
<th>Maximum concentration</th>
<th>Permitted in</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PABA derivatives</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Aminobenzoic acid</td>
<td>PABA</td>
<td>5%,* 15%†‡</td>
</tr>
<tr>
<td>Padimate O</td>
<td>2-Ethylhexyl 4-dimethylaminobenzate, octyldimethyl PABA, ED-PABA (OD-PABA)</td>
<td>8%</td>
</tr>
<tr>
<td>Ethoxylated ethyl 4-benzoic acid</td>
<td>PEG-PABA, Uvinul P25, Unipabol U17</td>
<td>10%</td>
</tr>
<tr>
<td>2,4,6-Trianilino-(p-carbo-2’-ethylyhexyl-1’-oxy)-1,3,5-triazine</td>
<td>Octyl triazone, ethylhexyl triazone, ET, Uvinul T150</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Cinnamates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethylhexyl methoxycinnamate</td>
<td>Octyl methoxycinnamate, EMC, (OMC), Escalo 557, Eusolex 2292, Neo Heliopan AV, Parsol, MCX</td>
<td>7.5–10%†</td>
</tr>
<tr>
<td>Cinoxate</td>
<td>2-Ethoxethyl p-methoxycinnamate</td>
<td>3%–6%‡</td>
</tr>
<tr>
<td>Isopentenyl-4-methoxycinnamate</td>
<td>Isopropyl 4-methoxycinnamate, IMC, Neo Heliopan E1000</td>
<td>10%</td>
</tr>
<tr>
<td><strong>Salicylates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Ethylhexyl salicylate</td>
<td>Octyl salicylate, octisalate, ES, (OS), Escalo 587, Neo Heliopan OS</td>
<td>5%</td>
</tr>
<tr>
<td>Homosalate</td>
<td>Homomethyl salicylate, HMS</td>
<td>10%–15%†‡</td>
</tr>
<tr>
<td>Trolamine salicylate</td>
<td>Triethanolamine salicylate</td>
<td>12%</td>
</tr>
<tr>
<td><strong>Camphors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benzylidene camphor sulfonic acid</td>
<td>BCSA, Mexoryl SD-20, Unisol S22</td>
<td>6%</td>
</tr>
<tr>
<td>Polymer of N-[2 and 4]-[(2-oxoborn-3-ylidene) methyl] benzyl acrylamide</td>
<td>Polymethylacrylamidobenzylidene camphor, PBC, Mexoryl SW</td>
<td>6%</td>
</tr>
<tr>
<td>N,N,N-Trimethyl-4-[(2-oxoborn-3-ylidenemethyl)lanilinium methyl sulphate</td>
<td>Camphor benzalkonium methosulfate, CBM, Mexoryl SK</td>
<td>6%</td>
</tr>
<tr>
<td>3’-(4’-Methylenylidene)-d-1 camphor</td>
<td>4-Methylbenzylidene camphor, MBC, Eusolex 6300, Neo Heliopan MBC</td>
<td>4%</td>
</tr>
<tr>
<td>3-Benzylidene camphor</td>
<td>3-Benzylidene camphor, BC, Mexoryl SD-20, Unisol S-22</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-cyano-3,3-diphenyl acrylic acid</td>
<td>2-ethyl hexyl ester, octocrylene, 2-ethylhexyl-2-cyano-3,3 diphenyl acrylate, Eusolex OCR, Neo Heliopan 303</td>
<td>10%</td>
</tr>
<tr>
<td>2-Phenybenzimidazole-S-sulfonic acid and its potassium, sodium and triethanolamine salts</td>
<td>Phenoxybenzimidazole sulfuric acid, PBSA, ensulizole, Eusolex 232, Neo Heliopan Hydro, Parsol HS</td>
<td>41%–8%*</td>
</tr>
</tbody>
</table>

### UVA filters

<table>
<thead>
<tr>
<th>Synonyms, abbreviations, and trade names</th>
<th>Maximum concentration</th>
<th>Permitted in</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benzophenones</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benzophenone-3§</td>
<td>Oxybenzone, BENZ-3, Eusolex 4360, Neo Heliopan BB, Uvinul M40, Escalo 567, UVAsorb METIC</td>
<td>61%–10%†‡</td>
</tr>
<tr>
<td>2-Hydroxy-4-methoxybenzenophene-5-sulfonic acid §</td>
<td>Sulisobenzone, BENZ-4, UVAsorb 55, Escalo 577, Uvinul MS 40</td>
<td>5%–10%†‡</td>
</tr>
<tr>
<td>2-Hydroxy-4-methoxybenzenophene-5-sulfonic acid §</td>
<td>Sulisobenzone, BENZ-4, UVAsorb 55, Escalo 577, Uvinul MS 40</td>
<td>5%–10%†‡</td>
</tr>
<tr>
<td>Dioxycinnamate</td>
<td>Benzophenone-8</td>
<td>3%</td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Menthyl anthranilate</td>
<td>Methyl-2-aminobenzoate, meradimate</td>
<td>5%</td>
</tr>
<tr>
<td>1-(4-tert-butylyphenyl)-3(4-methoxyphenyl)-propane-1,3-dione</td>
<td>Butyl methoxy dibenzoylmethane, avobenzone, BMDMB, Parsol 1789, Eusolex 9020</td>
<td>31%–5%†</td>
</tr>
<tr>
<td>2,2’-Methylene-bis-6-(2H-benzoizol-2-yl)-4-(tetramethylbutyl)-1,1,3,3-phenol</td>
<td>Methylene bis-benzotriazolyl tetrabutylphenol, MBIT, Tinosorb M</td>
<td>10%</td>
</tr>
<tr>
<td>Phenol-2-(2H-benzoizol-2-yl)-4-methyl-6(2-methyl-3-[1,3,3,3-tetramethyl-1-[(trimethylsilyloxy)disiloxanyl]propyl]</td>
<td>Drometrizole triisoxane, DTS, Mexoryl XL</td>
<td>15%</td>
</tr>
<tr>
<td>2,2’-(1,4-Phenylenebis-(1-H benzimidazol-4,6-disulfonic acid, monosodium salt §)</td>
<td>Disodium phenyl dibenzimidazol tosylsulfonate, bisimidazylate, DPDT, Neoheliopan AP</td>
<td>10%</td>
</tr>
<tr>
<td>Terephthalidilene dicamphor sulfonic acid</td>
<td>Ecamule, TDSA, Mexoryl SX</td>
<td>10%</td>
</tr>
<tr>
<td>4,4’-(1,1-dimethylthylamino)carbonylphenylamine-1,3,5-triazine-2,4-diylidimino]bis-[3’-ethylyhexyl]ester]</td>
<td>Diethoxyethylamidobenzalmalenate §</td>
<td>10%</td>
</tr>
<tr>
<td>Dimethico-diethylbenzalmalenate §</td>
<td>DDBM, Parsol SLX</td>
<td>10%</td>
</tr>
<tr>
<td>(1,3,5)-Trizene-2,4-bis[(4’-2-ethylhexyloxy)-2-hydroxy)-phenyl]-6-(4-methoxyphenyl)§</td>
<td>Bis-ethylhexyloxyphenol methoxyphenol triazine, bemotrizinol, BEMT, Tinosorb S</td>
<td>10%</td>
</tr>
</tbody>
</table>

### Inorganic absorbers

<table>
<thead>
<tr>
<th>Synonyms, abbreviations, and trade names</th>
<th>Maximum concentration</th>
<th>Permitted in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titanium dioxide § [Au: OK?]</td>
<td></td>
<td>25%</td>
</tr>
<tr>
<td>Zinc oxide § [Au: OK?]</td>
<td></td>
<td>25%</td>
</tr>
</tbody>
</table>


Table 2: Sunscreen agents permitted as active ingredients in listed products in Australia, the European Committee (EC), and the USA (Food and Drug Administration monograph)
Micronisation shifts the protective spectrum, via its property as an absorbing agent, towards shorter wavelengths. Microfine titanium dioxide and zinc oxide have been found to be highly protective against harmful ultraviolet rays and offer good protection against short-term UVB-induced immunomodulation in human trials.

Neither zinc oxide nor titanium dioxide have relevant skin irritating properties and sensitisation potential in human beings. Concerns related to the penetration of inorganic agents into the skin have been discussed, but in-vivo and in-vitro studies found no evidence of penetration of titanium dioxide and limited penetration only for zinc oxide, which is slightly soluble. Serum concentrations of zinc oxide after whole-body application were unchanged, and in-vitro studies have shown no penetration beyond the stratum corneum. Simple washing procedures are sufficient to remove microfine titanium dioxide from the skin. Microfine titanium dioxide has an absorption profile greater in UVB than does microfine zinc oxide; by contrast, zinc oxide was found to be more effective in UVA protection (up to 380 nm) than was microfine titanium dioxide. These effects vary largely due to the particle size of the substances. Microparticles tend to agglomerate and aggregate due to electrostatic effects, resulting in potentially greater loss in efficacy. Therefore, the micropigments have to be coated and kept in dispersion, which is still a major challenge for the cosmetic industry.

Sunscreens containing inorganic agents alone are generally recommended for children, because of their lack of penetration and subsequent degradation in the body, absence of photo-related effects (ie, phototoxicity), and no evidence of photogenotoxicity in vivo.

**Organic agents**

Organic sunscreens act by absorbing ultraviolet radiation. Ultraviolet radiation activates the agent’s electrons from the ground to an excited state. When returning to the stable condition, energy is emitted as insignificant quantities of warmth or fluorescent radiation. These agents are broadly divided into UVB, UVA, or broadband absorbers. To be effective, the filter should be stable photochemically in sunlight, dissolve or disperse easily and permanently in the vehicle, and remain in place after perspiration or swimming. Additionally, the agent should be non-toxic and not cause irritation or contact allergy. UVB absorbers have been commonly used worldwide for decades (table 2), whereas most UVA and broadband absorbers have been developed in recent years. Since a sunscreen has to protect against the entire ultraviolet spectrum, different filters have to be combined in the same product. The cinnamates (mostly 2-ethyl-4-methoxycinnamate; EMC) are by far the most popular UVB absorbers in the USA and Europe, and are used in combination with other UVB absorbers to achieve a high SPF. The second most popular filters in Europe during the recent past are the camphor derivatives, Salicylates and para-aminobenzoic acid (PABA) and its derivatives are among the oldest commercially available UVB filters and are still used worldwide.

The increasing need for broadband agents and improved photostability has led to the introduction of a new generation of filters, including methylene bis-benzotriazoyl tetrabenzylhexylbenzophenone (Tinosorb M) and bis-ethylhexylmethoxycamphor methoxyphenol triazine (Tinosorb S), both manufactured by Ciba Specialty Chemicals (Basel, Switzerland), as well as terephthalaldehydic dicamphor sulfonic acid (Mexoryl SX) and drometrizol trisiloxane (Mexoryl XL), produced by L’Oréal (Clichy, France). Mexoryl SX is a photostable broad-spectrum absorber and Mexoryl XL can absorb both UVB and UVA. Both protect against induction of pigmentation and show a synergistic effect when used in combination. Tinosorb S is an oil-soluble, highly photostable broad-spectrum ultraviolet filter with a good absorption in the UVA range, which can be used successfully to improve the photostability and efficiency of sunscreens containing avobenzone and EMC. Tinosorb M consists of microfine organic particles that are dispersed in the aqueous phase of the sunscreen emulsion, thus combining the benefits of an organic filter with those of an inorganic filter. Because tinosorb molecules are large, they are less likely to be absorbed through the skin. Absence of endocrine activity has been shown in vitro. The mexoryls and tinosorbs are not licensed in the USA and Japan. In Europe, the list of permitted ultraviolet filters recognised by the Council Directive of the European Committee has been regularly questioned. Currently five sunscreen agents are under review: benzophenone-3, camphor benzalkonium methosulfate, homosalate, PABA, and phenylbenzimidazole sulfonic acid.

The self-tanning product dihydroxyacetone has generally not been considered to provide sun protection. However, there is some evidence to suggest that it gives some protection, with an SPF of 2–3 and a durability of 5–6 days. The protective nature of the compound has been confirmed in a hairless mouse study that showed a delay of broad-spectrum ultraviolet carcinogenesis; the compound could therefore function as a base for other protection measures.

Organic and inorganic ultraviolet filter substances have been shown to act synergistically to increase the SPF. Inorganic agents increase the optical pathway of the photons in the topically applied absorbing formulation. In this way, more photons are absorbed, increasing the SPF.

**Controversies of sunscreens**

Recent controversy surrounding sunscreen efficacy and safety has stimulated a reassessment of their use and properties.
Efficacy

In terms of acute ultraviolet damage, actinic keratoses, and non-melanoma skin cancer, studies have shown a direct protective effect of sunscreen use in human beings. However, several studies have suggested that the use of sunscreens might be associated with increased nevus density, a strong predictor of risk of subsequent melanoma. These retrospective studies could have several shortcomings in controlling for the confounding effects of phenotype, prior sun exposure, and frequency of sunscreen use. A recent randomised controlled study in children showed attenuated new nevus development on intermittently sun-exposed body sites for white school children in Vancouver, Canada. However, whether sunscreens can reduce the risk for melanoma has not yet been proven definitively. A case-control study from southern Sweden of 571 patients with a first diagnosis of cutaneous malignant melanoma showed a significantly raised odds ratio for developing malignant melanoma after regular sunscreen use (odds ratio 1.8, 95% CI 1.1–2.9). A systematic review by Dennis and colleagues that examined 18 heterogeneous case-control studies failed to show any association between melanoma and sunscreen use. Different factors must be taken into account to explain the actual lack of proof of preventive effects. First, the median SPF of commonly used sunscreens before the early 1990s was 4–10, and these sunscreens incorporated active ultraviolet filters that were limited largely to the UVB waveband. Only by 1997 had median SPF risen to about 15. Second, higher protection has been anticipated to induce longer sun exposure by postponing warning signs such as sunburn, or by providing a false impression of safety in the sun. In this regard, two randomised trials in students during their holidays have shown that application of a high SPF prompted them to increase the duration of their sun exposure. A recent prospective study by Thieden and colleagues has shown that people use sunscreen when they know they are going to stay out for a long time and thus apply sunscreen as a tanning aid to avoid sunburn. Third, failure to apply sunscreen properly must be emphasised: sunscreens are frequently not applied before exposure or early after onset of exposure. Additionally, the SPF of a sunscreen is assessed after phototesting in vivo at an internationally agreed application thickness of 2 mg/cm². Several studies have shown that consumers apply much less than this, achieving only 10–25% of the protection expected from the product label. The uniformity of sunscreen application, failure to apply to all exposed skin, resistance to water immersion, and the number of applications per day are known to influence protection. Considering these factors, Diffey concludes that it is not surprising that case-control studies have failed to find any association between the use of sunscreens and the risk of melanoma, especially with such a limited effect of older sunscreens on modifying solar ultraviolet exposure. Therefore, to anticipate that newer formulations of sunscreens will lead to a benefit as a protective agent against melanoma is reasonable, although such a benefit might not be seen for several decades. However, their use in practice could be difficult, time consuming, and expensive, and they can, in general, be used only in addition to the more reliable measures of clothing protection and reduction of ultraviolet exposure during peak hours of solar radiation.

Safety

Adverse reactions from sunscreen ingredients—including allergic and irritant contact dermatitis, phototoxic and photoallergic reactions, contact urticaria, and even solitary cases of severe anaphylactic reactions—have been increasingly reported. Adverse reactions to sunscreens have been shown to occur in as many as 19% of individuals in one Australian study. Most of these adverse reactions were due to irritant rather than allergic reactions to either the sunscreen agents or the base. The prevalence of allergy to sunscreens in the general population is not known, but recent photopatch test results showed a low yield of positive reactions. Thus, despite the large increase in the use of ultraviolet filters over the past decade, the development of photoallergic reactions remains rare. Furthermore, most of the common ultraviolet filter photoallergens, including PABA, amyl dimethyl PABA, and benzophenone-10, are now rarely used in sunscreen manufacture; isopropyl dibenzoylmethane was voluntarily removed from the market in 1993. Currently, benzophenone-3 (a UVA filter) is the most common contact photoallergen still in widespread use. Other UVA filters—eg, avobenzone (Parsol 1789) and sulisobenzon (benzophenone-4)—and the UVB filters methylbenzylidene camphor, octyl methoxy-cinnamate, and ensulizole, are known to rarely induce contact allergic and photoallergic reactions. However, there is currently no evidence that allergic reactions represent a common clinical problem. Nevertheless, patients with photodermatoses such as polymorphous light eruption and chronic actinic dermatitis represent a group of patients at increased risk of developing photoallergy.
Gasparro and colleagues reviewed in-vitro and in-vivo studies of cytotoxicity and photogenotoxicity of sunscreens. Concerns from in-vitro studies that showed direct or indirect interactions from sunscreens—mainly PABA and its derivatives—with DNA following exposure to ultraviolet radiation could not be confirmed in vivo. Collectively, data from in-vivo studies seem to diminish, if not eliminate, photocarcinogenicity concerns, since sunscreens delay photocarcinogenesis in hairless mice. However, further in-vivo and in-vitro studies are needed to clarify mechanisms of immunosuppression, DNA repair, and mutagenesis.

Since sunscreens are increasingly included in diverse consumer products, questions regarding their long-term safety have been raised. 90% of all requisite vitamin D is formed within the skin through the action of ultraviolet radiation. Several studies have suggested a connection between vitamin D deficiency and over a dozen forms of cancer (eg, colon, breast, prostate) and recently also for malignant melanoma. Berwick and colleagues noted that subsequent mortality from melanoma was about half as high in those with signs of solar elastosis (assessed by means of a standardised physical examination) as in those without solar elastosis, indicating a possible link to vitamin D levels. These results should be interpreted cautiously since continued sun exposure would increase the risk of a second melanoma as well as squamous cell carcinoma. For 2004, the US economic burden due to vitamin D insufficiency from inadequate exposure to solar UVB irradiance, diet, and supplements is estimated to be $40–56 billion, whereas that for excess ultraviolet irradiance is estimated to be $6–7 billion; further research is required to confirm these estimates. However, clinical studies have shown that long-term use of sunscreen had little or no effect on vitamin D levels, and did not induce osteoporosis or secondary hyperparathyroidism. Only low levels of exposure to ultraviolet radiation are necessary to avoid vitamin D deficiency. Exposure of the hands, arms, and face two to three times a week to a third to a half of minimum erythemal dose (about 5 min for a skin type 2 adult in Boston at noon in July; panel 1) in the spring, summer, and autumn is more than adequate.

There are also concerns about the endocrine effect of ultraviolet filters. Five chemicals—benzophenone-3, homosalate, 4-methyl-benzylidene camphor, octyl-methoxycinnamate, and octyl-dimethyl-PABA—increased breast cancer cell proliferation in vitro, whereas butyl-methoxydibenzoylmethane was inactive. Oral application of 4-methylbenzilidene camphor and octyl-methoxy-cinnamate, and to a lesser degree benzophenone-3, increased uterine weight of immature Long-Evans rats. Dermal application of 4-methylbenzilidene camphor to immature hairless rats also increased uterine weight. Further data indicated that the ultraviolet filters benzophenone-3 and homosalate possess antiandrogenic activity in vitro in addition to oestrogenic activity. Thus, the conclusion was reached that daily exposure to sunscreen formulations might have oestrogenic effects in human beings. Questions have been raised about the methodological grounds used. In the animal experiments, topical exposure to ultraviolet filters was judged to be unrealistically high compared with potential human exposure scenarios. Furthermore, a study on 32 people, who had a sunscreen containing benzophenone-3, octyl-methoxycinnamate, and 3-(4-methylbenzilidene) applied to their whole body daily for 5 days did show uptake in the body but no effects on reproductive hormone levels. Therefore the biological relevance of the oestrogenic effect of the tested ultraviolet filters has not been established and additional long-term studies are required.

Clinical recommendations

In a skin cancer prevention strategy, behavioural measures—eg, wearing sun protective clothes and a hat and reducing sun exposure to a minimum—must be preferred to sunscreens. For improved protection, especially if midday summer exposure or tropical exposure is unavoidable, the use of clothing over as much of the skin surface as possible, and proper application of a highly protective sunscreen over the remainder of the exposed skin, is very effective. Since sun protection practices, especially in the western world, are still inadequate (figure 2) and since sunscreens will be used by many as the predominant mode of sun protection for various societal reasons (eg, healthiness of a tan, relaxation...
in the sun), the population has to be advised as to how to make the best use of sunscreens.\cite{12} The application of a liberal quantity of sunscreen is by far the most important factor for effectiveness of the sunscreen, followed by the uniformity of application (figure 3) and the specific absorption spectrum of the agent used.\cite{10} Application of organic sunscreens to exposed sites should be done 15–30 minutes before going out into the sun. Waterproof or water-resistant sunscreens should be used to diminish the need for reapplication after swimming followed by towelling, friction with clothing or sand, and sweating. Although the better protection against UVB that is provided by high SPF sunscreens (SPF >15) has not been clearly proven to further protect against skin cancer, the overall data has shown that a high SPF is preferable to low SPF sunscreen.\cite{14} Broad-spectrum sunscreens with adequate UVA protection should be used, but there is no clear definition of what is deemed to be adequate. Nevertheless, sunscreens should not be abused in an attempt to increase time in the sun to a maximum. The year-round daily use of sunscreen for people living in countries of low insolation—eg, the UK and northern Europe—cannot be recommended, and sunscreens are best avoided during October to March.\cite{13} There is some evidence to suggest that the year-round application of sunscreens can be beneficial in terms of prevention of cancer and solar elastosis in areas of high insolation, such as Queensland, Australia, and Texas, USA.\cite{14,127}

Conflicts of interest statement

We declare that we have no conflict of interest.

References


62 Thieden E, Philippsen PA, Sandby-Moller J, Wulf HC. Sunscreen use related to UV exposure, age, sex, and occupation based on personal dosimeter readings and sun-exposure behavior diaries. Arch Dermatol 2005; 141: 967–73.


79 Chatelein E, Gabard B, Photostabilization of butyl methoxydibenzoylemethane (Avobenzone) and ethylhexyl methoxycinnamate by bis-ethylhexyloxyphenyl methoxydibenzoylmethane (Tinosorb S), a new UV broadband filter. Photochem Photobiol 2001; 74: 401–06.


Hearing dysphasic voices

Daniela Hubl, Martinus Hauf, Claudia van Swam, René Muri, Thomas Diersk, Werner Strik

In August, 2006, a 63-year-old woman, with no medical or psychiatric history of note, fell off her bicycle, and was brought to our hospital with a head injury and loss of consciousness. MRI showed a subarachnoid haemorrhage, caused by a ruptured aneurysm of the left middle cerebral artery, and damage to a region encompassing part of the left frontal, temporal, and parietal lobes. 3 days after admission, the aneurysm was clipped, and the patient underwent a decompressive craniotomy. On regaining consciousness, 7 days after the injury, the patient was found to have right-sided hemiplegia and global aphasia. After 4 weeks of treatment in the intensive care unit, the hemiplegia was substantially reduced, but the speech deficit was still severe: the patient communicated using single words or incomplete short sentences only, and she understood only short phrases. Therefore, she was transferred to our neurorehabilitation ward. She developed restless-legs syndrome and mild depression, which were treated with pramipexole hydrochloride and venlafaxine, at daily doses of 250 μg and 75 mg, respectively. In late October, the dysphasia worsened suddenly. Electroencephalography (EEG) showed interictal left temporal spikes and spike-and-wave complexes, indicating epilepsy. The patient was prescribed levetiracetam, at a dose of 1500 mg daily, and the dysphasia had returned to its previous extent; it subsequently continued to improve, so the patient became able to converse using short sentences. She was aware of her difficulties with expression and comprehension.

In late December, the patient began to hear her own thoughts aloud, and echoes of earlier conversations. She also hallucinated the voices of the hospital staff, commenting on things that had just happened. The voices appeared several times a day, for periods of a few minutes at a time. The patient understood that the voices were caused by her head injury; curiously, she also believed them to be real, and heard them in external space. She reported that the voices sounded just like other voices, except that they said only very simple and short sentences, sometimes with interruptions, sometimes with words she had never heard before. Aside from her mild depression, she had no other psychiatric symptoms. The auditory hallucinations (AH) were thought likely to be caused by epilepsy; the pramipexole and venlafaxine were therefore stopped. Functional MRI confirmed that, despite the brain damage, both auditory cortices were active during listening, as were the left frontal and temporal lobes during speaking. Structural MRI showed no further brain damage to that identified by the previous MRI (figure). On discharge, just before Christmas, some residual dysphasia remained, but the patient had AH only about once a week. In March, 2007, the patient was readmitted when her dysphasia suddenly worsened. EEG showed focal seizures in the left frontotemporal region. She was prescribed sodium valproate, at a dose of 1800 mg daily, in addition to the levetiracetam. 3 days later, a repeat EEG showed no seizure activity; the dysphasia had returned to its previous level. Moreover, the hallucinations had stopped, and the patient was no longer depressed. When last seen, in June, 2007, the patient had no AH.

AH can arise not only in psychotic disorders, but also in neurological conditions affecting the parts of the brain used to process and monitor speech,1 which are thought to include frontal motor areas, the primary auditory cortex, and Wernicke’s area.2 In our patient, epilepsy seems to have caused AH. Brain injury can cause epilepsy; moreover, some psychotropic medications can increase neuronal excitability, and lower the threshold for seizures.3 AH are widely thought to be caused by epilepsy; the pramipexole and venlafaxine were therefore stopped. Functional MRI confirmed that, despite the brain damage, both auditory cortices were active during listening, as were the left frontal and temporal lobes during speaking. Structural MRI showed no further brain damage to that identified by the previous MRI (figure). On discharge, just before Christmas, some residual dysphasia remained, but the patient had AH only about once a week. In March, 2007, the patient was readmitted when her dysphasia suddenly worsened. EEG showed focal seizures in the left frontotemporal region. She was prescribed sodium valproate, at a dose of 1800 mg daily, in addition to the levetiracetam. 3 days later, a repeat EEG showed no seizure activity; the dysphasia had returned to its previous level. Moreover, the hallucinations had stopped, and the patient was no longer depressed. When last seen, in June, 2007, the patient had no AH.

References
1 Nicolson SE, Mayberg HS, Pennell PB, Nemeroff CB. Persistent auditory hallucinations that are unresponsive to antipsychotic drugs. Am J Psychiatry 2006; 163: 1153–59.