A strategic revolution in HIV and global health

Last week saw the conclusion of a landmark event in the recent history of AIDS. The two turning points took place in New York. The visible one was a high-level meeting on AIDS, which brought 3000 participants to the UN to review progress in defeating an epidemic 30 years into its devastating course. Ambitious new targets were agreed. Countries committed themselves to, by 2015: halving sexual transmission of HIV; halving HIV transmission among people who inject drugs; ensuring that no child will be born with HIV; getting 15 million people onto treatment; and halving deaths from tuberculosis among people living with AIDS.

But the invisible turning point was the realisation that simply strengthening the vertical programme that is AIDS has to end. The new opportunity is integration. As one senior UNAIDS scientist put it—AIDS is not an exceptional disease; it is an exceptional opportunity. Part of the reason for a change in strategy is a matter of brutal reality. Investment in AIDS is in decline relative to other spheres of global health. But the incredible success of the AIDS movement also means that it is in a strong position to embrace—warmly and generously—other sectors of global health. AIDS can be the engine that broadens a front to defeat the diseases of poverty.

A good example of the new integration opportunity is AIDS in children. There are around 400 000 new childhood HIV infections each year. But in the 68 countries where most child deaths occur, coverage with antiretroviral treatment for prevention of mother-to-child transmission (PMTCT) of HIV is painfully low. The independent Countdown to 2015 group estimated that PMTCT coverage was only 22% in these countries in 2010.

Led by a coalition of UN agencies, global health initiatives, and civil society organisations, a new commitment was sealed last week—to eliminate paediatric HIV infections. The Global Task Team put together to deliver this goal is not isolating AIDS, as perhaps it might have done a few years ago. Their objective is to eliminate new paediatric HIV infections and, at the same time, to improve maternal, newborn, and child health in the context of HIV. The monitoring arrangements for tracking progress in HIV in children will embed this broader perspective. For example, one cannot address paediatric HIV without tackling HIV in women. The 2009 baseline of 1·4 million HIV-positive women delivering a child must be cut to 700 000 by 2015. New HIV infections in women aged 15–49 years will be reduced from 1·04 million in 2009 to 520 000 in 2015. Unmet need for family planning must fall from 11% in 2009 to zero in 2015. And HIV-associated maternal deaths will be cut from 21 000 to 2100 by 2015. If these successes were to be achieved, there will be fewer than 40 000 new paediatric infections in 2015, a 90% reduction.

This new approach will require new money. Bernhard Schwartländer and colleagues recently set out their vision for the resources needed to finance the next phase of the AIDS response. Solving AIDS will only happen if health systems are strengthened too. The total investment required to fund a set of basic programme activities, together with what Schwartländer and colleagues call “critical enablers” and “synergies with development sectors”, is US$16·6 billion this year, rising to $22 billion in 2015. PMTCT is only a very small proportion of that total: $0·9 billion this year, rising to $1·5 billion in 2015. Eliminating paediatric AIDS over the next 4–5 years is entirely possible—but only if AIDS is attacked as part of a comprehensive programme of interventions, from strengthening maternal health to scaling up family planning services.

This strategic revolution in global health poses important questions for AIDS governance. The Global Fund to fight AIDS, Tuberculosis, and Malaria is already reinventing itself slowly, but successfully, as a financing mechanism with a broader remit. But it is UNAIDS, led by the politically astute and charismatic Michel Sidibé, that is perhaps in the best position to be a catalyst for integration. Unlike WHO, UNAIDS is not a member-state governed organisation. Indeed, UNAIDS was created precisely to fill gaps in the AIDS response left by countries, donors, and other UN and non-UN bodies. Its mandate is to be bold, to say and do what others cannot say and do. The forthcoming UN General Assembly meeting in New York in September will be an opportunity for UNAIDS to unveil a potentially new leadership role in global health—one complementing but distinctive from that of WHO, one that puts AIDS at the leading edge of a new movement for integrating health responses to disease. ■
Health concerns of adolescents who are in a sexual minority

Growing up is never easy. Among the many challenges that adolescents face is the development of sexual orientation and identity. Nonetheless, because young people are often perceived as healthy, they have been largely neglected in global public health, as detailed in a study by Fiona Gore and colleagues in The Lancet today.

Compared with their peers, adolescence is an even more difficult period for sexual minorities. In Morbidity and Mortality Weekly Report, research shows that students in grades 9–12 (normally 14–17 years) who are in a sexual minority—homosexual, lesbian, and bisexual students and students who were unsure of their sexual identity and had sexual contact with both sexes—were more likely to engage in risky health behaviours than were other students. The findings are based on a survey of 156,145 students in seven states and six large urban school districts over 2001–09.

The risky behaviours included: behaviours that contribute to unintentional injuries and violence; attempted suicide; tobacco and alcohol use; unprotected sexual behaviours; poor weight management; and physical inactivity.

Life is especially difficult for the 6 million drug addicts living in Russia because methadone is banned, and they are reluctant to use the few available needle and syringe exchange programmes for fear of being exposed. New drug laws are being drawn up by the Russian Government in its “total war on drugs”. These will go against the recommendations of the report by the Global Commission on Drugs Policy, on June 2, to “End the criminalization, marginalization and stigmatization of people who use drugs but who do no harm to others”, and the evidence-based treatments endorsed by organisations such as the UN Office on Drugs and Crime, UNAIDS, and WHO.

If the new laws are enacted, drug addicts will face imprisonment or be forced to undergo treatment for their addiction. And the treatment of drug dealers will be akin to that of serial killers. The new laws will increase the transmission of HIV. Prisoners are also exposed to tuberculosis, hepatitis C, violence, and abuse. According to Chris Beyrer, Center for Public Health and Human Rights, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, “...the narcologic and psychiatric establishments must join the mainstream of modern science and support and use opioid substitution therapy, with methadone, buprenorphine, or Suboxone. Continuing to make evidence-based addiction treatment illegal will only maintain Russia’s very high rates of drug addiction, albeit without compassion, the HIV epidemic will continue unabated if imprisonment and forced treatment are the only options given to addicts. Substitution therapy and needle and syringe exchange programmes will not encourage addiction, but will bring drug addicts into contact with people who will, hopefully, be able to help them. So rather than enacting punitive laws that will discourage addicts from seeking help for their chronic illness, the Russian Government needs to have a treatment infrastructure and reform its public health policy.”
HPV vaccine effect: is the glass half full or half empty?

In The Lancet, Julia Brotherton and colleagues1 report a decrease in precancerous cervical lesions in girls younger than 18 years after population-wide human papillomavirus (HPV) vaccination in Victoria, Australia, in 2007. The study suggests a 38% decrease in the rates of histologically confirmed high-grade cervical lesions in these girls during the 2 years after HPV vaccine introduction compared with the 3 years before vaccine introduction.1 This ecological finding might be an early sentinel of the potential real-life effect of the vaccine on the main outcome in the clinical trials: cervical intraepithelial neoplasia of grade 2 or worse (CIN2+). However, these results should be viewed with caution, in view of the well-known limitations of ecological studies. For example, guidelines that emphasise less aggressive management of low-grade cytology, which were published 9 months before HPV vaccine introduction, could have contributed to the reported CIN2+ decrease.2 Health-care providers might also have screened and managed vaccinated patients less aggressively, especially girls younger than the recommended screening age of 18 years. With the almost 40% decrease in the incidence of high-grade cervical abnormalities recorded in girls younger than 18 years, a similar though smaller decrease would be expected in girls in the next oldest age group (those aged 18–20 years), who were likely to benefit from the vaccine and in whom vaccine coverage was high. However, no decrease was observed in this age group.

Australia was the first country to launch a national HPV vaccination programme with a predominantly school-based strategy. The three-dose vaccine coverage for girls aged 14–15 years in 2007 was about 72%.3 Australia has also taken many initiatives in surveillance related to cervical cancer. In 1991, when the national screening programme was started, state-based Papanicolaou (Pap) test registries were established to monitor cytological and histological outcomes. And again in 2007, resources were dedicated to starting a national HPV vaccine registry to monitor coverage and safety and to allow for potential linkage with the Pap and cancer registries. These registries can enable identification and possibly genotyping of cervical disease.4

Active monitoring of CIN2+ incidence and associated HPV types is under way in various settings, including Nordic countries and the USA.5,6 A model developed by Cuzick and colleagues7 that used cervical intraepithelial neoplasia of grade 3 (CIN3) as the outcome (thought to be a more reproducible and proximal precursor to cervical cancer than is CIN2) predicted that, with 80% coverage, a 51% reduction in CIN3 would take 7 years in parts of the UK where screening begins at age 20 years (Scotland and Wales) and would take 12 years where screening begins at age 25 years (England). The reductions in incidence might take longer in the USA or Nordic countries because of lower vaccine coverage, or in countries where screening starts later, such as in England. Australia’s cervical screening programme begins screening at a younger age (18 years) than most other screening programmes. Although a population-wide vaccine effect has not been recorded, evidence8,9 lent support to beginning screening at later ages (≥21 years) because, in younger women, cervical cancer is rare, HPV infections and their effect on cervical cells are often transient, and aggressive treatment can result in adverse birth outcomes. In the USA, the specialty medical organisation for obstetricians and gynaecologists recommended against screening before age 21 years, irrespective of age of sexual initiation.10 The cervical screening strategy in Australia (age of initiation, screening interval, and possibly type of test) could change in view of the evidence that existed before vaccination8,9 and the more definitive data that future vaccine studies will generate.

See Articles page 2085

Cervical cells infected with HPV
The not-so-cautious optimist in us wants to hail this early finding as true evidence of vaccine effect. However, individual-level vaccine status was not considered—as it perhaps should have been in view of the availability of such data in Victoria—and, as stated by Brotherton and colleagues, linkage between vaccination and screening results is needed to confirm these findings independently of possible bias by screening policy or practice changes. Indeed, more rigorous epidemiological studies are needed—many are under way—to increase our understanding of HPV vaccine effectiveness against cervical disease.

A demonstrable reduction of the burden of cervical cancer—the main goal of HPV vaccines—will take several decades. A WHO expert consultation meeting on HPV vaccine monitoring produced unanimous agreement that systems to monitor vaccine effect are not needed for the introduction of an HPV vaccine. In deciding whether to undertake these activities, low-income countries must carefully consider the feasibility of, and resources needed for, monitoring vaccine effect. Showing the real-world effect of these highly efficacious vaccines is needed, and the responsibility is mainly with high-income countries to establish population-level and individual-level effect.

*Mona Saraiya, Susan Hariri
Centers for Disease Control and Prevention, Atlanta, GA 30341, USA
msaraiya@cdc.gov

Adolescence is characterised by low mortality. Young people entering adolescence have survived the ravages of infant and child mortality and most do not yet face the chronic health problems of middle and old age. The curve for age-specific global mortality is U-shaped—ie, high in infancy, declining into adolescence, before slowly rising with increasing age. However, several reports have broadened the understanding of health and mortality in young people.

In 2004, 2·6 million deaths occurred in 1·8 billion young people aged 10–24 years. 97% of these deaths were in low-income and middle-income countries. Intentional and unintentional injuries are the main causes of death in this group and such injuries are exacerbated by urbanisation, interpersonal violence, and access to weapons. HIV infection and tuberculosis are also prominent, contributing to 11% of deaths in the age group. Mortality rises sharply in those aged between 10 and 25 years. Additionally, pregnancy-related mortality (15% of deaths in girls and women aged 15–24 years) will take several decades. A WHO expert consultation meeting on HPV vaccination.

We declare that we have no conflicts of interest.

The not-so-cautious optimist in us wants to hail this early finding as true evidence of vaccine effect. However, individual-level vaccine status was not considered—as it perhaps should have been in view of the availability of such data in Victoria—and, as stated by Brotherton and colleagues, linkage between vaccination and screening registers is needed to confirm these findings independently of possible bias by screening policy or practice changes. Indeed, more rigorous epidemiological studies are needed—many are under way—to increase our understanding of HPV vaccine effectiveness against cervical disease.

A demonstrable reduction of the burden of cervical cancer—the main goal of HPV vaccines—will take several decades. A WHO expert consultation meeting on HPV vaccine monitoring produced unanimous agreement that systems to monitor vaccine effect are not needed for the introduction of an HPV vaccine. In deciding whether to undertake these activities, low-income countries must carefully consider the feasibility of, and resources needed for, monitoring vaccine effect. Showing the real-world effect of these highly efficacious vaccines is needed, and the responsibility is mainly with high-income countries to establish population-level and individual-level effect.

*Mona Saraiya, Susan Hariri
Centers for Disease Control and Prevention, Atlanta, GA 30341, USA
msaraiya@cdc.gov

We declare that we have no conflicts of interest.


The global burden of disease in 10–24-year-olds

Adolescence is characterised by low mortality. Young people entering adolescence have survived the ravages of infant and child mortality and most do not yet face the chronic health problems of middle and old age. The curve for age-specific global mortality is U-shaped—ie, high in infancy, declining into adolescence, before slowly rising with increasing age. However, several reports have broadened the understanding of health and mortality in young people.

In 2004, 2.6 million deaths occurred in 1.8 billion young people aged 10–24 years. 97% of these deaths were in low-income and middle-income countries. Intentional and unintentional injuries are the main causes of death in this group and such injuries are exacerbated by urbanisation, interpersonal violence, and access to weapons. HIV infection and tuberculosis are also prominent, contributing to 11% of deaths in the age group. Mortality rises sharply in those aged between 10 and 25 years. Additionally, pregnancy-related mortality (15% of deaths in girls and women aged 10–24 years) rises sharply as young women enter the childbearing period. In the past 50 years, the decline in mortality in children (particularly for communicable disease) has been greater than that in young people.

In 1955, mortality in children and young people (15–24 years) was highest in 1–4-year-olds. By 2004, mortality rates in male young people were 2–3 times higher than those in 1–4-year-olds; whereas in female young people and children aged 1–4 years, rates were the same. 15–24-year-olds have benefited less from the
epidemiological transition of the 20th century than have 1–4-year-olds. 3

In The Lancet, Fiona Gore and colleagues 4 use cause-specific disability-adjusted life-years (DALYs) calculated by WHO region for 2004 to provide new estimates for the global burden of disease in young people aged 10–24 years. The use of DALYs refines our understanding of mortality and morbidity by identification of years of life lost because of premature mortality (YLL) and those lost because of disability (YLD). We congratulate the investigators for obtaining a substantial amount of data to generate these new estimates. This work is important in generating data for health status by region and national income and by age of young people, and for identifying leading risk factors for incident DALYs, such as alcohol and drug misuse, neuropsychiatric disorders, unsafe sex, and iron deficiency. These data show that young women have a higher disease burden related to YLD than young men, who have a greater burden related to YLL. Although neuropsychiatric disorders, respiratory diseases, and iron deficiency contributed substantially to disability, their contribution to mortality was low. Unsurprisingly, mortality was greater in low-income than in high-income countries; however, the disease burden from disability was substantial even in high-income countries.

Gore and colleagues’ analysis may ultimately underestimate the disease burden in young people aged 10–24 years. Although DALYs provide a useful measure of the burden of disease at a specific age, they tell us little about the ages that contribute to the cause of these diseases. 5 This limitation is a general problem of DALYs, which becomes particularly important when young people are assessed. Epidemiological life course perspectives increasingly show that several disorders of adulthood, including cancers, cardiovascular disease, and neuropsychiatric disorders, start in childhood and young adults. 6 For example, tobacco use is most commonly started before age 20, but most tobacco-related disease burden is identified much later in life. Therefore population measures to assess the health of young people should account for the future importance of starting health-risk behaviours and disorders during this life stage. Alcohol and other drug misuse and unprotected sexual behaviours that start during adolescence might have short-term or long-term health consequences. 7 Initiation of smoking and establishment of unhealthy dietary and physical activity behaviours mostly have long-term consequences. 8

Health promotion and efforts for disease prevention in young people aged 10–24 years should recognise both the burden of disease in this age group and the influence of risk behaviours on health in later life. Interventions should address the behaviours and social conditions that have both short-term and long-term health consequences. Interventions that increase resilience—eg, efforts for increasing the connections of adolescents to communities, schools, and families—are crucial for health promotion in young people. 9 Furthermore, targeted public health interventions, including enforcement of seatbelt laws, redesign of cars, implementation of tobacco taxes, and distribution of condoms, are essential for reducing morbidity and mortality. Failure to act today will assuredly result in failure tomorrow.

*John S Santelli, Sandra Galea
Heilbrunn Department of Population and Family Health (JSS), and Department of Epidemiology (SG), Mailman School of Public Health, Columbia University, New York, NY, USA
js2637@columbia.edu

We declare that we have no conflicts of interest.

First-line treatment of advanced colorectal cancer

As we move into the era of stratified medicine, how should the Article by Timothy Maughan and colleagues1 in The Lancet shape treatment of patients with metastatic colorectal cancer? The findings were unexpected in that the antiepidermal growth factor receptor antibody cetuximab did not improve outcome when added to a chemotherapy doublet consisting of a fluoropyrimidine (capecitabine or fluorouracil) and oxaliplatin. This MRC COIN trial randomly assigned nearly 2500 patients in a prospective study in which analysis was enriched for patients most likely to benefit from cetuximab—those with KRAS wild-type tumours. Apart from a modest increase in the rate of tumour regression, cetuximab had no significant effect on progression-free survival or overall survival. The rationale to undertake the trial had been sound: cetuximab had previously been shown to be effective in chemorefractory colorectal cancer on its own or when combined with the topoisomerase inhibitor irinotecan.2 Moving the drug to earlier in the treatment course made sense—with each line of therapy the number of patients fit enough to receive treatment decreases, so, if cetuximab was effective, more patients would benefit from treatment.

So what went wrong? The overlapping side-effects between capecitabine and cetuximab, particularly skin toxicity and diarrhoea, are one explanation. This contention is supported in COIN by a reduction in dose intensity, and a subgroup analysis suggesting that the cohort of patients benefiting from cetuximab were those with KRAS wild-type tumours, had disease affecting 0–1 metastatic sites, and were treated with fluorouracil and oxaliplatin on clinician’s choice. Our own experience in the setting of neoadjuvant rectal cancer treatment is that the combination of oxaliplatin, capecitabine, and cetuximab, with capecitabine given at a lower dose to that commenced by most patients in COIN (1700 mg/m² per day), is well tolerated and efficacious.3 As an alternative to cetuximab in first-line treatment, other investigators have used panitumumab with fluorouracil and oxaliplatin, with an improvement in progression-free survival.4

The other consideration is that oxaliplatin is not the best chemotherapy drug for cetuximab; this notion fits with preclinical data, which are not particularly compelling.5 However, when cetuximab is combined with a fluorouracil-irinotecan doublet it can improve tumour control rates and overall survival in first-line treatment;6 and does appear to facilitate a higher rate of resection of metastatic disease in the liver. This is a key objective in patients with limited or so-called oligometastatic disease, because surgery could render them free of cancer in the long term. Indeed, these new data should prompt a review of present guidance from the UK’s National Institute of Health and Clinical Excellence, which recommends that patients with liver-only metastatic disease should receive an oxaliplatin-based chemotherapy doublet with cetuximab, and that irinotecan should be reserved for patients for whom oxaliplatin is deemed to be contraindicated or not tolerated.6

Further refinement of the biomarker selection criteria for patients who are most likely to benefit from cetuximab might be possible, including the presence of mutations in BRAF, NRAS, PIK3CA (exon 20),7 and KRASG13D,8 and levels of amphiregulin and epiregulin.9 In COIN, 43% of patients had mutations in KRAS, 8% in BRAF, and 4% in NRAS, resembling another cohort.7 However, BRAF mutations were not associated with a lack of benefit from cetuximab,7 but did predict a poor prognosis, which is similar to findings from another prospective trial.5 Other potential biomarkers include loss of PTEN expression,9 mutations in TP53,10 EGFR gene copy-number changes,9 and EGFR promoter hypermethylation.11 Micro-RNAs have been shown to modulate response to cetuximab in patients with KRAS mutations,12 whereas germline polymorphisms in FCGR3A and EGFR could affect response in patients with KRAS wild-type tumours.13 As our understanding of...
First-line treatment of advanced colorectal cancer

As we move into the era of stratified medicine, how should the Article by Timothy Maughan and colleagues in The Lancet shape treatment of patients with metastatic colorectal cancer? The findings were unexpected in that the antiepidermal growth factor receptor antibody cetuximab did not improve outcome when added to a chemotherapy doublet consisting of a fluoropyrimidine (capecitabine or fluorouracil) and oxaliplatin. This MRC COIN trial randomly assigned nearly 2500 patients in a prospective study in which analysis was enriched for patients most likely to benefit from cetuximab—those with KRAS wild-type tumours, had disease that the cohort of patients benefiting from cetuximab were those with KRAS wild-type tumours, had disease affecting 0–1 metastatic sites, and were treated with fluorouracil and oxaliplatin on clinician’s choice. Our own experience in the setting of neoadjuvant rectal cancer treatment is that the combination of oxaliplatin, capecitabine, and cetuximab, with capcitabine given at a lower dose to that commenced by most patients in COIN (1700 mg/m² per day), is well tolerated and efficacious. As an alternative to cetuximab in first-line treatment, other investigators have used panitumumab with fluorouracil and oxaliplatin, with an improvement in progression-free survival.

The other consideration is that oxaliplatin is not the best chemotherapy drug for cetuximab; this notion fits with preclinical data, which are not particularly compelling. However, when cetuximab is combined with a fluorouracil-irinotecan doublet it can improve tumour control rates and overall survival in first-line treatment, and does appear to facilitate a higher rate of resection of metastatic disease in the liver. This is a key objective in patients with limited or so-called oligometastatic disease, because surgery could render them free of cancer in the long term. Indeed, these new data should prompt a review of present guidance from the UK’s National Institute of Health and Clinical Excellence, which recommends that patients with liver-only metastatic disease should receive an oxaliplatin-based chemotherapy doublet with cetuximab, and that irinotecan should be reserved for patients for whom oxaliplatin is deemed to be contraindicated or not tolerated.

Further refinement of the biomarker selection criteria for patients who are most likely to benefit from cetuximab might be possible, including the presence of mutations in BRAF, NRAS, PIK3CA (exon 20), and KRASG13D, and levels of amphiregulin and epieregulin. In COIN, 43% of patients had mutations in KRAS, 8% in BRAF, and 4% in NRAS, resembling another cohort. However, BRAF mutations were not associated with a lack of benefit from cetuximab, but did predict a poor prognosis, which is similar to findings from another prospective trial. Other potential biomarkers include loss of PTEN expression, mutations in TP53, EGFR gene copy-number changes, and EGFR promoter hypermethylation. Micro-RNAs have been shown to modulate response to cetuximab in patients with KRAS mutations, whereas germline polymorphisms in FCGR3A and EGFR could affect response in patients with KRAS wild-type tumours. As our understanding of
modifiers of cetuximab response becomes increasingly complex, any new biomarkers will need rigorous, prospective assessment before being used to decide patients' care.

As an alternative to EGFR-targeted therapy, the antiangiogenic bevacizumab is routinely used outside the UK combined with chemotherapy doublets to treat advanced colorectal cancer. Indeed, a bevacizumab-containing group in COIN would have been a valuable comparator, although it would have added to the scale of the study. The negative results of COIN might reinforce the role for bevacizumab in this setting. Notably, bevacizumab cannot be used in combination with cetuximab or panitumumab.

**Figure:** Integration of radiological and molecular characteristics to formulate a treatment course for metastatic colorectal cancer

(A) The relative frequencies of radiological presentations in colorectal cancer (top panel) and tumour molecular characteristics (bottom panel) are shown. This information is used to establish a personalised approach to treatment (B). Consideration can be given to intensification of therapy to those tumours known to contain BRAF mutations in view of poor prognosis associated with this subset. RFA=radio frequency ablation. *Chemotherapy can be continued beyond 6 months if well tolerated and ongoing incremental response. If less than 6 months of treatment is chosen, patients should be carefully selected and closely monitored for disease progression. EGFR CA(n)<20: number of CA short-tandem repeats observed in this polymorphism; if either allele has ≥20 repeats, benefit is reduced.

<table>
<thead>
<tr>
<th>Potential biomarkers of efficacy for anti-EGFR receptor antibodies that need further validation</th>
<th>Predicting benefit</th>
<th>Predicting lack of benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutations</td>
<td>KRAS p.G13D, 7p31</td>
<td>PIK3CA (exon 20), NRAS, PTEN</td>
</tr>
<tr>
<td>EGFR receptor/ligand levels</td>
<td>Increased epiregulin, amphiregulin, and EGFR copy number</td>
<td>EGFR promoter methylation, other loss of PTEN function, micro-RNA levels</td>
</tr>
<tr>
<td>Other</td>
<td>EGFR CA(n)&gt;20 and FCGR3A 818AA germline polymorphisms, micro-RNA levels</td>
<td></td>
</tr>
</tbody>
</table>

BRAF mutant 8–9%  
KRAS mutant 40%
In the UK, the usual duration of first-line chemotherapy for advanced colorectal cancer is 6 months. However, treatment length varies substantially and some countries advocate treatment until disease progression, with patients proceeding directly to second-line therapy. The COIN investigators wondered whether the duration of initial chemotherapy could be shortened to 12 weeks, restarting on disease progression, to improve quality of life without compromising survival. To address this issue, one cohort was randomly assigned to an intermittent schedule, results now published in The Lancet Oncology.14 This group did not reach the primary endpoint of non-inferiority, with intermittent chemotherapy being associated with a marginally inferior outcome (median overall survival 14.4 months vs 15.8 months). This finding might be partly explained by failure to restart treatment on progression in 36% of patients. Those who did particularly poorly on intermittent treatment included patients with liver-only disease (22%) and those with a high platelet count (28%), the latter associated with a survival reduction of 5 months. Unsurprisingly, at 24 weeks, intermittent treatment was associated with fewer chemotherapy-related side-effects such as fatigue, nausea, or vomiting, but disappointingly no improvement in global quality of life and a significant increase in reporting of pain.

Thus we would urge caution in stopping treatment after 3 months. If clinicians and their patients choose this approach, we would recommend careful selection of patients and close monitoring for progressive disease so that chemotherapy can be reintroduced before the therapeutic window is closed. An alternative approach to complete discontinuation of chemotherapy could be the use of maintenance fluorouracil alone, which seemed to be beneficial in the French OPTIMOX2 trial,15 but patients would still be subject to some degree of ongoing toxicity. A shorter treatment duration is also acceptable in the neoadjuvant setting, before the surgical resection of liver metastases, but is usually followed by 3 months of further treatment in the postoperative period (figure).

For now, in our clinical practice most patients with advanced colorectal cancer are likely to receive a minimum of 6 months of chemotherapy, with the integration of targeted agents when possible and appropriate. Further research is needed to identify effective and relatively non-toxic maintenance treatment after initial chemotherapy. Cetuximab and panitumumab should be restricted to patients with KRAS wild-type tumours and cetuximab might be best combined with an irinotecan-based regimen, whereas bevacizumab-containing combinations can be considered irrespective of molecular profile.

Madeleine Hewish, *David Cunningham

The Royal Marsden Hospitals NHS Trust, London and Surrey, Sutton, Surrey SM2 5PT, UK

david.cunningham@rmh.nhs.uk

DC has received research funding from Amgen, Roche, Merck Sharpe and Dohme, Sanofi-Aventis, and Merck Serono, and has participated in uncompensated advisory boards for Roche and Amgen. MH declares that she has no conflicts of interest. We acknowledge National Health Service funding to the National Institute for Health Research Biomedical Research Center.

1 Maughan TS, Adams RA, Smith CG, et al. 1+1 on behalf of the MRC COIN Trial Investigators. Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for patients with advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. Lancet 2011; published online June 4. DOI:10.1016/S0140-6736(11)60613-2


3 Dewedey A, Cunningham D, Tabernero J, et al. EXPERT-C: a randomized phase II European multicenter trial of neoadjuvant chemotherapy (capecitabine/oxaliplatin) and chemoradiation (CRT) with or without cetuximab followed by total mesorectal excision (TME) in patients with MRI-defined high-risk rectal cancer. 29 (suppl 4): 350 (abstr).


14 Adams RA, Meade AM, Seymour MT, et al. 1+1 on behalf of the MRC COIN Trial Investigators. Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy for first-line treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. Lancet Oncol 2011; published online June 4. DOI:10.1016/S1470-2045(11)70102-4

Taxing tobacco profits to prevent the smoking epidemic

Around 1·5 billion people, most of them in developing countries, smoke cigarettes.1 Smoking kills about half of all lifelong smokers, so a massive global epidemic of death and disability from smoking, predominantly in the developing world, is inevitable. Already about 5 million people die from smoking every year,2 more than from HIV/AIDS, malaria, and tuberculosis combined, and this figure is predicted to rise to 8 million per year by 2030.3 This epidemic is man-made, resulting from global marketing of cigarettes by a few multinational tobacco companies. Prevention of further uptake of smoking and support to encourage existing smokers to stop smoking are obvious and immediate global health priorities.

WHO’s Framework Convention on Tobacco Control (FCTC)4 is an international treaty designed to guide an international response to the tobacco epidemic, and specifies key measures, such as advertising bans, taxation, smoke-free policy, health promotion, and cessation support, which governments should promote to prevent smoking. The treaty came into force in 2005, but progress with implementation worldwide is slow. Among developed countries in Europe, for example, implementation is at best incomplete,4 and in most developing countries, implementation is minimal.2 The FCTC has not even been ratified by the USA. Meanwhile, tobacco companies enjoy unrestricted freedom to apply immense financial resources and decades of marketing experience to promote tobacco smoking throughout the developing world, as, for example, by British American Tobacco in Africa.5

Recently, Callard6 has provided fresh insight into the magnitude of the financial resources at the industry’s disposal.6 In 2008, the five transnational tobacco companies that now dominate the world tobacco market generated a combined income of more than US$300 billion,6 a sum that exceeds the gross domestic product of most countries.7 These companies paid more than $160 billion in tax to governments, some of which also hold direct investments in tobacco, and realised $14 billion as profit.4 Tobacco-marketing budgets are not disclosed, but an estimated $12·4 billion was spent in 2006 in the USA alone.5 Against this vast financial power, combined funding for global tobacco-control activity in 2008 was about $240 million.8 The imbalance is massive.

Ironically, according to another recent study, by Gilmore and colleagues,7 the profits enjoyed by tobacco companies are partly an unintended consequence of tobacco-control policy. In rich countries, tobacco companies have exploited high tobacco taxes to conceal price increases that generate profit margins far greater than those of other companies that make consumer staple products. A means to reduce this extreme profitability and financial power is, therefore, urgently needed. Callard suggests using the FCTC to monitor, name, and shame governments that invest in tobacco—and hence have a conflict of interest in relation to tobacco control—and using taxes on sales crossing international boundaries to fund implementation of the Convention. Gilmore and colleagues suggest price capping to limit profitability at the outset. Both approaches present substantial regulatory challenges, but are examples of the radical thinking needed to create tax structures that reduce profitability while enabling more effective prevention of smoking in the poor countries that are increasingly the source of these profits.

The other major financial imbalance identified by Callard, which reflects poorly on international public health policy generally, is between developmental assistance for tobacco control and other major health problems. Callard estimates that such international spending in 2008 on each of HIV/AIDS, malaria, and tuberculosis exceeded that for tobacco control, and collectively amounted to more than $6 billion.7 This apparent failure to prioritise smoking prevention is symptomatic of the broad neglect of tobacco
control in international and national public health, and across governmental policy, that is also evident in medical research spending and in clinical practice.

Because urgent action to prevent the spread of smoking is crucially important, there is an argument for prioritising the promotion of key policy components of the FCTC in the developing world, and particularly against advertising, which is fundamental to the development and maintenance of new markets. An early advertising ban seems to have substantially delayed progression of the tobacco epidemic in at least one developing country, and advertising bans cost virtually nothing. Because quick implementation of such policies is vital to world health, responsibility to find innovative ways to do so lies particularly with leaders of the developed countries that benefit financially from the tobacco industry. Most of the fault for the evolving global tobacco epidemic lies with the tobacco industry, but not all.

John Britton
UK Centre for Tobacco Control Studies, Division of Epidemiology and Public Health, University of Nottingham, City Hospital, Nottingham NG5 1PB, UK
j.britton@virgin.net

I declare that I have no conflicts of interest.


6. Callard C. Follow the money. How the billions of dollars that flow from smokers in poor nations to companies in rich nations greatly exceed funding for global tobacco control and what might be done about it. Tob Control 2010; 19: 285-6-90.


Comment

China’s primary health-care reform

Published Online
March 30, 2011
DOI:10.1016/S0140-6736(11)60167-0

The proposal to rebuild a good primary health-care system occupies a central role in China’s health-care reform, which was announced in 2009. In the planned model, the functions of primary-care centres will include medical care, disease prevention, health promotion and education, birth control, and rehabilitation in the community. In time, these centres will also take on a gate-keeping role to reduce the cost burden arising from uncontrolled and inappropriate use of expensive hospital services.

To achieve these objectives, several measures have been introduced. First, an additional investment of CNY 850 billion (US$ 127 billion) to develop infrastructure and human resources in thousands of clinics across county, town, and village levels in rural areas, and community health centres and stations in cities. Second, the operational cost of primary-care centres will come from governmental subsidies and service charges, instead of relying on sales of drugs. Third, the separate roles of primary-care institutions and of secondary and tertiary hospitals will be clarified, and mechanisms for bidirectional referral between them will be established. The aim is that patients will be seen at primary-care centres for most common and minor illnesses, and will only be referred to hospitals for more complex problems.

Considerable challenges must be overcome to achieve these objectives. First and foremost is financing the planned changes. Funds earmarked for reform should come from both central and local governments, but local support has not always materialised. Some local governments do not fully recognise the importance of primary care, and others simply cannot afford the expected investment. Furthermore, in parts of western China, less than half of urban primary-care centres are within the local government-run system. To overcome difficulties in the local implementation of national policy, central government has substantially increased funding for central and western areas. However, from 2010 most new health-service investments by local governments
control in international and national public health, and across governmental policy, that is also evident in medical research spending and in clinical practice.

Because urgent action to prevent the spread of smoking is crucially important, there is an argument for prioritising the promotion of key policy components of the FCTC in the developing world, and particularly against advertising, which is fundamental to the development and maintenance of new markets. An early advertising ban seems to have substantially delayed progression of the tobacco epidemic in at least one developing country, and advertising bans cost virtually nothing. Because quick implementation of such policies is vital to world health, responsibility to find innovative ways to do so lies particularly with leaders of the developed countries that benefit financially from the tobacco industry. Most of the fault for the evolving global tobacco epidemic lies with the tobacco industry, but not all.

John Britton
UK Centre for Tobacco Control Studies, Division of Epidemiology and Public Health, University of Nottingham, City Hospital, Nottingham NG5 1PB, UK
j.britton@virgin.net

I declare that I have no conflicts of interest.

6 Callard C. Follow the money: how the billions of dollars that flow from smokers in poor nations to companies in rich nations greatly exceed funding for global tobacco control and what might be done about it. Tob Control 2010; 19: 285–90.
9 Gilmore AB, Branston JR, Swannar D. The case for OFSMOKE: how tobacco price regulation is needed to promote the health of markets, government revenue and the public. Tob Control 2010; 19: 423–30.

China’s primary health-care reform

The proposal to rebuild a good primary health-care system occupies a central role in China’s health-care reform, which was announced in 2009.1 In the planned model, the functions of primary-care centres will include medical care, disease prevention, health promotion and education, birth control, and rehabilitation in the community.2 In time, these centres will also take on a gate-keeping role to reduce the cost burden arising from uncontrolled and inappropriate use of expensive hospital services.

To achieve these objectives, several measures have been introduced. First, an additional investment of CNY 850 billion (US$ 127 billion) to develop infrastructure and human resources in thousands of clinics at county, town, and village levels in rural areas, and community health centres and stations in cities. Second, the operational cost of primary-care centres will come from governmental subsidies and service charges, instead of relying on sales of drugs.3 Third, the separate roles of primary-care institutions and of secondary and tertiary hospitals will be clarified, and mechanisms for bidirectional referral between them will be established.45

The aim is that patients will be seen at primary-care centres for most common and minor illnesses, and will only be referred to hospitals for more complex problems.6 Considerable challenges must be overcome to achieve these objectives. First and foremost is financing the planned changes. Funds earmarked for reform should come from both central and local governments, but local support has not always materialised. Some local governments do not fully recognise the importance of primary care, and others simply cannot afford the expected investment.7 Furthermore, in parts of western China, less than half of urban primary-care centres are within the local government-run system.8 To overcome difficulties in the local implementation of national policy, central government has substantially increased funding for central and western areas. However, from 2010 most new health-service investments by local governments

www.thelancet.com Vol 377 June 18, 2011

Published Online
March 30, 2011
DOI:10.1016/S0140-6736(11)60167-0
should be channelled into primary care. Finally, the Ministry of Health and provincial health departments have agreed on clear timetables, definite annual work tasks, and funding arrangements to implement national policy.

The second challenge is the serious shortage of trained general practitioners in the country. In response to this, in April, 2010, six relevant ministries jointly issued a plan for capacity development. The objectives include: to develop general practice as an academic discipline in universities; to establish a system for postgraduate training that includes hospital rotation for young medical graduates, and special programmes for more experienced doctors who wish to become general practitioners; and to strengthen continuing professional development for doctors in post already. Additionally, remuneration, prospects for promotion, and working conditions of general practitioners will be improved, to retain those who have been trained. With target figures of 60 000 and 300 000 general practitioners to be trained within 3 years and 10 years, respectively, the plan is ambitious but essential in view of the population’s size. We are also working with countries such as the UK and France to address the shortage of trainers for general practitioners.

Another difficulty is the deeply entrenched habit of patients to seek help from large hospitals. This behaviour is partly caused by a lack of public confidence in the quality of care provided in primary-care facilities. Improved training of general practitioners will be essential to modify this attitude. Meanwhile, many provinces have introduced preferential rates for reimbursement of costs incurred at primary-care centres compared with those at hospitals. Furthermore, in 16 cities where reform of public hospitals is being piloted, several hospitals will actually be converted into primary-care facilities.

The development of the primary-care system is an immutable priority. The inadequacies of the system that have arisen during the past two decades, compounded by the size of the population, make this task daunting. The planned changes should help to improve the situation, but close monitoring of their implementation is crucial. We also want to learn from international experiences in relevant areas, including the development of models of primary care and postgraduate training. Meanwhile, we are starting to see some encouraging changes at primary-care level. A review by the Ministry of Health found a continuing and substantial increase in the number of consultations in both urban and rural primary-care centres. By October, 2010, health records had been established for 36% of the urban population and 24% of the rural population. Nationally, more than 21 million patients with hypertension and nearly 6 million patients with diabetes are receiving care in primary-care centres. There is still a long way to go but we believe that, in time, primary care will re-emerge as a key and central component of the health service in China.

*Qian Liu, Bin Wang, Yuyan Kong, K K Cheng
Ministry of Health, People’s Republic of China, Beijing, China (QL, BW); Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, UK (KKC, YK)
qianliumoh@yahoo.cn

We declare that we have no conflicts of interest.

Health initiatives by Indigenous people in Australia

June Oscar is an Aboriginal woman from Marninwarntikura Women's Resource Centre, Fitzroy Crossing, in the Kimberley region of the Northern Territory, Australia, where about 4500 Aboriginal people live in more than 45 communities. She had attended 50 funerals in 1 year, many of which were for suicides related to alcohol abuse. She was also concerned about the one in four babies born with fetal alcohol spectrum disorder in the region, which she regarded as a particular disaster for an oral-based culture. After consultation with other women and their community Elders, the group made two films, Yajilarra (to dream) and Marulu (precious, worth nurturing, see webvideo), about how alcohol was destroying their lives through violence and crime. In 2009 they presented Yajilarra to the UN in New York and to the Australian Government. The result of their amazing initiative was that alcohol restrictions were introduced in the area: no drinks over 2.7% alcohol content could be sold in takeaways. The move was not universally popular, as the film depicts, but after the restrictions were introduced domestic violence fell by 43% and alcohol-related presentations to hospital were reduced by 55%—success indeed. The restrictions have now been taken up by some other rural communities.

Of the 8 billion Indigenous people in the world, about 570 000 live in Australia. In that country and other developed countries, such as New Zealand, the USA, and Canada, the health inequalities between Indigenous and non-Indigenous populations are enormous. In Australia the death rates from non-communicable diseases in Aboriginal and Torres Strait Islander people are shocking—those from diabetes are 13 times greater, those from kidney disease are five times greater, and those from heart disease are three times greater than in non-Indigenous people. There are also huge gaps in life expectancy. Aboriginal and Torres Strait Islander people die nearly 20 years younger than non-Indigenous Australians, by contrast with Indigenous people in Canada, the USA, and New Zealand, where this difference in life expectancy is between 3 and 7 years. Infant mortality rates in Aboriginal and Torres Strait Islander people are three times that of non-Indigenous children and about 50% higher than in Indigenous children in the USA and New Zealand.

These statistics can be directly attributable to several factors, the most important of which are social and economic disadvantage: poverty, poor housing, lack of education, poor access to medical care, and low income. Alcohol and smoking are particular problems in the Aboriginal population. Only 17% of Australians smoke, but 45% of Aboriginal people do so, indicating that the highly effective antismoking legislation in Australia has not filtered down to the Indigenous population. Alcohol-related deaths are disproportionately high and fetal alcohol syndrome is prevalent in Indigenous people.

A recurring theme in presentations at the Royal Australasian College of Physicians (RACP) 2011 Congress, in Darwin on May 22–26, was the importance of Indigenous communities, health workers, and...
Health initiatives by Indigenous people in Australia

June Oscar is an Aboriginal woman from Marninwarntikura Women’s Resource Centre, Fitzroy Crossing, in the Kimberley region of the Northern Territory, Australia, where about 4500 Aboriginal people live in more than 45 communities. She had attended 50 funerals in 1 year, many of which were for suicides related to alcohol abuse. She was also concerned about the one in four babies born with fetal alcohol spectrum disorder in the region, which she regarded as a particular disaster for an oral-based culture. After consultation with other women and their community Elders, the group made two films, Yajilarra (to dream) and Marulu (precious, worth nurturing, see webvideo), about how alcohol was destroying their lives through violence and crime. In 2009 they presented Yajilarra to the UN in New York and to the Australian Government. The result of their amazing initiative was that alcohol restrictions were introduced in the area: no drinks over 2.7% alcohol content could be sold in takeaways. The move was not universally popular, as the film depicts, but after the restrictions were introduced domestic violence fell by 43% and alcohol-related presentations to hospital were reduced by 55%—success indeed. The restrictions have now been taken up by some other rural communities.

Of the 8 billion Indigenous people in the world, about 570 000 live in Australia. In that country and other developed countries, such as New Zealand, the USA, and Canada, the health inequalities between Indigenous and non-Indigenous populations are enormous. In Australia the death rates from non-communicable diseases in Aboriginal and Torres Strait Islander people are shocking— those from diabetes are 13 times greater, those from kidney disease are five times greater, and those from heart disease are three times greater than in non-Indigenous people. There are also huge gaps in life expectancy. Aboriginal and Torres Strait Islander people die nearly 20 years younger than non-Indigenous Australians, by contrast with Indigenous people in Canada, the USA, and New Zealand, where this difference in life expectancy is between 3 and 7 years. Infant mortality rates in Aboriginal and Torres Strait Islander people are three times that of non-Indigenous children and about 50% higher than in Indigenous children in the USA and New Zealand.

These statistics can be directly attributable to several factors, the most important of which are social and economic disadvantage: poverty, poor housing, lack of education, poor access to medical care, and low income. Alcohol and smoking are particular problems in the Aboriginal population. Only 17% of Australians smoke, but 45% of Aboriginal people do so, indicating that the highly effective antismoking legislation in Australia has not filtered down to the Indigenous population. Alcohol-related deaths are disproportionately high and fetal alcohol syndrome is prevalent in Indigenous people.

A recurring theme in presentations at the Royal Australasian College of Physicians (RACP) 2011 Congress, in Darwin on May 22–26, was the importance of Indigenous communities, health workers, and
researchers having a major role in decisions about health care. The achievements of June Oscar and her colleague Maureen Carter, who presented the films at the Congress, are shining examples of the success that can be gained by the Indigenous community.

Another example of progress is the formation of the indigenously led Lowitja Institute in Melbourne.7 The aims of the institute are to build a national strategic research agenda to improve the health of Aboriginal and Torres Strait Islander people, who are represented by Dr Lowitja O’Donoghue, Ms Pat Anderson, and Professor Ian Anderson. Its charter will ensure that these groups will have a large say in the research process. Research projects are under way, and the institute offers scholarships to Aboriginal researchers. The expectation is that with Indigenous people under taking these projects they will win the Aboriginal and Torres Strait Islander peoples’ trust and engagement in health and education systems—something that has been scarce up to now. The overriding message from the RACP Congress and from the Lowitja Institute is that education and Indigenous leadership are key for the improvement of the lives of Australian Aboriginal people.

**Clinical Series and our clinical content**

Every journal has a personality. A journal’s content represents the mix of interests and skills in its editorial team. The Lancet tries, though surely fails on many occasions, to convey four aspects of our collective personality to readers. First, a commitment to the best international research that influences the ideas and practice of medicine. Second, a desire to put global health in the mainstream of modern medical thinking. Third, being a place for robust comment and opinion. And fourth, strengthening knowledge and understanding of the treatment and prevention of disease. As catalysts for bringing science to bear on practice, policy, and subsequently advocacy, our global health Series, such as those on stillbirths and vaccines, have sought to take the journal from being a passive recipient to an active participant in global health affairs. We now seek to do the same in the field of clinical medicine.

To further strengthen our clinical content, we are adding a programme of clinical Series to match our global health and country Series. We are publishing the first clinical Series in this issue—on arthritis. With these Series, we aim to provide a summary of pathophysiology, basic science, and—where important—genomic insights, as well as directly useful, practical, up-to-date clinical information on prevention, diagnosis, and management in a variety of important or neglected disorders. In addition, relevant new research will be critically examined and unanswered questions and new research areas will be highlighted to help define future research agendas. These Series will complement our Seminars, Reviews, and Clinical Core Collection, and serve as an opportunity to highlight and summarise in sufficient depth clinical areas in which an active and fast-moving body of research might make it difficult for busy clinicians to keep abreast of new findings and current best practice. Researchers might use these Series too as a basis for new research questions. We would be happy to hear about your ideas for topics that we should cover as a clinical Series.

**Stephanie Clark**

The Lancet, London NW1 7BY, UK

Offline: “The depth of the deception”?  

I mentioned a few weeks ago that a scandal might soon erupt at a leading London teaching hospital. It has now done so. The Home Office Equalities Minister, Lynne Featherstone, has called on the Chief Executive of Great Ormond Street Hospital for Children (GOSH), Dr Jane Collins, to resign. In a letter sent to the Secretary of State for Health, Andrew Lansley, as well as to the Chairs of the Care Quality Commission and the GOSH Board, Ms Featherstone writes that Dr Collins “appears [to have]... withheld vital information” from a serious case review into the death of Peter Connelly (also known as Baby P). On BBC television news, Featherstone said that Dr Collins “has to resign. I can think of no more serious charge”.*

Peter Connelly was born on March 1, 2006. He died on Aug 3, 2007. According to a January, 2008, report by Prof Jonathan Sibert and Dr Deborah Hodes, commissioned by Dr Collins, doctors who saw Baby P at North Middlesex Hospital found injuries, bruises, skin breakdown, and deep tissue damage on his body at the time of death. Prof Sibert and Dr Hodes agreed with the view of one witness they consulted that St Ann’s Hospital, where Baby P was seen just prior to his death and whose clinical staff were employed and managed by GOSH, was a “clinically risky situation”. The arrangements for child protection at St Ann’s “cause grave concern”, they wrote. When a serious case review of the events surrounding the care of Baby P took place, it might have been expected that Dr Collins would submit the full Sibert/Hodes report as evidence. But only “a partial and selective version” (Ms Featherstone’s words) was passed on: “important information was deliberately withheld.” In particular, because GOSH was responsible for “the dangerous conditions” under which Baby P was seen and eventually died, and since “it appears that Dr Collins has attempted to cover-up the fact that the situation was ‘clinically risky’”, Featherstone concluded in her letter to Lansley that “the depth of the deception that has been perpetrated is unbelievable”. “Dr Collins bears a share of responsibility”, the Member of Parliament writes, for what looks like “a deliberate attempt to hide the management failings highlighted in the Sibert Report.” BBC news says that GOSH denies a cover up. The hospital’s Board, chaired by Baroness Tessa Blackstone, has complete confidence in Dr Collins. But there are unanswered questions. Have the events that led to the death of Peter Connelly been fully and transparently investigated? Have the right lessons been learned? And have those who managed (and continue to manage) children’s services at GOSH and its associated facilities been held properly responsible for the quality of care they delivered? The answers to these three questions are the same—we don’t know. These uncertainties now rest with the Secretary of State for Health to resolve as a matter of urgency.  

* There remains one additional and very puzzling question. Why did an alleged “cover-up” take place at all? The reasons given are, first, that GOSH submitted a partial report based on legal advice and, second, that the hospital had employment obligations to staff mentioned in the report. There is a third reason that ought to be considered. GOSH is seeking Foundation Trust status, an objective it hopes to achieve by autumn, 2011. The hospital’s perfectly legitimate quest for greater independence, which brings more freedom to manage its affairs, has been ongoing over several years and has not always been a happy journey. But this prestigious repositioning is now within reach. What GOSH did not need at this late stage was a public debate about the integrity of its leadership or the quality of its management of services that failed Baby P. When the highly critical Sibert/Hodes report landed on the desks of GOSH’s managers, they clearly faced a difficult dilemma. If they made the findings public, the inevitable media scrutiny might have damaged their reputation and slowed the progress of their Foundation Trust application. If they edited out GOSH’s failings, they might leave themselves open to the claim of “cover up”. An unkind observer might conclude that GOSH’s board is still trying to reduce its exposure to public criticism. I have no evidence to prove that this was (or is) the calculus of GOSH’s managers. But as one person close to these events put it to me, if GOSH’s management team had been in Wigan they would almost certainly have departed by now. Perhaps GOSH is just too important to be seen to fail. Even when a child dies.

Richard Horton  
richard.horton@lancet.com
Scalpel solidarity: surgery in Palestine

In 2009, The Lancet Palestine Series called for increased international collaboration with Palestinian health professionals. David Holmes reports on how a group of UK surgeons are playing their part.

The tying of knots was much in vogue at the beginning of May. But few, with the exception of the newly wed British Royal couple, would have devoted more thought to the pastime than the 19 candidates at Augusta Victoria Hospital, East Jerusalem, who attended the first basic surgical skills course (BSSC) to be held in the occupied Palestinian territory. Like the Royal wedding, it all went smoothly. All the 19 candidates completed the training—vessel tying and all—to the exacting standards of the Royal College of Surgeons of Edinburgh (RCSE), UK. Unlike the Royal wedding it was done on a shoestring, and involved metres of sheep’s bowels.

The RCSE has taught the BSSC in the UK and internationally for about 17 years, but would never have made its way to occupied Palestinian territory had it not been for a family holiday. In December, 2009, Magdalena and Robin Kincaid, two surgeons at the Royal Cornwall Hospital in Truro, took their young family on a trip to Jerusalem. There they hired a car, and set about exploring the West Bank. What they found there would be familiar to readers of The Lancet Series on the occupied Palestinian territory: passionate, dedicated health professionals struggling to provide the best possible care for their patients against a backdrop of failing infrastructure and the paralysis caused by restrictions on the movement of patients and personnel throughout the territory. All they could think about, says Magdalena, was “what can anybody do to make things better?” One of the aspects that came through the conversations with [Palestinians] was “a lack of structure in anything”. This lack of structure is particularly acute when it comes to training.

As surgeons in the UK, the Kincaids both completed the BSSC. “It was the very basis of our training”, recalls Magdalena. “Equipping junior surgical trainees with good solid basic skills such as haemostasis, putting two bits of bowel together, incising the abdomen correctly so you don’t burst the bowel when you come in. These are all parts of the course”, she explains. Many of these skills are especially crucial for surgical trainees to master in the occupied Palestinian territory, where ambulances are often made to wait for hours at checkpoints. So, armed with nothing but their enthusiasm, the Kincaids resolved to take the BSSC to the occupied Palestinian territory.

What followed were months of “incredibly tedious but very necessary behind-the-scenes work” to secure the necessary sponsorship to fund the course. Instrumental to getting the idea off the ground were the links that the Kincaids established with the Juzoor Foundation for Health and Social Development (a Palestinian non-governmental organisation based in Jerusalem) and the UN Relief and Works Agency for Palestinian Refugees (UNRWA), both of which wrote letters to the RCSE in support of the Kincaid’s request that the college teach the BSSC in the occupied Palestinian territory. Links that the Kincaids hope will be exploited by other surgical colleges in future.

By September, 2010, the Kincaids had approval from the College, a team of faculty volunteers to teach the course (all of whom paid for all their own travel), and enough sponsorship to get everything they needed for the 16 training stations (including a full laparoscopic training kit) shipped to Jerusalem from as far afield as Dubai. All that was left was to negotiate Ben Gurion airport with an array of home-made abdominal incision kits, and buy up all the sheep and cow bowel they could find.

For the Palestinian candidates trying to get to the course, things were almost as complicated. As Dina Nasser, health adviser at the Juzoor Foundation, explains, “East Jerusalem is not accessible to all Palestinians, so permits had to be obtained from the Israeli military and two participants were denied permits”.

A pair of candidates practise suturing a vein patch using a school shirt and cow aorta
Despite some of his surgeons already being consultants, but were happy to attend.

The course itself was a resounding success, despite there being a far greater mix of ages and levels of experience than in a similar course in the UK. Some of the attendees were already consultants, but were happy to eat “humble pie”, said David Sedgwick, the course convener. For Ronza Salem—one of four female Palestinian surgical trainees—the course was invaluable (“especially the knots”) because it was so closely aligned with the work she does day in, day out. “It gave me more self confidence, because things like end-to-end anastomosis seemed so difficult, but after practicing it, it wasn’t that hard...it’s easy to understand and do it now”, she told The Lancet.

Tawfiq Nasser, Chief Executive Officer of Augusta Victoria Hospital, was also in no doubt as to the value of the course to his surgical staff. Despite some of his surgeons already practicing in the hospital’s cancer care centre, Nasser says that “the skills acquired in the course were very practical and added to improving their knowledge in updated new techniques in basic surgical skills. I rarely get such a unanimous opinion from all physicians and surgeons on anything.”

For Nasser, it is crucial that his staff are able to train inside the occupied territory. “Sending physicians and surgeons abroad is difficult because of medico-legal regulations that usually prevent our doctors from hands-on training abroad”, he explains. Equally important is that much of the available training abroad is linked to resources that simply do not exist in the occupied territory. “Hence, hosting experts who can train locally is a real privilege”, he says. In the future, Nasser is looking forward to having team training sessions “in which a full team is trained together (physicians or surgeons with nurses and others) so that we can develop skills uniformly among all clinical practitioners, serving the patient safely and with the highest possible standard of quality care”.

But, as valuable as the BSSC was for the surgeons who were able to take part, many will question whether initiatives like this can make a difference more broadly to the quality of care in the occupied territory. “They will”, says Juzoor’s Nasser, “once they become established and able to be implemented regularly”. Tony Waterston, who has helped the UK Royal College of Paediatrics and Child Health to run its own child health diploma course in the West Bank since 2005, agrees. “The early signs are that [the diploma course] is influencing practice in relation to communication skills, use of guidelines, emergency care and management of long-term conditions”, he explains.

Waterston feels that initiatives like the BSSC, which “directly improve practical skills without increasing brain drain, and generate partnerships which may extend to other areas of learning”, are intrinsically valuable, but making them sustainable will be essential if they are to make a difference in the longer term. Crucial to these long-term aims will be reducing the reliance on outside sources of funding, and making the courses as close to self-financing as possible. For the child health diploma, Waterston explains, “we are now about to achieve local accreditation, which would mean that the students can be charged for attendance, and this should cover local costs”. Waterston is also hopeful that the increasing adoption of internet video messaging will help to ease some of the access problems caused by the restrictions imposed on movement throughout the occupied territory, especially in Gaza.

For Juzoor’s Nasser, the lasting partnerships that schemes like the BSSC help form are crucial for the long-term development of the Palestinian health system. “When teams such as the one form the Royal College of Surgeons cooperate with local Palestinian organisations with a vision to build local capacities, working hand in hand with Palestinian surgeons who are responsible for teaching juniors, or are in charge of surgical residency programmes, they are contributing to the building blocks of a national professional education system in health”, she told The Lancet.

One of the most important aspects of the BSSC is its emphasis on techniques that are easy to disseminate more widely; something that Salem is keen to get to grips with. “Already my colleagues are noticing the difference and are interested in similar training”, she explains. “I’m planning to help them and pass on some of the things that we have learned.” Perhaps as important, says Magdalena, is that an end to medico-legal regulations that usually prevent our doctors from hands-on training abroad, and generate partnerships which may extend to other areas of learning”, are intrinsically valuable, but making them sustainable will be essential if they are to make a difference in the
Mental health care for US veterans heavily criticised

The US Department of Veterans Affairs has been taken to court for failing to provide adequate mental health care for American soldiers returning from active service. Sharmila Devi reports.

American military veterans are hoping for a better system of mental health care and benefits after an appeals court in May ordered a drastic overhaul following “unchecked incompetence” by the US Department of Veterans Affairs (VA).

About 18 veterans commit suicide every day and veterans groups have warned that the wars in Iraq and Afghanistan will lead to a further influx of psychologically and physically injured soldiers, further stretching the federal department that has requested a total budget of US$132·2 billion for 2012.

Active-duty troops are also a cause of concern with 25 American soldiers killing themselves in April alone, equal to about half the deaths in Afghanistan during that month.

“No more veterans should be compelled to agonise or perish while the government fails to perform its obligations”, wrote Judge Stephen Reinhardt of the US Court of Appeals for the Ninth Circuit in San Francisco in a strongly-worded, 104-page opinion released on May 10.

Veterans for Common Sense, a non-profit group that along with Veterans United for Truth brought the lawsuit against the VA, said recent war veterans had filed more than 550 000 disability claims. There are about 5 million veterans, including 1·6 million who served in Iraq and Afghanistan.

“The United States constitution confers upon veterans and their surviving relatives a right to the effective provision of mental health care and to the just and timely adjudication of their claims for health care and service-connected death and disability”, Reinhardt wrote.

But hundreds of veterans have reported delays and difficulties just to prove their injuries were incurred in combat. Reinhardt pointed out there were no suicide prevention officers at any of the 800 community-based outpatient clinics that are used most frequently by veterans and that 70% of health facilities did not have systems to track potentially suicidal veterans.

“Instead of using psychologists to conduct the evaluations of post-traumatic stress disorder, they’ve been using social workers and even pharmacologists in some offices...”

“Yet it takes an average of more than 4 years for a veteran to fully adjudicate a claim for benefits”, Reinhardt wrote. “During that time, many claims are mooted by deaths.”

The ruling found that the initial processing of disability claims took longer than the VA’s goal of 120 days. A veteran’s disability rating depends on the severity of injuries and the size of a veteran’s family and this determines the amount of benefits to be given, ranging up to $2600 a month. The VA says that its process is meant to detect fraudulent claims.

There are about 5 million veterans enrolled in the VA’s health-care system, which includes 153 hospitals and 800 clinics. A 2007 report by the Department of Veterans Affairs Office of the Inspector General found substantial delays in timely referrals from VA doctors for treatment of post-traumatic stress disorder and depression. Fewer than half the patients received same-day mental assessments, whereas others had to wait as long as 2 months for a counselling session.

“When George [W Bush] Two started his wars [in Afghanistan and Iraq], everyone had magnets on their car saying ‘Support Our Troops’ but that didn’t last very long”, said Bob Handy, chair of Veterans United for Truth and a former navy chief who retired after 22 years of active duty in 1970. “Bush made no plans for the massive numbers of veterans coming back and needing help.”

Handy noted that President Barack Obama and Eric Shinseki, a former army chief of staff who is Secretary of VA, had promised to reduce suicide rates among veterans and to shorten delays in disability applications.

“Going back to Reagan and before, Presidents always say they are going to improve things for veterans but within weeks, the budget people come back and say it’s going to cost too much. In the final analysis, there are always higher priorities”, he said.

“Now that we have an all-volunteer army, people tell me that servicemen and women know what they are getting into and if they are injured, that’s that.”

The Justice Department is reviewing the court’s ruling and has yet to decide whether the VA will appeal. Veterans groups hope the VA will...
Many veterans with disabilities do not have stable housing, hampering treatment compliance with the judge’s 2–1 ruling that it should enact a new mental health-care plan that implements a quicker process to appeal denied benefits, provides timely treatment, and ensures that suicidal veterans are seen immediately.

The VA said it had hired more than 3500 mental health professionals in recent years and created a new policy requiring an assessment within 24 h of veterans new to the system if they are considered to be “in crisis”. Mental health assessments of all veterans not in crisis are to be done within 14 days. The VA said it met this goal 95% of the time.

But Gordon Ersheimer, the lead attorney for the two veterans groups that brought the San Francisco lawsuit, said a lot of the new personnel hired by the VA were lower-level staff. “Instead of using psychologists to conduct the evaluations of post-traumatic stress disorder, they’ve been using social workers and even pharmacologists in some offices. And they’ve been moving away from individual to group therapy because it’s cheaper. You can argue about it but re-traumatising a soldier by getting him to tell his story for the first time in a group setting is not the best way to go”, he said. “Over the last decade, the VA has also drastically reduced the number of in-patient beds for mental health care and put veterans on a waiting list for treatment. It’s a budget thing.”

Another lawsuit against the VA was filed in California on June 9, and said many veterans were left homeless because of the failure to provide stable housing to those with severe mental disorders.

The number of veterans thought to be homeless on any given night has been reduced to 76 500 people from 131 000 2 years ago, according to numbers cited by Shinseki, the VA Secretary. The federal department has said it plans to end veteran homelessness by 2015.

The California lawsuit asks a federal judge to use empty buildings on the West Los Angeles Medical Centre campus to provide permanent supportive housing for veterans who need a stable home for treatment. “This case could be brought anywhere in the country there’s a homeless vet”, said Mark Rosenbaum, an attorney with the American Civil Liberties Union Foundation of Southern California. “There should be no such thing as a homeless veteran in America.”

The lawsuit argues that veterans with severe cases of post-traumatic stress disorder, brain trauma and other disorders must live in permanent housing to allow access to necessary treatment and services.

The American Legion, another veterans organisation, said in a statement after the San Francisco court ruling that members had cheered when Shinseki told its 2010 convention that he would “break the back” of the disability claims backlog. Soldiers injured today face a benefits waiting list of more than 650 000 veterans.

“But pledges are one thing, action is another”, said Verna Jones, director of the group’s Veterans Affairs and Rehabilitation division. “Rather than appeal [the ruling] or drag the legal case out further, the VA should be willing to sit down with the American Legion and other advocates to resolve this issue.

“Shhh! Our suicide prevention co-ordinators are identifying about 1000 suicide attempts per month among the veterans we see in our medical facilities”, wrote Ira Katz, a VA deputy chief, in an e-mail on Feb 13, 2008. “Is this something we should (carefully) address ourselves in some sort of release before someone stumbles on it?”

Another email written by Norma Perez, a VA psychologist, suggested that counsellors in Texas made a point to diagnose fewer post-traumatic stress disorder cases.

A 2008 study by the RAND Institute for Civil Justice found that about 18% of soldiers returning from Iraq and Afghanistan were diagnosed with post-traumatic stress disorder and that 30 000 troops currently deployed have the condition or major depression.

In the UK, Prime Minister David Cameron has called for a “national change in attitude” towards mental health problems among veterans. Combat Stress, a British veterans group, said it had seen a 66% increase in referrals for mental health treatment in 5 years.

Sharmila Devi
Christine Borland’s work is always understated, usually pale in hue, and very still, inviting contemplation. She has a knack of placing objects in such a way as to interest, intrigue, and provoke thought, without saying much herself.

This exhibition occupies a splendid room in the Camden Arts Centre, perfectly suited to the display: an elevated Victorian interior painted utterly white, with a high ceiling and many windows. Across the entire room, between two hefty rings set at a high diagonal, is hung a great length of cloth that has been dipped in plaster of Paris, in a long arc. On each side of this dramatic diagonal—hanging like an inverse monochrome rainbow—stands a single sculpture, also white, and composed of plaster, held aloft to a comfortable viewing level on metal industrial trestles. These are two identical forms, objects which take the form of a human body.

Borland first encountered a modern cast of this human body on display in the museum of the Royal College of Surgeons in Edinburgh: it’s an old body. Later, she found the even older plaster cast in the basement of the anatomical department of the University of Edinburgh, and it’s this solid copy that she gained permission to work from for Cast From Nature. The body is that of an anonymous Victorian man, laid in the position of Christ in Michaelangelo’s famous Pieta in St Peter’s at Rome, chest raised up, with head, arms, and legs hanging. He is semi-dissected, the skin peeled back to the fascia on his chest, neck, and part of his face. No-one now seems to know who this man was, but the casting is associated with the famous Edinburgh anatomist Sir John Good sir (1814–67).

The cast was originally titled From Nature and is said to date to 1845. I wonder if it isn’t more likely to be the casting described in Goodsir’s Anatomical Memoirs (1868) of a man dissected and cast in layers in the autumn of 1858, who was “an Edinburgh carter of intensely whisky habits, who in a drunken state fell from his cart and died on the spot [and] remained free from decomposition during thirty days”. This figure bears similarities to better-known castings of flayed men of an earlier era: “Smugglerius”, a flayed smuggler posed as the Dying Slave, and the corpse of James Legg crucified, both in the Royal Academy schools in Piccadilly. But there are differences, too, not least that this man was not a hanged criminal: his body was probably obtained legally under the 1832 Anatomy Act, being unclaimed in an institutional mortuary.

Goodsir was an unusual anatomist, being also an artist, and a man who thought about humanity. He had an interest in the proportions of the human body and their relation to musical intervals, in human dignity, and in the immaterial principle that animates the human. His obituarist said of Goodsir, that “He loved his art for his art’s sake, and longed to apply his science for the science and truth’s sake.” His work on casting this body was not simply for the creation of a “scientific” specimen, the entire body was cast, legs, feet with slight bunions, working hands, weary head. This imperfect human body was cast as the image of God on Earth, and placed in the invisible arms of Mary. There was already art in the first cast.

What Borland has done with the cast of this dead man’s body is transformative. The two casts in the exhibition were created during her residency at the Glasgow Sculpture Studios, in which she used the process of casting as the spectacle of work in progress—playing with the positive and negative of the cast and its mould to create a new mould, from which the new casts were made. Her castings are not solid, but a fibreglass skin, hollow within, a sort of thin three-dimensional silhouette of the contoured surface of
the original, taking the shape of the original man, dead so long ago.

One of these castings shows the man as he was originally cast in the pieta position, supported by the trestle from below as if the grieving mother who should be holding him is there in the empty space around him. Gravity works to drag his body into the encircling invisible arms we know should be there. The other casting is a perfect duplicate of this, but inverted, and held aloft on its trestle so that from the back we can see the plaster and fibreglass which fill the hollow, riven with a wriggling morass of Borland’s fingermarks, created as the plaster was pushed into the crevices of the mould, kneaded into the interstices of the fibreglass, making the whole fabric of the body cohere before the plaster set. Walking round the hollow form allows a full appreciation of the importance of the inversity—the man looks like a sky-diver, his arms and legs float above his shoulders and hips, the ends of his loin cloth here float upwards, and give the appearance of wings. The most wonderful thing is his face, which smiles.

Ruth Richardson
anatomyruth@gmail.com

In brief

Book  A “difficult” patient’s journey
Chloë Atkins is the type of patient that every doctor dreads—presenting with a plethora of symptoms that don’t offer any obvious medical explanation. There are multitudes of such patients in a general practitioner’s roster and most, thankfully, will not turn out to have a serious illness. But there are a few who do, and as Atkins’ book points out, this can be a harrowing experience.

Atkins turned out to have an atypical presentation of myasthenia gravis, and it took nearly 20 years to get this sorted out. As doctor after doctor was unable to find a diagnosis, her syndrome was labelled as psychosomatic, and Atkins chronicles the increasing hostility of the medical profession towards her. The book presents the stark reality of how medicine faltering when faced with uncertainty.

Doctors notoriously resent uncertainty, and this becomes quickly manifest in the doctor-patient relationship. We all have those “difficult” patients, and Atkins bluntly puts herself in that category, candidly admitting that she “burned people out”. That doctors are uncomfortable with ambiguity shouldn’t come as any surprise, based on our education that is grounded in the seeming solidity of facts. But there is also the element of fear: which of these many patients with vague and varied symptoms harbours a serious illness? This needle-in-the-haystack pressure is compounded by the unfortunate reality of short outpatient visits (or overflowing inpatient wards). It is impossible for even brilliant doctors to discern complicated, mysterious illnesses in 15 minutes or less.

By now, most physicians have abandoned the blatant “it’s all in your head” fall-back. Most of us feel that the chronic pain syndromes, irritable bowels, and fibromyalgias do indeed have a biological basis, even if poorly understood. But stress does wreak havoc on these illnesses, so it’s not unreasonable to work to ameliorate this. Most of us view these illnesses as “syndromes” of some sort, and try to manage both the biological and psychological sides.

Atkins is furious at the medical profession for not diagnosing her illness earlier and for not taking her symptoms seriously (she calls the book a “justice narrative”). Her anger is entirely understandable. But in reading the book, I had to honestly wonder whether her condition was actually possible to diagnose. Even the clinicians who took her seriously were stymied by uncharacterisable symptoms and conflicting test results. Even if every doctor had been the paragon of attentiveness, respect, and doggedness, it is quite likely that this rare disease with its atypical presentation would have been missed.

Atkins’ argument that her doctors’ attitudes were the cause of her misdiagnosis is in my view the most tangible shortcoming of the book. Certainly, these attitudes are shameful, harmful, and in need of addressing, but it’s not possible to draw the conclusion of causality in such a complex and inscrutable case.

This book is the first in a series entitled “How Patients Think”, a complement—or retaliation, depending on your bias—to Jerome Groopman’s How Doctors Think. Patients, however, are much more heterogeneous than doctors, so such a series will likely offer a host of individual experiences that may or may not be universal. But this still has immense value. Doctors easily fall into the trap of categorising patients, and we need reminding of their individuality.

My Imaginary Illness may not have the philosophical or literary reach of, say, Anatole Broyard’s Intoxicated by My Illness, but it is instructive nevertheless. Atkins’ case of a rare illness is itself not generalisable, however the flaws in the system that she illuminates certainly are.

Danielle Ofri
author@danielleofri.com
the original, taking the shape of the original man, dead so long ago.

One of these castings shows the man as he was originally cast in the pieta position, supported by the trestle from below as if the grieving mother who should be holding him is there in the empty space around him. Gravity works to drag his body into the encircling invisible arms we know should be there. The other casting is a perfect duplicate of this, but inverted, and held aloft on its trestle so that from the back we can see the plaster and fibreglass which fill the hollow, riven with a wriggling morass of Borland’s fingerprint, created as the plaster was pushed into the crevices of the mould, kneaded into the interstices of the fibreglass, making the whole fabric of the body cohere before the plaster set. Walking round the hollow form allows a full appreciation of the importance of the inversity—the man looks like a sky-diver, his arms and legs float above his shoulders and hips, the ends of his loin cloth here float upwards, and give the appearance of wings. The most wonderful thing is his face, which smiles.

Ruth Richardson
anatomyruth@gmail.com

In brief

Book  A “difficult” patient’s journey
Chloë Atkins is the type of patient that every doctor dreads—presenting with a plethora of symptoms that don’t offer any obvious medical explanation. There are multitudes of such patients in a general practitioner’s roster and most, thankfully, will not turn out to have a serious illness. But there are a few who do, and as Atkins’ book points out, this can be a harrowing experience.

Atkins turned out to have an atypical presentation of myasthenia gravis, and it took nearly 20 years to get this sorted out. As doctor after doctor was unable to find a diagnosis, her syndrome was labelled as psychosomatic, and Atkins chronicles the increasing hostility of the medical profession towards her. The book presents the stark reality of how medicine falters when faced with uncertainty.

Doctors notoriously resent uncertainty, and this becomes quickly manifest in the doctor-patient relationship. We all have those “difficult” patients, and Atkins bluntly puts herself in that category, candidly admitting that she “burned people out”. That doctors are uncomfortable with ambiguity shouldn’t come as any surprise, based on our education that is grounded in the seeming solidity of facts. But there is also the element of fear: which of these many patients with vague and varied symptoms harbour a serious illness? This needle-in-the-haystack pressure is compounded by the unfortunate reality of short outpatient visits (or overflowing inpatient wards). It is impossible for even brilliant doctors to discern complicated, mysterious illnesses in 15 minutes or less.

By now, most physicians have abandoned the blatant “it’s all in your head” fall-back. Most of us feel that the chronic pain syndromes, irritable bowels, and fibromyalgias do indeed have a biological basis, even if poorly understood. But stress does wreak additional havoc on these illnesses, so it’s not unreasonable to work to ameliorate this. Most of us view these illnesses as “syndromes” of some sort, and try to manage both the biological and psychological sides.

Atkins is furious at the medical profession for not diagnosing her illness earlier and for not taking her symptoms seriously (she calls the book a “justice narrative”). Her anger is entirely understandable. But in reading the book, I had to honestly wonder whether her condition was actually possible to diagnose. Even the clinicians who took her seriously were stymied by uncharacterisable symptoms and conflicting test results. Even if every doctor had been the paragon of attentiveness, respect, and doggedness, it is quite likely that this rare disease with its atypical presentation would have been missed.

Atkins’ argument that her doctors’ attitudes were the cause of her misdiagnosis is in my view the most tangible shortcoming of the book. Certainly, these attitudes are shameful, harmful, and in need of addressing, but it’s not possible to draw the conclusion of causality in such a complex and inscrutable case.

This book is the first in a series entitled “How Patients Think”, a complement—or retaliation, depending on your bias—to Jerome Groopman’s How Doctors Think. Patients, however, are much more heterogeneous than doctors, so such a series will likely offer a host of individual experiences that may or may not be universal. But this still has immense value. Doctors easily fall into the trap of categorising patients, and we need reminding of their individuality.

My Imaginary Illness may not have the philosophical or literary reach of, say, Anatole Broyard’s Intoxicated by My Illness, but it is instructive nevertheless. Atkins’ case of a rare illness is itself not generalisable, however the flaws in the system that she illuminates certainly are.

Danielle Ofri
author@danielleofri.com
Profile
Leonard Hayflick and the limits of ageing

Eponyms have long been a feature of medicine and science. The textbooks are dotted with eponymous syndromes and diseases, with laws and constants, all flattering the memory of their inventors or discoverers. But there is only one eponymous limit in biomedicine: the Hayflick Limit, the number of times (about 50) that normal human embryonic cells can divide before they succumb to senescence.

Leonard Hayflick, professor of anatomy at the University of California at San Francisco, advanced the concept 50 years ago. The Hayflick Limit, he contended, was both an explanation for the phenomenon of ageing and a demolition of the wishful view (of some) that the human lifespan need have no upper limit. But although he correctly identified the cell nucleus as the location of the responsible mechanism, it fell to others to discern the structures involved. It was biologists Nobel Prize winners Elizabeth Blackburn and Carol Greider who showed how the cell keeps a tally of the number of times it has divided during its progress towards the Hayflick Limit. The ends of chromosomes carry structures called telomeres. Every time a cell divides the telomeres become shorter, this loss being the basis of what Hayflick described not as a clock (the process is not dependent on measuring time) but as a counting device, a “replicometer”.

Now 83 years old, Hayflick no longer has his own laboratory, but is still closely involved with ageing. He writes, lectures, does consultancy work, and, in a satisfying manifestation of natural justice, reckons that the ageing process is treating him well. “I think I’m pretty much ahead of the game for two reasons”, he says. “First of all I’ve absolutely no chronic medical problems that I’m aware of. And, perhaps more important, my mother is about to celebrate her 105th birthday.” The genes, if not the telomeres, are on his side.

Life expectancy in most developed countries is, and has for many years, been creeping upwards: some 25 years over the past century. Much of this is the consequence of success at dealing with our traditional killer, infectious diseases. But the rate of increase, Hayflick contends, will diminish as we approach what he believes to be our natural average lifespan of around 92 years. America’s thriving anti-ageing movement annoys him greatly. “The invention of ways to increase human longevity is the world’s second oldest profession, or maybe even the first. Individuals are going to the bank at this moment with enormous sums of money gained by persuading people that they’ve found either a way to extend your life or to make you immortal.” To imagine that the current rate of life expectancy increase will continue indefinitely is as absurd as extrapolating the diminishing time taken to run a mile and concluding that it will sooner or later be done in one second. “Everything in the Universe changes or ages with time, and to think that you can reverse it is nonsense.” When asked how he responds to the idea in principle of finding some way to postpone death indefinitely, he responds with dry humour that any such development would be expensive, and therefore available only to the rich and powerful. “I don’t know how many rich and powerful friends you have, but some of the ones I have I certainly don’t want to live another decade or two beyond normal.”

Hayflick’s career has not been without struggle. That he looks back with no regrets surely reflects that he emerged from his two major disputes so triumphantly. The first concerned the discovery that bears his name. Ageing research generated little interest in the USA when he first entered it. It had been dogma for 60 years that the eventual death of normal human cells in culture was not due to some inherent property, but caused by ignorance of the proper conditions under which to culture them. It took what Hayflick describes as “10 or 15 painful years” for the scientific community to accept what he’d discovered. “To torpedo a half century old belief is not easy even in science.”

The other and potentially more damaging dispute was with the US Government. The conflict began in the early 1960s when Hayflick was working at the University of Pennsylvania’s Wistar Institute. He developed a cell line dubbed WI-38. Because it was living material it couldn’t then be patented, so the vaccine companies to which it was freely distributed (it had proved to be exceptionally good for growing viruses) were earning a commercial return while the institution and the individuals who had created it were not. Hayflick set up his own company to distribute WI-38 and to hold payments for packaging and shipping until ownership was settled. The National Institutes of Health (NIH), which had previously supported the distribution, then accused Hayflick of stealing government property. He in turn filed a lawsuit against the NIH. After 6 years the action was settled out of court on terms that allowed Hayflick to continue distributing WI-38. The subsequent passage of the 1980 Bayh-Dole Act specifically allowed researchers at universities to apply for patents on federally funded inventions. “What I had done was embraced as policy by the US Government”, says Hayflick, who seems to look back on the saga less in sorrow than in glee. “Had my ideas not prevailed there would be no biotechnology industry.” It was, after all, mostly scientists who’d developed new techniques and materials while supported by public funds who went on to set up the small companies which created that industry. “I consider it one of my most important achievements.” As a researcher, Hayflick has shaped form as well as content.

Geoff Watts
geoff@scileg.freeserve.co.uk
Advance decisions (advance directives) were placed on a statutory basis in English law in 2007. Such decisions can be an important and highly valued way in which a person can ensure that his wishes will be respected when, due to lack of capacity, he is no longer capable of expressing them himself. Health-care professionals, under English law, are obliged to follow an advance decision if they are satisfied that it is valid and applicable to the circumstances. Similar legal arrangements are in place in many other jurisdictions including the USA.

The theoretical reasons for respecting advance decisions are straightforward but the practice can be highly problematic. Significant concern has, for instance, been expressed in the context of suicidal behaviour in which the person, for example who takes a dangerous overdose of drugs, also makes clear through an advance decision (perhaps left as a “suicide note”) that he does not want resuscitation to be attempted. The concern is that in some situations health professionals, either fearing the law or believing their qualms to be old-fashioned paternalism, will wrongly allow patients to die. Most of the discussion around the problematic nature of advance decisions has been concerned with life-threatening situations. But their scope is far broader. They can be of great significance, and no less problematic, in day-to-day care, and particularly in the setting of chronic mental illness that affects capacity.

Consider the following rather ordinary situation that might arise in a home for residents with dementia. Mr A is 76 years of age and lives in a care home. He is Jewish and throughout his life has affirmed his religion, including a commitment to avoid eating pork. Before his dementia progressed he clearly expressed the wish to avoid eating pork. Recently, he ate some bacon and pork sausages from another resident’s plate and is demanding that he is also given these foods for breakfast. The staff are concerned because of the importance that he had previously attached to avoiding pork. He continues to demand bacon and sausages and attempts to take food from other residents’ plates.

There are two moral reasons given widely for respecting previously stated preferences. First, advance decisions have value because they ensure that what happens to a person is what is best for him. One way of expressing this point is in terms of “critical interests”. These are interests that give our lives as a whole meaning and significance. For example, an eye surgeon who devotes her life’s work to treating people with cataracts in the developing world has consciously adopted a project of life that gives her life as a whole meaning and significance. Mr A was committed to his faith throughout his life and his advance decision suggests that continuing to behave in accordance with this faith is important to how he understands his life as a whole. Second, advance decisions provide a way for patients’ choices to be respected when patients can no longer express autonomous wishes. This is morally important over and above protecting critical interests. An advance decision (like a last will and testament) has value to the extent to which it is an expression of the person’s competent choices, whether or not they are part of his critical interests: of what gives his life meaning as a whole.

Taken together, these appear compelling reasons for ensuring that Mr A does not eat pork, however much he wants and enjoys it. But is it obviously right that his present preferences should be put to one side? Although his present decision-making ability is seriously compromised he retains the ability to express preferences and clearly gained pleasure from eating sausages and bacon. All of us want positive experiences and believe they contribute to our lives going well. Some enjoy sailing.
or reading literature, others enjoy watching television and walking in the park. One way that this point has been expressed is to say that we have not only critical interests but also “experiential interests”. Mr A’s new liking for pork can be viewed as an experiential interest and one that contributes to his life going well.

Despite his dementia, Mr A continues to be a human being whose current interests deserve respect. Why should he be held hostage to what he wanted in the past? One simplistic analysis inherent in US and English law is to respect a person’s current wishes if he retains decision-making capacity, and to follow a valid advance decision if he does not. By ignoring current experiential interests, such an analysis seems to give a clear answer to the dilemma facing Mr A’s carers: prevent him from eating pork. But even on this analysis things are not clear-cut. We may have robust procedures for determining capacity to refuse medical treatment but how would we judge whether Mr A has capacity to choose to eat pork? Would the fact that he no longer understands the religious grounds for avoiding pork he once thought important show that he lacks capacity to choose to eat bacon? It is not clear that it does. There is a danger of circular reasoning. We follow his advance decision only if he lacks capacity to make the choice, but in considering what he needs to understand in order to have capacity we first decide whether we favour giving priority to his advance decision or his current wishes and pleasures. We then pick the criteria for assessing capacity to give us the answer we want. The legal analysis fails us: we cannot rely on the idea of competence to solve the dilemma. There are two reasons therefore for being wary of simply following Mr A’s previous wishes and advance decision: it gives no weight to his current experiential interests; and it is not clear that he should be regarded as unable to choose to eat bacon.

We cannot, therefore, avoid making judgments about how the types and strengths of reasons a person had for their previous beliefs affect the weight to be given to past, as opposed to present, interests. Consider the following four people who avoid eating pork: Mr P does not like the taste; Mr Q is vegetarian but has no moral objection to eating meat and has adopted his diet as a lifestyle choice; Mr R is a religious family and continued to be a member of that religious community where eating meat is proscribed by the community but he has never been evangelical about trying to convert others and thinks it is for others to make their own choices.

Mr P would have had little reason to make an advance decision or prevent his future self from eating pork. He has no critical interest in avoiding pork, and if the experiential interest now lies in eating pork, that is what he should be enabled to do. Do Mr Q and Mr R have equal interests in avoiding pork once their tastes change and they no longer understand or care about their previous values and lifestyle? If Mr R is allowed to eat pork his former self would have viewed this as contributing to animal suffering and killing. He has strong reasons for wanting others to protect him from being even an unwitting party to such cruelty. Mr Q would seem to have lesser reasons for wanting such protection. A change in lifestyle is not a concern of the same gravity as contributing to something that you believe is morally wrong. Mr S might have reasons to be protected from eating pork comparable with Mr R, some may say stronger. But if a religious belief requires active endorsement then, after dementia, Mr S might be considered to no longer hold the religion.

The carers of people with chronic mental disorder will, not infrequently, be called upon to balance the past values and wishes—including statements made in advance decisions—with the present wishes and interests of those in their care. This will be the case for family carers looking after their relative at home; the staff of care homes and nursing homes, as in the example of Mr A; and hospital staff many of whose patients suffer such chronic mental disorders. How are carers to make these decisions? As we have seen, relying on the “algorithm”—if the person has capacity for the decision follow her current wishes, and if she lacks capacity follow her advance decision—simply will not do. The mere fact of an advance decision does not clinch the matter; and given, in any case, that most people will not have completed an advance decision, carers will need to judge how much weight to put on previous beliefs and values. In theory the important issue may be the strength of those beliefs; in practice carers will be able to judge the strength only by careful consideration of the grounds. Furthermore, the grounds are also important in assessing capacity. Consider again Mr Q. Because of dementia he no longer cares about his former lifestyle choice, and perhaps no longer remembers it, but does that mean that he lacks capacity to choose now to eat pork? It may be some comfort to carers facing these difficult decisions and thinking that they might be missing some ethical skill to know that there is no tidy way to come to a right answer. Careful thought, compassion, and wise judgment may in the end be all we have. No algorithm resolves the enigma of the carer’s art.

Tony Hope, John McMillan
The Ethox Centre, University of Oxford, Oxford OX3 7LF, UK (TH); and School of Medicine, Flinders Medical Centre, Adelaide, SA 5001, Australia (JM)
tonyhope@doctors.org.uk

Further reading
Jürgen F J Kun

Molecular biologist who made contributions to human genetics and malaria research. Born in Moers, Germany, on March 6, 1959, he died of melanoma in Tübingen, Germany, on April 28, 2011, aged 52 years.

Some scientists are like gadflies, following their curiosity as it flits widely from one field to another. Others take a more focused approach, exploring different facets of a particular topic in the hunt for ever more complete answers. Jürgen Kun’s contribution speaks of such a career, focused primarily on understanding malaria.

Kun’s first thesis, written as an undergraduate at the Institute for Genetics at the University of Cologne, Germany, was on the isolation and characterisation of antigen-producing clones from a genomic expression library of *Plasmodium falciparum*. Over the next 4 years, he elaborated on this work for his PhD, constructing an expression library from genomic DNA of the same parasite, and then isolating and characterising antigenic epitopes that could serve as potential vaccine candidates. “He made his choice of subject very early in life. In those early years he got very close to the parasite”, remarks his collaborator and friend, Professor Peter Kremsner from the University of Tübingen, Germany. “Plasmodium falciparum malaria is one of the most important infectious diseases, so it was an obvious choice to make.”

Kun was fascinated by the interaction between this deadly parasite and the human immune system, Kremsner says. “He came very much from the molecular genetics side of things. His passion was basic science but he was also very happy to combine that interest with the chance to make a potentially life-saving difference in this disease.” For his postdoctoral work, Kun travelled to Australia where he worked in the immunoparasitology unit of the Walter and Eliza Hall Institute of Medical Research in Melbourne, and in the Malaria and Arbovirus Unit at the Queensland Institute of Medical Research. Professor Qin Cheng, now head of the drug resistance and diagnostics unit at the Australian Army Malaria Institute worked with Kun during those years in the early 1990s. “He was very intelligent and knowledgable”, she remembers. “He was passionate about research and got very excited whenever he had a new idea.”

Returning to Germany in 1996, to take up a post at the Department of Parasitology at the Institute for Tropical Medicine in Tübingen, Kun’s focus shifted from *P falciparum* itself towards understanding the human factors that influence disease. “Some children are more susceptible to malaria and some more protected against frequent re-infections and severe outcomes”, explains Kremsner. “That was the topic he was tackling for the past 15 years.” Since malaria is endemic in tropical Africa, the main focus of his group lay with the genetic variants present in African populations. Kun established a close relationship with the Albert Schweitzer Hospital in Lambaréné, Gabon, where he and his collaborators undertook clinical and epidemiological studies. Crucially, they detected important mutations in the promoter region of the genes coding for the nitric oxide synthase 2, interleukin-10, and interferon gamma receptor. The nitric oxide synthase 2 mutation, which the group named NOS2Lambaréné, resulted in higher expression of NOS2. They found that children who carried the mutation tended to be protected against severe malaria to about the same degree as people with sickle cell anaemia are protected. “This helps explain why some kids are more vulnerable than others”, Kremsner says.

Beyond malaria, Kun’s group explored the effect of host genetics on other infectious diseases, seeking to understand how complex cellular and molecular mechanisms regulate the host immune response to parasitic infections. One project involved a collaboration with the Tran Hung Dao Hospital in Vietnam, where the effects of human polymorphisms in hepatitis B virus infections was explored. In 2009, Kun was appointed professor and deputy director of the Institute for Tropical Medicine in the Department of Parasitology at the University of Tübingen. Kun wasn’t a man who was afraid of letting a little mess accumulate in his office, but his mind was never chaotic. “He planned his ideas very well, and was very organised in his mind and thoughts”, says Kremsner. As a group leader, he oversaw the work of some 50 masters and doctoral students, many of them from Asia and Africa, and some of whom travelled to Germany for his funeral. Kun is survived by his wife, Jutta, and their two children.

Stephen Pincock
stephen.pincock@journalist.co.uk
Open letter to Ulf Wiinberg, Chief Executive of Lundbeck Pharmaceuticals

As clinicians and prescribers of Lundbeck’s products, we are appalled at the inaction of Lundbeck to prevent the supply of their drug, Nembutal (pentobarbital), for use in executions in the USA. You have stated that, as a Danish multinational, Lundbeck is opposed to the death penalty and that if you had a simple solution you “would be extremely happy to implement it”.

Lundbeck’s code of ethics states that your corporate values are to be “imaginative, passionate and responsible” and that “Our standards for suppliers shall be aligned with internal standards”.

Yet, Lundbeck has failed to live up to these standards by not insisting that if you had a simple solution you would be extremely happy to implement it.

Lundbeck’s products, we are appalled that if you had a simple solution you would be extremely happy to implement it. Lundbeck’s code of ethics states that your corporate values are to be “imaginative, passionate and responsible” and that “Our standards for suppliers shall be aligned with internal standards”.

It is time for Lundbeck to stop issuing platitudes and ensure that its supply would have a significant negative effect on patients’ care. Pentobarbital is registered for the treatment of status epilepticus or acute convulsive episodes. The drug is used for emergency episodes in acute-care hospital settings or, if accessible, at specialised epilepsy centres that provide comprehensive diagnostic and treatment services primarily or exclusively to people with refractory epilepsy. A survey by Lundbeck reveals the importance of this drug: nearly 70% of recognised epilepsy centres and more than 700 hospitals purchase pentobarbital on a regular basis.

Lundbeck initially investigated the possibility of fully withdrawing the product from the US market. However, dialogue with medical experts indicated that the discontinuation of its supply would have a significant negative effect on patients’ care. Pentobarbital is registered for the treatment of status epilepticus or acute convulsive episodes. The drug is used for emergency episodes in acute-care hospital settings or, if accessible, at specialised epilepsy centres that provide comprehensive diagnostic and treatment services primarily or exclusively to people with refractory epilepsy. A survey by Lundbeck reveals the importance of this drug: nearly 70% of recognised epilepsy centres and more than 700 hospitals purchase pentobarbital on a regular basis. To further understand the opinions of clinicians on this issue, Lundbeck has commissioned an additional survey. The preliminary results again support the continued availability of pentobarbital for its legitimate uses.

Lundbeck is further examining opportunities for restricting the supply of pentobarbital within the USA. Our main obstacle is that we are unlikely to be able to prevent state authorities from obtaining our product, even with the significant restrictions on distribution suggested by David Nicholl and colleagues.

We are committed to ensuring that patients with life-threatening forms of epilepsy have access to treatment. We will continue the dialogue with human rights organisations, the medical community, and authorities, and we will more proactively disclose our efforts in this respect.

I am the Chief Executive Officer of Lundbeck.

Ulf Wiinberg

kjha@lundbeck.com

H Lundbeck A/S, 2500 Valby, Denmark


Response from Lundbeck

Lundbeck is deeply concerned and strongly opposes the use of pentobarbital for executions. We have communicated this position directly to the state authorities in question and we have protested publicly against the erroneous use of our product. Since becoming aware of this misuse, we have tried to find ways to prevent the unacceptable use of pentobarbital.

Lundbeck initially investigated the possibility of fully withdrawing the product from the US market. However, dialogue with medical experts indicated that the discontinuation of its supply would have a significant negative effect on patients’ care. Pentobarbital is registered for the treatment of status epilepticus or acute convulsive episodes. The drug is used for emergency episodes in acute-care hospital settings or, if accessible, at specialised epilepsy centres that provide comprehensive diagnostic and treatment services primarily or exclusively to people with refractory epilepsy. A survey by Lundbeck reveals the importance of this drug: nearly 70% of recognised epilepsy centres and more than 700 hospitals purchase pentobarbital on a regular basis.

To further understand the opinions of clinicians on this issue, Lundbeck has commissioned an additional survey. The preliminary results again support the continued availability of pentobarbital for its legitimate uses.

Lundbeck is further examining opportunities for restricting the supply of pentobarbital within the USA. Our main obstacle is that we are unlikely to be able to prevent state authorities from obtaining our product, even with the significant restrictions on distribution suggested by David Nicholl and colleagues.

Our dedication to stop this misuse remains strong, and at the same time we are committed to ensuring that patients with life-threatening forms of epilepsy have access to treatment. We will continue the dialogue with human rights organisations, the medical community, and authorities, and we will more proactively disclose our efforts in this respect.

I am the Chief Executive Officer of Lundbeck.

Ulf Wiinberg

kjha@lundbeck.com

H Lundbeck A/S, 2500 Valby, Denmark


Belimumab for systemic lupus erythematosus

The study by Sandra Navarra and colleagues (Feb 26, p 721) on the treatment of active systemic lupus erythematosus (SLE) with the BlyS-specific inhibitor belimumab represents a substantial advance in clinical trials of biological agents in this disease. However, we wish to raise several points.

On the disease-assessment indices used, about 20% of patients had renal involvement at study baseline. Of these, around half had clinically significant lupus nephritis (proteinuria ≥ 2 g per day), although severe active nephritis was a study exclusion criterion. In both belimumab and placebo groups, there was a similar clinically significant rise in proteinuria of more than 2 g per day in 14–18% of the patients by study end. Whether this increase resulted from progressive renal disease in patients with pre-existing nephritis or de-novo organ involvement is unclear. Data on renal...
function (serum creatinine concentrations or glomerular filtration rate) are not provided.

Renal involvement remains the single most common cause of morbidity and a major cause of mortality in SLE. Furthermore, the degree of proteinuria is closely associated with renal outcome and cardiovascular events. These data suggest that belimumab might have no therapeutic benefits in lupus nephritis, but specific studies to clarify this issue are clearly warranted.

We declare that we have no conflicts of interest.

*Neeraj Dhaun, David C Kluth
bean.dhaun@ed.ac.uk
Centre for Cardiovascular Sciences (ND) and Centre for Inflammation Research (DCK), University of Edinburgh, Queen’s Medical Research Institute, Edinburgh EH12 9TJ, UK


Sandra Navarra and colleagues report the results of a successful trial of belimumab biotherapcy for systemic lupus erythematosus (SLE). Their methods are robust and use a new composite index custom-designed from the results of a negative preliminary study, but the conclusions with regard to the place of this drug in clinical practice should be tempered. First, the narrow inclusion criteria (only patients with serologically active SLE, with no severe forms of the disease, and with a third of patients on less than 7.5 mg per day of prednisone at inclusion) prevent extrapolation to patients at greatest risk, in terms of control of activity and steroid-sparing, for whom other drugs have been unsuccessful.

Second, the effects of belimumab seem only modest. The choice of a reduction from the baseline score of more than 4 points on the Safety of Estrogens in Lupus Erythematosus National Assessment SLE Disease Activity Index (SELENA-SLEDAI) was chosen as clinically relevant, whereas a minimum of 7 points had been defined as such by an expert panel.

Finally, the number of patients for whom treatment failed as a result of the use of restricted drugs (including statins and angiotensin-converting-enzyme inhibitors) is higher in the placebo group than in the belimumab group, and no sensitivity analysis was done despite the 18% withdrawal. If a conservative analysis (withdrawal in the placebo group regarded as a success, withdrawal in the belimumab group regarded as a failure) were done, the conclusion would be very different: the success rate in the placebo group would rise to 65% (higher than the unchanged 58% and 51% in the belimumab groups).

We declare that we have no conflicts of interest.

*Laurent Chiche, Noémie Jourde, Julien Mancini
laurent.chiche@ap-hm.fr
Service de Médecine Interne, Hôpital de la Conception, 12005 Marseille, France


Author’s reply

Although patients with severe systemic lupus erythematosus (SLE) were included in BLISS-52, those with severe active lupus nephritis and central nervous system manifestations were excluded. This limitation of the trial is cited in the discussion section of the Article and is reflected in the recent US Food and Drug Administration label for belimumab.

With regard to the magnitude of effect, we did post-hoc analyses to look at even higher improvement thresholds of the Safety of Estrogens in Lupus Erythematosus National Assessment SLE Disease Activity Index (SELENA-SLEDAI) of 5-8 points instead of 4 points. Examination of the effect of these higher SELENA-SLEDAI thresholds when incorporated into the SLE responder index showed that the treatment effect remained significant for both belimumab groups over the course of the study and at week 52, with a greater relative magnitude of effect than with standard therapy alone (table). These data suggest that the belimumab treatment effect is more robust than that seen with the primary efficacy endpoint of the SLE responder index with the lower 4-point SELENA-SLEDAI improvement threshold.

We did four prespecified sensitivity analyses of the primary endpoint of the SLE responder index and all remained significant for the belimumab 10 mg/kg group. Additionally, we did three post-hoc sensitivity analyses of the SLE responder index at the request of the US Food and Drug Administration. The first was a last-observation-carried-forward analysis that did not consider any patient as having a treatment failure owing to use of rescue medications. The second used all available data at treatment week 52 and did not consider any patient as having a treatment failure owing to use of rescue medications. The third considered any belimumab-treated patient as a non-responder and any placebo-treated patient as a responder if they used new statins, angiotensin-
converting-enzyme inhibitors, or angiotensin-receptor blockers. All of these analyses showed that belimumab 10 mg/kg remained superior to standard therapy alone (table).

The proportions of patients entering BLISS-52 with renal disease were 18–21% as measured by SELENA-SLEDAI and 12–17% on the basis of British Isles Lupus Assessment Group A/B organ domain scores, with 7–9% having a proteinuria concentration of 2 g or more per 24 h. The comment by Neeraj Dhaun and David C Kluth that the proportion of patients with proteinuria of 2 g or more per 24 h increased to 14–18% (on the basis of data in table 3 of the original article) is not correct. The safety data presented in table 3 represent the incidence of any adverse event or the worst laboratory abnormality (reported in any of 13 assessments) that occurred at any point during the course of the study. The proportion of patients at week 52 with proteinuria of 2 g or more per 24 h actually decreased from baseline to 4·2% and 5·5% in the belimumab 10 mg/kg and 1 mg/kg groups, respectively, versus 7·6% with standard therapy alone; a 2-grade shift in proteinuria from baseline to week 52 or from 2 g/24 h at baseline to ≥3·5 g/24 h at any time.‡ Grade 2, >1·5–3·0 × upper limit of normal (ULN); grade 3, >3·0–6·0 × ULN.

Additional BLISS-52 data

<table>
<thead>
<tr>
<th></th>
<th>Placebo (n=305)</th>
<th>Belimumab 1 mg/kg (n=313)</th>
<th>p</th>
<th>Belimumab 10 mg/kg (n=316)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effect of different SELENA-SLEDAI improvement thresholds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>84 (29·3%)</td>
<td>108 (37·5%)</td>
<td>0·0027</td>
<td>130 (44·8%)</td>
<td>&lt;0·0001</td>
</tr>
<tr>
<td>6</td>
<td>81 (28·2%)</td>
<td>106 (36·8%)</td>
<td>0·0020</td>
<td>127 (43·8%)</td>
<td>&lt;0·0001</td>
</tr>
<tr>
<td>7</td>
<td>49 (22·1%)</td>
<td>64 (29·2%)</td>
<td>0·0135</td>
<td>77 (33·3%)</td>
<td>0·0034</td>
</tr>
<tr>
<td>8</td>
<td>47 (21·5%)</td>
<td>63 (29·2%)</td>
<td>0·0088</td>
<td>73 (31·7%)</td>
<td>0·0057</td>
</tr>
<tr>
<td><strong>Post-hoc sensitivity analyses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last observation carried forward</td>
<td>147 (51·2%)</td>
<td>161 (55·9%)</td>
<td>0·0806</td>
<td>190 (65·5%)</td>
<td>0·0003</td>
</tr>
<tr>
<td>Last available data for medication failure*</td>
<td>135 (47·0%)</td>
<td>154 (53·3%)</td>
<td>0·0297</td>
<td>175 (60·3%)</td>
<td>0·0010</td>
</tr>
<tr>
<td>Belimumab non-responders for statin/ACE/ARB</td>
<td>128 (44·6%)</td>
<td>148 (51·4%)</td>
<td>0·0257</td>
<td>167 (57·6%)</td>
<td>0·0014</td>
</tr>
<tr>
<td><strong>Renal disease data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteinuria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥2 g/24 h at week 52</td>
<td>17 (7·6%)</td>
<td>13 (5·5%)</td>
<td>0·3597</td>
<td>10 (4·2%)</td>
<td>0·1142</td>
</tr>
<tr>
<td>2-grade shift†</td>
<td>16 (5·6%)</td>
<td>16 (5·6%)</td>
<td>0·9919</td>
<td>14 (4·9%)</td>
<td>0·6857</td>
</tr>
<tr>
<td>Reduction at week 52, g/24 h (baseline)</td>
<td>-0·04 (0·6)</td>
<td>-0·09 (0·5)</td>
<td>0·5076</td>
<td>-0·16 (0·6)</td>
<td>0·0631</td>
</tr>
<tr>
<td>Creatinine‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any 2-grade worsening</td>
<td>4 (1·4%)</td>
<td>2 (0·7%)</td>
<td>0·0859</td>
<td>2 (0·7%)</td>
<td>0·2155</td>
</tr>
<tr>
<td>Grade 3/4</td>
<td>3 (1·0%)</td>
<td>2 (0·7%)</td>
<td>1·000</td>
<td>0</td>
<td>0·123</td>
</tr>
<tr>
<td>Renal flares§</td>
<td>13 (4·5%)</td>
<td>12 (4·2%)</td>
<td>0·7190</td>
<td>4 (1·4%)</td>
<td>0·0333</td>
</tr>
</tbody>
</table>

SELENA-SLEDAI=Safety of Estrogens in Lupus Erythematosus National Assessment SLE Disease Activity Index. ACE=angiotensin-converting enzyme inhibitor. ARB=angiotensin-receptor blocker. *Non-medication-failure dropout was considered a non-responder. †From <1 g/24 h at baseline to ≥2 g/24 h at week 52 or from ≥2 g/24 h at baseline to ≥3·5 g/24 h at any time.‡ Grade 2, >1·5–3·0 × upper limit of normal (ULN); grade 3, >3·0–6·0 × ULN.

Milestones in treatment: the tipping point and the ResQ Trial

Peter Nagele, in his Comment (Jan 22, p 276),1 recommends not to adopt active compression-decompression cardiopulmonary resuscitation with an impedance threshold device (ACDCPR + ITD) until there is further independent verification of the benefits. In fact there is already such verification.

In three previous European clinical studies of out-of-hospital cardiac arrest,2–4 this device combination was found to increase significantly both circulation during CPR and 24-h survival rates by more than 50% compared with standard CPR. Now the ResQ Trial,5 of which several of us were authors, reports a significant 50% increase in survival rates with good neurological function up to a year after out-of-hospital cardiac arrest in the USA. No study so far has shown harm with this new approach. If the new CPR method were to be widely applied in Europe and North America, each year many thousands of additional lives could be saved.

Of course further verification of the ResQ Trial would be ideal. However, another large-scale, multicentre study is unlikely owing to the immense costs and time needed. In the meantime, Nagele suggests that we should...
Correspondence

without this already well verified treatment because in his opinion more evidence is needed. Despite this recommendation, Nagele has provided no evidence that contradicts the beneficial effects of application of ACD CPR + ITD on outcomes.

Here is the crucial question: should we withhold ACD CPR + ITD therapy, shown to be beneficial and lifesaving in four clinical trials while we wait for yet another study that is unlikely to happen, or should we broadly implement the new CPR therapy with a high likelihood of saving additional thousands of lives?

We declare that we have no conflicts of interest.

*Demetris Yannopoulos, Tom P Auferheide, Ralph J Frascone, Marvin A Wayne, Brian D Mahoney, Robert A Swor, Robert M Domeier, Michael L Olinger, Richard G Holcomb, David E Tupper

yanno001@umn.edu

Department of Medicine—Cardiovascular Division, and Department of Neurology, University of Minnesota Medical Center, Minneapolis, MN 55455, USA (DY, DET); Department of Emergency Medicine, Medical College of Wisconsin, Milwaukee, WI, USA (TPA); Department of Emergency Medicine, Regions Hospital, St Paul, MN, USA (RJF); Whatcom County Emergency Medical Services, Department of Emergency Medicine, Peace Health, St Joseph Medical Center, Bellingham, WA, USA (MAW); Department of Emergency Medicine, Hennepin County Medical Center, Minneapolis, MN, USA (BDM); Department of Emergency Medicine, William Beaumont Hospital, Royal Oak, MI, USA (RAS); Department of Emergency Medicine, St Joseph Mercy Hospital, Ann Arbor, MI, USA (RMD); Department of Emergency Medicine, Indiana University School of Medicine, Indianapolis, IN, USA (MLO); and Quintiles Consulting, Rockville, MD, USA (BGH)


In his Comment1 on the ResQ Trial,2 Peter Nagele suggests that enrolment of a larger cohort could have changed the findings of the ResQ Trial. However, figure 5 of the study shows results with an evident temporal consistency of the beneficial effect, making it extremely unlikely that the results would change.

Nagele recommends not to change cardiopulmonary resuscitation (CPR) practice on the basis of the results of the ResQ Trial and on the positive results of the previous four studies on active compression-decompression CPR with an impedance threshold device, including three by our group, claiming that there is not enough evidence.3 To support his statement, Nagele indicates that two components of the new CPR method when applied independently have not shown definitive benefit. He misses the point of the substantial physiological synergy with the combination of the two devices that led to the current study.4,5 Actually, the new clinical trial results were consistent with the results from all the European studies.

Nagele seems to suggest there are two approaches for assessing therapies for patients in cardiac arrest. One is extreme caution with any novelty even if it has gone through vigorous scrutiny in highly ranked journals; if adopted, this type of evidence-based innovation could save thousands of lives. The other is for adoption of established standards that have little or no clinical evidence to support their broad application.6 Clearly the evidence-based approach, such as is provided in the ResQ Trial publication, is what we should insist on.

We declare that we have no conflicts of interest.

Nicolas Segal, *Patrick Plaisance

secretariat.plaisance@lrh.aphp.fr

Hôpital Lariboisière, Service d’Accueil des Urgences, 75010 Paris, France

Author’s reply

Without a doubt, the ResQ Trial is a significant accomplishment for which the authors should be congratulated. Doing rigorous research and clinical trials in the out-of-hospital setting is tremendously difficult, particularly in the USA, and as a researcher myself I have nothing but the utmost respect for what the members of the Resuscitation Outcomes Consortium (and other involved researchers) have accomplished over the past years. Without hesitation, I fully support that only a rigorous, hypothesis-driven, evidence-based approach should guide therapeutic interventions (not only) in the prehospital setting.

As a Comment writer, however, the expectation of the reader and the journal is to do a thorough and unbiased assessment of a research paper and to put the findings into perspective. This was my goal. As much as I understand the enthusiasm of the involved researchers, the reader needs to be informed that, by themselves, neither active compression-decompression nor the use of an impedance-threshold device improved survival after out-of-hospital cardiac arrest. Most scientists would agree that a claim that only the combination of two treatments is effective when each treatment by itself is not, requires
very strong evidence—and quite often independent replication. This reasoning was behind my statement that independent replication of the findings would be ideal. I understand that such a trial might never happen owing to costs and logistical challenges, but one can at least suggest it. Again, I would like to congratulate the ResQ trialists for their substantial accomplishment and look forward for many high-quality clinical trials to come.

I declare that I have no conflicts of interest.

Peter Nagele
nagelep@wustl.edu

Department of Anesthesiology, Washington University School of Medicine, Box 8054, St Louis, MO 63110, USA


Social networking and health

By 2010, 60% of adults in the UK were using the internet every day or almost every day, and this figure continues to increase.1 Although the use of the internet is becoming increasingly common in areas such as e-commerce and social networking, health systems continue to lag in their use of such technology to communicate with patients.2 The conflicting opinions expressed in the World Report by Sharmila Devi (April 2, p 1141)3 only extend the confusion around the use of social networking for communication with patients. We need a more pragmatic approach to the introduction of new technologies. As well as seeking to produce new evidence, we should be using current evidence on how social networking might be used to improve communication with patients.

An alternative approach could involve considering the use of social networking in terms of wider clinical behaviour. Would you consider searching for additional information on your patient if the internet did not make it easy to do so with one click? Would you socialise with patients out of choice in a personal capacity? Would you feel threatened by your patients talking in the waiting room about their differing treatments? Most clinicians would feel confident in answering these questions.

Concerns about the effect of new technology on the doctor-patient relationship were probably being expressed when telephones were first introduced more than 100 years ago. Rather than viewing new technology as a threat, we should use the opportunities it offers to improve the efficiency and effectiveness of health systems and to improve people’s knowledge of their health and illnesses.

We declare that we have no conflicts of interest.

*Helen Atherton, Azem Majed
helen.atherton07@imperial.ac.uk

Department of Primary Care and Public Health, Imperial College London, London W6 BRP, UK


Sharmila Devi4 airs the ethical dilemmas surrounding the use of social networking sites by health professionals. Last year we formed a group of representatives from the Australian Medical Association, the New Zealand Medical Association, and the Australian and New Zealand Medical Students’ Associations, and produced guidelines on the use of social media by medical professionals.5

Our guidelines explore the issues that social media present for doctors, such as confidentiality and doctor-patient boundaries. Rather than prescriptive advice, we present scenarios and discuss the potential ethical and practical implications to educate readers. Unlike our American counterparts, we have not explicitly advocated the formal reporting of unprofessional online behaviour; instead we encourage medical practitioners to notify colleagues discreetly themselves.

Looking ahead, we believe that more research is required, particularly into the changing relationship between health-care providers and patients, as social media acquires a more prominent role in society, and into potential uses for social media in the delivery of health care. Although change is inevitable, maintenance of professional and ethical standards is essential to protect health professionals and patients. We will update our guidelines as more evidence emerges, and look forward to what transpires from other groups working in this area.

We declare that we have no conflicts of interest.

*Sarah J Mansfield, Michael A Bonning, Stewart G Morrison, Hugh O Stephens, Sheng-Hui Weng, Andrew W Perry, Rob C Oliver
s.mansfield1@gmail.com

Australian Medical Association, PO Box 6090, Kingston, ACT 2604, Australia (SJM, MAB, SGM, AWP); Australian Medical Students Association, Kingston, ACT, Australia (HOS, ROC); and New Zealand Medical Association, Wellington, New Zealand (SHW)


Facebook use leads to health-care reform in Taiwan

Social networking services are transforming the delivery of health care. Facebook, for example, is now commonly used by medical students, patients, and other stakeholders in the health-care system.6 7 We describe how Facebook enabled collaboration between stakeholders
Correspondence

in emergency-medicine policy in Taiwan, which led to reforms.

The Taiwan Society of Emergency Medicine1 has been in slow-moving negotiation with the Department of Health for the past several years over an appropriate solution to emergency-room overcrowding. A turning point was reached on Feb 8, 2011, when an emergency physician who was an active social network user and popular blogger among the emergency-room staff created a Facebook group called “Rescue the emergency room”. Within a week about 1500 people—most of the emergency department staff around Taiwan—became members of this group and started discussing actively and sharing their experiences. One of the members then posted the group’s concerns and problems on the Facebook profile of the Taiwanese Minister of Health. This caused the minister to join the group and get engaged in the discussion. A multiparty dialogue involving many different stakeholders and perspectives was suddenly possible.

Early on one of the members posted “Is there any use of these posts? Does our minister have time to read Facebook?” The Minister replied by posting “every message is read by me and my staff”. This modest gesture satisfied the emergency-room staff that their concerns were being taken seriously by the Department of Health, and further motivated them to engage in discussing the issue. By March 11, there were about 1800 members, 455 posts, and 3745 comments and “likes” on these posts. The number of members and posts is increasing every hour.

After monitoring these discussions, the Minister and his team decided to make a surprise visit to emergency departments in ten different cities. The next day, in a press release, he also promised to initiate dialogue with the Bureau of National Health Insurance on organisational issues affecting emergency departments and vowed to spend more resources for hospitals to improve emergency-room overcrowding and quality of care.

This case has implications for the future of health care, since it shows how social networking can break down the rigid social and professional hierarchical structures that can hinder reform.

We declare that we have no conflicts of interest.

Shabbir Syed Abdul, Che-Wei Lin, Jeremiah Scholl, Luis Fernandez-Luque, Wen-Shan Jian, Min-Huei Hsu, Der-Ming Liou, Yu-Chuan Li

jaak88@gmail.com

Institute of Biomedical Informatics, National Yang Ming University, Taipei, Taiwan (SSA, DML); Emergency Department, Taipei Medical University Hospital, Taipei, Taiwan (CWL); Health Informatics Centre, Karolinska Institutet, Stockholm, Sweden (JS); Northern Research Institute, University of Tromsø, Norway (LFL); School of Health Care Administration, Taipei Medical University, Taipei, Taiwan (WSJ); Department of Health, Taipei Hospital, Taipei, Taiwan (MHH); and Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei 110, Taiwan (YCL)

Early effect of the HPV vaccination programme on cervical abnormalities in Victoria, Australia: an ecological study

Julia M L Brotherton, Masha Fridman, Cathryn I May, Genevieve Chappell, A Marion Saville, Dorota M Gertig

Summary

Background Australia introduced a human papillomavirus (HPV) vaccination programme with the quadrivalent HPV vaccine for all women aged 12–26 years between 2007 and 2009. We analysed trends in cervical abnormalities in women in Victoria, Australia, before and after introduction of the vaccination programme.

Methods With data from the Victorian Cervical Cytology Registry between 2003 and 2009, we compared the incidence of histopathologically defined high-grade cervical abnormalities (HGAs, lesions coded as cervical intraepithelial neoplasia of grade 2 or worse or adenocarcinoma in situ; primary outcome) and low-grade cytological abnormalities (LGAs) in five age groups before (Jan 1, 2003, to March 31, 2007) and after (April 1, 2007, to Dec 31, 2009) the vaccination programme began. Binary comparisons between the two periods were done with Fisher’s exact test. Poisson piecewise regression analysis was used to compare incident rate trends.

Findings After the introduction of the vaccination programme, we recorded a decrease in the incidence of HGAs by 0·38% (95% CI 0·61–0·16) in girls younger than 18 years. This decrease was progressive and significantly different to the linear trend in incidence before introduction of the vaccination (incident rate ratio 1·14, 1·00–1·30, p=0·05). No similar temporal decline was recorded for LGAs or in older age groups.

Interpretation This is the first report of a decrease in incidence of HGAs within 3 years after the implementation of a population-wide HPV vaccination programme. Linkage between vaccination and screening registers is needed to confirm that this ecological observation is attributable to vaccination and to monitor participation in screening among vaccinated women.

Funding None.

Introduction Since the first prophylactic vaccine against human papillomavirus (HPV) was licensed in mid-2006, the quadrivalent vaccine (which provides protection against high-risk HPV types 16 and 18, and low-risk types 6 and 11, which cause 90% of genital warts) or bivalent vaccine (targeting HPV types 16 and 18) have been implemented in more than 28 countries as part of their national immunisation programmes and implemented at a sub-national level through donations in at least 17 developing countries.1 Persistent infection with high-risk genital HPV types is needed for the development of cervical cancer, and HPV types 16 and 18 are detected in 70% of cervical cancers, half of high-grade cervical abnormalities (HGAs), and a quarter of low-grade cervical abnormalities (LGAs) worldwide.2 Although the target age groups vary in different countries, the vaccine is aimed mainly at girls between the ages of 9 and 12 years because it is most effective when given before the onset of sexual activity, because it has no effect against HPV infection—which is transmitted sexually—once it has been acquired. Various countries have also chosen to implement short-term catch-up programmes aimed at older age groups, ranging from 13–18 years to 26 years.3

Australia was the first country to roll out an extensive, funded national HPV vaccination programme with the quadrivalent vaccine GARDASIL (Merck, Whitehouse Station, NJ, USA) in April, 2007, within the context of an already intensive and successful national cervical screening programme. The vaccination programme consists of a continuing component that targets 12–13-year-old girls in schools and two catch-up programmes, one for 13–17-year-old school girls, and one for 18–26-year-old women through general practice and community settings delivered between July, 2007, and December, 2009. In Victoria, the second most populous Australian state, the HPV vaccine programme in secondary schools began on April 16, 2007. Girls in school years 7 (ages 12–13 years), 10, 11, and 12 (ages 15–18 years) were offered vaccination in 2007, with the remaining two catch-up cohorts (aged 13–14 and 14–15 years in 2007) offered vaccine in 2008.4 Vaccine coverage estimates from the National HPV Vaccination Program Register for the school programme in Victoria show a three-dose coverage of 79% in first-year high-school students and 71% in final-year high-school students.5 A population-based telephone survey done in Victoria in early 2009 noted self-reported coverage rates of 74% for one dose, 69% for two doses, and 56% for three doses in young women aged 18–28 years.3 These data indicate that the programme probably achieved high coverage. Australia’s HPV vaccination programme includes the broadest funded catch-up age range in the world4 and

See Comment page 2057

Victorian Cytology Service
Registries, Victorian Cytology Service, East Melbourne, VIC, Australia
J M L Brotherton B Med,
C I May BAppSc,
G Chappell BAppSc,
A M Saville MB ChB,
D M Gertig ScD; and
Biosciences Research Division,
Department of Primary Industries, Bundoora, VIC, Australia (M Fridman PhD)

Correspondence to: Dr Julia M L Brotherton, Victorian Cytology Service Registries, PO Box 310, East Melbourne, VIC 8002, Australia
jbrother@vcs.org.au

www.thelancet.com Vol 377 June 18, 2011 2085
overlaps with the age cohort presently eligible for cervical screening in Australia. The National Cervical Screening Program policy recommends one cervical cytology test every 2 years, beginning at age 18 years (or 2 years after onset of sexual activity, whichever is later) until age 69 years. The National Cervical Screening Program was established in 1991, and since that time both cervical cancer incidence and mortality have halved. Participation rates in the programme are 61-2% of women every 2 years, 73-9% every 3 years, and 86-3% every 5 years. Monitoring of the early effect of the vaccine in Australia is helped by the existence of state and territory Papanicolaou (Pap) test registers that record nearly all cervical cytology and histology results and the National HPV Vaccination Program Register, which was established to support and monitor the HPV vaccination programme.

A rapid effect on infection with vaccine-targeted HPV types is predicted after the implementation of population-based HPV vaccination programmes. Indeed, early data from sexual health clinics in Australia suggest that the incidence of genital warts in Victoria began to decrease in the first year of the vaccination programme. However, because of the long lead-time between infection and development of malignant disease, the programme’s effect on cancer incidence will take decades to assess. Hence monitoring of cervical abnormality rates in a country such as Australia, with a longstanding high-quality cervical screening programme, is especially important because the effect on these abnormalities is more proximal than, but closely related to, the development of cervical cancer, and the treatment of such lesions is associated with morbidity and cost.

Here we present data from Victoria, reporting cervical abnormality rates in young women for the first 3 years (2007–09) after the introduction of a widely targeted population-based HPV vaccination programme.

Methods

Data collection

The Victorian Cervical Cytology Registry (VCCR) is one of eight Pap test registries in Australia and promotes regular participation of women in the National Cervical Screening Programme by sending reminder letters and enables the follow-up of women with abnormal Pap tests. In brief, follow-up of cervical abnormalities detected by screening programmes in Australia is guided by national recommendations, with incident LGAs generally followed up with another smear test after 12 months to establish whether the abnormality has resolved or whether colposcopy is needed. Patients with HGAs or possible HGAs are immediately referred for colposcopy. The VCCR compiles statistics for the purpose of monitoring and research.

The VCCR receives timely data for almost all cervical cytology and cervical histopathology taken in Victoria, with a population of more than 2.7 million girls and women. Less than 1% of women request for their test results not to be held on the VCCR. Cervical cytology results, coded by reporting laboratories with the Australian Standard Modified Bethesda coding schedule, are forwarded to the VCCR. Copies of relevant histopathology results are received from reporting laboratories and coded according to an in-house coding schedule, with most coding checked by a second staff member for quality assurance purposes.

De-identified data were extracted from the VCCR for all screening-related episodes between Jan 1, 2001, and Dec 31, 2009. To minimise the prevalent pool effect, which would result in prevalent lesions being regarded as incident because of an absence of preceding data, a clearance period of 2 years was applied to the data. We therefore analysed LGA and HGA incidence rates between 2003 and 2009.

The process of data exclusion from the analytical dataset is shown in the webappendix (p 1). Episodes that were not related to cervical diagnoses (eg, vaginal and non-cervical diagnoses) were excluded. Other exclusions included HPV DNA tests, non-diagnostic episodes (describing clinical procedures or treatment), and diagnoses obtained through colposcopy alone.

Data analysis

We aimed to find out whether the incidence of cervical abnormalities detected by screening has changed since

See Online for webappendix

<table>
<thead>
<tr>
<th></th>
<th>Before vaccination (Jan 1, 2003, to March 31, 2007)</th>
<th>After vaccination (April 1, 2007, to Dec 31, 2009)</th>
<th>Difference in proportions (95% CI)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of women screened</strong></td>
<td>13620</td>
<td>5538</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>≥18 years</td>
<td>13620</td>
<td>5538</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>18–20 years</td>
<td>86,356</td>
<td>50,644</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>21–25 years</td>
<td>237,599</td>
<td>152,531</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>26–30 years</td>
<td>281,767</td>
<td>177,776</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>≥31 years</td>
<td>1,798,842</td>
<td>1,178,351</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><strong>LGA incidence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18 years</td>
<td>1658 (12.2%)</td>
<td>691 (12.5%)</td>
<td>0.3% (–0.8 to 1.4)</td>
<td>0.6</td>
</tr>
<tr>
<td>18–20 years</td>
<td>9465 (11.0%)</td>
<td>5506 (10.9%)</td>
<td>–0.1% (–0.5 to 0.3)</td>
<td>0.6</td>
</tr>
<tr>
<td>21–25 years</td>
<td>18,671 (7.9%)</td>
<td>11,067 (7.3%)</td>
<td>–0.6% (–0.8 to –0.4)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>26–30 years</td>
<td>14,049 (5.0%)</td>
<td>7,810 (4.4%)</td>
<td>–0.6% (–0.7 to –0.5)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>≥31 years</td>
<td>44,408 (2.5%)</td>
<td>23,106 (2.0%)</td>
<td>–0.5% (–0.47 to –0.54)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>HGA incidence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18 years</td>
<td>109 (0.80%)</td>
<td>23 (0.42%)</td>
<td>–0.38% (–0.61 to –0.16)</td>
<td>0.003</td>
</tr>
<tr>
<td>18–20 years</td>
<td>1035 (1.20%)</td>
<td>593 (1.17%)</td>
<td>–0.03% (–0.15 to 0.09)</td>
<td>0.7</td>
</tr>
<tr>
<td>21–25 years</td>
<td>3639 (1.53%)</td>
<td>2699 (1.71%)</td>
<td>0.18% (0.10 to 0.26)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>26–30 years</td>
<td>3561 (1.26%)</td>
<td>2542 (1.43%)</td>
<td>0.17% (0.10 to 0.24)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>≥31 years</td>
<td>6320 (0.35%)</td>
<td>4397 (0.37%)</td>
<td>0.02% (0.01 to 0.04)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Data are number and percentage of women screened, unless otherwise stated. HGA=high-grade abnormality. LGA=low-grade abnormality. NA=not applicable.

Table 1: Number of individuals screened and incidence of low-grade cervical cytological abnormalities and high-grade cervical histopathological abnormalities before and after introduction of the national human papillomavirus vaccination programme, by age group
the introduction of the HPV vaccination programme in April, 2007, compared with the 4 years before its introduction. The incidence of histopathologically defined HGAs was our primary outcome measure, and the incidence of cytologically defined LGAs was our secondary outcome measure.

An LGA was defined according to the results of Pap tests, coded with the Bethesda system. Low-grade squamous intraepithelial lesions and atypical squamous cells of undetermined significance were classified as cases of LGA.

Histopathology results were used to define cases of HGA and cancer. HGA included all lesions coded as cervical intraepithelial neoplasia of grade 2 or worse or adenocarcinoma in situ, with invasive cancers grouped separately, and according to the national data dictionary and Australian Institute of Health and Welfare classification system. Cancer data are not presented in this paper because Victorian Cancer Registry data are not available for 2008 and 2009.

An LGA or HGA outcome was regarded as incident if it was a woman’s first LGA or HGA diagnosis, or a woman’s first abnormality that occurred at least 2 years (730 days) after a previous abnormality, with at least two negative tests in the intervening period.

A woman’s first HGA diagnosis was also regarded as incident if it occurred after an LGA diagnosis, irrespective of test results in the intervening period. No event was defined as incident if it occurred after a cancer diagnosis, meaning that these records were excluded from the analysis. Incidence rates were defined as the number of incident events per 100 women tested within 3 months.

**Statistical analysis**

HGA and LGA incidence rates were estimated for 3-month periods and stratified by five age groups,
which had different exposures to the vaccination programme (individuals aged ≤17 years, 18–20 years, 21–25 years, 26–30 years, and ≥31 years) and two periods: before vaccination (Jan 1, 2003, to March 31, 2007) and after vaccination (April 1, 2007, to Dec 31, 2009). Binary comparisons between the two periods for each age group were done with Fisher’s exact test.

Temporal trend analysis was used to test the hypothesis that HGA would decrease more in younger age groups than in older age groups after the introduction of HPV vaccination in April, 2007, and that this decrease would be detected at a population level as a progressive decrease (negative slope) in HGA incidence. Lowess smoothing (bandwidth 0·5) was used to show incidence trends over time. A quantitative comparison of HGA temporal trends before and after vaccination was done with piecewise Poisson regression analysis.16–18 In the context of a constant trend, the incidence rate ratio (IRR) was used as a measure of proportional change in incidence rate within a 3-month period. In the piecewise comparison of trends, IRR was used to estimate the
ratios of slopes for temporal trends before and after vaccination. StataSE (version 10) was used to do all statistical analyses.

**Role of the funding source**
There was no funding source for this study. MF, JMLB, and DMG had full access to data and JMLB had final responsibility for the decision to submit for publication.

**Results**
Table 1 shows the number of individuals included in the analysis and incidence rates for LGA and HGA diagnoses before and after introduction of the vaccination programme. Although a decrease in LGA incidence was recorded in age groups 21–25 years, 26–30 years, and 31 years and older, analysis of temporal trends suggests that these changes are a continuation of long-term trends that began before vaccination (figure 1). Figure 1 also indicates no decrease in LGA incidence in individuals aged younger than 18 years or those aged 18–20 years after the introduction of the HPV vaccination programme.

We recorded a significant decrease of 0·38% (95% CI 0·61–0·16; p=0·003) in HGA incidence in women younger than 18 years, beginning shortly after introduction of the HPV vaccination programme (figure 2), with a reduction from 0·85% in 2006 (the year before vaccination) to 0·22% in 2009 (p=0·003). We recorded no significant change in incidence in women aged 18–20 years although figure 2 shows a non-linear decline in incidence. Small increases in incidence were recorded in women aged 21–30 years (0·17–0·18%, 95% CI 0·10–0·26; p=0·0001) and in those aged 31 years or older (0·02%, 0·01–0·04; p=0·002; figure 2). Trends in the prevalence of LGA and HGA by age group and time are shown in the webappendix (pp 2–3), and accord with trends shown in figure 2.

A quantitative comparison of linear trends also showed a significant decrease in HGA incidence after introduction of the vaccination programme in individuals aged 17 years or younger but no significant decrease in those aged 18–20 years (table 2). Figure 3 shows predicted HGA incidence trends from piecewise regression models for the two youngest age groups. In girls aged younger than 18 years, there is a progressive linear decrease in the HGA incidence rate after the introduction of the vaccination programme; in those aged 18–20 years, the HGA incidence trend after introduction of vaccination is non-linear, and the decline is smaller and seems delayed (figure 3).

**Discussion**
This ecological analysis reports a decrease in the incidence of high-grade cervical lesions in girls aged younger than 18 years in the 3 years after the start of the HPV vaccination programme in Victoria. This decrease began soon after the introduction of the vaccination programme. In women aged 18–20 years, a decrease in incidence seems to have begun about 1·5 years after vaccine introduction. Our finding that the decrease in HGA incidence occurred in the youngest vaccination cohort before it occurred in the older, catch-up cohorts (who were more likely to have been previously sexually experienced) reinforces the appropriateness of the targeting of prophylactic HPV vaccines to pre-adolescent girls.

The strengths of our analyses are that we have almost complete population-based data about cervical-screening-related outcomes on the VCCR. Coding of histopathological abnormalities was done with the national standard classification, and a 6-month period was allowed for reporting of histology to the register and checking of data. Our definition of incident abnormalities was conservative, requiring both an extended time interval and two negative tests after a previous abnormality for new lesions to be defined as incident. Prevalence trends in our study were similar to the incidence trends and support the robustness of the findings. Our definition of the period after vaccination was also conservative because we defined this phase as starting at the introduction of the vaccination programme, rather than after the first date (4 months after its introduction) when women could...
have completed the three-dose course. This starting point allows for some vaccine effectiveness after receipt of one to two doses of prophylactic HPV vaccines, which is biologically plausible.15

The main limitation of our analysis is that it is ecological in nature, and therefore a causal link between the recorded decrease in incidence and the vaccination programme cannot necessarily be ascribed. To substantiate these findings, cervical cytology data should be linked to HPV vaccination register data to enable analysis of cervical abnormality rates and participation rates by vaccination status. Monitoring of the effect of the vaccine is complex and needs data from several sources regarding cancer and abnormality rates, participation in screening, adverse events, and HPV typing of cancers and abnormalities.26 However, we believe that our findings have strong biological plausibility and that the specific temporal association, differential by age (which is related to both coverage and likelihood of sexual activity and therefore HPV exposure before vaccination), suggests that the vaccination programme caused the decrease. Data from cohort studies and HPV vaccine trials indicate that the time from incident infection with HPV types 16 or 18 to development of cervical intraepithelial neoplasia of grade 2 or worse is often less than 12 months.21,27

New guidelines for the management of abnormalities detected by screening were adopted in Australia in 2006.13 These new guidelines were more conservative than the previous guidelines in the management of women with LGAs and are unlikely to have had an effect on the reported incidence of HGAs specifically in younger women; neither guidelines have specific recommendations targeting women aged 20 years or younger. Little is known about the characteristics of women who attend Pap screening before the recommended starting age in Victoria. However, few young women were screened—on average, 2000–3000 per year between 2003 and 2009. These women could have been screened because of a misinterpretation of the screening policy or they could have been at higher-than-average risk for HPV infection and cervical intraepithelial neoplasia. Some individuals could have been screened too early because they were sexually active early in mid-adolescence, meaning they would have received vaccination after they had become sexually active. However, this possibility could not explain our findings because the vaccine would be less effective for such individuals. Similarly, if they were screened early because they were deemed at high risk, we would expect the lesion prevalence to be higher not lower in those women. One scenario that could contribute to a decrease in incidence is if young women at high risk are preferentially no longer being screened. We believe such a scenario is unlikely for the following reasons: no significant decrease was apparent in the older catch-up cohorts; all vaccinated cohorts were targeted with the same information about the need for screening after vaccination; and the decrease in screening rates in younger women occurred before the introduction of the vaccination programme.

Understanding of the possible effect of vaccination on screening behaviour is important to exclude differential screening in vaccinated and unvaccinated women as an explanation for recorded changes in lesion prevalence. Widespread publicity that accompanied the vaccination roll-out emphasised the importance of continued screening, and a Victorian population-based telephone survey in 2009 found that 96% of women aged 18–28 years knew that Pap tests were still needed after vaccination.1 In Victoria, as in the rest of Australia, overall cervical screening participation by the target group of women aged 20–69 years has been stable for about a decade. However, in women younger than 35 years, a gradual decrease in participation has been recorded in the past decade.7 In Victoria, 58% of women aged 20–24 years and 70% of women aged 25–29 years had a Pap test between 2007 and 2009, compared with 62% of women aged 20–24 and 74% of women aged 25–29 years between 2004 and 2006.12 Reasons for this decrease are unclear but reported barriers to screening for young women include young women having a low awareness of the purpose of cervical screening, perceiving that the test would be embarrassing or painful, and reporting a lack of time or not even having thought of having a Pap test.13 There has also been an increase in the population of eligible women in Victoria, and a delay in health-service use in young women newly migrated to Victoria could be a contributing factor. A gradual decrease in the number of women who were screened too early (before 18 years of age) is evident in Victoria, perhaps as a result of increased efforts in education for practitioners; this improvement in compliance with screening recommendations is not temporally related to the introduction of the vaccination programme.
programme and is unlikely to explain our findings, because the denominator is screened women. As the cohorts vaccinated before becoming sexual active enter screening, data linkage between the vaccine and Pap registers will provide information about screening participation in both vaccinated and unvaccinated women, and will be crucial to confirm the emerging trends in the incidence of cervical abnormalities reported in this study.

We recorded no significant decrease in incidence of LGAs, which are a subset of acute HPV infections. Although HGAs are strongly associated with the detection of HPV types 16 and 18 (detected in >50% of all patients, probably more in young women),22 LGAs are associated less strongly with detection of HPV types 16 and 18 (about 25%; HPV types 6 and 11 are detected in about 10%).25 All 40 genital HPV types can lead to low-grade Pap test abnormalities, and most young women have concurrent infections with more than one type.26 Furthermore, physiological changes such as inflammation and atrophy can closely mimic the appearance of LGAs.27 Therefore, a reduction in infection with HPV types 16 or 18 might not result in a demonstrable decrease in the detection of LGAs on Pap tests. An intention-to-treat analysis from the phase 3 quadrivalent vaccine trials of more than 17’000 women aged 15–26 years recorded a statistically significant 19% reduction of any HGAs (with an average follow-up of 3–6 years), but a non-significant reduction in any Pap abnormality (11·3% reduction; difference 1·32 per 100 person-years at risk, 95% CI 0·74–1·90).28 Although an eventual decrease in LGAs because of vaccination in HPV-naive cohorts is predicted,29 these data emphasise that cervical abnormalities will continue to occur in vaccinated women in the future.30

We are aware of no other study to document the possible effect of a national HPV vaccination programme on cervical abnormalities at a population level (panel). We have shown a decrease in the incidence of HGA in young women after the implementation of the vaccination programme, and that this decrease occurred soon after vaccination. This finding suggests an urgent need to review the age at which cervical screening is begun in Australia and in other countries with national vaccination programmes that begin screening of women at a young age, because cost-effectiveness of screening will decrease for the youngest age groups screened. In countries that screen women at an older age, the effect of the vaccination will take longer to be seen. During the study period, we recorded no decrease in incidence of LGAs in women younger than 21 years (in whom LGA incidence was greater than 10%), which was to be expected because of the lower proportion of abnormalities that are due to vaccine preventable types. Long-term gradual decreases in LGA rates were, however, noted in women older than 21 years. Although more time and linked data analyses by vaccination status are now needed to substantiate these ecological results, our findings are a timely reminder that cervical screening programmes will need to adapt and respond to a post-vaccination environment in which lesion prevalence will decrease, accelerating the need to define workable screening algorithms, especially in vaccinated populations.30

Contributions
JMLB, DMG, and MF designed and were principal investigators of the study, with assistance from MS, GC, and CLM. MF prepared and analysed incidence data. Data interpretation was led by DMG. JMLB, and MS, with statistical interpretation by MF. JMLB and DMG wrote the first draft and all authors contributed to the final report.

Conflicts of interest
JMLB, DMG, and MS are investigators on an Australian Research Council Linkage Grant, for which CSL Biotherapies is a partner organisation. JMLB was an investigator on a national HPV prevalence study that received partial, equal, and unrestricted funding from CSL Biotherapies and GlaxoSmithKline. GC, MF, and CLM declare that they have no conflicts of interest.

Acknowledgments
We thank Grace Zampogna (VCCR) for her assistance with variable definitions and appreciate the methodological advice received from Ian Gordon (Statistical Consulting Centre, University of Melbourne, Victoria, Australia), Karen Canfell (NSW Cancer Council, NSW, Australia), and Carolyn Nickson (University of Melbourne, Victoria, Australia). The VCCR is fully funded by the Victorian Government and operated by the Victorian Cytology Service.

References
Global burden of disease in young people aged 10–24 years: a systematic analysis

Fiona M Gore, Paul J N Bloem, George C Patton, Jane Ferguson, Véronique Joseph, Carolyn Coffey, Susan M Sawyer, Colin D Mathers

Summary

Background Young people aged 10–24 years represent 27% of the world’s population. Although important health problems and risk factors for disease in later life emerge in these years, the contribution to the global burden of disease is unknown. We describe the global burden of disease arising in young people and the contribution of risk factors to that burden.

Methods We used data from WHO’s 2004 Global Burden of Disease study. Cause-specific disability-adjusted life-years (DALYs) for young people aged 10–24 years were estimated by WHO region on the basis of available data for incidence, prevalence, severity, and mortality. WHO member states were classified into low-income, middle-income, and high-income countries, and into WHO regions. We estimated DALYs attributable to specific global health risk factors using the comparative risk assessment method. DALYs were divided into years of life lost because of premature mortality (YLLs) and years lost because of disability (YLDs), and are presented for regions by sex and by 5-year age groups.

Findings The total number of incident DALYs in those aged 10–24 years was about 236 million, representing 15·5% of total DALYs for all age groups. Africa had the highest rate of DALYs for this age group, which was 2·5 times greater than in high-income countries (208 vs 82 DALYs per 1000 population). Across regions, DALY rates were 12% higher in girls than in boys between 15 and 19 years (137 vs 153). Worldwide, the three main causes of YLDs for 10–24-year-olds were neuropsychiatric disorders (45%), unintentional injuries (12%), and infectious and parasitic diseases (10%). The main risk factors for incident DALYs in 10–24-year-olds were alcohol (7% of DALYs), unsafe sex (4%), iron deficiency (3%), lack of contraception (2%), and illicit drug use (2%).

Interpretation The health of young people has been largely neglected in global public health because this age group is perceived as healthy. However, opportunities for prevention of disease and injury in this age group are not fully exploited. The findings from this study suggest that adolescent health would benefit from increased public health attention.

Funding None.

Introduction In 2008, the worldwide population of young people aged between 10 and 24 years was more than 1·8 billion, the largest cohort ever, representing 27% of the population. This number is projected to peak in 2032 at about 2 billion, with 90% of these people in this age group living in low-income and middle-income countries. The size of this population makes their health status of interest, not only as a determinant of future population health, but also for social and economic development. Adolescence is generally thought to be a time of good health, when disease burden is low. Although risk factors and the lifestyles that young people adopt might not affect their health during this period, they can have a substantial effect in later life and can potentially affect the health of future generations. For example, high patterns of physical activity that are adopted during youth and sustained thereafter are thought to have protective effects against the onset of cardiovascular diseases and type 2 diabetes.

A report of the global and regional patterns of mortality for young people aged between 10 and 24 years recorded 2·6 million deaths in 2004 from a worldwide population of 1·8 billion in this age group. However, data for mortality only partly indicate disease burden because they do not show the conditions and behaviours that can lead to premature mortality and future disability—eg, the large burden that is associated with non-lethal mental disorders, which are common in adolescents and young adults. Therefore, mortality data alone probably underestimate the potential importance of the contribution of adolescence to overall population health. An example is the onset of tobacco use and dependence, which typically occur during this period,10 the prevention of even a small number of adolescents from smoking could substantially reduce the burden on future health and health systems. Other important determinants of health risk emerging during adolescence relate to eating patterns, physical activity and weight, sexual behaviours, use of addictive substances, and the use of motorised transportation.

This paper describes the global and regional burden of disease arising in young people aged 10–24 years, and the contribution of risk factors to that burden. It aims to provide policy makers with comparative data by cause, sex, and different age ranges from early adolescence to young adulthood. We have several specific aims: to describe all-cause and cause-specific disability-adjusted

www.thelancet.com Vol 377 June 18, 2011 2093

Lancet 2011; 377: 2993–102
Published Online June 7, 2011
DOI: 10.1016/S0140-6736(11)60512-6
See Comment page 2058
Department of Health Statistics and Informatics (F M Gore MSc, V Joseph MSc, C D Mathers PhD) and Department of Child and Adolescent Health and Development (P J N Bloem MBA, J Ferguson MSc), WHO, Geneva, Switzerland; Centre for Adolescent Health, Royal Children’s Hospital and Murdoch Children’s Research Institute, Melbourne, VIC, Australia (Prof G C Patton MD, J Ferguson MSc, Prof S M Sawyer MD); and Department of Paediatrics, The University of Melbourne, Melbourne, VIC, Australia (Prof G C Patton, C Coffey, Prof S M Sawyer MD). Correspondence to: Ms Fiona M Gore, Department of Health Statistics and Informatics, WHO, 20 Avenue Appia, 1211 Geneva 27, Switzerland; goref@who.int
life-years (DALYs) across global regions for people aged 10–24 years with breakdown by 5-year age bands (10–14, 15–19, and 20–24 years); to describe incident disability caused by years lost because of disability (YLDs) across regions for this age group with breakdown by 5-year age bands as above; and to describe the main global and regional risk factors that contribute to incident DALYs for 10–24-year-olds (10–14 and 15–24 years).

**Methods**

**Data collection**

The data used in this analysis were from WHO’s 2004 Global Burden of Disease (GBD) study. This study uses several data sources to quantify global and regional effects of disease, injuries, and risk factors on public health, and to provide a comprehensive and comparable assessment of worldwide mortality and loss of health attributable to these causes. Population data for 2004 were from the UN population division. Data were divided into three 5-year groups (10–14 years, 15–19 years, and 20–24 years) for the estimation of DALYs, and into two groups (10–14 years and 15–24 years) for the estimation of DALYs attributable to health risks. Changes in DALYs in early adolescence (10–14 years), late adolescence (15–19 years), and young adulthood (20–24 years) were investigated and analysed by sex and region. WHO member states were classified into several groups based on their income and region. WHO member states were classified into several groups based on their income and region (webappendix pp 1–2). High-income countries were those in the Americas (n=3), the Mediterranean region (n=5), Europe (n=25), and the western Pacific (n=6). Low-income and middle-income countries were those from Africa (n=46); the Americas (n=31); the Mediterranean region (n=16), including Afghanistan, Iran, and Pakistan; Europe (n=27), including central Asia; southeast Asia (n=11); and the western Pacific (n=21) (webappendix p 1).

**Estimates of DALYs and risk factors**

The overall burden of disease was assessed with the DALY—a summary measure combining years of life lost because of premature mortality (YLLs) with YLDs for incident cases of the disease or injury. One DALY represents the loss of the equivalent of 1 year of full health. YLLs were calculated from the number of deaths at each age and multiplied by a global standard life expectancy for each age. The standard DALYs reported in this analysis use 3% discounted and non-uniform age weights, and differ from the discounted but non-age-weighted DALYs that are used in the Disease Control Priorities Project. YLDs for a specific cause for 2004 were calculated by multiplication of the estimated number of incident cases in that period by the average duration of the disease, and then by a weight factor. The weight factor showed the disease severity on a scale ranging from 0 (optimum health status) to 1 (death). The disability weights used for YLD calculations are mostly the same as those used in the 2000 and 2002 versions of the GBD studies. Further details about data sources and cause-specific methods are available from previous studies. Patton and colleagues’ study provides a summary of the data sources, coverage, compilation methods, and modelling for the estimates of all-cause mortality and cause-of-death estimates for the 2004 GBD dataset.

Although we calculated mortality and YLLs for 5-year age groups, WHO’s 2004 GBD study calculated estimates of incidence, prevalence, and YLDs for broad age ranges—namely, 0–4 years, 5–14 years, 15–29 years, 30–44 years, 45–59 years, and older than 60 years. For causes in which the ratio of YLDs to YLLs was less than 5, we assumed that the incidence (and YLD rate) had the same age pattern as for the mortality rate. This assumption applies for maternal conditions, injuries, and other fatal causes. For the other non-fatal causes, we assumed that incidence and YLD rates per 1000 population were constant across the 5-year age groups within each global burden of disease age group. Detailed information from UNAIDS about the age pattern for HIV incidence was used to calculate YLDs for HIV in 5-year age groups.

Information about the risk factors that cause or are associated with disease and injury is an important part of the GBD study. On the basis of the framework that was published in the comparative risk assessment, data and information for 24 global risk factors were obtained from WHO programmes and scientific studies of both for data exposure and for the causal associations of risk exposure to outcomes of disease and injury. The most current risk analyses were applied to the latest regional estimates of mortality and disease burden for a comprehensive set of diseases and injuries for 2004. Comparative risk assessment estimates of disease burden and injuries attributed to a risk factor or group of risk factors are based on a comparison with a counterfactual distribution of exposure that would result in the lowest population.
risk, irrespective of whether attainable in practice, which is referred to as the theoretical minimum-risk exposure distribution. Furthermore, many diseases can be caused by more than one risk factor—the sum of the mortality or burden of disease attributable to each of the separate risk factors is often more than the combined mortality and burden of disease attributable to the groups of these risk factors.

Statistical analysis

Data sources and uncertainty of estimates have been previously discussed for estimates of deaths and YLLs. A previous analysis of 95% uncertainty ranges for regional estimates of cause-specific mortality from the GBD study ranged from 1% for high-income countries to 15–20% for sub-Saharan Africa, which shows differential availability of data. Uncertainty ranges were generally larger for deaths from specific diseases. For example, the relative uncertainty for deaths from road-traffic accidents ranged from 3% for high-income countries to 25% for sub-Saharan Africa, and from 10% for high-income countries to 30% for sub-Saharan Africa for stroke. For the analyses reported here, which are restricted to the adolescent and young adult age ranges, uncertainty ranges are almost certainly larger than the ranges quoted above for all ages combined.

The YLD estimates from WHO’s GBD study are based on systematic assessments of the available data for incidence, prevalence, duration, and severity of several disorders. However, these assessments are often based on inconsistent, fragmented, and partial data from different studies; therefore, substantial data gaps and uncertainties remain. Uncertainties in YLD estimates are determined mainly by the uncertainty in epidemiological estimates for the prevalence and incidence of specific diseases and injuries, and in the distribution of disability severity that is associated with these factors. Previous assessments of YLD uncertainty for specific causes have accounted not only for typical values of measurement error in the input datasets, but also for expert judgment about the degree of uncertainty arising from the scarcity of available data for each region. The ranges for YLD uncertainty will generally be larger than those for mortality uncertainty, particularly with the additional assumptions that are used to impute estimates for adolescent age groups from the GBD analyses. However, for some causes there are specific and complete sources of information.

Although 2004 GBD estimates have similarly large ranges of uncertainty for some causes and regions, they provide useful information about broad relativities of disease burden, the importance of mortality and disability, and regional patterns of disease burden. Previous analyses of the levels of uncertainty in the GBD estimates reinforce the need for caution when global comparative epidemiological assessments are interpreted for adolescents. Although our results provide useful information about the relativities and inequalities in the burden of disease and risk factors in adolescents, care should be taken to not overinterpret small differences.

Role of the funding source

There was no funding source for this study. FMG, VJ, and CDM had full access to the all the data in the study, and FMG had final responsibility for the decision to submit for publication.

Results

The overall burden for both sexes was much higher for children younger than 5 years (700 DALYs per 1000 population) than for other age groups (figure 1). We noted a substantial decrease in the burden for children aged 5–9 years. The burden slightly fell in those aged 10–14 years when compared with those aged 5–9 years, before increasing steadily from late adolescence to early adulthood and later life (figure 1). Overall, DALY rates were equal between sexes until adulthood, except for the 15–19-year-olds for whom rates were higher in young women than in young men (152 DALYs per 1000 vs 136 DALYs per 1000). For those aged 25 years and older, rates were noticeably higher for men than for women and remained so into old age (figure 1).

The high rate of DALYs for girls and young women aged 10–24 years in Africa and southeast Asia, compared with boys and young men in these regions (table 1), is attributable to a female excess burden in 15–19-year-olds (figure 1). In Africa, the difference between males and females was only small in those aged 25 years and older (458 DALYs per 1000 vs 460 DALYs per 1000) when regarded across the lifespan. Worldwide, DALYs tended to be higher in males than in females (table 1). DALYs were lower in 10–24-year-olds than in those aged 25 years and older (134 DALYs per 1000 vs 232 DALYS per 1000) (table 1). Total DALYs for young people aged 10–24 years were about 236 million, representing 15·5% of the total DALY burden for all age groups. Of these DALYs, 93·4% were in low-income and middle-income countries, and more than half of these were in Africa (21·5%) and southeast Asia (31·9%) (webappendix p 3). The 10–24-year-olds are the only age

<table>
<thead>
<tr>
<th>Region</th>
<th>Male</th>
<th>Female</th>
<th>Both</th>
<th>Male</th>
<th>Female</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-income countries</td>
<td>85</td>
<td>78</td>
<td>82</td>
<td>159</td>
<td>135</td>
<td>147</td>
</tr>
<tr>
<td>Low-income and middle-income countries</td>
<td>184</td>
<td>232</td>
<td>208</td>
<td>458</td>
<td>460</td>
<td>459</td>
</tr>
<tr>
<td>Africa</td>
<td>143</td>
<td>109</td>
<td>126</td>
<td>228</td>
<td>176</td>
<td>201</td>
</tr>
<tr>
<td>Americas</td>
<td>147</td>
<td>144</td>
<td>145</td>
<td>274</td>
<td>232</td>
<td>253</td>
</tr>
<tr>
<td>Europe</td>
<td>130</td>
<td>93</td>
<td>112</td>
<td>328</td>
<td>207</td>
<td>263</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>140</td>
<td>170</td>
<td>154</td>
<td>290</td>
<td>275</td>
<td>283</td>
</tr>
<tr>
<td>Western Pacific</td>
<td>103</td>
<td>85</td>
<td>95</td>
<td>196</td>
<td>160</td>
<td>178</td>
</tr>
</tbody>
</table>

Table 1: Estimated total number of DALYs per 1000 population for those aged between 10 and 24 years, and 25 years and older.
group for whom worldwide DALYs were higher in women than in men in some regions. This difference is driven by higher female rates in Africa because of maternal mortality and disability, and, similarly, in southeast Asia (table 1).

Compared with the disease burden in those aged 15–19 years and 20–24 years, the burden in young people aged 10–14 years was low for both sexes across regions (figure 2). Africa had the highest all-cause rate of DALYs of all the regions, except for boys aged 15–19 years in the Americas who have a similar number of DALYs (figure 2). Males in this age group also had the smallest difference between regions (120 DALYs per 1000 in the western

Figure 2: Major causes of disease burden in DALYs in adolescents per 1000 population
(A) 10–24-year-olds. (B) 10–14-year olds. (C) 15–19 year-olds. (D) 20–24 year-olds. DALYs=disability-adjusted life years. TB=tuberculosis.
Pacific and 160 per 1000 in Africa) compared with, for example, women aged 20–24 years in whom the DALYs ranged from 110 per 1000 in the western Pacific, to almost 350 per 1000 in Africa (figure 2, webappendix p 3). The disease burden more than doubled across regions in groups aged 10–14 years and 20–24 years, with a clear

| Males | 10–24 years | | 10–14 years | | 15–19 years | | 20–24 years |
|--------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Cause | Total DALYs (x1000) (%) | Cause | Total DALYs (x1000) (%) | Cause | Total DALYs (x1000) (%) | Cause | Total DALYs (x1000) (%) |
| 10–24 years | | | | | | | |
| 1 Road traffic accidents | 93 (7.8%) | Unipolar depressive disorders | 115 (9.8%) | Unipolar depressive disorders | 193 (8.2%) | | |
| 2 Unipolar depressive disorders | 78 (6.6%) | Schizophrenia | 46 (4.0%) | Road traffic accidents | 127 (5.4%) | | |
| 3 Violence | 69 (5.8%) | Bipolar disorder | 44 (3.7%) | Schizophrenia | 96 (4.1%) | | |
| 4 Alcohol use | 62 (5.3%) | Abortion | 43 (3.7%) | Bipolar disorder | 88 (3.8%) | | |
| 5 Schizophrenia | 50 (4.2%) | HIV/AIDS | 38 (3.2%) | Violence | 81 (3.5%) | | |
| 6 Bipolar disorder | 45 (3.8%) | Road traffic accidents | 34 (2.9%) | Alcohol use | 71 (3.0%) | | |
| 7 Self-inflicted injuries | 35 (3.0%) | Self-inflicted injuries | 32 (2.7%) | HIV/AIDS | 70 (3.0%) | | |
| 8 HIV/AIDS | 32 (2.7%) | Maternal sepsis | 32 (2.7%) | Self-inflicted injuries | 67 (2.8%) | | |
| 9 Tuberculosis | 32 (2.7%) | Lower respiratory infections | 30 (2.6%) | Tuberculosis | 60 (2.6%) | | |
| 10 Asthma | 32 (2.7%) | Panic disorder | 30 (2.6%) | Lower respiratory infections | 60 (2.6%) | | |
| 10–14 years | | | | | | | |
| 1 Road traffic accidents | 15 (6.0%) | Lower respiratory infections | 15 (6.3%) | Unipolar depressive disorders | 28 (5.7%) | | |
| 2 Unipolar depressive disorders | 14 (5.4%) | Unipolar depressive disorders | 14 (6.1%) | Lower respiratory infections | 28 (5.6%) | | |
| 3 Lower respiratory infections | 13 (4.9%) | Asthma | 12 (5.1%) | Road traffic accidents | 26 (5.2%) | | |
| 4 Asthma | 10 (4.1%) | Migraine | 11 (4.8%) | Asthma | 23 (4.6%) | | |
| 5 Drownings | 10 (3.7%) | Road traffic accidents | 10 (4.2%) | Refractive errors | 19 (3.8%) | | |
| 6 Refractive errors | 10 (3.7%) | Refractive errors | 9 (3.8%) | Iron-deficiency anaemia | 17 (3.4%) | | |
| 7 Falls | 9 (3.4%) | Iron-deficiency anaemia | 8 (3.5%) | Falls | 16 (3.2%) | | |
| 8 Iron-deficiency anaemia | 9 (3.4%) | Falls | 7 (2.9%) | Migraine | 16 (3.2%) | | |
| 9 Schizophrenia | 6 (2.5%) | Diarrhoeal diseases | 6 (2.7%) | Drownings | 14 (2.9%) | | |
| 10 Lymphatic filariasis | 6 (2.5%) | Fires | 6 (2.5%) | Diarrhoeal diseases | 12 (2.4%) | | |
| 15–19 years | | | | | | | |
| 1 Unipolar depressive disorders | 34 (8.0%) | Unipolar depressive disorders | 53 (11.7%) | Unipolar depressive disorders | 86 (9.9%) | | |
| 2 Road traffic accidents | 33 (7.8%) | Schizophrenia | 23 (5.2%) | Schizophrenia | 46 (5.3%) | | |
| 3 Alcohol use | 30 (7.2%) | Bipolar disorder | 22 (4.9%) | Road traffic accidents | 46 (5.3%) | | |
| 4 Schizophrenia | 23 (5.3%) | Abortion | 17 (3.8%) | Bipolar disorder | 44 (5.1%) | | |
| 5 Bipolar disorder | 23 (5.3%) | Maternal sepsis | 14 (3.1%) | Violence | 26 (3.0%) | | |
| 6 Violence | 19 (4.9%) | Self-inflicted injuries | 13 (3.0%) | Self-inflicted injuries | 24 (2.8%) | | |
| 7 Drug misuse | 17 (4.3%) | Asthma | 10 (2.3%) | Asthma | 18 (2.0%) | | |
| 8 Asthma | 11 (2.6%) | Road traffic accidents | 13 (2.9%) | Panic disorder | 23 (2.7%) | | |
| 9 Self-inflicted injuries | 11 (2.6%) | Chlamydia | 10 (2.3%) | Asthma | 18 (2.0%) | | |
| 10 Drownings | 10 (2.5%) | Iron-deficiency anaemia | 9 (2.1%) | HIV/AIDS | 17 (2.0%) | | |
| 20–24 years | | | | | | | |
| 1 Road traffic accidents | 44 (8.7%) | Unipolar depressive disorders | 48 (9.9%) | Unipolar depressive disorders | 79 (7.9%) | | |
| 2 Violence | 41 (8.1%) | HIV/AIDS | 24 (5.0%) | Road traffic accidents | 56 (5.6%) | | |
| 3 Unipolar depressive disorders | 31 (6.0%) | Abortion | 24 (4.9%) | Violence | 47 (4.7%) | | |
| 4 Alcohol use | 28 (5.6%) | Schizophrenia | 21 (4.4%) | HIV/AIDS | 44 (4.4%) | | |
| 5 Self-inflicted injuries | 19 (4.0%) | Bipolar disorder | 20 (4.1%) | Schizophrenia | 42 (4.2%) | | |
| 6 Schizophrenia | 21 (4.0%) | Maternal sepsis | 18 (3.2%) | Bipolar disorder | 40 (3.1%) | | |
| 7 Bipolar disorder | 20 (4.0%) | Tuberculosis | 15 (3.2%) | Tuberculosis | 35 (3.5%) | | |
| 8 HIV/AIDS | 20 (3.9%) | Self-inflicted injuries | 14 (2.9%) | Self-inflicted injuries | 35 (3.5%) | | |
| 9 Tuberculosis | 20 (3.9%) | Panic disorder | 14 (2.9%) | Alcohol use | 32 (3.2%) | | |
| 10 War | 14 (2.7%) | Road traffic accidents | 11 (2.3%) | Abortion | 24 (2.4%) | | |

DALY=disability-adjusted life-year.

Table 2: Main causes of DALYs for 10–24-year-olds and for 5-year age groups

www.thelancet.com Vol 377 June 18, 2011
increase in disease burden from injuries in men aged 20–24 years (figure 2). In this age group, the disease burden in women is increased by maternal conditions particularly in Africa, southeast Asia, and the eastern Mediterranean, and by communicable disease in Africa (figure 2). With more than 200 DALYs per 1000 population, Africa is the region with the highest rate of DALYs in 10–24-year-olds. The lowest rates were in high-income countries, with only 82 DALYs, followed by the western Pacific with 95 DALYs (table 1, webappendix pp 1–2).

Figure 2 provides an overall picture of the distribution of major causes of DALYs by disease group in 10–24-year-olds across region and by sex. Further division of age groups showed an increasing burden of disease and disability for those aged 10–14 years, 15–19 years, and 20–24 years (figure 2). In the 10–24-year age group, injuries affected males more than females and were high across all regions (table 2, figure 2). Overall, neuropsychiatric disorders were the main cause of burden in high-income countries, especially in those aged 15–24 years (50 DALYs per 1000 males and 52 DALYs per 1000 females) (webappendix pp 4–5). The burden of disease from these disorders was also high in low-income and middle-income countries (figure 2). In southeast Asia and the eastern Mediterranean regions, injuries were as important as neuropsychiatric disorders (figure 2).

Disaggregation of DALYs into YLLs and YLDs showed that more all-cause DALYs in late adolescence and early adulthood were caused by incident disability rather than by mortality for those aged 10–24 years (figure 3). YLD rates doubled for males between 10–14 years and 20–24 years, and increased more than 2·5 times for females. YLD rates for females aged 15–19 years were more than double those for girls aged 10–14 years (figure 3). YLD rates for young people varied less between regions than did YLL rates (figure 4). YLLs ranged from 22 to 99 per 1000 males across the different regions (figure 4), and from 9 to 117 per 1000 females. YLDs ranged from 63 to 88 YLDs per 1000 males and from 67 to 115 YLDs per 1000 females (figure 4).

The overall burden of disabling disorders was dominated by causes that contributed more than 80% of the burden of YLDs. The six main causes of worldwide disability in both sexes were neuropsychiatric disorders (including substance misuse), unintentional injuries, infectious and parasitic diseases, maternal conditions, diseases of the sense organs, and respiratory disease (webappendix pp 4–5). Neuropsychiatric disorders were the main cause of YLDs in all regions (webappendix pp 4–5). These disorders ranged from mostly unipolar major depression (20%) and alcohol use (11%) in high-income countries, to unipolar major depression (12%) and schizophrenia (7%) in the eastern Mediterranean region, and unipolar major depression (7%) and bipolar disorder (5%) in Africa. Unintentional injuries—from mainly road-traffic accidents—are the second leading cause of YLDs worldwide and range from 6% of total YLDs in high-income countries to 16% in both southeast Asia and the eastern Mediterranean. Intentional injuries, which are mostly from self-inflicted injuries and violence, are the second leading cause of disability in the Americas with 8% of total YLDs and 5% of total YLDs in the eastern Mediterranean (webappendix pp 4–5). Infectious and parasitic disorders were a leading cause of YLDs in all regions except for in high-income countries. Worldwide, these disorders represented 10% of total disabilities and were the cause of more than a fifth of all disabilities in Africa (22% of total YLDs) (webappendix pp 4–5). Maternal conditions ranked fourth worldwide and within the top

---

**Figure 3:** DALY breakdown by YLLs and YLDs
DALY breakdown is per 1000 population for males and females for the age groups 10-14 years, 15-19 years, and 20-24 years. YLLs=years of life lost due to premature mortality. YLDs=years lost due to disability. DALY=disability-adjusted life-year.

**Figure 4:** DALY breakdown by YLLs and YLDs by region and sex in those aged 10-24 years
DALY=disability-adjusted life-year. YLDs=years lost due to disability. YLLs=years of life lost due to premature mortality.

---

**Table:**
<table>
<thead>
<tr>
<th>Region</th>
<th>YLLs (per 1000 males)</th>
<th>YLDs (per 1000 males)</th>
</tr>
</thead>
<tbody>
<tr>
<td>World</td>
<td>57</td>
<td>75</td>
</tr>
<tr>
<td>Western Pacific</td>
<td>37</td>
<td>66</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>Europe</td>
<td>52</td>
<td>78</td>
</tr>
<tr>
<td>Eastern Mediterranean</td>
<td>69</td>
<td>77</td>
</tr>
<tr>
<td>Americas</td>
<td>55</td>
<td>88</td>
</tr>
<tr>
<td>High income</td>
<td>22</td>
<td>63</td>
</tr>
<tr>
<td>Africa</td>
<td>99</td>
<td>85</td>
</tr>
</tbody>
</table>

---

**Table:**
<table>
<thead>
<tr>
<th>Region</th>
<th>YLLs (per 1000 females)</th>
<th>YLDs (per 1000 females)</th>
</tr>
</thead>
<tbody>
<tr>
<td>World</td>
<td>88</td>
<td>49</td>
</tr>
<tr>
<td>Western Pacific</td>
<td>67</td>
<td>19</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>101</td>
<td>68</td>
</tr>
<tr>
<td>Europe</td>
<td>74</td>
<td>19</td>
</tr>
<tr>
<td>Eastern Mediterranean</td>
<td>95</td>
<td>50</td>
</tr>
<tr>
<td>Americas</td>
<td>87</td>
<td>22</td>
</tr>
<tr>
<td>High income</td>
<td>69</td>
<td>9</td>
</tr>
<tr>
<td>Africa</td>
<td>115</td>
<td>117</td>
</tr>
</tbody>
</table>

---

**Figure:**
- **Figure 3:** DALY breakdown by YLLs and YLDs
- **Figure 4:** DALY breakdown by YLLs and YLDs by region and sex in those aged 10-24 years
six main causes of disability in each region. The percentage of total YLDs in females ranged from 2% of total YLDs in high-income countries to 16% in Africa (webappendix pp 4–5). Disorders of sense organs ranked fifth worldwide, contributing 5% of total YLDs. More than two-thirds of the YLDs were caused by refractive errors that can be corrected with the provision of glasses or other visual correction. Respiratory disorders were the sixth main cause of disability worldwide. Most of these YLDs were caused by asthma. The burden was highest in high-income countries where these disorders contributed to 7% of total YLDs (webappendix pp 4–5).

We calculated results of risk factors contributing to DALYs for 10–14-year-olds and 15–24-year-olds only. More detailed age distributions of exposure were not available. We selected the risk factors on the basis that they have a global spread, that data were available to estimate population exposures or distributions, and that known means exist to reduce them.5 The main risk factors for all ages (0–80 years and older) worldwide are underweight, unsafe sex, alcohol use, and unsafe water, sanitation, and hygiene.5 By contrast, the main risk factors contributing to DALYs in 10–24-year-olds were alcohol use, unsafe sex (increasing risk of disease transmission), iron deficiency,
Panel: Research in context

Systematic analysis
This paper presents information for a specific population (young people aged 10–24 years) on the basis of WHO's 2004 Global Burden of Disease (GBD) study. No previous published studies were available that provided a comprehensive picture of the global burden of disease among young people. Murray and Lopez describe the GBD approach as a metasynthesis—a construction of a comprehensive and comparable view of health problems using all available sources of information. Available data have many gaps and uncertainties because they are from regions with limited, incomplete, and poor quality data.

Interpretation
This is the first systematic description of global disease burden arising during adolescence and young adulthood. One important aspect of the Article is the consideration of the contributions of mortality and morbidity to disease burden with use of disability-adjusted life-years (DALYs) as a measure. It provides the most complete overview until now of disease burden for this age group and is seen as complementary to Patton and colleagues' 2009 paper, which provided an overview of global mortality patterns. Although mortality contributes of mortality and morbidity to disease burden with use of disability-adjusted life-years (DALYs) as a measure. It provides the most complete overview until now of disease burden for this age group and is seen as complementary to Patton and colleagues' 2009 paper, which provided an overview of global mortality patterns. Although mortality improves with economic development, the disease burden from disability is relatively high in all regions; therefore, non-fatal disease burden should be a main determinant of policy in young people. This finding in turn has implications for the types of data and health information systems that countries should use when informing youth health policies.

Discussion
Worldwide, young people bear a substantial burden of DALYs, both for YLLs and for YLDs, representing 15·5% of the total DALY burden for all age groups versus 18·5% in children younger than 5 years. This age group is the only one for whom DALYs were higher in women than men, notably in Africa and in southeast Asia. Africa had the highest regional rate of DALYs for those aged 10–24 years—2·5 times greater than in high-income countries. Differences in the causes of disease burden between high-income, middle-income, and low-income countries were substantial with very low rates of contribution from communicable diseases and maternal conditions in high-income regions. The contribution of YLLs and YLDs to overall disease burden also changed across country grouping. Improvements in levels of health (ie, low DALY rates) decreased the proportion of YLLs to overall DALYs to less than 25% but was close to 50% in the worst affected regions. This finding suggests that to improve health, policy makers should increasingly base decisions about young people's health on the disability-related proportion of DALYs because they represent the highest burden for the health system throughout the life-course.

The previous study of patterns of global mortality in young people concluded that investment in injury prevention—the main cause of death in this age group—had fallen behind investments in areas such as reproductive health and HIV/AIDS, a finding that is also supported in this analysis. DALYs for road-traffic accidents ranked second and violence was the fifth leading cause (table 2, webappendix pp 4–5). Importantly our analysis emphasises the causes of disease burden that rarely lead to death. Neuropsychiatric disorders, which have largely been overlooked in public health, are the leading cause of disability in young people in each region. However, this area tends to be poorly measured, making it a challenge to obtain a realistic estimate for the extent of the problem, especially in low-income and middle-income countries where communicable disease is often the research priority. Mental health is a mostly overlooked area in public health programmes of low-income and middle-income countries for all ages, but especially so in young people. Poor mental health in adolescence is associated with a high prevalence of adult emotional, behavioural, and severe psychiatric problems, and a large proportion of all adult mental health disorders start in adolescence.

The disease burden arising in early adolescence from major risk factors is low. However, rates rise sharply in late adolescence and early adulthood for both alcohol use and unsafe sex. For other risk factors that commonly start in adolescence such as tobacco use, low physical activity, high blood pressure, and overweight and obesity, their contribution to disease becomes apparent only in mid-to-late adulthood. The rising burden of non-communicable diseases is an increasing focus of global public health. Our risk factor data suggest that preventive strategies should adopt a life-course approach whereby the focus on the adolescent and young-adult years is prominent.
The epidemiological transition occurring with economic development has profoundly affected patterns of health but, until now, has not been substantially explored in adolescence and young adulthood. Our findings suggest that one consequence of this transition is the need to increase focus on non-communicable and non-fatal causes of disease burden both in adolescence and in later adult life. In turn, this focus is likely to shift attention to lifestyle risk factors and their social and environmental determinants.

This shift will present challenges in the political willingness among key stakeholders to invest in programmes that take many years to show their full effects. For example, the extent to which comprehensive tobacco-control programmes can prevent young people from becoming persistent smokers will affect mortality rates in the middle or second half of the 21st century. The introduction of vaccination of adolescent girls for human papillomavirus for the prevention of cervical cancers in high-income, low-income, and middle-income countries might indicate that willingness is growing to invest in prevention when new technologies and strong evidence for effectiveness of interventions become available.

Although methodological and data developments in the past decade have improved the empirical base for assessment of disease burden, substantial data gaps and uncertainties still remain, particularly for causes of death and levels of adolescent and adult mortality in Africa and parts of Asia. Improvements in population-level information about causes of death and the incidence, prevalence, and health states that are associated with causes of major disease and injury are still a main priority for national and international health and statistical agencies. Better information for young people than is currently available will need improved health-information systems, notably in efforts towards improving death-registration data and data obtained through household surveys and research studies. Such data systems and surveys should report results for more detailed age categories that are relevant to young people, rather than only for broad age ranges, as is often the case.

Until improved data become available, systematic assessments and syntheses of the available evidence will continue to provide important inputs for global health planning. Several areas still need further research development to address these data gaps, such as innovative methods involving sample registration and the use of verbal autopsy questionnaires in surveys. Research of strategies to improve comparability of cause-of-death certification and coding practices across countries is also a high priority. Nonetheless, for the first time, a systematic analysis of the global disease burden for young people aged 10–24 years is now available (panel).

**Contributors**
PJB, BJF, FMG, and GCP conceived the idea for the study. FMG, CDM, GCP, SMS, and CC compiled, analysed, and summarised the study estimates. Tables, graphs, and figures were prepared by FMG and VI, with input from all other authors. FMG led the writing of the paper with contributions from all other authors.

**Conflicts of interest**
We declare that we have no conflicts of interest.

**References**
Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial

Timothy S Maughan, Richard A Adams, Christopher G Smith, Angela M Meade, Matthew T Seymour, Richard H Wilson, Shelley Ildiaszczyczyk, Rebecca Harris, David Fisher, Sarah L Kenny, Edward Kay, Jenna K Mitchell, Ayman Madi, Bharat Jasani, Michelle D James, John Bridgewater, M John Kennedy, Bart Claes, Diether Lambrechts, Richard Kaplan, Jeremy P Cheadle, on behalf of the MRC COIN Trial Investigators

Summary

Background In the Medical Research Council (MRC) COIN trial, the epidermal growth factor receptor (EGFR)-targeted antibody cetuximab was added to standard chemotherapy in first-line treatment of advanced colorectal cancer with the aim of assessing effect on overall survival.

Methods In this randomised controlled trial, patients who were fit for but had not received previous chemotherapy for advanced colorectal cancer were randomly assigned to oxaliplatin and fluoropyrimidine chemotherapy (arm A), the same combination plus cetuximab (arm B), or intermittent chemotherapy (arm C). The choice of fluoropyrimidine therapy (capecitabine or infused fluorouracil plus leucovorin) was by intention to treat. Further analyses with respect to arm A and B, for which the primary outcome was overall survival in patients with KRAS wild-type tumours. Analysis was by intention to treat. Further analyses with respect to NRAS, BRAF, and EGFR status were done. The trial is registered, ISRCTN27286448.

Findings 1630 patients were randomly assigned to treatment groups (815 to standard therapy and 815 to addition of cetuximab). Tumour samples from 1316 (81%) patients were used for somatic molecular analyses; 565 (43%) had KRAS mutations. In patients with KRAS wild-type tumours (arm A, n=367; arm B, n=362), overall survival did not differ between treatment groups (median survival 17·9 months [IQR 10·3–29·2] in the control group vs 17·0 months [9·4–30·1] in the cetuximab group; HR 1·04, 95% CI 0·87–1·23, p=0·67). Similarly, there was no effect on progression-free survival (8·6 months [IQR 5·0–12·5] in the control group vs 8·6 months [5·1–13·8] in the cetuximab group; HR 0·96, 0·82–1·12, p=0·60). Overall response rate increased from 57% (n=209) with chemotherapy alone to 64% (n=232) with addition of cetuximab (p=0·049). Grade 3 and higher skin and gastrointestinal toxic effects were increased with cetuximab (14 vs 114 and 67 vs 97 patients in the control group vs the cetuximab group with KRAS wild-type tumours, respectively). Overall survival differs by somatic mutation status irrespective of treatment received: BRAF mutant, 8·8 months (IQR 4·5–27·4); KRAS mutant, 14·4 months (8·5–24·0); all wild-type, 20·1 months (11·5–31·7).

Interpretation This trial has not confirmed a benefit of addition of cetuximab to oxaliplatin-based chemotherapy in first-line treatment of patients with advanced colorectal cancer. Cetuximab increases response rate, with no evidence of benefit in progression-free or overall survival in KRAS wild-type patients or even in patients selected by additional mutational analysis of their tumours. The use of cetuximab in combination with oxaliplatin and capecitabine in first-line chemotherapy in patients with widespread metastases cannot be recommended.

Funding Cancer Research UK, Cancer Research Wales, UK Medical Research Council, Merck KGaA.

Introduction

The introduction and biomarker refinement of treatments targeting the epidermal growth factor receptor (EGFR) has been one of the most promising developments in oncology treatment in the past 5 years. The benefit of EGFR tyrosine kinase inhibitors (gefitinib and erlotinib) in lung cancer is limited to patients whose tumours contain a mutation at the drug-binding site in the ATP-binding domain of the receptor, and has been seminal in the biomarker-defined enrichment of the responsive populations.12 In colorectal cancer, no such mutations in EGFR occur, but clinical benefit has been shown with monoclonal antibodies, which bind to the extracellular receptor domain inhibiting ligand binding (notably epidermal growth factor, amphiregulin, and epiregulin) and receptor dimerisation.7 This clinical benefit, apparently limited to patients whose tumours contain no evidence of a mutation in KRAS, was first noted in non-randomised studies13 and was subsequently confirmed in randomised trials of antibody monotherapy in patients...
who were refractory to chemotherapy.5,6 This finding is plausible because KRAS encodes a G protein, which is a key link in the signal transduction pathway (RAS–RAF–MAP kinase) from receptor to nucleus, and the observed mutations result in constitutive activation of the pathway unlikely to be affected by cell surface receptor binding. Other activating mutations such as those in BRAF and NRAS in colorectal cancers might have similar negative effects on the efficacy of EGFR-targeted therapy.7

We present the results of the Medical Research Council (MRC) COIN trial, which was the largest trial of the addition of an EGFR-targeted monoclonal antibody (cetuximab) to chemotherapy (in this case a regimen of oxaliplatin and a fluoropyrimidine) in the first-line treatment of advanced colorectal cancer, and in which the effect was prospectively analysed primarily in relation to the mutational status of KRAS in tumour tissue, and secondarily in relation to the mutational status of BRAF, NRAS, and KRAS in tumour tissue. The COIN trial also assessed the effect of preplanned treatment interruptions in oxaliplatin and fluoropyrimidine combination chemotherapy on overall survival; these results are reported in a companion paper.8

Methods

Trial design and participants

The COIN trial protocol is available on the MRC Clinical Trials Unit website. Patients were eligible to participate in the trial if they had given written informed consent, had histologically confirmed adenocarcinoma of the colon or rectum, inoperable metastatic or locoregional measurable disease, had received no previous chemotherapy for metastatic disease, and had WHO performance status 0–2 and good organ function.

COIN was approved by national research ethics committees in the UK and Ireland and both the Medicines and Healthcare Regulatory Agency and Irish Medicines Board. The trial was undertaken by the MRC Clinical Trials Unit, following the principles of Good Clinical Research Practice, and overseen by an independent trial steering committee. Confidential interim analyses were reviewed at least annually by an independent data monitoring committee.

Randomisation and masking

Patients were randomly assigned with minimisation by the MRC Clinical Trials Unit via telephone (1:1:1 ratio) to receive continuous chemotherapy (control, arm A) or one of two research interventions: continuous chemotherapy plus cetuximab (arm B) or intermittent chemotherapy (arm C; figure I). The minimisation factors were hospital, WHO performance status, chemotherapy regimen, previous adjuvant chemotherapy, liver metastases, and peritoneal metastases. Treatment allocation was not masked. The arm A versus B results are reported in this paper and the arm A versus C results are reported separately.9

Procedures

Oncologists chose between two chemotherapy regimens according to local hospital policy or patient preference: oxaliplatin plus capecitabine, or oxaliplatin plus fluorouracil and folinic acid. Oxaliplatin plus capecitabine was given as a 3-weekly regimen of intravenous oxaliplatin 130 mg/m² over 2 h followed by oral capecitabine twice a day for 2 weeks. The dose of capecitabine was 1000 mg/m² orally twice a day, but was reduced to 850 mg/m² twice a day in a protocol amendment for patients in arm B only after 1775 (73%) patients had been randomly assigned to all groups, when an analysis of toxic effects showed that the rate of grade 3 or 4 diarrhoea was higher than expected (30%).9 Oxaliplatin plus fluorouracil and folinic acid was given as a 2-weekly regimen of intravenous L-folic acid 175 mg or D,L-folic acid 350 mg over 2 h given concurrently with oxaliplatin 85 mg/m² over 2 h, followed by intravenous bolus fluorouracil 400 mg/m² then fluorouracil 2400 mg/m² infusion over 46 h administered via an ambulatory pump and a central venous line. In arm B, cetuximab was given as an initial intravenous dose of 400 mg/m² over 2 h and subsequently at 250 mg/m² over 1 h once a week. Treatment was continued until disease progression, development of cumulative toxic effects, or patient choice. Patients were allowed to discontinue one or more agents within the regimen as a result of toxic effects, while continuing on the remaining agent or agents.

CT scans were done within 4 weeks before start of treatment, and were repeated every 12 weeks and assessed on the basis of RECIST (version 1.0) criteria.10 Because overall survival was the primary outcome measure of the trial, responses were not confirmed by repeat scans and external radiological review was not undertaken. Symptoms were scored with National Cancer Institute Common Terminology Criteria for Adverse Events (version 3.0).11 Serious adverse events and deaths, together with an assessment of causality, were continuously reported and assessed by an expert practising oncologist on behalf of the MRC.

In view of the emerging consensus that EGFR immunohistochemistry was not a reliable predictor of response to EGFR-targeted therapy,12 all patients irrespective of EGFR status were eligible for the COIN trial. However, all patients were asked to provide a tumour sample for future analysis of EGFR status. Consultation with the UK Human Tissue Authority and the research ethics committee concluded that the KRAS test could be done on all patients in addition to the EGFR test. Patients were also asked to provide additional consent for future bowel cancer research. Research staff at randomising sites requested patients’ tumour samples, which were stored at the Wales Cancer Bank.

Sections of the formalin-fixed, paraffin-embedded tumour blocks were stained with haematoxylin and eosin, optical images were stored, and tumour
material was macrodissected, cored, or laser-capture microdissected for the DNA analyses. Tissue microarrays were constructed for immunohistochemical analysis of EGFR. DNA was extracted with QIAmp DNA Microkits (Qiagen, Hilden, Germany). KRAS mutations in codons 12, 13, and 61 and BRAF mutations in codon 600 were screened by pyrosequencing. Additionally, KRAS (all three codons), BRAF (codons 594 and 600), and NRAS (codons 12 and 61) mutations were screened with MALDI-TOF mass array (Sequenom, San Diego, CA, USA).7 For KRAS, more than 1000 samples were successfully analysed by both techniques with greater than 99% genotype call concordance, and for BRAF more than 850 samples were analysed by both methods and greater than 98% of genotype calls were consistent. For discordant genotype calls, Sanger sequencing was used to establish genotype (webappendix p 1–3).

EGFR immunohistochemistry analysis was done on triplicate 0·6 mm cores in 25 tissue microarrays with validated control tissues stained with the standard US Food and Drug Administration approved EGFR pharmDx assay (DAKO, Glostrup, Denmark) at the University College London advanced diagnostics reference laboratory (London, UK). The cutoff points examined for positive versus negative tumours were 0 versus the rest, less than 10% versus 10% or more, and less than 20% versus 20% or more of total tumour cells showing membrane staining (webappendix p 3).

The original objective was to establish whether the addition of cetuximab improved overall survival in patients with advanced colorectal cancer. However, shortly after COIN completed recruitment, external evidence showed that anti-EGFR antibodies were unlikely to benefit patients with this disease whose tumours carry KRAS mutations.5,6 The decision was

---

**Patients with advanced colorectal cancer, measurable disease, fit for chemotherapy, no previous EGFR immunohistochemistry**

- 2445 consented
- 815 assigned to arm A
- 815 assigned to arm B
- 815 assigned to arm C
- Intermittent therapy

**KRAS**
- 527 BRAF wild type
- 613 BRAF mutant
- 18 NRAS wild type
- 18 NRAS mutant
- 13 test failed
- 268 KRAS mutant
- 367 KRAS wild type
- 362 KRAS wild type
- 297 KRAS mutant
- 9 test failed
- 616 BRAF wild type
- 45 BRAF mutant
- 637 NRAS wild type
- 32 NRAS mutant
- 9 did not start therapy
- 358 started therapy
- 35 started therapy
- 5 did not start therapy

**Capecitabine**
- 165 received full dose
- 67 received reduced dose
- 232 capecitabine-based
- 126 fluorouracil-based
- 137 fluorouracil-based
- 240 capecitabine-based
- 176 received full dose
- 64 received reduced dose

**Fluorouracil**
- 249 died
- 229 colorectal cancer
- 5 treatment-related
- 13 other
- 2 unknown
- 33 no data received for ≥6 months
- 76 on follow-up
- 78 on follow-up
- 26 no data received for ≥6 months
- 253 died
- 230 colorectal cancer
- 3 treatment-related
- 17 other
- 3 unknown

See Online for webappendix

---

**Figure 1: Trial profile**
EGFR=epidermal growth factor receptor.
taken to revise the primary research hypothesis before any analysis was done and before KRAS mutation data for COIN patients became available. The revised statistical plan was reviewed and approved by the independent data monitoring committee and independent trial steering committee. The revised primary objective thus became to determine whether the addition of cetuximab to continuous chemotherapy resulted in improved overall survival in patients with KRAS wild-type tumours. The secondary objectives were to evaluate whether the research intervention resulted in improved overall survival in four groups: (1) patients with tumours wild-type for all of KRAS, BRAF, and NRAS; (2) patients with KRAS mutant tumours; (3) patients with tumours mutant for any of KRAS, BRAF, or NRAS; and (4) all patients randomly allocated to treatment groups. Progression-free survival, response, and toxic effects were all evaluated in each of these patient groups. At the time of analysis for both overall and progression-free survival, survivors were censored at the date they were last known to be alive.

**Statistical analysis**

Originally the sample size for the arm B versus A comparison was 1614 patients, 807 in each group to parallel the arm C versus A comparison (see protocol for details of

<table>
<thead>
<tr>
<th>All patients</th>
<th>KRAS wild-type group</th>
<th>KRAS mutant group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>525 (64%)</td>
<td>543 (67%)</td>
</tr>
<tr>
<td>Female</td>
<td>290 (36%)</td>
<td>272 (33%)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (years)</td>
<td>63 (56–69)</td>
<td>63 (58–70)</td>
</tr>
<tr>
<td>&gt;75 years</td>
<td>74 (9%)</td>
<td>72 (9%)</td>
</tr>
<tr>
<td>WHO performance status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>375 (46%)</td>
<td>376 (46%)</td>
</tr>
<tr>
<td>1</td>
<td>378 (46%)</td>
<td>377 (46%)</td>
</tr>
<tr>
<td>2</td>
<td>62 (8%)</td>
<td>62 (8%)</td>
</tr>
<tr>
<td>Site of primary tumour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>453 (56%)</td>
<td>444 (54%)</td>
</tr>
<tr>
<td>Rectum</td>
<td>243 (30%)</td>
<td>262 (32%)</td>
</tr>
<tr>
<td>Rectosigmoid junction</td>
<td>113 (14%)</td>
<td>106 (13%)</td>
</tr>
<tr>
<td>Status of primary tumour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resected</td>
<td>445 (55%)</td>
<td>420 (52%)</td>
</tr>
<tr>
<td>Unresected</td>
<td>331 (41%)</td>
<td>345 (42%)</td>
</tr>
<tr>
<td>Local recurrence</td>
<td>39 (5%)</td>
<td>49 (6%)</td>
</tr>
<tr>
<td>Timing of metastases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metachronous</td>
<td>249 (31%)</td>
<td>239 (29%)</td>
</tr>
<tr>
<td>Synchronous</td>
<td>552 (68%)</td>
<td>569 (70%)</td>
</tr>
<tr>
<td>No metastases</td>
<td>7 (1%)</td>
<td>6 (1%)</td>
</tr>
<tr>
<td>Type of metastases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver only</td>
<td>174 (21%)</td>
<td>194 (24%)</td>
</tr>
<tr>
<td>Liver plus others</td>
<td>436 (53%)</td>
<td>418 (51%)</td>
</tr>
<tr>
<td>Non-liver</td>
<td>198 (24%)</td>
<td>197 (24%)</td>
</tr>
<tr>
<td>Number of metastatic sites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>283 (35%)</td>
<td>305 (37%)</td>
</tr>
<tr>
<td>2</td>
<td>326 (40%)</td>
<td>311 (38%)</td>
</tr>
<tr>
<td>&gt;2</td>
<td>199 (24%)</td>
<td>193 (24%)</td>
</tr>
<tr>
<td>Previous treatment for metastases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiotherapy</td>
<td>26 (3%)</td>
<td>24 (3%)</td>
</tr>
<tr>
<td>Surgery</td>
<td>142 (17%)</td>
<td>130 (16%)</td>
</tr>
<tr>
<td>Alkaline phosphatase &lt;300 U/L</td>
<td>670 (82%)</td>
<td>696 (85%)</td>
</tr>
<tr>
<td>Platelet count &gt;400 000 per μL</td>
<td>564 (69%)</td>
<td>549 (67%)</td>
</tr>
<tr>
<td>White blood cell count &gt;10 000 per L</td>
<td>577 (71%)</td>
<td>574 (70%)</td>
</tr>
</tbody>
</table>

Data are n (%) or median (IQR).

**Table 1: Baseline characteristics for all patients and by KRAS mutation status**
arm C vs A sample size calculation). With 1614 patients, a 6·4% advantage in overall survival at 2 years (from 20% to 26·4%; hazard ratio [HR] 0·828) could be detected with 90% power. When the primary objective of the arm B versus A comparison was prospectively revised to focus on the KRAS wild-type population, the primary analysis of arm A versus B was planned to take place when 511 overall survival events had occurred in patients with KRAS wild-type tumours. In this molecularly selected cohort, a higher HR of 0·76 could be detected at 87% power with a two-sided α of 0·05.

Analyses were undertaken according to a predefined statistical analysis plan, which was approved in advance by the COIN trial management group before the database was locked (Sept 2, 2009). All patients randomly assigned to treatment group were included in the analyses, on the basis of the intention-to-treat principle. All p values are two-sided and were not adjusted for multiple testing.

Time-to-event curves for analysis of overall and progression-free survival were estimated with the Kaplan-Meier method. HRs, confidence intervals, and p values were estimated with the log-rank method.

We compared worst toxic effects experienced overall between treatment groups using a χ² test, or Fisher’s exact test in case of low event rates (n<5). Exploratory analyses to identify predictive factors were done with a Cox proportional hazards model entering treatment group, the potential predictive factor, and a treatment-predictive factor interaction term. Interaction tests were done with likelihood-ratio tests of the null hypothesis that the interaction coefficient is zero. Stata (version 11.1) was used for all analyses.

The trial is registered, ISRCTN27286448.

Role of the funding source
The trial was conceived and developed by the National Cancer Research Institute advanced colorectal clinical studies group. The MRC was the overall sponsor of the study, with some responsibilities for the sites in Ireland delegated to the Irish Clinical Oncology Research Group.

**Figure 2:** Kaplan-Meier overall survival curves for patients with (A) KRAS wild-type, (B) KRAS mutant, (C) BRAF mutant and KRAS wild-type, and (D) KRAS, NRAS, and BRAF all wild-type tumours. HR=hazard ratio.
The MRC, through its employees in the MRC Clinical Trials Unit, are authors on the paper and were integral to the collection and analysis of the paper and the writing of the report. The corresponding author had full access to the data and had final responsibility for the decision to submit for publication.

**Results**

Between March 9, 2005, and May 9, 2008, 2445 patients were randomly assigned to treatment groups at 111 centres in the UK and Ireland, of whom 1630 patients were assigned to the arm A versus B comparison (figure 1). In the UK and Ireland, of whom 1630 patients were randomly assigned to treatment groups at 111 centres between March 9, 2005, and May 9, 2008, 2445 patients were assigned to the arm A versus B comparison (figure 1).

Between March 9, 2005, and May 9, 2008, 2445 patients were randomly assigned to treatment groups at 111 centres in the UK and Ireland, of whom 1630 patients were assigned to the arm A versus B comparison (figure 1). Their baseline characteristics are shown in table 1 and were well balanced between the two trial groups and across the different subgroups defined by KRAS mutation status.

Samples from 1316 (81%) patients were suitable for analysis of KRAS, BRAF, and NRAS mutation status (samples from 141 patients [9%] were not available and 173 [11%] samples contained insufficient tumour material for processing). Mutations were detected in KRAS in 565 (43%) patients, BRAF in 102 (8%), and NRAS in 50 (4%) with greater than 98% success rates (webappendix p 3). Figure 1 shows the full results by mutation site and by trial group. 706 (54%) patients carried a KRAS, BRAF, or NRAS mutation (no patients had KRAS and BRAF mutations or BRAF and NRAS mutations, and 581 (44%) patients were all wild type.

37 patients (2%) did not start trial therapy because of clinical deterioration after randomisation, patient choice, or because they were subsequently found to be ineligible. Figure 1 shows choice of chemotherapy (capecitabine-based or fluorouracil-based). 153 (19%) patients in the cetuximab group (arm B) started therapy with the lower dose of capecitabine of 850 mg/m². The median duration of follow-up among surviving patients with KRAS wild-type tumours randomly assigned to the control group was 21 months (IQR 18–29) and to the cetuximab group was 23 months (17–29), and was similar for the whole population.

At the time of analysis, there was no evidence of a difference in overall survival between treatment groups. 257 (71%) patients with KRAS wild-type tumours had died in each group, with a median survival of 17·9 months (IQR 10·3–29·2) in the control group and 13·6 months (IQR 9·5–22·8) in the cetuximab group (HR 1·04, 95% CI 0·87–1·23; p=0·67; figure 2). Of these deaths, 467 (90·9%) were due to colorectal cancer, eight (1·6%) were treatment-related, 32 (6·2%) were due to other causes, and the causes for seven (1·4%) are still unknown. There was no evidence of an effect of the addition of cetuximab to oxaliplatin-based chemotherapy on overall survival in any of the prespecified cohorts—eg, median survival in patients with KRAS mutant tumours was 14·8 months (IQR 9·5–22·8) in the control group and 13·6 months (8·0–26·1) in the cetuximab group (HR 0·98, 95% CI 0·81–1·17, p=0·80; figure 2, webappendix p 7).

We noted no evidence of an effect of cetuximab on the risk of progression in the KRAS wild-type group (median progression-free survival was 8·6 months in both groups, IQR 5·0–12·5 in the control group, 5·1–13·8 in the cetuximab group; HR 1·04, 95% CI 0·87–1·23; p=0·67; figure 2). Of these deaths, 467 (90·9%) were due to colorectal cancer, eight (1·6%) were treatment-related, 32 (6·2%) were due to other causes, and the causes for seven (1·4%) are still unknown. There was no evidence of an effect of the addition of cetuximab to oxaliplatin-based chemotherapy on overall survival in any of the prespecified cohorts—eg, median survival in patients with KRAS mutant tumours was 14·8 months (IQR 9·5–22·8) in the control group and 13·6 months (8·0–26·1) in the cetuximab group (HR 0·98, 95% CI 0·81–1·17, p=0·80; figure 2, webappendix p 7).

We noted no evidence of an effect of cetuximab on the risk of progression in the KRAS wild-type group (median progression-free survival was 8·6 months in both groups, IQR 5·0–12·5 in the control group, 5·1–13·8 in the cetuximab group; HR 1·04, 95% CI 0·87–1·23; p=0·67; figure 2). Of these deaths, 467 (90·9%) were due to colorectal cancer, eight (1·6%) were treatment-related, 32 (6·2%) were due to other causes, and the causes for seven (1·4%) are still unknown. There was no evidence of an effect of the addition of cetuximab to oxaliplatin-based chemotherapy on overall survival in any of the prespecified cohorts—eg, median survival in patients with KRAS mutant tumours was 14·8 months (IQR 9·5–22·8) in the control group and 13·6 months (8·0–26·1) in the cetuximab group (HR 0·98, 95% CI 0·81–1·17, p=0·80; figure 2, webappendix p 7).

We noted no evidence of an effect of cetuximab on the risk of progression in the KRAS wild-type group (median progression-free survival was 8·6 months in both groups, IQR 5·0–12·5 in the control group, 5·1–13·8 in the cetuximab group; HR 1·04, 95% CI 0·87–1·23; p=0·67; figure 2). Of these deaths, 467 (90·9%) were due to colorectal cancer, eight (1·6%) were treatment-related, 32 (6·2%) were due to other causes, and the causes for seven (1·4%) are still unknown. There was no evidence of an effect of the addition of cetuximab to oxaliplatin-based chemotherapy on overall survival in any of the prespecified cohorts—eg, median survival in patients with KRAS mutant tumours was 14·8 months (IQR 9·5–22·8) in the control group and 13·6 months (8·0–26·1) in the cetuximab group (HR 0·98, 95% CI 0·81–1·17, p=0·80; figure 2, webappendix p 7).
0·51–2·24, p=0·86) compared with 1·03 (0·74–1·44, p=0·85) in patients with other KRAS mutations.

Irrespective of treatment received, median overall survival was shorter in patients who had mutations in any of the three oncogenes (n=706, 13·6 months, IQR 7·9–23·1) than among those whose tumours were wild type for all of KRAS, NRAS, and BRAF (n=581, 20·1 months, IQR 11·5–31·7; log-rank test p=0·0001, figure 3). If the mutational type was then segregated, median overall survival was shorter in patients whose tumours had mutations in BRAF (n=102, 8·8 months, IQR 4·5–16·1) than in those with BRAF wild-type tumours but a mutation in KRAS (n=548, 14·4 months, 8·5–24·0) or NRAS (n=38, 13·8 months, 8·2–24·1). A global test for differences was highly significant (p<0·0001), as was a test for trend (p=0·0001, figure 3). The overall survival for the primary analysis cohort, patients with KRAS wild-type tumours, was 17·5 months (n=729, IQR 9·6–30·3). This finding is not shown in either figure because it contains prognostically mixed cohorts—ie, it also includes some patients with BRAF and NRAS mutant tumours.

A prognostic effect of patients’ tumour mutation status was also noted on progression-free survival, with median progression-free survival ranging from 5·6 months (IQR 3·8–16·0) for patients with wild-type tumours to 9·0 months (5·7–14·1) for those tumours wild type for all of KRAS, NRAS, and BRAF (n=102, 8·8 months, IQR 4·5–16·1) than in those with BRAF wild-type tumours but a mutation in KRAS (n=548, 14·4 months, 8·5–24·0) or NRAS (n=38, 13·8 months, 8·2–24·1). A global test for differences was highly significant (p<0·0001), as was a test for trend (p=0·0001, figure 3). The overall survival for the primary analysis cohort, patients with KRAS wild-type tumours, was 17·5 months (n=729, IQR 9·6–30·3). This finding is not shown in either figure because it contains prognostically mixed cohorts—ie, it also includes some patients with BRAF and NRAS mutant tumours. For progression-free survival, all three mutations remained independently significant (p=0·0001, p=0·0001, and p=0·0088 for KRAS, BRAF, and NRAS, respectively), whereas for overall survival, KRAS and BRAF remained significant (p=0·0001 for both), but NRAS did not (p=0·15). Among the other variables in the multivariate model were the four prognostic indicators suggested by Köhne and colleagues—ie, white blood cell count, number of metastatic sites, alkaline phosphatase, and WHO performance status. For both overall and progression-free survival, all four remained independently significant at the 5% level.

An agreed score for EGFR immunohistochemistry was available for 1065 (65%) patients. Reasons for unavailability of scores included no tissue block received, no satisfactory score achievable, and tissue regarded as unsatisfactory for analysis by immunohistochemistry. The cutoff point of less than 10% versus 10% or more of cells showing EGFR membrane staining was selected as being the most reproducible and widely used index for scoring of EGFR positivity. This approach identified 310 (58%) of 533 patients in the control group and 329 (62%) of 532 patients in the cetuximab group as having EGFR-positive tumours. Positive EGFR immunohistochemistry was a poor prognostic factor in the KRAS wild-type cohort (for progression-free survival, HR 1·27, 95% CI 1·07–1·52, p=0·0078 in univariate analysis and HR 1·25, 95% CI 1·05–1·50, p=0·015 after adjustment for prognostic factors found to be significant for progression-free survival in a separate

<table>
<thead>
<tr>
<th>Arm A</th>
<th>Fluorouracil-based therapy (N=281)</th>
<th>Capcitabine-based therapy (N=534)</th>
<th>p value Fluorouracil-based therapy vs Capcitabine-based therapy</th>
<th>Arm B</th>
<th>Fluorouracil-based therapy (N=279)</th>
<th>Capcitabine-based therapy (N=536)</th>
<th>p value Fluorouracil-based therapy vs Capcitabine-based therapy</th>
<th>p value for A vs B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platelet count</td>
<td>7 (3%)</td>
<td>16 (3%)</td>
<td>0·70</td>
<td>7 (2%)</td>
<td>16 (3%)</td>
<td>0·68</td>
<td>0·99</td>
<td>0·99</td>
</tr>
<tr>
<td>Haemoglobin</td>
<td>6 (2%)</td>
<td>7 (1%)</td>
<td>0·36</td>
<td>21 (7%)</td>
<td>17 (3%)</td>
<td>0·005*</td>
<td>0·0033†</td>
<td>0·038†</td>
</tr>
<tr>
<td>White blood cell count</td>
<td>28 (10%)</td>
<td>6 (1%)</td>
<td>0·0001*</td>
<td>33 (12%)</td>
<td>3 (1%)</td>
<td>0·0001*</td>
<td>0·52</td>
<td>0·32</td>
</tr>
<tr>
<td>Neutrophil count</td>
<td>86 (31%)</td>
<td>21 (4%)</td>
<td>0·0001*</td>
<td>88 (31%)</td>
<td>13 (2%)</td>
<td>0·0001*</td>
<td>0·90</td>
<td>0·17</td>
</tr>
<tr>
<td>Nausea</td>
<td>13 (5%)</td>
<td>37 (7%)</td>
<td>0·21</td>
<td>17 (6%)</td>
<td>48 (9%)</td>
<td>0·14</td>
<td>0·58</td>
<td>0·21</td>
</tr>
<tr>
<td>Vomiting</td>
<td>10 (4%)</td>
<td>27 (5%)</td>
<td>0·34</td>
<td>18 (6%)</td>
<td>38 (7%)</td>
<td>0·70</td>
<td>0·17</td>
<td>0·16</td>
</tr>
<tr>
<td>Diarrhoea</td>
<td>31 (11%)</td>
<td>82 (15%)</td>
<td>0·10</td>
<td>55 (20%)</td>
<td>141 (26%)</td>
<td>0·030†</td>
<td>0·0055†</td>
<td>0·0001†</td>
</tr>
<tr>
<td>Hand-foot syndrome</td>
<td>7 (3%)</td>
<td>25 (5%)</td>
<td>0·13</td>
<td>18 (6%)</td>
<td>67 (13%)</td>
<td>0·0064†</td>
<td>0·026†</td>
<td>0·0001†</td>
</tr>
<tr>
<td>Nail changes</td>
<td>0</td>
<td>0</td>
<td>0·94</td>
<td>6 (2%)</td>
<td>11 (2%)</td>
<td>0·94</td>
<td>0·0141</td>
<td>0·00045†</td>
</tr>
<tr>
<td>Skin rash</td>
<td>0</td>
<td>1 (1%)</td>
<td>0·99</td>
<td>56 (20%)</td>
<td>108 (20%)</td>
<td>0·92</td>
<td>0·0001†</td>
<td>0·0001†</td>
</tr>
<tr>
<td>Peripheral neuropathy</td>
<td>63 (23%)</td>
<td>86 (16%)</td>
<td>0·022*</td>
<td>38 (14%)</td>
<td>73 (14%)</td>
<td>0·95</td>
<td>0·0053*</td>
<td>0·28</td>
</tr>
<tr>
<td>Hypomagnesaemia</td>
<td>0</td>
<td>0</td>
<td>0·99</td>
<td>16 (6%)</td>
<td>16 (3%)</td>
<td>0·059</td>
<td>0·0001†</td>
<td>0·0001†</td>
</tr>
<tr>
<td>Anorexia</td>
<td>12 (4%)</td>
<td>31 (6%)</td>
<td>0·37</td>
<td>28 (10%)</td>
<td>43 (8%)</td>
<td>0·36</td>
<td>0·0141</td>
<td>0·14</td>
</tr>
<tr>
<td>Alopoeia</td>
<td>1 (1%)</td>
<td>2 (1%)</td>
<td>0·99</td>
<td>0</td>
<td>1 (1%)</td>
<td>0·99</td>
<td>0·32</td>
<td>0·57</td>
</tr>
<tr>
<td>Pain</td>
<td>34 (12%)</td>
<td>75 (14%)</td>
<td>0·47</td>
<td>34 (12%)</td>
<td>72 (13%)</td>
<td>0·58</td>
<td>0·98</td>
<td>0·81</td>
</tr>
<tr>
<td>Stomatitis</td>
<td>13 (5%)</td>
<td>4 (1%)</td>
<td>0·00040*</td>
<td>27 (10%)</td>
<td>18 (3%)</td>
<td>0·00021*</td>
<td>0·023†</td>
<td>0·0023†</td>
</tr>
<tr>
<td>Lethargy</td>
<td>52 (18%)</td>
<td>98 (18%)</td>
<td>0·99</td>
<td>81 (29%)</td>
<td>128 (24%)</td>
<td>0·13</td>
<td>0·0033†</td>
<td>0·023†</td>
</tr>
<tr>
<td>Vein pain</td>
<td>0</td>
<td>8 (1%)</td>
<td>0·056</td>
<td>0</td>
<td>3 (1%)</td>
<td>0·56</td>
<td>0·13</td>
<td>0·13</td>
</tr>
</tbody>
</table>

Data are n (%) or p value. *More toxic effects in arm A than in arm B, or in patients on fluorouracil-based therapy than on capcitabine-based therapy (p<0·05). †More toxic effects in arm B than arm A, or on capcitabine-based therapy than on fluorouracil-based therapy (p<0·05).

Table 2: Grade 3 or higher toxic effects in all patients randomly assigned to treatment groups, by chemotherapy regimen over entire treatment period.
HR=hazard ratio.

### Exploratory predictive factor analyses

<table>
<thead>
<tr>
<th>Sex</th>
<th>N</th>
<th>HR (95% CI)</th>
<th>Interaction p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>408</td>
<td>1.02 (0.74–1.41)</td>
<td>p=0.38</td>
</tr>
<tr>
<td>Female</td>
<td>173</td>
<td>1.00 (0.80–1.26)</td>
<td>p=0.22</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤65</td>
<td>338</td>
<td>0.81 (0.62–1.06)</td>
<td>p=0.02</td>
</tr>
<tr>
<td>&gt;65</td>
<td>241</td>
<td>0.84 (0.66–1.07)</td>
<td>p=0.68</td>
</tr>
<tr>
<td>Site of primary tumour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>301</td>
<td>0.96 (0.71–1.39)</td>
<td>p=0.0026</td>
</tr>
<tr>
<td>Rectum</td>
<td>196</td>
<td>0.90 (0.78–1.06)</td>
<td>p=0.12</td>
</tr>
<tr>
<td>Number of metastatic sites</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–1</td>
<td>230</td>
<td>0.73 (0.55–0.97)</td>
<td>p=0.036</td>
</tr>
<tr>
<td>≥2</td>
<td>351</td>
<td>1.07 (0.86–1.33)</td>
<td>p=0.38</td>
</tr>
<tr>
<td>Liver metastases only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>428</td>
<td>1.02 (0.84–1.25)</td>
<td>p=0.044</td>
</tr>
<tr>
<td>Yes</td>
<td>153</td>
<td>0.68 (0.48–0.97)</td>
<td>p=0.044</td>
</tr>
<tr>
<td>Synchronous metastases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>201</td>
<td>1.01 (0.75–1.35)</td>
<td>p=0.12</td>
</tr>
<tr>
<td>Yes</td>
<td>374</td>
<td>0.90 (0.72–1.11)</td>
<td>p=0.22</td>
</tr>
<tr>
<td>Fluoropyrimidine therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cetuximab-based</td>
<td>391</td>
<td>1.02 (0.82–1.26)</td>
<td>p=0.10</td>
</tr>
<tr>
<td>Fluorouracil-based</td>
<td>190</td>
<td>0.72 (0.53–0.98)</td>
<td>p=0.10</td>
</tr>
<tr>
<td>WBC count</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;100000 (per L)</td>
<td>428</td>
<td>0.88 (0.73–1.08)</td>
<td>p=0.41</td>
</tr>
<tr>
<td>≥100000 (per L)</td>
<td>153</td>
<td>1.05 (0.75–1.46)</td>
<td>p=0.12</td>
</tr>
<tr>
<td>All patients</td>
<td>581</td>
<td>0.92 (0.78–1.10)</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4:** Exploratory predictive factor analyses

HR=hazard ratio.

### Analysis

In the KRAS wild-type cohort, there was no evidence of a predictive effect of EGFR immunohistochemistry irrespective of cutoff point. However, in the KRAS mutant cohort, patients with more than 10% of tumour cells EGFR positive had a detrimental effect on progression-free survival from the addition of cetuximab (HR 1.28, 95% CI 1.00–1.63, p=0.047, n=285). By contrast, no such effect was identified in the EGFR less than 10% cohort (HR 0.87, 0.64–1.29, p=0.37, n=169).

A complete or partial response was reported in 209 (57%) KRAS wild-type patients in the control group and in 232 (64%) KRAS wild-type patients receiving cetuximab (OR 1.35, 95% CI 1.00–1.82, p=0.049). The addition of cetuximab resulted in reduced dose intensity in KRAS wild-type patients over the first 24 weeks (for fluorouracil-based therapy: median 78% in the control group [IQR 70–87] vs 73% [66–82] in the cetuximab group, p=0.031; for capecitabine-based therapy: 85% [74–92] vs 79% [67–88], p=0.0021). In patients receiving fluorouracil-based therapy, this difference was noted mainly in the dose intensity of infused fluorouracil (p=0.02, webappendix p 4). In those receiving capecitabine-based therapy, dose intensity was different for both oxaliplatin and capecitabine, but these differences largely disappeared after the capecitabine dose reduction in arm B.

The addition of cetuximab (arm B) did not have a noticeable effect on time on treatment in either the capecitabine-based or fluorouracil-based therapy groups. In both arms combined, the overall median duration of treatment was 29 weeks (IQR 16–41) in patients receiving fluorouracil-based therapy, but 25 weeks (IQR 13–35) in those given capecitabine-based therapy (p=0.0028 adjusting for treatment group; webappendix p 4).

Addition of cetuximab increased the incidence of skin (rash, nail changes, and hand–foot syndrome), gastrointestinal (diarrhoea, stomatitis), and other toxic effects (anaemia, lethargy, and hypomagnesaemia; table 2). In both treatment groups there were increased toxic effects with fluorouracil-based therapy for neutropenia, peripheral neuropathy, and stomatitis compared with the patients on capecitabine-based therapy. With the addition of cetuximab (arm B), diarrhoea and skin toxic effects were reported more often in patients on capecitabine-based therapy than on fluorouracil-based therapy, and peripheral neuropathy was less common in patients treated with fluorouracil-based therapy. In the KRAS wild-type cohort, similar differences in toxic effects were noted, but were not statistically significant (webappendix p 5). After the dose reduction in the subgroup of patients receiving capecitabine-based therapy and cetuximab, the incidence of grade 3 or higher diarrhoea fell from 30% (116/381; p<0.0001 vs no cetuximab) to 16% (25/153; p=0.25 vs no cetuximab). Other toxic effects were also lowered (data not shown).

Among all patients randomly assigned to treatment groups, treatment-related deaths were reported in ten patients in the control group and nine in the cetuximab group. Of the nine patients taking cetuximab, eight were in the capecitabine-based therapy plus cetuximab subgroup. Seven of the deaths occurred before the capecitabine dose reduction and were predominantly related to gastrointestinal toxic effects. In the control group, the ten deaths were split evenly between capecitabine-based and fluorouracil-based therapy, with no pattern to the causative toxic effects noted.

16 additional factors were explored for predictive value of progression-free survival among the all-wild-type patients, because this subset was judged the most likely to be responsive. Improved progression-free survival with cetuximab was seen in patients treated with fluorouracil-based therapy (HR 0.72, 95% CI 0.53–0.98, p=0.037), but not in those treated with capecitabine-based therapy (HR 1.02, 0.82–1.26, p=0.88; p=0.10 for interaction; see webappendix p 9 for data for KRAS wild-type patients). Additionally, patients with no or one metastatic site had improved progression-free survival with cetuximab (HR 0.73, 0.55–0.97, p=0.030) whereas those with two or more metastatic sites did not (HR 1.07, 0.86–1.33, p=0.56; p=0.036 for interaction; figure 4).
Similar results were obtained for patients with liver metastases only versus more widespread metastatic disease. Further analyses showed that both number of metastatic sites or liver-only metastases and choice of fluoropyrimidine therapy had significant (at the 10% level) interactions with treatment group. Fluoropyrimidine therapy was the weaker of the two, but neither was substantially affected by adjustment for important prognostic factors, or for each other. Thus, progression-free survival benefit was restricted to patients with KRAS wild-type tumours and zero or one metastatic site treated with fluorouracil-based therapy (n=96, HR 0.55, 95% CI 0.35–0.87; p=0.011).

In analysis of surgery for metastases, no increase in potentially curative liver resections was identified, with resection rates among KRAS wild-type patients who had liver-only metastases at baseline of 13% (n=12/91) in the control group and 15% (n=13/87) in the cetuximab group (p=0.74). At the time of analysis, second-line therapy was administered to significantly fewer patients in the cetuximab group (386 [56%] of 695) than in the control group [448 [62%] of 724; p=0.015] in patients who were eligible to receive further treatment (patients who were alive, had completed COIN protocol treatment, and were not lost to follow-up). Patients treated with fluorouracil-based rather than capecitabine-based therapy were also more likely to receive second-line therapy (table 3).

### Discussion

COIN is the largest trial of the addition of an EGFR-targeted monoclonal antibody to first-line combination chemotherapy in patients with advanced colorectal cancer. No benefit could be shown with the addition of cetuximab to oxaliplatin-based chemotherapy. This finding holds true both for the primary outcome measure of overall survival and for the secondary outcome measure of progression-free survival for the KRAS wild-type cohort. In a predefined secondary analysis, we postulated that the group of patients with no mutation in the genes tested within the RAS–RAF–MAP kinase pathway would be the most likely to show a benefit. However, even patients with tumours wild type for all three genes did not show any evidence of a benefit from the addition of cetuximab. Conversely, we have recorded no evidence of a detrimental effect for patients with KRAS mutant tumours. De Roock and colleagues have suggested that the KRAS Gly13Asp mutation might not predict lack of benefit with use of EGFR inhibitors. In COIN, this mutation was identified in 110 patients, but was not associated with any difference in outcome with the addition of cetuximab. We do, however, show that these mutations have a strong prognostic effect, with median survival ranging from 8–8 months for patients with BRAF mutant tumours, about 14 months for patients with KRAS or NRAS mutant tumours, to 20–1 months for patients with tumours that were all wild type. These factors remain strongly prognostic in a multivariate analysis, and consideration should therefore be given in future trials to their use as stratification factors or inclusion or exclusion criteria.

The overall survival of patients in COIN was inferior to other trials of similar design in this setting for chemotherapy versus chemotherapy plus anti-EGFR therapy. Evidence that patients with colorectal cancer in the UK (and Denmark) have 5–10% inferior survival compared with patients from Canada, Australia, and Scandinavia has recently been published. This effect was ascribed to more advanced disease stage at presentation, reflected especially in inferior 1-year survival. In COIN, the patients were drawn from 111 hospitals across the UK and Ireland and thus are more broadly representative of patients with advanced colorectal cancer than are those in other trials in which recruitment is from more selected centres. The COIN trial was not intended for patients who receive first-line chemotherapy in anticipation of possible resection of metastases, and the trial cohort therefore had fairly advanced disease with a substantial proportion of patients having unresected primary tumours and synchronous progression.

### Table 3: Second-line therapy received by KRAS wild-type patients, by chemotherapy received

<table>
<thead>
<tr>
<th>Therapy received</th>
<th>Arm A</th>
<th>Arm B</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any second-line therapy</td>
<td>210 (65%)</td>
<td>169 (54%)</td>
<td>0.0061</td>
</tr>
<tr>
<td>Irinotecan</td>
<td>171 (53%)</td>
<td>132 (42%)</td>
<td>0.0082</td>
</tr>
<tr>
<td>Fluoropyrimidine</td>
<td>155 (48%)</td>
<td>121 (39%)</td>
<td>0.021</td>
</tr>
<tr>
<td>Oxaliplatin</td>
<td>60 (19%)</td>
<td>46 (15%)</td>
<td>0.20</td>
</tr>
<tr>
<td>EGFR-targeted therapy</td>
<td>16 (5%)</td>
<td>21 (7%)</td>
<td>0.33</td>
</tr>
<tr>
<td>Other therapy</td>
<td>10 (3%)</td>
<td>8 (3%)</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Data are n (%) or p value. EGFR=epidermal growth factor receptor. *Patients deemed eligible for second-line therapy were those that survived at least 14 days after coming off trial.
Articles

metastatic disease. These factors affect survival in the control group as well as the experimental group. However, the characteristics of the population were not fundamentally different from those in other trials of EGFR-targeted antibodies so, although contributing to the shorter overall survival, they do not adequately account for the failure to detect the expected improvement with addition of cetuximab. Less aggressive use of effective treatments has been a criticism of the UK health system. However, combination chemotherapy with or without cetuximab used until disease progression is thought to be aggressive first-line therapy. The fact that bevacizumab is not approved for reimbursement in the UK National Health Service means that the small increase in overall survival from the use of this agent will not be reflected in the survival of COIN patients. A significant reduction in the use of second-line therapy was also noted in the cetuximab group (56% vs 62%). This finding could be a consequence of the increased toxic effects noted with addition of cetuximab, but by itself is unlikely to account for the absence of survival benefit in view of the lack of overall benefit in progression-free survival and the small difference shown.

One major factor that could affect the benefit of the addition of cetuximab to chemotherapy is the precise nature of the agents used in combination. The addition of bevacizumab to cetuximab and combination chemotherapy seems to be detrimental.15,16 The only phase 3 trial in first-line therapy showing an overall survival benefit to date used irinotecan and infusional fluorouracil as the chemotherapy backbone.17 By comparison, the trials using oxaliplatin have not shown improved overall survival and this failure has raised the possibility of a negative interaction between oxaliplatin and cetuximab.19–22

A broad set of predefined exploratory analyses have been done in an attempt to understand the results of the COIN trial. The only group for which some evidence of a potential benefit was suggested were those patients who have three coincident factors: KRAS wild-type tumours, treatment with infused fluorouracil rather than capecitabine, and a limited distribution of metastatic disease (either zero or one metastatic site vs two or more sites, or liver metastases only vs more widespread disease). This cohort generally conforms to those patients identified in guidance from the UK National Institute for Health and Clinical Excellence for the use of cetuximab, which was issued shortly before the trial was analysed.24 For patients with KRAS wild-type tumours treated with infused fluorouracil, the benefit in progression-free survival (HR for fluorouracil-based therapy was 0.77, p=0.06, compared with HR for capecitabine-based therapy of 1.06, p=0.56, p for interaction 0.07; webappendix p 9) was consistent with other trials.25–22 This finding again suggests the potential importance of the agents used in combination when using EGFR-targeted therapies.

Within the COIN trial, after identification of increased toxic effects in patients treated with capecitabine, oxaliplatin, and cetuximab, a capecitabine dose reduction was mandated in the cetuximab group only.1 This change successfully reduced levels of toxic effects to rates similar to those in the control group and improved the dose of oxaliplatin delivered, but obviously reduced the exposure of those 19% of patients in arm B to fluoropyrimidine. The lack of benefit in patients treated with capecitabine could be accounted for at least in part by the increase in toxic effects recorded, and the resulting reduction in dose intensity of the chemotherapy administered or by other undetermined factors. However, we emphasise that these subgroup findings need validation.

This trial was one of the first not to require positive EGFR immunohistochemistry as a patient selection characteristic. 22% of patients were completely negative and 40% had less than 10% EGFR membrane staining. In COIN, there is no evidence to suggest that EGFR immunohistochemistry staining is a predictive factor for clinical benefit of the addition of cetuximab in the KRAS wild-type population. This finding does not support the requirement for EGFR testing in the current cetuximab licence. Outcomes in patients with KRAS mutant tumours were heterogeneous with respect to EGFR staining, such that those with greater than 10% EGFR staining experienced a detrimental effect on addition of cetuximab as shown in some other studies,27 but this effect was not recorded in patients with less than 10% EGFR staining.

The results of the COIN trial are unexpected and add to the variance seen in other trials evaluating the use of EGFR monoclonal antibodies in combination with

Panel: Research in context

Systematic review

Several phase 3 trials have assessed the benefit of the addition of epidermal-growth-factor-receptor-targeted therapies to standard therapy for advanced colorectal cancer with varying results.5,6,15–18,21,22 Systematic review suggests that this therapy is associated with improved progression-free survival, but with the greatest advantage seen in patients who have received previous chemotherapy. The benefits are confined to patients with no evidence of mutation in the KRAS oncogene.

Interpretation

The outcome of this study is contrary to expectation showing no benefit from the addition of cetuximab in patients with KRAS wild-type tumours. There is an interaction with the choice of chemotherapy used, such that patients choosing the oral fluorouracil produrg capecitabine plus oxaliplatin gained no benefit in any subgroup. This finding could be attributable to overlapping gastrointestinal toxic effects and as a result this triple combination cannot be recommended. By contrast, the results with infusional fluorouracil plus oxaliplatin treatment mirrored those of other studies showing a small benefit with the addition of cetuximab.
chemotherapy in the first-line treatment of advanced colorectal cancer (panel). The overall lack of benefit of the addition of cetuximab to oxaliplatin and fluoropyrimidine combinations seen in the COIN trial, even in the absence of bevacizumab, is likely to be attributable to the specific toxic effect profile of the combination of the oxaliplatin and cetapicabine chemotherapy backbone, with which no benefit was seen in any subgroup when cetuximab was added. By contrast, in patients treated with oxaliplatin and infusional fluorouracil, similar benefits in progression-free survival were seen as in other studies. The potent prognostic effect of *BRAF*, *KRAS*, and *NRAS* mutations on the outcome of patients with advanced colorectal cancer shows the fundamental importance of these changes and emphasises the need to stratify future trials for these factors and to seek specific therapeutic approaches within these molecular subgroups.

**COIN Trial Investigators**

**Trial management group** T S Maughan (chair), T M Seymour, R H Wilson, R A Adams, H Wasan, A Madi, J Cassidy, M J Kennedy, E Hodgkinson, B Rogers, M Pope.


**Genetic analysis** C G Smith, R Harris, S Iddles, B Claes, D Lambrechts, J P Cheadle.

**EGFR immunohistochemistry** B Jasan, S Sterkel, R Adams, M James.

**Data monitoring committee** J Northover (chair), J Brown, M Aapro, R Stout.

**Trial steering committee** C Parker (current chair), M Mason, R Rudd, P W M Johnson, J Whelan.

**Sponsor** Medical Research Council.

**Clinical investigators** (Institution [number of patients contributed])

C Lowdell, R H Phillips, R Ahmad, P Riddle, A Crewe (Charing Cross Hospital, London, UK [83]); J Cassidy, A McDonald, A Waterston, N Mohammed, J White, H Yosef, A Hennessy (Beatosn Oncology Centre, Glasgow, UK [81]); E M Bessell, V Potter (Nottingham City Hospital, Nottingham, UK [72]); M Seymour, A Anthony, A Melcher, R Cooper, D Sebag-Montefiore (St James’ University Hospital, Leeds, UK [69]); P Ross, M Leslie, N Maisey, A Gaya, G Michael (Guy’s and St Thomas’s Hospitals, London, UK [67]); J Summers, M Hill, R James (Maidstone Hospital, Maidstone, UK [66]); T S Maugham, A Brewer, T Crosby, S Mukherjee, N Iqbal (Velindre Cancer Centre, Cardiff, UK [63]); J Wadley, D Furniss, S Pledge, J Hornblucke, S Clinton (Weston Park Hospital, Sheffield, UK [63]); R Glynne-Jones, M Harrison, S Mawdsley, N Anyamele, R Hughes (Mount Vernon Hospital, UK [57]); D Proper, C Cottrill (St Bartholomew’s Hospital, London, UK [57]); D Smith, S Myint, N Ali, M Iqbal, P Clark, B Haylock (Clatterbridge Centre for Oncology, UK [50]); T Ivesson, A Bateman, C Baughan (Southampton University Hospital, UK [46]); S Falk, R Hopkins (Bristol Oncology Centre, Bristol, UK [45]); J Bridgewater, S Karp (North Middlesex Hospital, London, UK [44]); C Topham, G Middleton, S Essapen, S Cummins (Royal Surrey County Hospital, UK [44]); B Sizer (Essex County Hospital, UK [42]); S Collins (Glen Cwyd Hospital, UK [42]); A Maraveyas (Castle Hill Hospital, Hull, UK [41]); S Tahir (Broomfield Hospital, Essex, UK [36]); J Bridgewater (Princess Alexandra Hospital, UK [36]); C Blessing (Great Western Hospital, Swindon, UK [34]); H Wasan (Hammersmith Hospital, London, UK [34]); M Tighe, S Falk (Musgrove Park Hospital, Taunton [34]); R Ellis (Royal Cornwall Hospital, UK [34]); J Dent, J K Jaffe (Huddersfield Royal Infirmary, UK [33]); S Shepherd, R Benson, D Farrugia (Cheltenham General Hospital, UK [32]); P Mack, A Hamid, M Butt, R Roy (Diana, Princess of Wales Hospital, Grimsby, UK [31]); T Hickish (Royal Bournemouth Hospital, UK [31]); H Yosef (Hartmeyer Hospital, UK [29]); D Ting, P Leonard, I Prejibisz (Southend Hospital, UK [28]); J Ledermann, J Bridgewater (University College London Hospital, London, UK [27]); T Hickish, R Osborne (Poole Hospital, Dorset, UK [26]); A O’Callaghan, S R Muthuraman (Portsmouth Oncology Centre, Queen Alexandra Hospital, UK [25]); F Adab (University of North Staffordshire NHS Trust, UK [25]); N J Wadd, J Van der Voet, D Wilson (James Cook University Hospital, UK [34]); J Wagstaff, C Askild (Singleton Hospital, UK [24]); I Pedley, A Azzabi (Sunderland Royal Hospital, UK [24]); E Marshall (Whiston Hospital, UK [24]); R James (Kent and Canterbury Hospital, UK [23]); F Daniel (Derriford Hospital, Plymouth, UK [22]); R Osborne (Dorset County Hospital, UK [22]); T J Ivesson (Salisbury District Hospital, UK [22]); D Jodrell, C McLean, S Clive, L Dawson, H A Phillips (Western General Hospital, Edinburgh, UK [22]); S Falk (Yeovil District Hospital, UK [22]); R James (Queen Elizabeth, The Queen Mother Hospital, UK [21]); P Chakraborti, R Kulkarni (Royal Derby Hospital, UK [21]); I M Samuel, G MacDonald (Aberdeen Royal Infirmary, UK [20]); C Bradley, C Twelves (Bradford Royal Infirmary, UK [20]); S Gridharzani, F Adab (Staffordshire General Hospital, UK [20]); S Myint (Southport & Formby District General Hospital, UK [18]); S Collins (Wrexham Maelor Hospital, UK [18]); N Stuart, C Bale (Wshyt Gwynedd Hospital, UK [18]); (William
Articles

Harvey Hospital, UK [17]; J Valle, M Saunders, G Wilson (Christie Hospital, Manchester, UK [16]; A Weaver, A Jones (Churchill Hospital, Oxford, UK [16]); M K McAdam, C Jeppcott (Peterborough District Hospital, UK [16]); S Cleator (St Mary’s Hospital, London, UK [16]); A Webb (Worthing Hospital, UK [16]); J Glaholm (Good Hope Hospital, UK [15]); R Hughes, P M Mulholland (Lister Hospital, UK [15]); N Steven (Queen Elizabeth, University Hospital, Birmingham, UK [15]); A Mayer, T Meyer (Royal Free Hospital [15]); N Lo, G Cogill, L Toy (Torbay Hospital, UK [13]); M Wilkins (West Wales General Hospital, UK [13]); C Wilson, C Palmer (Addenbrooke’s Hospital, UK [14]); J Geh (Birmingham Heartlands Hospital, UK [14]); R Thomas (Bedford Hospital NHS Trust [13]); J Nicoll (Cumberland Infirmary, UK [13]); P Chakrabarti (Queens Hospital, Burton Upon Trent [13]); A Azzali (South Tyneside Hospital, UK [13]); R Wilson, M Eatoek (Belfast City Hospital, UK [12]); A Hartley (Manor Hospital, Walsall, UK [12]); J Gildersleeve, A Freehain (Berkshire Cancer Centre, Berkshire, UK [12]); G P Deutsch, A Webb, M Wilkins (Royal Sussex County Hospital, Brighton, UK [12]); F McKenna (Eastbourne District General Hospital, UK [11]); D Cunningham, J Chau (Royal Marsden, Sutton, UK [11]); S Susnerwala, M Wise, A Birtle (Royal Preston Hospital, UK [11]); A Hamid (Scunthorpe General Hospital, UK [11]); F J Lofs (St George’s Hospital, London, UK [11]); J Kennedy (St James’ Hospital, Dublin, Ireland [11]); D Whillis (Raimore Hospital, UK [10]); S Susnerwala (Blackpool Victoria Hospital, UK [9]); I T Tan, C Palmer (Hinchingbrooke Hospital, UK [9]); N Mason (Queen Elizabeth Hospital, UK [9]); M Keane (University College Hospital, Galway, Ireland [9]); C Macmillan, K Patel (Northampton General Hospital, UK [8]); D Farrugia (Worcestershire Royal Hospital, UK [8]); G Bertelli, V Vigneswaran (Wythbury General Hospital, UK [7]); D Smith (Aintree University Hospital, UK [6]); R McDermott (AMNCH, Tallaght Hospital, Dublin, Ireland [6]); S O'Reilly (Mercy Hospital, Cork, Ireland [6]); V Hall (Royal Hampshire County Hospital, Winchester, UK [6]); C Hamilton, A Dhadda (Scarborough Hospital, UK [6]); C Baughan (St Mary’s Hospital, Isle of Wight, UK [6]); H Ford, M Moody (West Suffolk Hospital, UK [6]); M Tomlinson (Weston General Hospital, Weston Super Mare, UK [6]); I Grogan, O Breathnach (Beaumont Hospital, Dublin, Ireland [4]); R Soo mal (Ipswich Hospital NHS Trust, UK [4]); J McCaffrey (Mater Misericordiae Private Hospital, Dublin, Ireland [4]); J McCaffrey (Mater Misericordiae Public Hospital, Dublin, Ireland [4]); C Rees (North Hampshire Hospital, Basingstoke, UK [4]); P J Alberton (North Tyneside General Hospital, UK [4]); S Beeley (Conquest Hospital, UK [3]); W Dobrovsky (Wansbeck General Hospital, UK [3]); G Leonard (Waterford Regional Hospital, Waterford, Ireland [3]); R Gupta (Midwestern Regional Hospital, Limerick, Ireland [2]); J Stewart (Milton Keynes General Hospital, UK [2]); D Cunningham (Royal Marsden Hospital, London, UK [2]).

References

Arthritis 1

Osteoarthritis: an update with relevance for clinical practice

Johannes W J Bijlsma, Francis Berenbaum, Floris P J G Lafeber

Osteoarthritis is thought to be the most prevalent chronic joint disease. The incidence of osteoarthritis is rising because of the ageing population and the epidemic of obesity. Pain and loss of function are the main clinical features that lead to treatment, including non-pharmacological, pharmacological, and surgical approaches. Clinicians recognise that the diagnosis of osteoarthritis is established late in the disease process, maybe too late to expect much help from disease-modifying drugs. Despite efforts over the past decades to develop markers of disease, still-imaging procedures and biochemical marker analyses need to be improved and possibly extended with more specific and sensitive methods to reliably describe disease processes, to diagnose the disease at an early stage, to classify patients according to their prognosis, and to follow the course of disease and treatment effectiveness. In the coming years, a better definition of osteoarthritis is expected by delineating different phenotypes of the disease. Treatment targeted more specifically at these phenotypes might lead to improved outcomes.

Introduction

Epidemiology

The prevalence of osteoarthritis is dependent on the precise definition used and on the site of interest. The knee, hip, and hand are most affected by the disease (figure 1). Osteoarthritis becomes more common with age, and after age 50 years more women than men are affected. For example, the Rotterdam study of a population-based cohort of 3906 people 55 years or older reported that 67% of women and 55% of men had radiographic osteoarthritis of the hand. In people older than 80 years, 53% of women and 33% of men had radiographic osteoarthritis of the knee. The age-standardised and sex-standardised incidence of osteoarthritis of the hand is 100 per 100 000 person-years, for the hip is 88 per 100 000 person-years, and for the knee is 240 per 100 000 person-years.

Osteoarthritis in general develops progressively over several years, although symptoms might remain stable for long periods within this period. The diagnosis of the disease relies on clinical and radiological features (panel). Nearly half of patients with radiological features of osteoarthritis have no symptoms and vice versa. Risk factors for occurrence and progression of osteoarthritis have been identified, and differ on the basis of the joints involved (table 1).

Pathology

In addition to the involvement of several joint tissues, osteoarthritis has long been mainly characterised by a failure of the repair process of damaged cartilage due to biomechanical and biochemical changes in the joint. Cartilage is non-vascularised, so this restricts the supply of nutrients and oxygen to the chondrocytes—the cells that are responsible for the maintenance of a very large amount of extracellular matrix. At an early stage, in an attempt to effect a repair, clusters of chondrocytes form in the damaged areas and the concentration of growth factors in the matrix rises. This attempt subsequently fails and leads to an imbalance in favour of degradation. Increased synthesis of tissue-destructive proteinases (matrix metalloproteinases and aggrecanases), increased apoptotic death of chondrocytes, and inadequate synthesis of components of the extracellular matrix, lead to the formation of a matrix that is unable to withstand normal mechanical stresses. Consequently, the tissue enters a vicious cycle in which breakdown dominates synthesis of extracellular matrix. Since articular cartilage is aneural, these changes do not produce clinical signs unless innervated tissues become involved. This is one reason for the late diagnosis of osteoarthritis.

Although the pathophysiology of osteoarthritis has long been thought to be cartilage driven, recent evidence shows an additional and integrated role of bone and synovial tissue, and patchy chronic synovitis is evident in the disease. Synovial inflammation corresponds to clinical symptoms such as joint swelling and inflammatory pain, and it is thought to be secondary to cartilage debris and catabolic mediators entering the synovial cavity. Synovial macrophages produce catabolic and proinflammatory mediators and inflammation starts negatively affecting the balance of cartilage matrix degradation and repair. This process in turn amplifies synovial inflammation, creating a vicious cycle. Synovial inflammation happens in early as well as late phases of osteoarthritis and is seldom as severe as in rheumatoid

Search strategy and selection criteria

The information in our paper is primarily based on PubMed searches with the terms “osteoarthritis” in combination with “cartilage”, “bone”, “synovitis”, “imaging”, “biomarker”, and “treatment”. We mainly included papers from the past 5 years, with the addition of highly regarded older papers. We also included some review articles and book chapters as comprehensive overviews, the details of which are beyond the scope of our report.

See Comment page 2067

This is the first in a Series of three papers about arthritis

Correspondence to:

Prof Johannes W J Bijlsma, University Medical Centre Utrecht, Department of Rheumatology and Clinical Immunology, PO Box 85 500, Utrecht, 3508 GA, Netherlands

j.w.j.bijlsma@umcutrecht.nl

www.thelancet.com Vol 377 June 18, 2011 2115
arthritis, but it might add to the vicious cycle of progressive joint degeneration.

The main characteristics of osteoarthritis are changes in the subchondral bone. Osteophyte formation, bone remodelling, subchondral sclerosis, and attrition are crucial for radiological diagnosis. Several of these bone changes take place not only during the final stage of the disease, but also at the onset of the disease—possibly before cartilage degradation.10,11 This finding led to the suggestion that subchondral bone could initiate cartilage damage.

Clinical features and diagnosis

Pain is the first and predominant symptom of osteoarthritis that causes patients to visit their family doctor. The pain experienced is intermittent, typically worst during and after weight-bearing activities. Inflammatory flares can happen during the course of the disease. Patients with osteoarthritis also experience stiffness: in the morning, after a period of inactivity, or particularly in the evening. This stiffness generally resolves in minutes, unlike the prolonged (usually >30 min) stiffness caused by rheumatoid arthritis.

Loss of movement and function is another reason patients visit their family doctor. Patients report symptoms that limit their day-to-day activities, such as stair climbing, walking, and doing household chores. Symptomatic osteoarthritis might be associated with depression and disturbed sleep, which additionally contribute to disability. The symptoms of osteoarthritis diminish the patients’ quality of life.10

Physical examination is needed to confirm and characterise joint involvement, and to exclude pain and functional syndromes with other causes—eg, inflammatory arthritis.11 Joint enlargement results from joint effusion, bony swelling, or both. A synovial effusion might not only be identified during osteoarthritis flares, but also during chronic phases as a persistent feature. Restricted passive movement can be the first and sole physical sign of symptomatic disease. Bursitis, tendinitis, muscle spasm, and tissue response to, for instance, damaged meniscus can cause the same pain syndrome and must be carefully sought during examination. Crepitus, a sensation of crunching or crackling, is commonly felt on passive or active movement of a joint with osteoarthritis. Joint deformities relate to advanced disease with joint damage that involves cartilage, periaricular bone, synovium, articular capsule, ligaments, and muscles (figure 2). A joint can lock if loose bodies or fragments of cartilage (or meniscus) get into the joint space. Caution should be exercised to correctly attribute pain to the correct site—eg, patients with osteoarthritis of the hip might report knee pain because of referred pain or anserine bursitis. Additional neurological and spine examination is often needed.

Imaging investigations are seldom needed to confirm the diagnosis; they might be useful to establish the severity of joint damage and to monitor disease progression. However, some sites and clinical scenarios need imaging assessment (including MRI or scintigraphy) to exclude other diseases, including avascular osteonecrosis, Paget’s disease, complex regional pain syndrome, inflammatory arthropathies, and stress fractures. Also, blood tests are not routinely needed in cases of uncomplicated chronic pain arising from clearly defined osteoarthritis. ESR and C-reactive protein are usually within the normal range. Some laboratory tests might be done to exclude other diseases, such as anticyclic citrullinated peptide antibodies for rheumatoid arthritis and uric acid for gout. Synovial fluid should be assessed if another arthropathy or septic arthritis is suspected. In patients with osteoarthritis, synovial fluid is sterile, without crystals, and a white-cell count of less than 1500 cells per μL.

Figure 1: Osteoarthritic joints of the hand, hip, and knee

(A) Osteoarthritis is predominantly identified in the distal interphalangeal and proximal interphalangeal joints—deformations of the distal interphalangeal joints are clearly visible. (B) Plain radiograph of an osteoarthritic hip joint showing the narrowing of the joint space and clearly visible osteophytes. (C) MRI of an osteoarthritic knee with clear medial cartilage loss and osteophyte formation, with minor synovial swelling.
Markers of tissue damage
Why there is little relation between clinical characteristics and structural tissue changes in osteoarthritis remains unclear. Insensitivity of available monitoring methods for damage to joint tissue combined with slow progression of this damage might underlie the discrepancy. Assessment of structural changes is a challenge in studying the disease and improving treatment modalities.

Early and minimum tissue damage is difficult to assess in vivo. Biopsies for detailed histochemical and biochemical assessment of cartilage, bone, and synovial tissue in osteoarthritis are not feasible and are often contraindicated. Also, tissue changes are often focal and can be missed by random biopsy procedures.

The exterior of the cartilage can be seen through arthroscopic procedures. However, these procedures involve invasive techniques, and there is doubt whether the various stages of degeneration or regeneration processes of the cartilage can be reliably detected.

Therefore, at present only surrogate markers, as indirect measures of the actual destructive process, can be used for diagnosis and follow-up of tissue damage. There has been much effort to develop new biomarkers, in the hope that they will improve early diagnosis and treatment of the disease. However, although promising in research settings, there is little use for these markers in daily practice.

Plain radiography is the gold standard in imaging of osteoarthritic joints, since the technique is inexpensive, fast, and easily available. Radiography has the advantage that high-resolution images can be obtained quickly and routinely under weight-bearing conditions. Restrictions are radiation exposure and that only calcified bone can be visualised, which provides an indirect measure of cartilage thickness without providing information about synovial tissue. Regulatory agencies (US Food and Drug Administration, European Medicines Agency) recommend joint-space narrowing on radiographs, in addition to pain and function, as coprimary endpoints, to establish the effectiveness of disease-modifying drugs.

Kellgren and Lawrence classification has been developed as a radiological grading of osteoarthritis for several joints, including knees, hips, and hands. The classification focuses on a sequence of osteophyte formation, joint-space narrowing, and bone sclerosis, and provides simple and practical ordinal scales for each joint. Additional scores have been developed to provide a simple and practical ordinal scale for each classification. Focuses on a sequence of osteophyte area. Standardisation of radiographs is now more time consuming they are and the more complex analysis becomes. Improvement of scoring methods by making them more objective and reproducible is hampered by a lack of standardisation of the image acquisition. For instance, the position of the joint in the x-ray beam is crucial for visualisation of the joint-space width and for the estimation of bone density and osteophyte area. Standardisation of radiographs is now the crucial step in the reproducibility of radiographic scoring. Reasonably, several radiographic views, including the patellofemoral joint for the knee and the so-called faux profile image for hip (with backwards rotated pelvis) improve the relation between clinical and radiographic changes.

Generally, clinically significant changes in radiographic scores take at least 1 or even 2 years. For the knee, the smallest detectable difference of joint-space width is about 0.20 mm by an expected average annual decrease of about 0.15 mm. More subtle changes can be detected in a shorter time through the use of advanced standardisation methods during image acquisition and more complex analyses, preferably of several images.

Panel: American College of Rheumatology radiological and clinical criteria for osteoarthritis of the knee and hip

**Hand (clinical)**

**Osteoarthritis if 1, 2, 3 or 4 or 1, 2, 3, 5 are present:**

1. Knee pain for most days of previous month
2. Hard tissue enlargement of two or more of ten selected joints
3. Swelling in two or more metacarpophalangeal joints
4. Hard tissue enlargement of two or more distal interphalangeal joints
5. Deformity of two or more of ten selected hand joints

**Hip (clinical and radiographic)**

**Osteoarthritis if 1, 2, 3 or 1, 2, 4 or 1, 3, 4 are present:**

1. Hip pain for most days of previous month
2. Erythrocyte sedimentation rate of less than 20 mm in the first hour
3. Femoral or acetabular osteophytes on radiographs
4. Hip joint space narrowing on radiographs

**Knee (clinical)**

**Osteoarthritis if 1, 2, 3 or 1, 4 or 1, 3, 4 are present:**

1. Knee pain for most days of previous month
2. Crepitus on active joint motion
3. Morning stiffness lasting 30 min or less
4. Age 38 years or older
5. Bony enlargement of the knee on examination

**Knee (clinical and radiographic)**

**Osteoarthritis if 1, 2, 3, 4 or 1, 2, 5 or 1, 4, 5 are present:**

1. Knee pain for most days of previous month
2. Osteophytes at joint margins on radiographs
3. Synovial fluid typical of osteoarthritis (laboratory)
4. Age 40 years or older
5. Crepitus on active joint motion
6. Morning stiffness lasting 30 min or less

**Hand (clinical)**

**Osteoarthritis if 1, 2, 3, 4 or 1, 2, 3, 5 are present:**

1. Hand pain, aching, or stiffness for most days of previous month
2. Hard tissue enlargement of two or more of ten selected joints
3. Swelling in two or more metacarpophalangeal joints
4. Hard tissue enlargement of two or more more distal interphalangeal joints
5. Deformity of two or more of ten selected hand joints

**Hip (clinical and radiographic)**

**Osteoarthritis if 1, 2, 3, 4 or 1, 2, 5 or 1, 4, 5 are present:**

1. Hip pain for most days of previous month
2. Erythrocyte sedimentation rate of less than 20 mm in the first hour
3. Femoral or acetabular osteophytes on radiographs
4. Hip joint space narrowing on radiographs

**Knee (clinical)**

**Osteoarthritis if 1, 2, 3 or 1, 4 or 1, 3, 4 are present:**

1. Knee pain for most days of previous month
2. Crepitus on active joint motion
3. Morning stiffness lasting 30 min or less
4. Age 38 years or older
5. Bony enlargement of the knee on examination

**Knee (clinical and radiographic)**

**Osteoarthritis if 1, 2, 3, 4 or 1, 2, 5 or 1, 4, 5 are present:**

1. Knee pain for most days of previous month
2. Osteophytes at joint margins on radiographs
3. Synovial fluid typical of osteoarthritis (laboratory)
4. Age 40 years or older
5. Crepitus on active joint motion
6. Morning stiffness lasting 30 min or less

*Ten selected joints include bilateral second and third interphalangeal proximal joints, second and third proximal interphalangeal joints, and first carpometacarpal joint.*
dominate the disease, and as such should be targeted for treatment. In the interplay between cartilage, bone, and synovial tissue, one of the tissues might be the only tissue not innervated. On the right, the bidirectional interplay between cartilage, bone, and synovial tissue is shown. Note that the different tissues involved in clinical and structural changes of the disease are shown on the left. Figure 2: Schematic drawing of an osteoarthritic joint.

The different tissues involved in clinical and structural changes of the disease are shown on the left. Note that cartilage is the only tissue not innervated. On the right, the bidirectional interplay between cartilage, bone, and synovial tissue is shown, and the two-way interaction between this interplay and the ligaments and muscles. In the interplay between cartilage, bone, and synovial tissue, one of the tissues might dominate the disease, and as such should be targeted for treatment.

Apart from plain radiography, other imaging techniques have been further developed (table 2): CT, ultrasound, and MRI. Regular CT has similar disadvantages to plain radiography with clearly higher radiation exposure, but the advantage is a three-dimensional image and the option of contrast agents (contrast enhanced CT) to visualise cartilage in addition to bone. Bone is innervated and evidence is accruing that bone changes might be an important source of pain in osteoarthritis. Assessment of CT has shown a strong relation between the dissolving of cystic bone areas and pain relief after treatment of end-stage osteoarthritis. Techniques are still improving, but are unlikely to become the standard.

Ultrasound has the advantages that it also images soft-tissue structures (such as synovial tissue) in several planes, it does not need contrast agents, and it allows the visualisation of movement. Limitations exist in the depth that the signal can penetrate and the sites (tissues) that can be assessed. Most importantly, ultrasound is very dependent on the experience and skills of the user. The use of power doppler signal to image vascularisation and specific integrated techniques to assess cartilage thickness enhances its applications. Although the use of ultrasound to detect osteoarthritic pathological changes (specifically in hand joints) is increasing, its ultimate role in osteoarthritis is not certain.

MRI provides objective quantitative assessment of morphology (volume, area, and thickness) and integrity (quality) of articular cartilage. A broad range of sequences and scoring systems allow for sensitive analyses of periarticular soft tissues in addition to cartilage and bone. Important limitations are cost, acquisition time (on average 45 min), complexity of the more advanced techniques, and time for whole-organ analyses. These limitations hamper the use of MRI for the imaging of osteoarthritis, although its value in identifying bone marrow and meniscal lesions is well established. The use of fat-suppressed spoiled gradient echo sequences produces a high cartilage signal and low signal from adjacent joint fluid, and at present is the standard for quantitative morphological imaging of cartilage. The availability of higher field strengths, up to 3 tesla, makes these measurements even more accurate.

Several semiquantitative scoring systems (table 2) have been developed that focus on the size and location of the lesions, and on subchondral, cartilage, bone, and other abnormalities. Apart from tissue-specific scores, whole-organ scores have been developed, such as the knee osteoarthritis scoring system, the whole-organ magnetic resonance imaging score, and the Boston Leeds osteoarthritis knee score, each with their own advantages.

More complex acquisition sequences have been developed that focus on cartilage quality; T2 MRI relaxation time is related to collagen orientation and density of articular cartilage; a possible relation with cartilage degeneration has been shown. Also, the T1 MRI technique provides information that allows proteoglycan distribution in articular cartilage to be mapped. The negatively charged proteoglycans are responsible for the fixed charged density of the cartilage matrix that makes sodium MRI and delayed gadolinium-enhanced MRI of cartilage useful for visualising proteoglycan content. When given intravenously, gadolinium-diethyleneetriamine penta-acetic acid homes in on regions of cartilage with low proteoglycan content. Some clinical applications have shown its value, but variables such as body-mass index, age, sex, intensity of sport activities, subchondral bone oedema on MRI, obesity, intense sport activities, quadriceps strength, bone density, previous injury, hormonal replacement therapy, and congenital deformities have been shown to be significant factors.

Table 1: Selected risk factors for the occurrence and progression of osteoarthritis in knees, hips, and hands

<table>
<thead>
<tr>
<th>Occurrence</th>
<th>Knee</th>
<th>Hip</th>
<th>Hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, sex, physical activity, body-mass index (including obesity), intense sport activities, quadriceps strength, bone density, previous injury, hormone replacement therapy (protective), vitamin D, smoking (protective or deleterious), malalignment (including varus and valgus), genetics</td>
<td>Age, physical activity, body-mass index (including obesity), intense sport activities, genetics</td>
<td>Age, grip strength, occupation, intense sport activities, genetics</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progression</th>
<th>Knee</th>
<th>Hip</th>
<th>Hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, body-mass index (including obesity), vitamin D, hormone replacement therapy (protective), malalignment (including varus and valgus), chronic joint effusion, synovitis, intense sport activities, subchondral bone oedema on MRI</td>
<td>Age, symptomatic activity, sex, intense sport activities</td>
<td>Unknown</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Schematic drawing of an osteoarthritic joint.
severity of synovitis, and subchondral bone alterations, make use of the technique complex.60–62 Most of the developments in MRI involve the knee, much less research has been done on hips63 and hands.64

In general, MRI sequences and scoring systems provide good quantitative analyses of several joint structures, with more advanced techniques providing information about cartilage quality. Unfortunately, cost, acquisition, and analytical time restrict developments of these techniques in research settings and their use in daily clinical practice. In the future, MRI assessments in larger clinical trials might become standard; in today’s practice they have value only for specific diagnostic questions.

Biochemical markers of joint metabolism, disease, or both are molecules or molecular fragments that are released into biological fluids (synovial fluid, blood, and urine) from extracellular matrix turnover (synthesis and breakdown), such as collagen or proteoglycan fragments (or neo-epitopes) and cellular metabolism (eg, proteases or cytokines) of articular cartilage, subchondral bone, and synovial tissue. Biochemical markers seemed to help understand the pathophysiology of osteoarthritis and in the prediction of structural changes. However, breakthroughs have been sparse and there is doubt about how these markers might be used.65 We lack sufficient knowledge about molecular validity, systemic origin, metabolism, and kinetics (absorption, distribution, and excretion) from many biochemical markers.66,67 The same marker might increase as well as decrease, dependent on the point in the degradation process.

Urine and blood are the most relevant compartments in which to assess biomarkers. There are few studies of biomarkers reporting on their diagnostic and prognostic properties, their relation to burden of disease, and their relation to effectiveness of intervention. The relation of biomarkers with structural changes is in general better understood than their relation with clinical characteristics.68

Table 3 lists the most reported biomarkers and their performance. Markers of cartilage degradation, such as CTXII in urine and COMP in serum, have been assessed extensively and show a moderate to good relation with clinical and radiographic variables of osteoarthritis. Markers of bone metabolism are less effective, presumably because of the size of the bone compartment (mostly outside the joints) and the high turnover of bone. Not enough is known about markers of bone metabolism, which might have an important role in osteoarthritis since bone changes might be an important source of pain.69–72 Markers of synovial tissue metabolism are the least studied, but produce positive results, underscoring a role for inflammation in osteoarthritis. Homogeneity of the studied population and standardisation of sample collection might improve the relation between a biomarker and clinical or radiographic characteristics, since diurnal rhythms and effects of exercise have been described for several markers.70–72

Table 2: Imaging techniques for assessment of tissue-structure changes in osteoarthritis

<table>
<thead>
<tr>
<th>Primary use</th>
<th>Analyses</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain radiograph</td>
<td>Cartilage thickness</td>
<td>(Semi)quantitative</td>
<td>Low cost, easy applicable</td>
</tr>
<tr>
<td>CT</td>
<td>Bone characteristics</td>
<td>Semiquantitative</td>
<td>Three dimensional</td>
</tr>
<tr>
<td>CECT</td>
<td>As standard plus cartilage volume</td>
<td>Semiquantitative</td>
<td>Three dimensional, information on cartilage</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>Standard</td>
<td>Inflammation</td>
<td>Impression</td>
</tr>
<tr>
<td>Power doppler</td>
<td>Vascularisation</td>
<td>Semiquantitative</td>
<td>Direct measure</td>
</tr>
<tr>
<td>MRI</td>
<td>Standard SPGR</td>
<td>Cartilage morphology</td>
<td>Quantitative</td>
</tr>
<tr>
<td>T2 MRI relaxation</td>
<td>Collagen distribution</td>
<td>Semiquantitative</td>
<td>Information on cartilage quality</td>
</tr>
<tr>
<td>T1ρ</td>
<td>Proteoglycan distribution</td>
<td>Semiquantitative</td>
<td>Information on cartilage quality</td>
</tr>
<tr>
<td>31Na MRI</td>
<td>FCD/proteoglycan content</td>
<td>Semiquantitative</td>
<td>Information on cartilage quality</td>
</tr>
<tr>
<td>dGEMRIC</td>
<td>FCD/proteoglycan content</td>
<td>Semiquantitative</td>
<td>Information on cartilage quality, early changes</td>
</tr>
<tr>
<td>MRI whole-organ scoring</td>
<td>KOSS</td>
<td>–</td>
<td>Semiquantitative</td>
</tr>
<tr>
<td>WORMS</td>
<td>–</td>
<td>Semiquantitative</td>
<td>Whole-organ score</td>
</tr>
<tr>
<td>BLOKS</td>
<td>–</td>
<td>Semiquantitative</td>
<td>Whole-organ score</td>
</tr>
</tbody>
</table>

CECT=contrast-enhanced CT. SPGR=spoiled gradient echo. FCD=fixed charge density. dGEMRIC=delayed gadolinium-enhanced MRI of cartilage. KOSS=knee osteoarthritis scoring system. WORMS=whole-organ magnetic resonance imaging score. BLOKS=Boston Leeds osteoarthritis knee score. *Techniques that have a more common clinical and research applications for the assessment of cartilage (and bone), bone, and synovial inflammation, as well as quantitative cartilage morphology (at present the most used MRI modality in clinical trials).
function as an outcome in clinical trials. More needs to be understood about biochemical markers, and combinations thereof, to make these markers of use in general clinical practice.

### Treatment

In early osteoarthritis, pain and stiffness dominate the other symptoms.77 Treatment should therefore focus on the reduction of pain and stiffness and on the maintenance and improvement of functional capacities. Furthermore, prevention of progression of joint damage and improvement of quality of life are long-term goals. There are three treatment modalities: non-pharmacological, pharmacological, and surgical. In many patients these modalities are combined, tailored to individual needs and risk factors. The European League Against Rheumatism and the Osteoarthritis Research Society International have published evidence-based guidelines for the treatment of osteoarthritis.76–78 Daily practice is based on these guidelines and updates from published work.

Self-management interventions can be defined as patient centred and as designed to foster active participation of patients to promote wellbeing and to manage symptoms. These programmes in chronic diseases are now thought key elements of good-quality care.79 In long-term disease management these interventions seem to be effective and necessary for the compliance of patients, although there are few reported benefits.80

Symptoms can be reduced by providing the patient with information about osteoarthritis, its symptoms, the objectives of its treatment, and the importance of changes in lifestyle—although the effect size of these interventions is small (<0.20).77,78 Pain has many components, and is also affected by comorbidities, such as sleeping problems, loneliness, and mood disorders,83 improvement of mental and social wellbeing is therefore also a target in some patients.82

There is evidence for a positive effect of exercise, pacing of activities, joint protection, weight reduction, and other measures to unload damaged joints (effect size 0.20–0.50).76–78 It is unclear if particular exercises are more beneficial than others for specific joints. Probably the best exercises should be established through personalised advice, which takes into account individual factors. Exercises that strengthen muscles and improve aerobic condition are most effective, at least for osteoarthritis of the hip and knee.81

Weight reduction is not easy, but quite effective, especially in osteoarthritis of the knee. Randomised controlled trials have shown that weight reduction has led to lessening of pain and improvement of physical function,84 and recent research has also shown structural improvement of cartilage85 and positive changes in biomarkers of cartilage and bone.86

Unpopular measures such as braces, cranes, and other forms of joint protection might have a slight effect and are generally cost effective.87,88 These measures should be discussed with the individual patient.

Commonly used treatment modalities are insoles, lasers,89 transcutaneous electrical nerve stimulation,90 ultrasound,91 electrotherapy,92 or acupuncture,93 but evidence is scarce, as is the effect size. However, applications of heat and ice are easy to use and quite effective.94

Paracetamol is the first-choice oral analgesic for osteoarthritis because of its safety and effectiveness,76–78 but patients have often used paracetamol with little effect before they visit their physician. Sometimes a dose increase to an optimum regimen for the individual patient is a therapeutic option, but often a non-steroidal anti-inflammatory drug (NSAID) is added or substituted. The use of stronger analgesics, such as weak opioids and narcotic analogues, is contraindicated.77,78

NSAIDs can be used in patients with symptomatic osteoarthritis of the hand, hip, or knee, preferably at the lowest effective dose and for the shortest duration.95,77 In patients with cardiovascular risk factors all NSAIDs,

### Table 3: Overview of published work on biomarkers over the past 5 years for knee and hip osteoarthritis

<table>
<thead>
<tr>
<th></th>
<th>Diagnostic value</th>
<th>Relation to burden of disease</th>
<th>Prognostic value</th>
<th>Relation to efficacy of treatment</th>
<th>Overall positive proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cartilage degradation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTXII in urine†</td>
<td>1/23</td>
<td>16/25</td>
<td>1/23</td>
<td>4/5</td>
<td>74%</td>
</tr>
<tr>
<td>COMP in serum†</td>
<td>9/12</td>
<td>15/26</td>
<td>6/17</td>
<td>1/2</td>
<td>54%</td>
</tr>
<tr>
<td>Coll 2-1 (NO2)† in urine and serum</td>
<td>7/8</td>
<td>2/6</td>
<td>2/4</td>
<td>··</td>
<td>61%</td>
</tr>
<tr>
<td>KS in serum</td>
<td>1/2</td>
<td>3/8</td>
<td>3/5</td>
<td>1/2</td>
<td>47%</td>
</tr>
<tr>
<td>YKL-40 in serum</td>
<td>1/3</td>
<td>5/12</td>
<td>0/4</td>
<td>1/1</td>
<td>35%</td>
</tr>
<tr>
<td>C2C in urine and serum</td>
<td>1/1</td>
<td>3/9</td>
<td>0/4</td>
<td>1/3</td>
<td>29%</td>
</tr>
<tr>
<td>CL2C in urine and serum</td>
<td>–</td>
<td>1/6</td>
<td>0/4</td>
<td>0/2</td>
<td>8%</td>
</tr>
<tr>
<td><strong>Cartilage synthesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PINP in serum*</td>
<td>2/2</td>
<td>1/4</td>
<td>2/3</td>
<td>0/1</td>
<td>50%</td>
</tr>
<tr>
<td>PIIINP in serum</td>
<td>–</td>
<td>3/7</td>
<td>0/4</td>
<td>··</td>
<td>27%</td>
</tr>
<tr>
<td>CSB-46 in serum</td>
<td>0/1</td>
<td>1/7</td>
<td>0/3</td>
<td>··</td>
<td>9%</td>
</tr>
<tr>
<td><strong>Bone degradation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTX-I in urine and serum*</td>
<td>1/2</td>
<td>1/1</td>
<td>2/5</td>
<td>2/2</td>
<td>60%</td>
</tr>
<tr>
<td>(D)PYR† in urine</td>
<td>2/3</td>
<td>6/15</td>
<td>0/10</td>
<td>2/2</td>
<td>33%</td>
</tr>
<tr>
<td>CTXII in urine and serum</td>
<td>2/4</td>
<td>1/16</td>
<td>1/6</td>
<td>0/1</td>
<td>15%</td>
</tr>
<tr>
<td><strong>Bone synthesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OC in serum*</td>
<td>1/5</td>
<td>2/12</td>
<td>2/6</td>
<td>1/2</td>
<td>24%</td>
</tr>
<tr>
<td>BSP in serum</td>
<td>2/2</td>
<td>1/3</td>
<td>0/2</td>
<td>··</td>
<td>43%</td>
</tr>
<tr>
<td>PINP in serum</td>
<td>0/1</td>
<td>1/4</td>
<td>0/4</td>
<td>··</td>
<td>11%</td>
</tr>
<tr>
<td><strong>Synovium degradation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HA in serum†</td>
<td>7/9</td>
<td>7/22</td>
<td>8/11</td>
<td>1/3</td>
<td>51%</td>
</tr>
<tr>
<td>Glc-Gal-PYR in urine</td>
<td>2/2</td>
<td>3/4</td>
<td>–</td>
<td>0/1</td>
<td>71%</td>
</tr>
<tr>
<td><strong>Synovial synthesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PINP in serum</td>
<td>1/2</td>
<td>2/4</td>
<td>0/2</td>
<td>··</td>
<td>43%</td>
</tr>
</tbody>
</table>

Data are n/N unless otherwise stated. Biomarkers with less than five reports are not included. Data from van Spil and colleagues.68 *The most relevant and best performing commercial biomarkers. †Combined biomarkers: Coll 2-1 with Coll 2-1 NO2 and PYR with D-PYR.

www.thelancet.com Vol 377 June 18, 2011
Including both non-selective and cyclo-oxygenase-2 selective drugs should be used with caution and are sometimes contraindicated; the individual drug characteristics seem to be more relevant than the class of drug.9 In patients with high gastrointestinal risk, either a cyclo-oxygenase-2 selective drug or a non-selective NSAID with co-prescription of a proton pump inhibitor for gastroprotection, might be considered. A possible additional argument for the use of selective cyclo-oxygenase-2 drugs was reported in a trial comparing celecoxib versus omeprazole and diclofenac in patients with osteoarthritis and rheumatoid arthritis.95 Both drugs for gastroprotection, might be considered. A possible additional argument for the use of selective cyclo-oxygenase-2 drugs was reported in a trial comparing celecoxib versus omeprazole and diclofenac in patients with osteoarthritis and rheumatoid arthritis.95 Both drugs were equally effective for the treatment of upper gastrointestinal problems, but celecoxib was better than diclofenac and omeprazole in the reduction of all gastrointestinal events (especially clinically significant anaemia of presumed gastrointestinal origin). Another attempt to reduce the gastrointestinal and cardiovascular side-effects of NSAIDs is the linking of an NSAID with a nitric-oxide-donating group, which creates a cyclo-oxygenase-inhibiting nitric-oxide donor. Nitric oxide thus might help to maintain gastric integrity and cardiovascular homoeostasis.96,97 Topical NSAIDs are recommended as alternative or adjunctive treatment and have been reported to be as effective as and possibly safer than oral NSAIDs.98

The use of opioid analgesics for the treatment of osteoarthritis has risen, but a real improvement in osteoarthritic pain that has not responded to NSAIDs has been noted only with strong opioids (oxymorphone, oxycodone, oxytrex, fentanyl, morphine sulphate).99 This use is reserved for exceptional circumstances, such as patients awaiting planned surgery; there is a high (over 30%) withdrawal rate of patients treated, because of nausea, constipation, dizziness, somnolence, and vomiting.99 Whether the efficacy of weaker opioids (tramadol or codeine) has not been assessed in long-term trials. Paracetamol–codeine combinations provide a small (5%), but statistically significant (p<0.05), benefit over paracetamol alone, but are associated with more adverse events.100 In the absence of convincing evidence for their safe and effective use, concerns about risks of dependence or addiction to opiates affects the prescription of these drugs.77

Patients sometimes use a group of symptomatic slow-acting drugs for osteoarthritis—ie, glucosamine sulphate, chondroitin sulphate, hyaluronic acid—and, less commonly, avocado soybean unsaponifiable, and diacerein. Randomised trials with glucosamine sulphate have been debated heavily—there is concern about bias, heterogeneity of outcomes, and effect size.98 Most published studies show that glucosamine sulphate has a beneficial effect on pain, with effect size ranging between 0-30 and 0.87,97 but no effect on function and controversial effects on structure modification.101,102 Whether glucosamine sulphate is effective in osteoarthritis remains undetermined.101 In the USA, glucosamine hydrochloride has been assessed thoroughly, but no beneficial effect has been reported.

There is less, but still conflicting, evidence for the effectiveness of chondroitin sulphate on pain and function.104 Avocado soybean unsaponifiables have been assessed for the treatment of osteoarthritis of the knee and hip, but not of the hand. This treatment was effective in relieving pain and improving function in hip more than in knee, osteoarthritis (effect size 0·01–0·76).105 Avocado soybean unsaponifiables are used in some regions of the world, but are unknown in others. Diacerein is reported to have slow-acting, but persistent, symptomatic relief in patients with osteoarthritis (effect size for pain 0·24; 95% CI 0·08–0·39).79

Intra-articular injection of long-acting glucocorticoids is an effective treatment of inflammatory flares of osteoarthritis (effect size for pain relief 0·58); the effect is greatest after 1 week, and diminishes thereafter.106 After injection of the weight-bearing large joints (ankle, knee, hip) the effectiveness of the injection can be enhanced by complete bed rest of the treated joint for 72 h.107

Hyaluronic acid has varying effectiveness when used for intra-articular injections for the treatment of osteoarthritis of the knee. Different products with different injection regimens (up to five consecutive weekly injections) have been used (effect size up to 0·39).108 Investigators have suggested that the high molecular-weight products (even cross-linked components, such as Hylan G-F 20) need to be injected less often and improve effectiveness.79,109,110

A Cochrane review of surgical lavage and debridement in osteoarthritis of the knee111 showed no benefit in the short or long term compared with placebo; in general this procedure is not advised. Other surgical interventions include osteotomy, joint fusion, joint distraction, and joint replacement. Joint replacement is very cost effective in patients with severe symptoms or functional limitations associated with a reduced quality of life, despite conservative treatment.77

New developments

New discoveries about the pathophysiology of osteoarthritis prompt the division of the disease into distinguishable phenotypes. Delineating the different clinical and structural phenotypes of the disease will improve understanding—of disease in patients with pain, trauma, or obese-dominated clinical phenotypes (table 4 lists our attempt)—and will also allow specific targeted treatment in those in whom structural changes in either cartilage, bone, or synovial tissue dominate the disease. Although these phenotypes are not yet fully characterised, distinguishing different phenotypes could herald the start of further discussions. Lack of in-depth understanding of disease pathogenesis and the misconception that all forms of osteoarthritis are the same and have the same clinical and structural characteristics might restrict further development of
Table 4: Proposal for differentiation of clinical phenotypes of osteoarthritis

<table>
<thead>
<tr>
<th>Post-traumatic (acute or repetitive)</th>
<th>Metabolic</th>
<th>Ageing</th>
<th>Genetic</th>
<th>Pain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>Young (&lt;45 years)</td>
<td>Middle-aged (45-65 years)</td>
<td>Old (&gt;65 years)</td>
<td>Variable</td>
</tr>
<tr>
<td><strong>Main causative feature</strong></td>
<td>Mechanical stress</td>
<td>Mechanical stress, adipokines, hyperglycaemia, oestrogen/progesterone imbalance</td>
<td>AGE, chondrocyte senescence</td>
<td>Gene related</td>
</tr>
<tr>
<td><strong>Main site</strong></td>
<td>Knee, thumb, ankle, shoulder</td>
<td>Knee, hand, generalised</td>
<td>Hip, knee, hand</td>
<td>Hand, hip, spine</td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td>Joint protection, joint stabilisation, prevention of falls, surgical interventions</td>
<td>Weight loss, glycaemia control, lipid control, hormone replacement therapy</td>
<td>No specific intervention, sRAGE/AGE breakers</td>
<td>No specific intervention, gene therapy</td>
</tr>
</tbody>
</table>

Osteoarthritis is not one disease, and might benefit from the recognition of its different phenotypes. AGE=advanced glycation endproducts. sRAGE=soluble receptor for advanced glycation endproducts.

Diagnosis, treatment, and monitoring of the many forms of the disease. A consensus on subgrouping osteoarthritis into such phenotypes will take time.

In the structure phenotype, after entering a point of no return in which damage of the cartilage matrix over-rides synthesis, a vicious cycle of progressive damage ensues in which impaired biomechanical properties result in further damage. Autocrine loops of soluble factors released by the triggered chondrocytes trigger an inflammatory response that accelerates the breakdown process. This inflammatory activity is enhanced by the accelerated release of catabolic cartilage constituents that provide an additional vicious cycle in the process of tissue destruction. The stiffening of the subchondral bone (sclerosis) reported in the more advanced stages of the disease increases stresses in the overlying cartilage and adds to the damage. Also, in the early phase, subchondral bone changes might cause cartilage damage and might even precede it. Soluble factors produced locally in subchondral bone are potential candidates to act on deep-zone articular chondrocytes to promote abnormal remodelling and metabolism of deep cartilage, leading to its breakdown. This breakdown is facilitated by the interaction between bone and cartilage that was originally thought to be a tight interface, but is now recognised as allowing soluble factors to migrate between bone and cartilage. Moreover, angiogenesis has been identified at the junction of articular hyaline cartilage and adjacent subchondral bone and therefore tissue damage of the whole joint is also biochemical and not merely mechanical.

In the age phenotype, chondrocytes sense alterations in mechanical stresses and, dependent on the context, respond with anabolic or catabolic biochemical processes. The cartilage degeneration is not merely mechanically induced wear and tear, but a complex of biochemical interactions.

Ageing alters the response of chondrocytes: aged chondrocytes produce more inflammatory cytokines, tissue degrading enzymes, and growth factors. Moreover, advanced glycation endproducts (AGES), which accumulate in cartilage, can bind to specific receptors (receptor of advanced glycation endproducts; RAGE) expressed on chondrocytes, increasing their catabolic activity. AGE induces alteration of biomechanical properties by stiffening the cartilage, which makes it brittle and more prone to damage. There is no clear clinical evidence of how AGE contributes to the development and progression of osteoarthritis.

Development of soluble AGE receptors, sRAGE, that bind to AGEs and thereby inhibit the activation of cell-surface RAGE, showed efficacy in the treatment of vascular complications in animal models of diabetes. Also, prevention or reversal of AGE formation by diet or specific cleavage of AGE-crosslinks is subject to study and might become feasible.

In the obesity phenotype, the overload effect on joint cartilage might, in part, explain the greater risk of osteoarthritis in overweight people. Advances in the physiology of adipose tissue provide further information about the relation between obesity and osteoarthritis. Indeed, a positive association between obesity and osteoarthritis has been reported for non-weight-bearing joints, such as those of the hands, and not only knee joints. These reports suggest that joint damage might be caused by systemic factors such as adipose factors, the so-called adipokines, which might provide a metabolic link between obesity and osteoarthritis, and which, in addition to weight loss, could become a specific therapeutic target.

Growing knowledge of the pathogenetic mechanisms involved in osteoarthritis will lead to the development of new classes of drugs for targeted treatment; many new pharmacological approaches in the management of osteoarthritis are under development. Calcitonin, a hormone of calcium homeostasis, inhibits osteoclast activity and also has a direct effect on cartilage by the inhibition of matrix metalloproteinase activity. In a pilot study in patients with osteoarthritis, CTXII, a degradation marker, decreased after patients were given oral calcitonin. At present, calcitonin is under investigation in a long-term randomised controlled trial.

Nitric oxide is one of the catabolic mediators in cartilage and synovium. Inducible nitric oxide synthase is upregulated in osteoarthritis and in various pain states. At present, a study is assessing a specific inducible nitric...
oxide synthase inhibitor (SD-6010) in patients with knee osteoarthritis. Studies with bisphosphonates were done with the aim of inhibiting increased bone turnover in osteoarthritis; however, they were negative with regard to symptoms and radiological progression, although biochemical markers of cartilage turnover decreased.116

Advances in pain neurobiology have shown the role of supraspinal pathways and downstream neurotransmitters and effectors in chronic pain. Antibodies to nerve growth factor have been developed by several companies. The benefit-to-risk ratio of these compounds is not yet clear and needs to be assessed further.117 Initial studies in patients with osteoarthritis have shown a slight clinical benefit of the centrally acting compound duloxetine in patients with chronic painful osteoarthritis.118 Duloxetine is a serotonin–norepinephrine reuptake inhibitor used in the treatment of depression, and also assessed in fibromyalgia and diabetic peripheral neuropathy.

Contributors
All authors contributed equally to the search of published work, the discussions, and writing. All authors gave their final approval for the decision to submit for publication.

Conflicts of interest
We declare that we have no conflicts of interest.

References


94 Brosseau L, Y
97 Lohmander LS, McKeith D, Svensson O, et al. A randomised,
90 Osiri M, Welch V, Brosseau L, et al. Transcutaneous electrical nerve
88 Lafeber FP, Intema F, van Roermund PM, Marijnissen AC.
87 Brouwer RW, Jakma TS, Verhagen AP, Verhaar JA,
86 Anandacoomarasamy AJ, Leibman S, Smith G. Weight loss in
85 Richette PJ, Pointou C, Garnero P. Beneficial effects of massive
83 Roddy E, Zhang W, Doherty M. Aerobic walking or strengthening
70: 883–87.
67: 20–27.
66: 139–44.
65: 373–84.
63: 20–27.
62: 139–44.
61: 1704–11.
60: 524–33.
59: 398–408.
58: 254–60.
57: 138–146.
56: 107–16.
54: 876–86.
53: 2: CD005328.
52: 254–60.
51: 1704–11.
50: 69–119.
48: 871–89.
47: 19–27.
44: 871–89.
41: 151–56.
40: 876–86.
39: 181–86.
38: 181–86.
37: 7: CD005328.
36: 871–89.
34: 16: 391–402.
33: 254–60.
32: 138–146.
31: 151–56.
28: 181–86.
27: 871–89.
26: 151–56.
24: 871–89.
23: 151–56.
22: 871–89.
21: 181–86.
20: 871–89.
19: 181–86.
18: 871–89.
17: 871–89.
16: 871–89.
15: 871–89.
14: 871–89.
13: 871–89.
12: 871–89.
11: 871–89.
10: 871–89.
9: 871–89.
8: 871–89.
7: 871–89.
6: 871–89.
5: 871–89.
4: 871–89.
3: 871–89.
2: 871–89.
1: 871–89.
0: 871–89.


Arthritis 2

Spondyloarthritis

Maxime Dougados, Dominique Baeten

Spondyloarthritis is a group of several related but phenotypically distinct disorders: psoriatic arthritis, arthritis related to inflammatory bowel disease, reactive arthritis, a subgroup of juvenile idiopathic arthritis, and ankylosing spondylitis (the prototypic and best studied subtype). The past decade yielded major advances in the recognition of spondyloarthritis as an entity, the classification of the disease, and understanding of the genetic and pathophysiological mechanisms of disease-related inflammation and tissue damage. In parallel, new clinical and imaging outcomes have allowed the assessment of various therapeutic modalities. Blockers of tumour necrosis factor are a major therapeutic advance, but the exact roles of physiotherapy, and treatment with non-steroidal anti-inflammatory drugs and other biological treatments are unknown. The major challenges with direct relevance for clinical practice for the next decade are the development of techniques for early diagnosis, therapeutic modulation of structural damage, and, ultimately, induction of long-term, drug-free remission.

Introduction

In 1974, Moll and colleagues1 established the concept of a group of inter-related disorders originally termed seronegative spondarthritides. The group of diseases now called spondyloarthritis consists of psoriatic arthritis, reactive arthritis, arthritis related to inflammatory bowel disease, a subgroup of juvenile idiopathic arthritis, and ankylosing spondylitis—the prototype of spondyloarthritis.2 The various clinical forms include spinal (axial) features, peripheral arthritis, enthesopathy, and extra-articular features such as uveitis, psoriasis, and inflammatory bowel disease. The clinical rationale for grouping these diseases is that they are simultaneously or sequentially identified in the same patient or in a family member. Furthermore, clinical characteristics such as eye involvement and enthesopathy are similar whatever the diagnosis.3 A strong argument, based on work in animals, in favour of grouping these diseases is that HLA-B27 transgenic rats develop the various clinical features that are noted in human beings with spondyloarthritis.4

One subject of debate at present is whether the clinical approach, including diagnosis, classification, and management, should be focused on a specific disease subtype (eg, ankylosing spondylitis) or on the overall group of spondyloarthritis. In the 1970s, several sets of criteria were proposed to classify patients with a specific spondyloarthritis subtype, such as the modified New York criteria for ankylosing spondylitis.4 These criteria have important restrictions in clinical practice: they focus exclusively on the axial features, omitting the other clinical features of the disease. In 1990, Amor and colleagues5 proposed the first set of classification criteria for the entire group of spondyloarthritis, allowing a patient to be classified as having spondyloarthritis whatever the presenting symptoms. A different set of criteria for the entire group of spondyloarthritis was developed by the European Spondyloarthritis Study Group,6 with inflammatory back pain and peripheral arthritis as major entry criteria. Recognition of the drawbacks of criteria focused on a specific subtype, the Assessment of Spondyloarthritis International Society (ASAS) did a large cross-sectional study to propose new criteria on the basis of the two main clinical features identified in daily practice—eg, axial symptoms and peripheral involvement.

In the first set of criteria focusing on patients presenting with axial symptoms (panel),7 the term axial spondyloarthritis was proposed for the entire range of axial diseases irrespective of structural damage. These criteria emphasise three important points: the relevance of the clinical features identified whatever the presenting symptoms, the value of new imaging techniques to detect sacroiliac changes, and the contribution of HLA-B27 typing.

One important advance is the use of MRI to assess sacroiliac changes. Plain radiographs can detect only structural changes such as joint erosion and

Search strategy and selection criteria

We searched The Cochrane Library and Medline for work published in the past 5 years (2005–10), as well as the abstracts of the American (American College of Rheumatology) and European (European League Against Rheumatism) congresses of Rheumatology published during the past 2 years (2009–10). We used the search terms "spondyloarthropathy", "spondylarthropathy", "spondyloarthritis", "spondylarthritides", "ankylosing spondylitis", and "psoriatic arthritis". We limited our search to published work in English. We also searched the reference lists of articles identified by this search strategy and in particular the articles that summarised systematic research on a specific topic. Review articles and book chapters are cited to provide readers with more details and more references than we can accommodate in this paper.

See Comment page 2067

This is the second in a Series of three papers about arthritis

Paris-Descartes University, Medicine Faculty, UPRES EA 4058, AP-Hôpital Cochin, 27 Rue du Faubourg Saint Jacques, Paris, 75014, France

Correspondence to: Prof Maxime Dougados, René Descartes University, Department of Rheumatology, Hôpital Cochin, 27 Rue du Faubourg Saint Jacques, Paris, 75014, France

m.doug@chc.aphp.fr
Panel: ASAS classification criteria for axial spondyloarthritis in patients with back pain for 3 months or more and age at onset younger than 45 years

Sacroiliitis on imaging* plus one or more features of spondyloarthritis† or HLA-B27 plus two or more other features of spondyloarthritis‡

ASAS=Assessment of Spondyloarthritis International Society. *Active (acute) inflammation on MRI highly suggestive of sacroiliitis associated with spondyloarthritis or definite radiographic sacroiliitis according to modified New York criteria. †Inflammatory back pain, arthritis, enthesitis (heel), uveitis, dactylitis, psoriasis, Crohn’s disease or ulcerative colitis, good response to non-steroidal anti-inflammatory drugs, family history for spondyloarthritis, HLA-B27, or elevated C-reactive protein (a spondyloarthritis feature in the context of chronic back pain).

subchondral-bone sclerosis seen at the late stage of the disease; this restriction is also the case for CT, although with higher sensitivity and specificity but greater exposure to radiation. Unfortunately, the medical term used to describe such chronic changes focuses on inflammation—eg, sacroiliitis—despite the fact that plain radiographs cannot detect inflammation. By contrast, MRI allows the visualisation of synovial fluid, synovitis within the sacroiliac joint, and subchondral-bone oedema. The relevant abnormalities detected with MRI have been described and clearly defined, allowing inclusion of active inflammatory lesions of sacroiliac joints with definite bone-marrow oedema and osteitis on MRI in the new criteria for axial spondyloarthritis. Whether such a definition—eg, MRI findings at the sacroiliac joints—is optimum remains an open question since data suggest that inflammatory lesions of the posterior structures of the spine as well as the spinal fatty Romanus lesions (fatty changes at vertebral corners) are also suggestive of spondyloarthritis. More importantly, however, these criteria were developed in a well defined cross-sectional study population (eg, age <45 years and with back pain for at least 3 months) and have not yet been validated for diagnostic use in prospective studies in clinical practice.

The second set of criteria proposed by ASAS is focused on patients presenting with peripheral rheumatological involvement (eg, peripheral arthritis, enthesopathy, dactylitis) without axial symptoms. These criteria (figure 1) also emphasise the importance of the different clinical features, HLA-B27 typing, and imaging of sacroiliac joints despite the absence of spinal symptoms. Sacroiliac abnormalities at imaging raise the question of which of the investigations should be done when spondyloarthritis is suspected, whatever the presenting symptoms. Clinicians agree on the use of HLA-B27 typing, although it is only useful in cases with an a-priori high suspicion, and a negative result does not preclude the presence of spondyloarthritis. The findings of the study used to develop the criteria also suggest that in a case of peripheral rheumatological presentation, the systematic radiological (eg, plain radiographs and MRI) assessment of the sacroiliac joints might be of interest even in the absence of any axial features. Similarly, a systematic assessment of different entheses allows differentiation between spondyloarthritis patients and controls even in the absence of clinical enthesopathy.

Enthesopathy, inflammation at the bone insertion sites of ligaments and tendons, is an important ASAS criterion. The main peripheral clinical location is the heel (inferior part at the insertion of plantar fascia on the calcaneus and posterior part at the insertion of Achilles tendon on the calcaneus). The recognition of spondyloarthritis and the use of these new criteria should allow clinical trials in patients with early disease and thereby the assessment of treatments to alter the course of the disease. Whether these criteria will also shorten the diagnostic delay remains to be investigated prospectively. Another interesting approach to reduce the diagnostic delay is the development of early referral strategies, since patients with back pain are usually first seen by primary care physicians. Defining better strategies and techniques for early diagnosis remains one of the major challenges in spondyloarthritis for the next decade.

Pathophysiology

Advances in the classification of spondyloarthritis show that progress in the understanding of genetics (eg, the gene for HLA-B27), the pathophysiology of inflammation (eg, lesions on MRI), and structural damage (eg, sacroiliitis on plain radiographs) affect clinical practice in the context of classification and diagnosis. Basic understanding of the pathophysiology of the disease is even more relevant for outcome measurement and targeted treatment.

Through familial aggregation studies investigators have estimated that genetic risk factors contribute to 80–90% of the susceptibility to ankylosing spondylitis. The stronger concordance rates between monozygotic (50–75%) versus dizygotic (15%) twins confirms that familial aggregation is related to genetic rather than environmental factors.
The major genetic risk factor is HLA-B27, an MHC class I molecule. This association is present in many genetically diverse populations and across all major HLA-B27 subtypes. Whereas HLA-B27*06 and HLA-B27*09 have long been thought to be protective, the finding of ankylosing spondylitis in carriers of these alleles that encode these molecules suggests a hierarchy of association of different HLA-B27 subtypes with ankylosing spondylitis. Whether the effect of the specific aminoacid substitutions in the peptide binding groove of HLA-B27*06 and HLA-B27*09 can explain the differential association in vivo remains to be established.

The presence of HLA-B27 in 80–90% of patients with ankylosing spondylitis and the spontaneous spondyloarthropathy-like disease in HLA-B27 transgenic rats suggest a direct and dominant effect of the gene encoding this molecule. However, only a small proportion of people in the general population who harbour HLA-B27 (5–6% in white people) develop ankylosing spondylitis, and HLA-B27 explains only 20–40% of the genetic susceptibility to ankylosing spondylitis—suggesting the contribution of additional genes. Genome-wide association studies (GWASs) have allowed the identification of several of these additional genes (table).

A definite association has been identified with the genes for endoplasmic reticulum aminopeptidase 1 (ERAP1), interleukin 23 receptor (IL23R), and the gene deserts on chromosomes 2p15 and 21q22. Besides these definite associations, GWAS findings suggested potential associations with genes for tumour necrosis factor (TNF) receptor 1 (TNFSF1A), the signalling molecule TNF receptor 1-associated death domain protein (TRADD), the TNF superfamily cytokine TNFSF15, interleukin 1α (IL1A), interleukin 1 receptor 2 (IL1R2), the vascular morphogenesis protein gene antrax toxin receptor 2 (ANTXR2), and the innate immune receptor caspase recruitment domain family, member 9 (CARD9). Other candidate genes such as non-B27 MHC genes, the familial Mediterranean fever-related MEFV, and signal transducer and activator of transcription 3 (STAT3) need additional confirmation.

The strong genetic predisposition also applies to other spondyloarthritides subtypes as suggested by a recurrence rate of disease in 12% of the first-degree relatives of spondyloarthritides patients. Accordingly, genes encoding HLA-B27 and interleukin 23 receptor are associated with different spondyloarthritides subtypes. Additionally, genes such as IL23R also confer risk for spondyloarthritides-associated disorders such as Crohn’s disease and psoriasis (table). The absence of familial clustering of distinct phenotypic features of the subtypes suggests a dominant shared genetic factor in all spondyloarthritides forms, with additional genetic and environmental factors contributing to the phenotypic diversity. Reinforcing this idea, HLA-B27 transgenic rats develop not only spondylitis but also the full spondyloarthritides clinical range with peripheral arthritis, colitis, uveitis, and skin disease, with environmental factors such as the gut flora and additional genetic factors determining the exact phenotype.

The traditional pathophysiological framework for spondyloarthritides is the arthritogenic-peptide theory, which proposes that HLA-B27 presents self-peptides that resemble pathogen-derived peptides to CD8-restricted T lymphocytes. Circumstantial evidence for this hypothesis is provided by the triggering of spondyloarthritides by gastrointestinal or urogenital infections, and the presence of HLA-B27-restricted CD8-T-cell clones that are reactive against bacterial antigens as well as against self-proteins from cartilage in the inflamed joint. However, this hypothesis has been seriously challenged by two independent reports that CD8 T cells are not needed for disease in HLA-B27 transgenic rats. Alternatively, the anticartilage responses in human beings are not disease-specific, suggesting common secondary autoimmune responses rather than primary pathophysiological processes. In more general terms, the scarce evidence for HLA-B27-restricted autoimmune T-cell responses, the absence of shared genetic risk factors for autoimmune diseases such as PTPN22 polymorphisms, and the absence of known disease-specific autoantibodies question whether spondyloarthritides is a genuine autoimmune disease driven by T-cell or B-cell reactivity towards self-antigens.

Two additional hypotheses have emerged to explain the role of HLA-B27 (figure 2). Both hypotheses argue for an autoinflammatory rather than autoimmune origin since HLA-B27 has a role in triggering innate immune responses rather than its canonical role of antigen presentation. If correct, this hypothesis might have important implications because it predicts, for example, that inflammation will happen at sites of

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Associated with</th>
</tr>
</thead>
<tbody>
<tr>
<td>6p21.3</td>
<td>HLA-B</td>
<td>Antigen presentation</td>
</tr>
<tr>
<td>5q15</td>
<td>ERAP1</td>
<td>Aminopeptidase</td>
</tr>
<tr>
<td>1p31.2</td>
<td>IL23R</td>
<td>Cytokine receptor</td>
</tr>
<tr>
<td>2p15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21q22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12p13.2</td>
<td>TNFSF1A</td>
<td>Cytokine receptor</td>
</tr>
<tr>
<td>16q22</td>
<td>TRADD</td>
<td>Signalling</td>
</tr>
<tr>
<td>9q32</td>
<td>TNFSF15</td>
<td>Inflammatory cytokine</td>
</tr>
<tr>
<td>2q14</td>
<td>IL1A</td>
<td>Inflammatory cytokine</td>
</tr>
<tr>
<td>2q12</td>
<td>IL1R2</td>
<td>Cytokine receptor</td>
</tr>
<tr>
<td>9q34</td>
<td>CARD9</td>
<td>Innate immune defence</td>
</tr>
<tr>
<td>4p21.3</td>
<td>ANTXR2</td>
<td>Vascular morphogenesis</td>
</tr>
</tbody>
</table>

*Some factors are also associated with psoriasis or inflammatory bowel disease.
bacterial or mechanical stress, and that T-cell or B-cell directed treatments might not be effective in spondyloarthritis.

The first hypothesis proposes that β2 microglobulin-free HLA-B27 heavy chains can assemble into disulphide-linked homodimers expressed at the cell surface that can be directly recognised by the killer immunoglobulin-like receptors KIR3DL2 independently of the bound peptide. The resulting unfolded-protein response (UPR) induces an altered responsiveness and cytokine production of inflammatory cells to a range of innate immune stimuli. However, overexpression of human β2 microglobulin to reduce the UPR in HLA-B27 transgenic rats exacerbated rather than prevented arthritis and spondylitis, whereas colitis was unchanged. Although investigators still debate to what extent β2 microglobulin overexpression really down-regulates UPR, this discrepancy emphasises that the non-mutually exclusive functions of HLA-B27 might differ between models and between distinct features of spondyloarthritis.

The altered cellular responsiveness induced by the UPR accords with the predilection of spondyloarthritis for tissues exposed to either bacterial or mechanical stress. Bacterial stress is shown by the association with inflammatory bowel disease, gastrointestinal infections, and abnormal Toll-like receptor expression and function. The role of mechanical stress is emphasised by imaging and pathological findings that inflammation happens mainly at the synovio-entheseal complex. Taken together with the prominent infiltration with innate immune cells at affected sites, the stress hypothesis proposes that inflammation in spondyloarthritis is induced by abnormal innate immune responsiveness to mechanical or bacterial danger signals and should thus be seen as an autoinflammatory rather than autoimmune disorder.

Two cytokines are of particular interest in the propagation and perpetuation of inflammation in spondyloarthritis. First, a key role for TNF has been shown through the effectiveness of TNF blockers. This role fits with the genetic associations with TNFRI and the TNFRI signalling molecule TRADD; however, how TNF drives spondyloarthritis is unclear. Many models of TNF overexpression lead to sacroiliitis, with one model giving TNFRI signalling to stromal cells a prominent role. However, these models differ fundamentally from spondyloarthritis by their polyarticular, erosive character without osteoproliferation. The low titres of soluble TNF in spondyloarthritis synovitis and spinal deformities in mice overexpressing transmembrane TNF warrant further investigation of the forms of TNF and TNF receptors in the disease process.

The second cytokine of interest is interleukin 23. Besides the genetic association with IL23R, evidence is emerging that the HLA-B27 induced UPR augments the production of interleukin 23. Altered interleukin-23 production or signalling in spondyloarthritis could lead to abnormal interleukin-17 responses, certainly in view of the data that TNFSF15, CARD9, and the DR3-TRADD pathways can also affect responses of T-helper-17 (Th17) cells. Early demonstration of interleukin-17 overexpression in spondyloarthritis could, however, not be confirmed by independent studies on blood, synovial fluid, and gut. Keeping in mind that interleukin 23 has several functions and targets many cells besides Th17 cells, these expression studies need to be extended to functional studies in vitro and in vivo. Investigators have yet to clarify the potential role of interleukin 1, as suggested by the genetic associations, and interleukin 6 in the induction of a Th17 response, and more generally in the pathophysiology of spondyloarthritis. Emerging data from clinical trials aiming to block interleukins 1, 6, 17, or 23 will be crucial to understand the role of these cytokines.

Genetic risk factors, and the related hypotheses, fall short of explaining the second major feature of spondyloarthritis: the prominent tissue remodelling that leads to osteoproliferation and ankylosis. Three major hypotheses have emerged.

First, the typical structural features cannot be explained by the presumption that the disease is non-erosive. Imaging and histological studies clearly show that bone destruction and erosions are prominent

---

**Figure 2: Potential roles of HLA-B27 in triggering the pathogenesis of spondyloarthritis**

The three main hypotheses relate to the presentation of arthritogenic peptides to autoreactive T lymphocytes, the formation of heavy-chain homodimers (which activate natural killer cells), and the misfolding of HLA-B27 in the endoplasmic reticulum leading to an unfolded-protein response. The role of ERAP1 has not been assessed. Upon bacterial or mechanical stress, these pathways can lead to the abnormal production of proinflammatory cytokines such as tumour necrosis factor and interleukin 12. Investigators do not completely understand the role of additional genetic associations. ERAP1=endoplasmic reticulum aminopeptidase 1. TRADD = TNF receptor 1-associated death domain protein. IL23R = interleukin 23 receptor. IL1A = interleukin 1α. IL1R2 = interleukin 1 receptor 2. TNFSF15 = TNF superfamily cytokine 15. CARD9 = caspase recruitment domain family, member 9.
features of both axial and peripheral spondyloarthritis.\textsuperscript{55–58} Accordingly, cellular and molecular pathways of cartilage and bone destruction are activated at the sites of pathology and, as in rheumatoid arthritis, are largely dependent on TNF.\textsuperscript{55–58}

In line with these findings, the second emerging hypothesis is that the structural features of spondyloarthritis relate to important pathways of endochondral-bone formation. In a model of spontaneous ankylosing enthesitis, signalling by bone morphogenetic proteins was the key pathway driving the structural changes and active signalling of the proteins was identified in target tissues of human spondyloarthritis.\textsuperscript{59} In TNF transgenic mice, activation of Wnt signalling by targeting the inhibitor Dickkopf-related protein 1 reversed the process of bone destruction and induced fusion of sacroiliac joints.\textsuperscript{60,61} Several inhibitors of the Wnt pathway seem to be dysfunctional in human spondyloarthritis and are associated with new bone formation.\textsuperscript{62,63} Further functional analyses of bone morphogenetic proteins, Wnt, and other tissue-remodelling pathways are of paramount importance because they could be attractive targets for treatment.

The third emerging possibility is that osteoproliferation in spondyloarthritis is, at least partly, uncoupled from inflammation. Two hypotheses have been proposed to account for this uncoupling. The first hypothesis claims that osteoproliferation can be explained by the intermittent nature of the inflammation.\textsuperscript{64} In an early disease phase, TNF would simultaneously drive destruction and inhibit remodelling by the Wnt pathway by upregulating Dickkopf-related protein 1 (figure 3). On downregulation of TNF in a later phase, the brake on Wnt-mediated remodelling would be released and the early erosions would trigger reactive osteoproliferation. The relation between early inflammation and subsequent new bone formation is, however, still highly debatable in human ankylosing spondylitis because although inflammation is associated with a greater likelihood new bone formation, most syndesmophytes are located at sites without detectable inflammation.\textsuperscript{65,66} Moreover, this hypothesis cannot explain why new bone formation is independent of osteoclasts in various models\textsuperscript{67,68} or why TNF blockade does not prevent ankylosing enthesitis.\textsuperscript{69} The second hypothesis proposes that direct activation of stromal pathways, including the pathways of bone morphogenic protein, leads to new tissue formation independent of inflammation or early erosive changes.\textsuperscript{70} Mechanical stress at synovio-enthesal complexes might then induce distinct and unrelated pathways of inflammation and tissue remodelling.

Although both hypotheses are not mutually exclusive (figure 3), the relative contribution of both mechanisms and the exact relation between inflammation and stromal-cell activation has major clinical implications: the first hypothesis predicts that early anti-inflammatory treatment will prevent structural damage whereas the second hypothesis predicts that separate assessment and therapeutic targeting of stromal pathways is needed for optimum management of spondyloarthritis.

**Outcome assessment**

The optimum management of patients necessitates systematically addressing five points related to the possible clinical presentations (axial, peripheral, enthesopathy, and extra-articular): does the patient really have the disease, is the disease active, is the disease severe, is the disease potentially severe, and is the disease refractory? One of the major challenges in spondyloarthritis remains the development of sensitive and specific imaging or biological markers for early diagnosis.

Activity in spondyloarthritis is a reference to the inflammation caused by the disease, which is commonly assessed in daily practice with the Bath ankylosing spondylitis disease activity index (BASDAI).\textsuperscript{71} This index consists of questions related to the patient’s self-assessment (eg, fatigue, axial symptoms, peripheral symptoms, enthesopathy, and duration and intensity of morning stiffness). To improve the objective properties of such an index, an ankylosing spondylitis disease activity score (ASDAS) has been developed that includes not only four questions from the BASDAI, but also the level of acute phase reactants.\textsuperscript{72} Preliminary data suggest that the ASDAS is more discriminative than BASDAI when in assessment of TNF blockers.\textsuperscript{73} However, clinicians must further assess the usefulness of this new composite index in daily practice.
Apart from clinical assessment, the activity of the disease could also be assessed by MRI of the spine and sacroiliac joints. Several scoring systems at present assess the reliability, validity, and responsiveness of the technique. Investigators are yet to clarify if the inflammatory abnormalities of the posterior elements of the spine should be included or even if a whole-body MRI should be preferred. Whatever the scoring system, axial disease activity measured by MRI as well as other imaging modalities (eg, ultrasonography for enthesopathy) are useful additional outcome measures in clinical trials but their added value in daily clinical practice is unknown.

The severity of spondyloarthritis is a reference to irreversible structural damage caused by the disease, often due to tissue remodelling and its functional consequences. For clinical studies, several outcomes have been proposed to show severity: death, job loss, functional impairment, range of motion, and hip involvement. Radiological scoring systems assess structural damage at the axial level (eg, mainly new bone formation because of syndesmophytes). The scoring system recommended at present is the modified stoke ankylosing spondylitis scoring system, which consists of cervical and lumbar assessments. The addition of the thoracic spine might improve the sensitivity to change. This new system is very useful in clinical research but it remains unclear whether it should be used routinely in clinical practice.

A further factor in the optimum management of the disease relates to prediction of the natural course of the disease at an early stage in an individual patient. This notion is clinically highly relevant because structural damage and functional impairment in spondyloarthritis are largely irreversible. If the hypothesis that early inflammatory and erosive lesions trigger subsequent osteoproliferation is correct, highly efficient anti-inflammatory treatments should be started as early as possible. However, preliminary data suggest that axial inflammatory lesions detected with MRI are not highly predictive for subsequent ossification. If the structural damage is independent of inflammation but relates to stromal remodelling pathways, it would make a case for additional treatments targeting stromal remodelling irrespective of disease activity and before irreversible structural damage. Long-term follow-up of patients presenting with recent inflammatory back pain in different European cohorts will help to delineate which patients are at risk for long-term structural damage.

Whether the disease is refractory is important to guide the decision for second-line treatments such as anti-TNF. For axial spondyloarthritis, the present recommendation from ASAS and the European League Against Rheumatism is to define a patient as refractory when active disease persists despite the intake of at least two courses of non-steroidal anti-inflammatory drugs (NSAIDs) taken at an optimum dose for at least 2 weeks without needing to be classified a failure of a disease-modifying anti-rheumatic drug. For peripheral arthritis, a refractory disease is defined by an active disease despite current or past intake of disease-modifying anti-rheumatic drugs. To disseminate and facilitate the implementation of such methods, ASAS has recently published a guide to assess spondyloarthritis.

**Treatment**

The objectives of treatment of spondyloarthritis are to improve the condition of the patient (eg, pain, functional disability) as well as to prevent subsequent clinical deterioration. ASAS has provided recommendations for both the management of spondyloarthritis in general and the use of TNF blockers in particular.

A recent Cochrane systematic review of published work concluded that an individual home-based or supervised exercise programme is better than no intervention, that supervised group physiotherapy is better than home exercises, and that combined in-patient spondyloarthritis-exercise therapy with subsequent group physiotherapy is better than group physiotherapy alone. Despite the modality of physiotherapy, another important question is related to the characteristics of the patients who should benefit most from this therapy. In particular, the benefit of such therapy during the painful inflammatory flares of the disease or at a very early stage has not been investigated.

NSAIDs are the cornerstone of pharmacological intervention for ankylosing spondylitis, rapidly reducing pain and stiffness after 48–72 h. Despite this dramatic symptomatic effect, NSAIDs might be also effective on some other outcome measures. These drugs might substantially reduce the level of acute-phase reactants compared with placebo, but with questionable relevance of the recorded size of effect. The investigators of one study also suggest that NSAIDs can delay radiological progression of spine disease when given continuously as a daily dose over 2 years, compared with an on-demand treatment schedule. Although this finding suggests that a systematically continuous daily intake of NSAIDs might be of benefit, the converse argument is the potential long-term gastrointestinal and cardiovascular toxic effects of such therapy, in particular in patients recognised as having more comorbidities than the general population.

Conventional disease-modifying antirheumatic drugs such as sulfasalazine, methotrexate, and leflunomide which have been shown to be effective in the treatment of rheumatoid arthritis, have no proven efficacy for either the axial or enthesopathic features of spondyloarthritis, but some efficacy for peripheral arthritis and other extra-articular features such as psoriasis, uveitis, and inflammatory bowel disease.

Thalidomide has some efficacy in axial spondyloarthritis in open uncontrolled studies, possibly because of its anti-TNF effect, but is thought too toxic for widespread
use.100–102 Although some clinical effectiveness of pamidronate, a bisphosphonate with potential anti-inflammatory and antierosive effects, has been reported,100 further placebo-controlled studies are needed before this treatment can be recommended.

The major clinical and therapeutic advance in spondyloarthritis care is the successful use of TNF blockade in active, refractory disease.104–107 Registration studies in early, preradiographic axial spondyloarthritis as well as in undifferentiated peripheral spondyloarthritis are ongoing.

A couple of issues are of particular relevance for daily clinical practice. First, TNF blockade is highly effective in targeting the different disease features—eg, not only axial disease but also peripheral arthritis, enthesitis, and extra-articular features such as psoriasis or uveitis.108–110 TNF blockade also has a substantial effect on general symptoms such as fatigue and substantially improves the overall function and quality of life. Long-term follow-up studies suggest that effectiveness is maintained for several years of treatment. Second, short-term and long-term studies suggest a safety profile of TNF blockade in spondyloarthritis that is similar to that of rheumatoid arthritis and inflammatory bowel disease, and reactivation of tuberculosis remains the major concern.111 Third, the various TNF blockers seem to be equally potent for the treatment of axial, peripheral, and extra-articular features, with the exception that etanercept has no proven efficacy in inflammatory bowel disease. As to safety, large registries suggest that the risk for tuberculosis and possibly also herpes zoster might be lower with etanercept than with the monoclonal anti-TNF antibodies infliximab and adalimumab.112–115 Finally, by contrast with rheumatoid arthritis and Crohn’s disease, there is no recommendation at present in axial spondyloarthritis to combine TNF blockers with drugs such as methotrexate or azathioprine.

Despite its major therapeutic effectiveness, TNF blockade also has important limitations. First, 20–40% of the patients do not respond well to treatment and clinical, biological, or imaging findings that predict a better response at the group level lack specificity to make reliable predictions in individual patients.116,117 In case of failure of a first TNF blocker, trying a second drug is justified since many patients do still respond to a different anti-TNF.118,119 Second, TNF blockade does not induce longlasting remission since almost all patients relapse within 6–12 months of interruption of treatment.120 Third, TNF blockade seems to halt joint destruction,121 but fails to substantially slow new bone formation in spondyloarthritis.122,123 It remains unclear whether this effect is related to the fact that TNF blockade was started too late in the disease course in these studies or to the fact that new bone formation is uncoupled from TNF-driven inflammation in spondyloarthritis. These three caveats suggest that there is still an important unmet need for highly effective anti-inflammatory treatments as well as for remission-inducing and structure-modifying therapies.

Several other biological agents have been tested in small proof-of-concept trials in ankylosing spondylitis and psoriatic arthritis on the basis of their efficacy in related diseases such as rheumatoid arthritis and psoriasis. B-cell depletion by the anti-CD20 antibody rituximab did not show similar efficacy in ankylosing spondylitis as in rheumatoid arthritis, although some response was noted in TNF-blocker naive patients in an open study.124,125 T-cell targeted therapies have also not been very successful in spondyloarthritis. Despite the role of T cells in psoriasis, psoriatic arthritis did not respond to the anti-CD11a monoclonal antibody efalizumab126 and showed only slight clinical improvement with the anti-LFA3 antibody alefacept127 and the CTLA4-Ig construct abatacept.128 These findings are thus consistent with the pathophysiological concept that spondyloarthritis might be an autoinflammatory rather than a T-cell or B-cell driven autoimmune disease.

The genetic associations have raised interest in blockade of cytokines other than TNF. Interleukin-1 blockade with the interleukin-1 receptor antagonist anakinra did not show consistent efficacy in ankylosing spondylitis,129 the human anti-interleukin 12/1 interleukin 23 monoclonal antibody ustekinumab had slight but significant (p=0.0002) efficacy in psoriatic arthritis.130 Trials with interleukin-6 and interleukin-17 blockade are underway.

**Future prospects**

From the present state of the art in spondyloarthritis, several important clinical and pathophysiological issues seem unsolved: the development and validation of better clinical or biological markers for early diagnosis and for prognosis, clarification whether the subtypes of spondyloarthritis are driven by different pathophysiological processes or rather represent different phenotypes of a single pathological entity, deciphering the functional role and the interaction of genes emerging from GWAS, in-depth understanding of the cellular and molecular mechanisms of tissue remodelling and their interaction with inflammation, and the development of newer anti-inflammatory therapies, including the clinical assessment of interleukin-6 blockade and interleukin-17 blockade, as well as treatments targeting tissue remodelling. In view of the present research efforts, our understanding of these issues will probably develop rapidly over the coming years.

**Contributors**

The authors contributed equally to the preparation of this paper.

**Conflicts of interest**

MD has received consultancy fees, speaker’s fees, and his department has received research grants from Abbott, Aventis, Pfizer, Roche, Sanofi, Shering Plough, and UCB. DB has received consultancy fees, speaker’s fees, and research grants from Abbott, Centocor, Novartis, Pfizer, Shering Plough, and UCB.
Series

Acknowledgments
We thank Kris A Reedquist for critical reading of our paper.

References

www.thelancet.com Vol 377  June 18, 2011


Madsen KB, Jurik AG. Magnetic resonance imaging grading system for active and chronic spondylarthritides changes in the sacroiliac joint. Arthritis Care Res 2010; 62: 11–18.


Madsen KB, Jurik AG. Magnetic resonance imaging grading system for active and chronic spondylarthritides changes in the sacroiliac joint. Arthritis Care Res 2010; 62: 11–18.


Arthritis 3

Juvenile idiopathic arthritis

Berent Prakken, Salvatore Albani, Alberto Martini

Juvenile idiopathic arthritis is a heterogeneous group of diseases characterised by arthritis of unknown origin with onset before age of 16 years. Pivotal studies in the past 5 years have led to substantial progress in various areas, ranging from disease classification to new treatments. Gene expression profiling studies have identified different immune mechanisms in distinct subtypes of the disease, and can help to redefine disease classification criteria. Moreover, immunological studies have shown that systemic juvenile idiopathic arthritis is an acquired autoimmune inflammatory disease, and have led to successful studies of both interleukin-1 and interleukin-6 blockade. In other forms of the disease, synovial inflammation is the consequence of a disturbed balance between proinflammatory effector cells (such as T-helper-17 cells), and anti-inflammatory regulatory cells (such as FOXP3-positive regulatory T cells). Moreover, specific soluble biomarkers (S100 proteins) can guide individual treatment. Altogether these new developments in genetics, immunology, and imaging are instrumental to better define, classify, and treat patients with juvenile idiopathic arthritis.

Introduction

Juvenile idiopathic arthritis is not a single disease, but a term that encompasses all forms of arthritis that begin before a patient is aged 16 years that persist for more than 6 weeks and are of unknown origin.1 It is the most common childhood chronic rheumatic disease and causes much disability. In high-income countries it has a yearly incidence of 2–20 cases per 100 000 population and a prevalence of 16–150 cases per 100 000 population.1 In this Seminar we focus on developments in the understanding of pathogenesis and in the diagnosis and treatment, and discuss how translational research and new imaging modalities and biomarkers are expected to improve diagnostic and treatment options.

Clinical manifestation and classification

Disorders described by the term juvenile idiopathic arthritis have been grouped on the basis of clinical and laboratory features to try and identify homogeneous, mutually exclusives categories.1 Clinical and laboratory findings have improved the understanding of the different forms of chronic childhood arthritis.4 Although some categories identify definite disease entities, others still include heterogeneous disorders.7

Well-characterised categories

Systemic juvenile idiopathic arthritis is characterised by prominent systemic features, such as fever, rash, and serositis, and is much like adult-onset Still’s disease.1 Pronounced activation of a patient’s innate immune system and the absence of any consistent association with autoantibodies or human leucocyte antigen (HLA) have led to the hypothesis that this systemic form of disease is a polygenic autoimmune inflammatory syndrome.8 Findings from previous studies suggesting a major pathogenic role for interleukin-6 in the disease have been substantiated by the reported effectiveness of treatment with tocilizumab, an anti-interleukin 6 receptor antibody.9,10 Moreover, the finding that anti-interleukin 1 treatment can also be effective11 has led to the subsequent delineation of two subpopulations of this form of disease: one with a pronounced, complete response to interleukin-1 blockade (much the same as seen in cryopyrin-associated autoinflammatory syndromes) and another that is resistant to treatment or has an intermediate response. These two subpopulations do not differ in interleukin-1 in vitro production or in serum cytokine concentrations, but only in the number of joints affected and in neutrophil counts; patients with fewer joints affected or with a higher neutrophil count have an increased probability of responding to anti-interleukin-1 treatment. Thus, systemic juvenile idiopathic arthritis can be stratified into at least two subgroups on the basis of responsiveness to inhibition of interleukin-1 and therefore possible pathogenic relevance of—interleukin-1. Once the findings from a phase 3 trial of canakinumab (an anti-interleukin-1 antibody; NCT00889863) and a second phase 3 trial of tocilizumab (NCT00642460) are available, the pattern of response might provide new clinical or laboratory biomarkers useful to further understand the
heterogeneity between these two subgroups and the reciprocal roles of interleukin-1 and interleukin-6 in disease pathogenesis.

Rheumatoid factor positive polyarthritis, a small subcategory of juvenile idiopathic arthritis (affecting 5% of patients), is thought to be much like adult rheumatoid factor (RF)-positive rheumatoid arthritis; indeed, it is the only form of juvenile idiopathic arthritis with positive antibodies to cyclic citrullinated peptides. Major differences with the adult form of disease are in the effect that the disease can have on a growing skeleton, leading to either general growth retardation or accelerated growth of an affected joint.

Enthesitis-related arthritis is a form of undifferentiated spondyloarthropathy. Most patients are HLA-B27 positive, and, in about 30–40% of patients, the disease can progress to affect sacroiliac joints. Although the category oligoarthritis, as a whole, is probably heterogeneous, most patients—at least in developed countries—have a well-defined disease that is seen only in children. Arising more often in girls than in boys, oligoarthritis has early onset (before 6 years of age), has consistent HLA associations, and is characterised by asymmetric arthritis that affects mainly large joints. Patients have high concentrations of positive antinuclear antibodies (ANA) and a high risk of developing chronic iridocyclitis. Juvenile idiopathic arthritis classification criteria distinguish two categories of oligoarthritis: persistent oligoarthritis, in which the disease affects four joints or fewer, and extended oligoarthritis, in which more than four joints are affected after the first 6 months of disease. However, patients with either persistent or extended oligoarthritis who test positive for ANA have similar clinical characteristics (eg, age at onset, sex ratio, asymmetry of articular involvement, and frequency of iridocyclitis), which strongly suggests that these two categories of oligoarthritis are the same disease, differing only in severity.

Less well-characterised categories
Rheumatoid factor-negative polyarthritis is a heterogeneous category of juvenile idiopathic arthritis. At least two subcategories can be identified: one that is similar to adult-onset RF-negative rheumatoid arthritis, characterised by a symmetric synovitis of large and small joints, onset at school age, and the absence of ANA expression, and another that resembles oligoarthritis, apart from the number of joints affected in the first 6 months of disease. Similarities between this second subset and early-onset oligoarthritis led to the hypothesis that they are the same disease, with a more rapid spread of arthritis in the second subset than in early-onset arthritis. This view has been lent support by the finding that ANA-positive oligoarthritis share the same features with ANA-positive RF-negative polyarthritis, but not with ANA-negative RF-negative polyarthritis or with ANA-negative oligoarthritis. The view is also lent support by the fact that ANA-positive RF-negative polyarthritis is seldom seen in countries in which ANA-positive oligoarthritis is rare.

If psoriatic arthritis is defined according to the presence of arthritis and psoriasis or some psoriatic features (as it is in the Vancouver criteria), two disease entities exist: one in the enthesitis-related arthritis category, which is therefore, like adult psoriatic arthritis, a form of spondyloarthropathy, and a second that is very similar to ANA-positive oligoarthritis with only small differences such as it affects small joints more often than large joints, a feature that could be attributable to psoriatic diathesis in the ANA-positive oligoarthritis phenotype. The association of psoriasis with arthritis seems to lead to the identification of two subsets of patients—one with disease that is similar to adult psoriatic arthritis and another with disease that has only minor differences with ANA-positive oligoarthritis.

Indeed, most patients who meet the present classification criteria for psoriatic arthritis, in which patients with enthesitis are by definition excluded, have features of ANA-positive oligoarthritis.

Perspectives for a new classification
To improve our understanding of the cause and development of the various forms of childhood chronic arthritis and find more suitable treatments, the identification of categories that, at least from a clinical point of view, seem as homogeneous as possible is essential to enable immunological, gene expression, and genome-wide association studies. If more homogeneous groups within juvenile idiopathic arthritis are to be identified, which seems likely in view of the heterogeneity within the present subcategories, then some classification criteria need to be reconsidered.

In 2003, we suggested that the number of joints affected and the presence of psoriasis are not suitable criteria with which to identify homogeneous disease entities, and that children with clinical features that are strongly suggestive of a common cause (eg, asymmetric arthritis, early onset, sex ratio, ANA positivity, high risk for iridocyclitis) are wrongly classified into three different disease categories—oligoarthritis, RF-negative polyarthritis, and psoriatic arthritis. We postulated that the grouping of patients into these three categories according to criteria (alone or in combination) such as ANA positivity, age at disease onset, or pace (asymmetrical or symmetrical) at which the disease affects joints could lead to the definition of more homogeneous categories.

The analysis of gene expression profiles has confirmed the heterogeneity of polyarticular juvenile idiopathic arthritis. Moreover, Barnes and colleagues recorded a B-cell signature that characterises patients with early-onset arthritis independently from the number of joints affected. Their study accords with previous findings showing that plasma cell infiltration of the synovium was more common in the early phase of joint inflammation but not related to disease activity or severity. Additionally,
high-resolution HLA class I and class II typing has shown similarities between early-onset polyarticular and oligoarticular forms of disease. All these findings corroborate the suggestion that ANA-positive, early-onset arthritis is a homogeneous entity that is classified into different disease categories because of differences in the spread of arthritis or the association with psoriasis or psoriatic features.

Ultrasonography has shown much discrepancy between imaging and clinical examination in the assessment of the number of affected joints, making a classification on the basis of the number of joints affected even more complicated. However, ultrasonography allows better differentiation between tendon and articular involvement, thus providing information that could also be of relevance for classification purposes.

Because in children, as in adults, several different diseases exist that all cause chronic arthritis, the terms juvenile idiopathic arthritis and juvenile idiopathic arthritis onset-forms will probably become outdated as more is learnt about each disease. These terms wrongly suggest that juvenile idiopathic arthritis is a single disease (as was thought many years ago) and that the various onset-forms (or categories) are only phenotypic variants.

**Cause and pathogenesis**

One of the most intriguing questions in the study of human autoimmune diseases is what determines the phenotype and organ specificity of a disease. In juvenile idiopathic arthritis, for example, the occurrence of uveitis is related to explicit risk factors, such as age at disease onset, sex, the presence of ANA auto-antibodies, and the subtype of juvenile idiopathic arthritis. Despite these associations, the immune pathogenesis underlying the link between arthritis and uveitis is unknown. However, studies done in the past 5 years have provided information about the immune pathogenesis of juvenile idiopathic arthritis, confirming, among other things, that the systemic form is a different disease from other types of juvenile idiopathic arthritis, with a distinct immune pathogenesis, and should be treated as would an acquired autoinflammatory disease.

**Gene expression profiling**

In addition to searching for susceptibility genes, gene expression profiling can identify any genes for which expression patterns change during the course of disease, providing a disease signature that can lead to the identification of unique biomarker patterns. Although this approach has disadvantages (such as the high costs, the complexity of the composition of the DNA microarrays, and the analysis of the results), its potential is underscored by various studies in patients with juvenile idiopathic arthritis—studies of patients with different forms of disease showed a consistent, though not fully elucidated, association with interleukin-10-regulated genes in all forms of the disease. Also, peripheral blood mononuclear cells from patients with recent-onset disease have a different gene expression profile dependent on their subtype.

In another study, researchers compared early samples of either peripheral blood mononuclear cells or synovial fluid mononuclear cells from patients who developed either persistent or widespread oligoarticular disease and identified patterns that were different between both groups. However, the most striking finding was an increased CD4 to CD8 ratio in patients with remitting disease, underscoring the important part played by T cells in disease course. Also, as discussed above, gene expression profiles in peripheral blood mononuclear cells of patients with recent-onset juvenile idiopathic arthritis identified expression patterns that differed with age of onset but not with the numbers of joints affected, suggesting different biological mechanisms in patients with either early-onset and late-onset disease.

Gene expression patterns in systemic juvenile idiopathic arthritis are different from those of other subtypes, and include upregulation of genes related to
innate immunity and complement systems, and of a group of mostly haemopoietic genes. The next challenge will be to relate these molecular profiles to specific immune pathways and the clinical manifestations to individual pathways. Thus gene expression profiling can provide a basis for a novel molecular method of classification of patients at disease onset, which could help to better predict treatment response and thus guide individual treatment regimens.4,44

Environmental triggers
In a healthy immune system, effector and regulator mechanisms are kept in balance, assuring a tailor-made response that adequately protects against an invading pathogen while preventing unwarranted damage to an individual and preserving immune tolerance. To achieve this balance, the innate and adaptive immune systems closely interact. Much the same as with most human autoimmune diseases, the cause of juvenile idiopathic arthritis is assumed to be multifactorial. A genetically susceptible individual might develop a deleterious and uncontrolled response towards a self-antigen on exposure to an unknown environmental trigger.12 This response causes a self-perpetuating loop of activation of both innate and adaptive immunity that causes tissue damage. In juvenile idiopathic arthritis, infections and vaccinations have been suggested as two candidate triggers, but neither has been confirmed as a trigger because of a scarcity of proper controlled, prospective studies.4,32,45 A prospective study did not show a relation between vaccination with meningococcus C6 and disease exacerbation in juvenile idiopathic arthritis, and a retrospective analysis did not show any association with measles, mumps, and rubella vaccination and disease exacerbation.47 Patients with polyarticular disease, however, showed an increased proinflammatory immune response towards vaccine antigens after vaccination with meningococcus C, suggesting a potential risk for aggravating inflammation.48 Again, much larger studies including both genetic susceptibility and up-to-date immunological analysis will be needed to define the role of environmental triggers in juvenile idiopathic arthritis.

Balancing tolerance and inflammation
The autoreactive immune response in juvenile idiopathic arthritis is assumed to be initially triggered by an adaptive (T cell or B cell) response towards a self-antigen. This assumption is underscored by the fact that joint inflammation in juvenile idiopathic arthritis is characterised by selective accumulation of activated memory T cells in the synovium, which are clustered around antigen-presenting (dendritic) cells.24,53–59 However, soon after the initial autoreactive insult, almost all the immune system is taking part in the immune response (figure 1).33–35 Therefore, the question of what initially triggers the disease is of minor importance. Even if in individual patients such a disease-causing antigen could be identified, at the time the patient has developed juvenile idiopathic arthritis, the inflammatory process would have already spread and both innate and adaptive immunity would have been activated.33–35

Moreover, human autoimmune diseases such as juvenile idiopathic arthritis are complex genetic diseases with a combination of genetic susceptibility and largely unknown environmental triggers leading to disease. Such complexity hinders the study of the underlying immunological mechanisms that contribute to such diseases. However, unlike other autoimmune diseases, juvenile idiopathic arthritis has distinct subtypes, including subtypes with a remitting form of arthritis, which allows the comparison not only between subtypes, but also of immune responses at different stages of disease (in active disease and during remission).

The idea that juvenile idiopathic arthritis can have a remitting course has led to speculation about whether so-called regulatory T cells play a part in the disease. Regulatory T cells (Tregs), first described by Sakaguchi, are part of a unique population of T cells that can specifically suppress other immune cells. They express the transcription factor FoxP3 and are crucial for the regulation of inflammation and the maintenance of immune tolerance in both animals and people. Although most information about the role of Tregs in autoimmunity comes from animal studies, they are probably important for the regulation of inflammation in human autoimmune diseases. The first description of the role of Tregs in people comes from studies in juvenile idiopathic arthritis, showing increased numbers of Tregs in remitting disease and after autologous stem cell transplantation, and co-expression of CD27 or CD39 in the synovial fluid of patients with juvenile idiopathic arthritis. Both natural Tregs (those directly derived from the thymus) and antigen-induced Tregs are present in increased numbers in the synovial fluid and peripheral blood of patients with remitting forms of juvenile idiopathic arthritis, with heat shock proteins (heat shock proteins (hsp)60 and dnaJ) being the most well-defined antigens.

Because Tregs are present in increased numbers at the site of inflammation not only in juvenile idiopathic arthritis, but also in other autoimmune disease such as rheumatoid arthritis, questions have been raised as to whether they are insufficient in quality, quantity, or both, to affect disease. Alternatively, the effector response could be simply too strong to be counter-regulated by regulatory T cells. T-helper (Th)17 cells have also been recorded in the joints of patients with juvenile idiopathic arthritis. These cells are characterised by the expression of transcription factor RORc and have a reciprocal relation with FOXP3-positive Tregs that could prove crucial for the regulation of joint inflammation. Growing evidence suggests that the different subtypes of T cells in the synovial fluid (not only Tregs and Th17, but also, for example, Th1, Th2, and T regulatory-1 cells) should not
be thought of as separate entities, but as a community of T-cell subsets that are in a constantly changing, dynamic balance. If this suggestion holds true, future treatment options should not focus on only cytokines, Tregs, or Th17 cells, but perhaps on all potential targets.

Causes and pathogenesis of systemic disease
Prominent systemic clinical features, the absence of auto-antibodies, and a relation with HLA alleles have led to the realisation that systemic juvenile idiopathic arthritis is a specific disease entity, with more similarities with autoinflammatory diseases than with classic autoimmune diseases. This realisation is emphasised by the fact that patients with the systemic form of disease have a much less favourable response to anti-TNF treatment than have patients with other forms of the disease, despite also having raised concentrations of TNF α in synovial fluid.

Systemic juvenile idiopathic arthritis is related to an overproduction of the proinflammatory cytokine interleukin-6, which is the basis of successful treatment with tocilizumab. Pascual and co-workers made major steps forward thanks to well-designed studies of the biological and transcriptional effects of a patient’s serum on healthy blood cells. Altogether, these studies showed that patients with systemic disease have a unique interleukin-1 signature. Whether early treatment with anti-interleukin-1 drugs can prevent the involvement of adaptive immunity and the onset of severe arthritis is unknown.

The cytokine expressed most abnormally in plasma or synovial fluid is neither interleukin-6 or interleukin-1, but interleukin-18. Interleukin-18 is in the same family as interleukin-1, and is essential for natural killer cell activation, which is compromised in patients with systemic juvenile idiopathic arthritis because of an
interleukin-18-receptor signalling defect. Another distinctive feature of the systemic form of disease is the strong association with macrophage activation syndrome, a major and life-threatening complication that occurs in 5–10% of patients with juvenile idiopathic arthritis. Macrophage activation syndrome has strong resemblance with either genetic-haemophagocytic or acquired-haemophagocytic lymphohistiocytosis. Genetic (familial) disease is associated with mutations in the perforin gene and MUNC genes. Despite the absence of gene mutations in either gene in systemic juvenile idiopathic arthritis, there are clear indications of an association with acquired functional or genetic defects in the perforin and MUNC genes in the disease. 

Another major step forward in the understanding and management of systemic juvenile idiopathic arthritis was the identification of phagocyte-specific S100 proteins as biomarkers for disease and treatment response. These proteins belong to a novel group of so-called damage-associated molecular pattern molecules (DAMP), and are also known as myeloid-related proteins (MRPs). They have clear proinflammatory effects on other immune cells and can act as endogenous activators of toll-like receptor 4. Several well-designed studies have shown that two of these proteins, MRP-8 (S100A8) and MRP-14 (S100A9), are excellent biomarkers in several autoimmune and inflammatory diseases, especially systemic juvenile idiopathic arthritis. Indeed, the concentration of MRP-8 and MRP-14 is greatly increased in the serum of patients with systemic disease, and can distinguish such patients in an early phase not only from healthy individuals, but also from patients with infections or malignant disease. Increased concentrations of MRP-8 or MRP-14 are even helpful as biomarkers to predict disease relapse after stopping methotrexate treatment. Because they are closely related to disease activity, these molecules can also be attractive targets for immunological treatment of arthritis. Thus, this group of calcium-binding proteins can have both benefits and disadvantages in juvenile idiopathic arthritis. They can be biomarkers for the disease and are potential targets for immune treatment. More generally, experience with S100 proteins as biomarkers and the growing insight in the mechanisms of disease should help in the design of phase 0 studies aimed at understanding the mechanisms of disease and the basis for response to treatment (figure 2).

Imaging
Imaging has not been used to its full potential in childhood arthritis. Findings from studies in adults are not applicable to children because of the unique features of the growing skeleton, such as age-related variations in the thickness of the articular cartilage, incomplete ossification, and bone growth anomalies induced by the disease. Indeed, for conventional radiology, scoring systems specific for juvenile idiopathic arthritis has been devised. 

As well as being useful in the assessment of subclinical synovitis and in the differentiation of tendon from articular involvement, ultrasonography can be used to assess cartilage thickness, although more data are needed for the normal thickness of cartilage in children of different ages. In patients with juvenile idiopathic arthritis, ultrasonography is as good as or better than conventional radiology for the detection of cortical erosions in sonographically accessible areas. Therefore, although more data for reproducibility and sensitivity to change are needed, important information will be acquired in the near future when serial ultrasonography is used to supplement clinical assessment in the monitoring of disease activity and in the assessment of response to treatment. Despite the great diagnostic potential of MRI, few studies of juvenile idiopathic arthritis, all of which have differing methodologies, have been done.

---

Figure 2: Established and emerging immunological treatments
Intervention are targeted at T cells in a specific (1) or non-specific (2–4) way, B cells (5), cytokines (6), and other soluble mediators (7–8). TNF = tumour necrosis factor. HLA = human leucocyte antigen. TNF = tumour necrosis factor. MMP = matrix metalloproteinase.
Figure 3: Imaging of recent-onset juvenile idiopathic arthritis

Delayed gadolinium-enhanced MRI cartilage imaging in a 9-year-old patient, showing early cartilage degradation in their radiolunate, lunocapitate, and capitate third metacarpal base joints. T1 mapping shows the concentration of gadolinium (diethylene triamine pentaacetic acid), which is anionic and distributes in articular cartilage inversely to the concentration of negatively charged glycosaminoglycans. Areas with glycosaminoglycan loss have a higher concentration of gadolinium and a shorter T1 relaxation time (seen in red).

The feasibility of a quantitative measurement of synovial volume and of dynamic contrast-enhanced MRI to provide a quantitative assessment of inflammation on the basis of analysis of the time course of signal changes after gadopentetic acid injection has been reported. These approaches could provide useful information to assess responsiveness to treatment and predict joint destruction, as already reported in adult rheumatoid arthritis. The prognostic value of bone marrow oedema has not been established for juvenile idiopathic arthritis, but studies done in adults with rheumatoid arthritis strongly suggest that its assessment would also be very useful in children.

MRI can also detect cartilage damage; dedicated studies are needed to define the variation of cartilage volume with age in healthy individuals to better understand the importance of changes in cartilage volume over time in patients with juvenile idiopathic arthritis. In the near future, molecular imaging could also provide information on cartilage injury before morphological damage is apparent. For instance, delayed gadolinium-enhanced MRI cartilage imaging (figure 3) is a sensitive technique to assess cartilage proteoglycan content with the negative charge of the paramagnetic MRI contrast agent, which distributes into the cartilage, inversely to the fixed charge density of negatively charged glycosaminoglycans. Finally, consistent with the results of studies in adults, the findings of a pilot multi-imaging study have shown that MRI is the most sensitive imaging modality for the detection of bone erosions in wrists of patients with juvenile idiopathic arthritis, detecting more than twice as many erosions as did conventional radiography or ultrasonography.

In the near future, ultrasonography and MRI are therefore likely to provide much essential new information useful for disease management, including the early identification of patients at high risk of erosive disease. Indeed, at present, even within a given category of juvenile idiopathic arthritis, difficulties exist in the identification of patients at risk for a poor outcome early enough to allow aggressive treatment at disease onset.

Treatment

The management of juvenile idiopathic arthritis is based on a combination of pharmacological interventions, physical and occupational therapy, and psychosocial support. Until a decade ago, very few randomised controlled trials (RCT) were done in children with this disease. This situation changed completely with the implementation of the so-called paediatric rule by the Food and Drug Administration (FDA) and the European Medicines Agency (EMA). According to this rule, a company that wishes to register a new treatment for a given disease in adults has to also test their product in children if there is a paediatric counterpart of the illness. The paediatric rule opens the door for more targeted studies, which are crucial for the safety of paediatric patients such as those with juvenile idiopathic arthritis. Paediatric rheumatologists were able to quickly take advantage of this rule thanks to the existence of two very large networks that have worked in a highly integrated and synergistic way, the Pediatric Rheumatology Collaborative Study Group (PRCSG) in North America and the Paediatric Rheumatology International Trial Organisation (PRINTO) in Europe and the rest of the world. Moreover, these two networks had already defined validated criteria to assess response to treatment in juvenile idiopathic arthritis.

In the past decade, great advances have been made in the treatment of juvenile idiopathic arthritis, which will probably greatly improve the long-term prognosis of the disease and alleviate some of the heavy burden it imposes on children, their parents, and society. The widespread use of intra-articular triamcinolone hexacetonide joint injections has played an important part in the prevention of deformities secondary to joint contractures, and the dose of methotrexate with the greatest effectiveness has been established in a randomised trial. Although results from a trial have shown that leflunomide is an effective treatment for polyarticular juvenile idiopathic arthritis, clinical experience with this drug in children is scarce. The efficacy of several anti-TNF drugs and of abatacept has been shown in randomised controlled trials in the polyarticular form of disease—a controlled trial with tocilizumab (NCT00642460) is taking place and is due to end in October, 2014. As mentioned above, anti-interleukin-6 and anti-interleukin-1 treatments have shown much efficacy in the treatment of systemic juvenile idiopathic arthritis—a disease that has little sensitivity to anti-TNF drugs. Once the results of the phase 3 trials that are in progress are reported, these
two therapeutic approaches, if their safety is confirmed with long-term follow-up, will probably substantially change the treatment of systemic juvenile idiopathic arthritis and its long-term prognosis.

In uveitis, early diagnosis is essential for the success of treatment. Topical corticosteroids and mydriatics are effective in most cases. In patients with resistant disease, periocular and systemic corticosteroids are indicated. In moderate-to-severe uveitis, several drugs have been claimed to be effective, including methotrexate, ciclosporin, mycophenolate mofetil, and alkylating agents. However, no controlled trials have been done. The effectiveness of etanercept is controversial, whereas infliximab, adalimumab, and abatacept have been anecdotally reported to be of benefit.

The effectiveness of biological agents has also changed physicians’ expectations; much improved treatment response is now needed to show efficacy in clinical trials. Furthermore, definitions of minimum disease activity and remission have been established. Treatment strategies for patients in clinical remission, including the withdrawal of treatment, will probably be guided in the future not only by clinical data, but also by imaging assessment of residual synovitis and normalisation of those biomarkers (such as gene expression patterns or serum proteins), the presence of which might prove to be associated with a risk of relapse.

Biological agents seem to have a good safety profile. The most consistent experience is with anti-TNF agents and is congruent with findings in adults with the only exception of a possible small increased risk of lymphoma and other types of cancer in children. However, a clear causal relation cannot be established because of the concomitant use of immunosuppressive drugs and the absence of information on the potential risk of malignant disease when used with or without methotrexate.

In a population-based cohort study done in Sweden, children diagnosed with juvenile idiopathic arthritis between 1969 and 2007 did not have an increased risk of cancer. However, when the analysis was restricted to children who were diagnosed from 1987 onwards, those with juvenile idiopathic arthritis were more than twice as likely to have a malignant disease and more than four times as likely to have a lymphoproliferative malignant disease than were children without juvenile idiopathic arthritis. This increased risk could not be explained by the introduction of biological treatment because the association was similar in analyses ending in 1999. However, use of methotrexate for the treatment of juvenile idiopathic arthritis was first reported in 1986.

Safety is assessed through passive surveillance of adverse events, which substantially underestimates the incidence rate of adverse events, or through product-specific observational registries done by pharmaceutical companies, which result in competing and duplicate efforts to gather similar information from small populations of patients. Future collaboration between regulatory agencies, industry, and research networks could provide an opportunity to set up larger and more effective international pharmacovigilance registries. Such registries will be implemented by the Paediatric Rheumatology European Society (PRES) and PRINTO in Europe and by the PRCSG and the Childhood Arthritis Rheumatology Research Alliance (CARRA) in North America.

In most randomised controlled trials of juvenile idiopathic arthritis, the use of the double-blind, randomised withdrawal study design proposed by Giannini and Lovell has been very useful. According to this design, all children are treated with the active drug and the responders are then randomly allocated to receive either the active drug or a placebo (double-blinded phase); patients who have a pre-defined definition of disease worsening (eg, flare) are withdrawn from the double-blind withdrawal phase and usually unmasked and given the experimental treatment. An added benefit of this design is that it keeps the time that children are exposed to placebo to a minimum.

Researchers designing future randomised trials of treatments for juvenile idiopathic arthritis will face challenges because of both the heterogeneity of the disorder and the availability of effective drugs. Parallel trials with an active comparator (the available effective treatment) could be the design of choice, but would need a very large population of patients to show a significant difference. If only patients for whom a previous biological treatment had failed are included, the number of available individuals will be much reduced. This problem, as with any rare disease or rare subcategories of common disorders, will need to be solved with the improvement of methodologies for randomised controlled trials in small populations and the identification of sensitive imaging techniques that can quickly assess the disease-modifying potential of new drugs, and will take advantage of the existence of PRCSG and PRINTO.

However, once the efficacy and safety of a class of drugs (eg, anti-TNF agents) have been established for the treatment of juvenile idiopathic arthritis, then approval for use in children of a new drug in that class could need proof of safety and efficacy in adults only, with pharmacokinetic and pharmacodynamic dose-finding studies in a small sample of patients with juvenile idiopathic arthritis, followed by adequate phase 4 registries to assess safety and effectiveness.

Future perspectives

Major advances during the past 5 years have led to a novel outlook for the care of patients with juvenile idiopathic arthritis. Thanks to pivotal genetic and molecular-immunology studies, immune pathogenesis can be much better linked with clinical phenotypes of disease, which should lead to a revision of the original disease classification criteria. Moreover, these studies have led to novel treatment strategies for systemic juvenile idiopathic...
arthritis and the identification of immune biomarkers. The next challenge will be to combine genetic and immunological mechanistic studies with new imaging modalities to better define subgroups of patients, individual risk profiles, and response to treatment. This should lead to the next generation of clinical trials that aim to restore the immune balance in patients with juvenile idiopathic arthritis while restricting potential long-term side-effects of treatment.

**Contributors**

BP, SA, and AM designed the paper. BP and AM wrote the paper. SA provided critical input and composed figure 1 and figure 2.

**Conflicts of interest**

We declare that we have no conflicts of interest.

**References**

16. Marini A. Are the number of joints involved or the presence of psoriasis still useful tools to identify homogeneous disease entities in juvenile idiopathic arthritis? *J Rheumatol* 2003; 30:1900–03.


In October, 2010, a 76-year-old woman born in the USA presented to our emergency department complaining of a severe sore throat and painful swallowing. She reported having had a non-productive cough without haemoptysis 1 month previously. The cough resolved, but a sore throat then developed. At the time she was assessed by her general practitioner and specialists in the emergency and otolaryngology departments, who diagnosed pharyngitis and recommended conservative management. Her odynophagia progressed until she was unable to eat solids, prompting her to present to our hospital. She denied fevers, chills, night sweats, myalgias, dysphonia, dysphonia, or dysphagia but did report a 7 kg weight loss. She also reported occasional tobacco and alcohol use. On examination she was afebrile and thin but appeared well. She had dry mucous membranes and mild erythema of the posterior pharynx without tonsilar enlargement or exudate. No oral thrush, cervical lymphadenopathy, rales, splenomegaly, or rash were noted. Leucocyte concentration was $7.9 \times 10^9/L$ with a predominance of neutrophils. She was admitted for further assessment.

Rapid streptococcal antigen test, monospot test, human immunodeficiency virus (HIV) antibody test, and rapid antigen test for influenza A and B were negative. A temperature of 38·7°C prompted a chest radiograph, which showed bilateral airspace disease and an area of dense consolidation in the right upper lung with a rounded lucency likely to be cavitation (figure A), and rapid antigen test for influenza A and B were negative. A temperature of 38·7°C prompted a chest radiograph, which showed bilateral airspace disease and an area of dense consolidation in the right upper lung with a rounded lucency likely to be cavitation (figure A), subsequently confirmed on chest CT (figure B). CT of her neck showed non-specific epiglottis oedema. Purified protein derivative skin test was negative. A few acid fast bacilli were present in the sputum, and she was started on empiric therapy for presumed pulmonary tuberculosis. DNA probe performed on sputum confirmed *Mycobacterium tuberculosis* which was sensitive to all antituberculous drugs tested. Subsequently, her fever and odynophagia resolved. Laryngeal tuberculosis was the most likely cause of our patient’s sore throat as evidenced by her clinical symptoms and presumed active pulmonary tuberculosis, the non-specific oedema of the epiglottis noted on CT, and the resolution of her sore throat and odynophagia with antituberculous therapy. However, a definitive diagnosis cannot be made without laryngoscopy and histopathological confirmation; laryngoscopy was deferred because of the substantial infectious risk associated with the procedure in the setting of active tuberculosis. She was discharged with instructions for tuberculosis testing of close contacts and plans for direct observed therapy. At last contact in February, 2011, she was well, with no recurrence of symptoms.

Laryngeal tuberculosis, which accounted for 25% of all cases of tuberculosis in the early 20th century, now accounts for less than 1% of cases and is typically associated with pulmonary infection. Whereas extrapulmonary tuberculosis occurs frequently in persons with HIV, cases of laryngeal tuberculosis have been reported in immunocompetent patients. Our patient presented with odynophagia, a common presenting symptom in patients with laryngeal tuberculosis, but she denied hoarseness, the most common presenting symptom. She denied any history of tuberculosis or exposure to known contact with persons with tuberculosis, but she did report having worked as a home health aide during the 1970s. Laryngeal tuberculosis results from either direct bronchogenic spread from infected sputum or, less commonly, haematogenous spread. Laryngeal tuberculosis is highly infectious, and a delay in diagnosis can lead to disease progression and longer exposure to contacts. Therefore, physicians should consider laryngeal tuberculosis in the differential diagnosis of patients presenting with a persistent sore throat and odynophagia, especially in patients such as ours who also report weight loss. A chest radiograph should be considered before laryngoscopy to minimise unnecessary risk to health-care providers.

**Contributors**

All authors cared for the patient and wrote the report. Written consent to publish was obtained.

**References**


