

## Impact of a new national screening policy for Down's syndrome in Denmark: population based cohort study

Charlotte K Ekelund,<sup>1</sup> Finn Stener Jørgensen,<sup>2</sup> Olav Bjørn Petersen,<sup>3</sup> Karin Sundberg,<sup>1</sup> Ann Tabor,<sup>1,4</sup>  
Danish Fetal Medicine Research Group

**EDITORIAL** by Alfirevic  
**RESEARCH** p 453

<sup>1</sup>Department of Fetal Medicine, Rigshospitalet, Copenhagen University Hospital, 2100 Copenhagen, Denmark

<sup>2</sup>Fetal Medicine Unit, Department of Obstetrics and Gynaecology, Hvidovre University Hospital, Denmark

<sup>3</sup>Department of Obstetrics and Gynaecology, Skejby University Hospital, Aarhus, Denmark

<sup>4</sup>Faculty of Health Sciences, University of Copenhagen

Correspondence to: A Tabor [ann.tabor@rh.regionh.dk](mailto:ann.tabor@rh.regionh.dk)

Cite this as: *BMJ* 2008;337:a2547  
doi:10.1136/bmj.a2547

### ABSTRACT

**Objectives** To evaluate the impact of a screening strategy in the first trimester, introduced in Denmark during 2004-6, on the number of infants born with Down's syndrome and the number of chorionic villus samplings and amniocenteses, and to determine detection and false positive rates in the screened population in 2005 and 2006.

**Design** Population based cohort study.

**Setting** 19 Danish departments of gynaecology and obstetrics and a central cytogenetic registry 2000-7.

**Participants** 65 000 pregnancies per year.

**Main outcome measures** Primary outcomes measured were number of fetuses and newborn infants with Down's syndrome diagnosed prenatally and postnatally and number of chorionic villus samplings and amniocenteses carried out. Secondary outcomes measured were number of women screened in 2005 and 2006, screen positive rate, and information on screening in 2005 and 2006 for infants with a postnatal diagnosis of Down's syndrome.

**Results** The number of infants born with Down's syndrome decreased from 55-65 per year during 2000-4 to 31 in 2005 and 32 in 2006. The total number of chorionic villus samplings and amniocenteses carried out decreased from 7524 in 2000 to 3510 in 2006. The detection rate in the screened population in 2005 was 86% (95% confidence interval 79% to 92%) and in 2006 was 93% (87% to 97%). The corresponding false positive rates were 3.9% (3.7% to 4.1%) and 3.3% (3.1% to 3.4%).

**Conclusion** The introduction of a combined risk assessment during the first trimester at a national level in Denmark halved the number of infants born with Down's syndrome. The strategy also resulted in a sharp decline in the number of chorionic villus samplings and amniocenteses carried out, even before full implementation of the policy.

### INTRODUCTION

In 2004 the Danish National Board of Health issued guidelines that pregnant women should be offered information about screening methods and, if desired, a combined risk assessment for Down's syndrome in the first trimester based on a combination of maternal age, nuchal translucency scanning, and a biochemical test (the double test).<sup>1</sup> On the basis of this assessment

women were to be told their risk of carrying a fetus with Down's syndrome. Women with a risk above a defined cut-off were to be offered chorionic villus sampling or amniocentesis. The new policy was expected to detect 90% of fetuses with Down's syndrome at a 5% false positive rate.

We evaluated the impact of the screening strategy on the number of both infants born with Down's syndrome and chorionic villus samplings and amniocenteses. We also assessed whether the detection and false positive rates in the screened population for 2005 and 2006 were as expected.

### METHODS

From Statistics Denmark we retrieved data on the number of liveborn infants per year (about 65 000), the distribution of maternal age at delivery, and the mean maternal age at delivery for 2000-6. Using the maternal age specific risk of delivering an infant with Down's syndrome we calculated the expected number of affected liveborn infants.<sup>2</sup>

From the Danish central cytogenetic registry we obtained information on the number of chorionic villus samplings and amniocenteses during 2000-6, the indications for these, and chromosome analyses prenatally and postnatally on fetuses and newborn infants. We also obtained information on the number of infants with Down's syndrome born during 2000-4 and the identification numbers of infants with Down's syndrome born during 2005-7 and their mothers.

All 19 obstetrics and gynaecology departments in Denmark use the same fetal medicine software program (Astraia, Germany) for calculating risk based on formulas derived by the Fetal Medicine Foundation in London.

### Evaluation of screening performance in 2005 and 2006

From the obstetrics and gynaecology departments we collected information on the number of women who had a risk assessment in the first trimester in 2005 and 2006, either as the combined test (age, nuchal translucency, and biochemistry) or the combination of age and nuchal translucency or biochemistry. The departments reported the number of women given a

risk assessment of 1:300 or more at the time of screening.

In the calculation of screening performance we included fetuses and newborn infants with Down's syndrome when first trimester screening had been done in 2005 or 2006. Information about gestational age at delivery for all infants with Down's syndrome born during 2005-7 was obtained from the Danish National Board of Health.

We cross checked the identification numbers of women delivering an infant with Down's syndrome during 2005-7 with Astraia database servers in Denmark to determine whether screening had been done in the first trimester. Information on screening was also requested when Down's syndrome was diagnosed prenatally by an invasive procedure carried out for indications other than an increased risk of Down's syndrome.

## RESULTS

In January 2005 nine of the 15 Danish counties offered screening, and by June 2006 the whole country was covered. The yearly number of deliveries decreased slightly during 2000-6 (see *bmj.com*), whereas mean maternal age at delivery increased from 29.7 in 2000 to 30.3 in 2006. Based on the distribution of age and if no screening or invasive procedure had been done, the estimated expected number of infants with Down's syndrome increased from 121 in 2000 to 132 in 2005 and 135 in 2006.

The number of newborn infants with Down's syndrome decreased from 55-65 per year in 2000-4 to 31 in 2005 and 32 in 2006. The total number of fetuses and

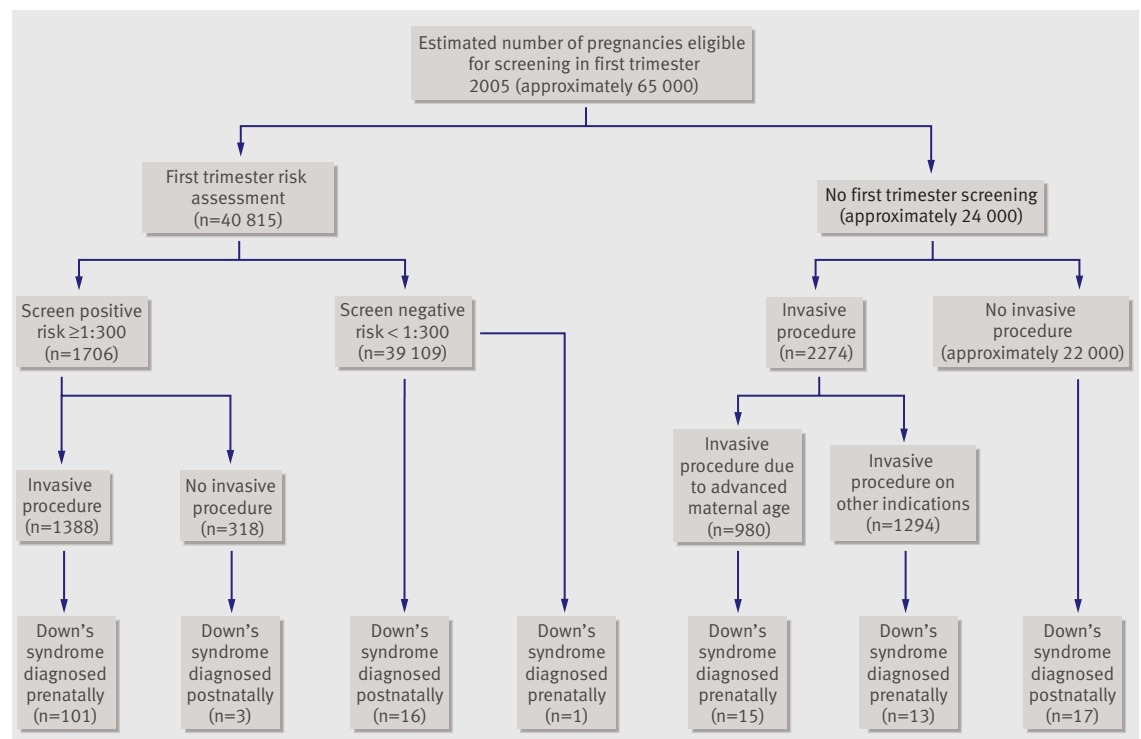
newborn infants with Down's syndrome diagnosed prenatally or postnatally in 2000-3 was stable at around 135-140 per year, with an increase to 157 in 2004, 161 in 2005, and 149 in 2006 (see *bmj.com*). The proportion of cases diagnosed prenatally increased from 53-61% during 2000-4, to 81% in 2005 and 79% in 2006.

The number of prenatal invasive procedures decreased from 7524 in 2000 to 3510 in 2006 (see *bmj.com*). The number of chorionic villus samplings decreased from 3322 in 2000 to 2302 in 2006, while the number of amniocenteses carried out decreased from 4202 to 1208 in the same years. This corresponds to an increase in the proportion of chorionic villus samplings from 44% to 66%.

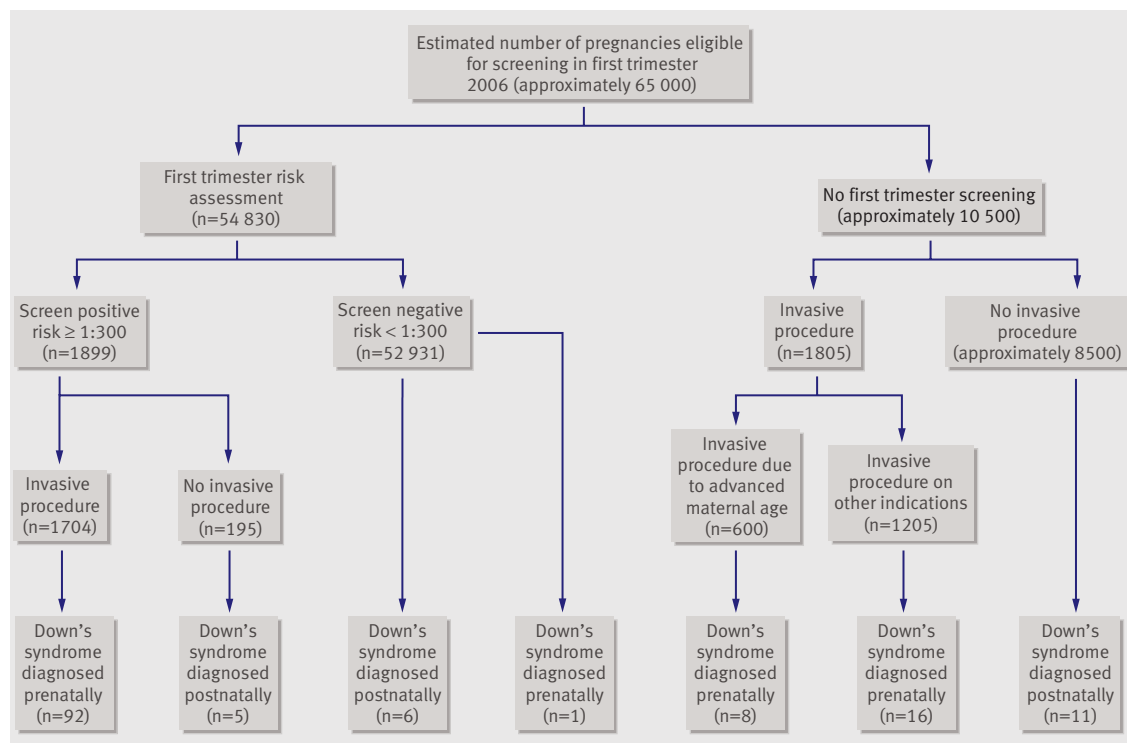
About 65 000 women were pregnant in Denmark during 2005-6. In 2005 40 815 women (62.8%) had a risk assessment in the first trimester, increasing to 54 830 (84.4%) in 2006. Figures 1 and 2 show the distribution of women eligible for screening and the groups in which infants with Down's syndrome were diagnosed prenatally and postnatally.

In 2005, 1706 women (4.2%) had a risk of 1:300 or more (screen positive rate) and among these, 1388 women (81.4%) had a diagnostic test (fig 1). Seventy two per cent of the procedures were chorionic villus samplings, the rest amniocenteses. In 2006, 1899 women (3.5%) had a risk of 1:300 or more and 1704 (89.7%) had a diagnostic test (fig 2). Seventy six per cent of the diagnostic procedures were chorionic villus samplings.

In the population screened in 2005 the detection rate of Down's syndrome was 86% (95% confidence interval 79% to 92%); 104 of 121 women were screened



**Fig 1** Number of fetuses and newborn infants with Down's syndrome diagnosed prenatally or postnatally according to screening results in Denmark, 2005. Invasive procedures are chorionic villus samplings or amniocenteses



**Fig 2** | Number of fetuses and newborn infants with Down's syndrome diagnosed prenatally or postnatally according to screening results in Denmark, 2006. Invasive procedures are chorionic villus samplings or amniocenteses

true positive (fig 1). Thus 17 women received a false negative result. An adjusted detection rate taking into account fetal loss from screening to birth (estimated as 25%<sup>3</sup>) was 82% (95% confidence interval 73% to 90%). The false positive rate was 3.9% (3.7% to 4.1%).

In 2006 the detection rate was 93% (87% to 97%) as only seven women received a false negative result. The adjusted detection rate taking fetal loss into account was estimated at 92% (83% to 97%). The false positive rate was 3.3% (3.1% to 3.4%).

The odds of being affected (carrying a fetus with Down's syndrome) after a screen positive risk assessment in the first trimester were 1:16 in 2005 and 1:20 in 2006. The odds of being affected after a screen negative result were 1:2301 in 2005 and 1:7562 in 2006. The odds of being affected after an invasive diagnostic procedure because of maternal age were similar in 2005 and 2006 (1:65 and 1:75); 15 fetuses with Down's syndrome were diagnosed among 980 women in 2005 and eight fetuses among 600 women in 2006. Thirty infants with Down's syndrome were born to mothers who had a risk assessment in the first trimester during 2005 and 2006 (see [bmj.com](#)).

## DISCUSSION

Even before full implementation of the policy for combined risk assessment during the first trimester in Denmark, the number of infants born with Down's syndrome decreased by about 50% and the number of cases diagnosed prenatally increased by around 30%.

The number of fetuses and newborn infants with Down's syndrome diagnosed prenatally or postnatally

increased during 2000-5, with a slight decline in 2006 (see [bmj.com](#)). This was partly due to increasing maternal age, but was as expected because more fetuses with Down's syndrome are lost spontaneously than those that are chromosomally normal. This increased rate has been estimated at around 25% from week 14 to term.<sup>3</sup> Based on the known distribution of maternal age at delivery in 2005 and 2006, 132 and 135 infants with Down's syndrome would have been expected in our population of 65 000 liveborn infants if the mothers had no prenatal intervention. Down's syndrome was diagnosed in 31 infants postnatally and 130 prenatally in 2005 and in 32 infants postnatally and 117 prenatally in 2006. Given a rate for fetal loss of 25%, this corresponds to 129 infants with Down's syndrome diagnosed postnatally in 2005 and 120 diagnosed postnatally in 2006. In 2005 the expected numbers were close to those reported, whereas in 2006 the reported number was lower than expected. This may be due to chance fluctuation, as we believe follow-up is complete.

In 2005 one third of women were not offered screening or declined and gave birth to a total of 17 infants with Down's syndrome. In 2006 the proportion of non-screened women decreased to 15%. About 8500 women who were not offered screening or declined screening or a diagnostic test in 2006 gave birth to a total of 11 infants with Down's syndrome. The national guidelines on screening emphasise that risk assessment for Down's syndrome should be done only if women choose the test on the basis of an informed choice. Therefore despite the programme now being

**WHAT IS ALREADY KNOWN ON THIS TOPIC**

Many countries are currently trying to achieve national screening strategies for Down's syndrome

None has described how a combined screening strategy in the first trimester affects numbers of infants born with Down's syndrome or rate of invasive procedures

Detection rates and false positive rates for the combined first trimester risk assessment have been reported only from specialised centres or from regional experience

**WHAT THIS STUDY ADDS**

After implementation of a national screening policy in Denmark, the number of infants born with Down's syndrome and the rate of invasive procedures was noticeably reduced

The screening strategy achieved high detection rates and low false positive rates

accessible to all pregnant women in Denmark, it is expected that a proportion will still choose not to be screened.

We found that the number of invasive procedures carried out yearly decreased by more than 50% during 2000-6. A decrease in the number could be seen even before the policy was changed, probably because pregnant women became aware of alternative investigations such as nuchal translucency scanning (see [bmj.com](http://bmj.com)).

In 2005 and 2006 about 3% of women still had an invasive procedure because of indications other than a screen positive result, with a tendency towards a reduced number of tests from 2005 to 2006 (2274 women in 2005, 1805 in 2006). The decrease was mainly due to fewer women choosing invasive tests on the basis of age (980 in 2005, 600 in 2006). The relatively high number of women choosing an invasive procedure was probably partly due to lack of implementation of the new screening programme. It is also possible that women who had a diagnostic procedure for a previous pregnancy because of age may have requested one again. Overall, 10-20% of women with a screen positive result did not have an invasive test. This was probably because of the risk of miscarriage.

The difference in odds of carrying a fetus with Down's syndrome for those who were tested because of a screen positive result (1:16 in 2005, 1:20 in 2006) compared with that of being tested because of age (1:65 in 2005, 1:75 in 2006) illustrates the rationale in screening using a combined risk assessment in the first trimester. As expected, this strategy reduced the number of unnecessary diagnostic procedures. The procedure related risk of miscarriage after chorionic villus sampling or amniocentesis is reported to be 1%.<sup>4</sup> In the group of women having an invasive diagnostic test done because of advanced maternal age in 2005 and 2006 16 chromosomally normal fetuses would then have been miscarried, to diagnose 23 cases of Down's syndrome. This should be compared with the 31 fetuses possibly miscarried to diagnose 193 cases of

Down's syndrome in the group of women with a screen positive result. Thus the false positive rate of prenatal diagnostic testing has been reduced by changing the selection criterion from age to risk assessment in the first trimester. The false negative rate has also changed: previously those women who chose to have an invasive procedure because of age had a diagnostic test. Currently women choose to have a screening test; 0.4 per 1000 women in 2005 and 0.1 per 1000 in 2006 delivered a child with Down's syndrome, despite having a risk assessment below the 1:300 cut-off. This emphasises the importance of informing all women about the limitations of screening.

For false positive rates of 3.9% and 3.3% in the screened populations we found detection rates for Down's syndrome of 86% in 2005 and 93% in 2006. This is in accordance with the screening performance expected by the Danish National Board of Health and may be considered high. Furthermore, we report the result of routine clinical practice, where not all risk assessments are based on the optimal variables (combination of maternal age, nuchal translucency, and biochemistry) as some are given only on age and nuchal translucency or biochemistry. In a large prospective multicentre study the detection rate using a combined screening programme in the first trimester was 92.6% for a false positive rate of 5.2%.<sup>5</sup> Our data show that it is possible to introduce this screening strategy in as many as 19 different centres and still obtain national detection and false positive rates similar to those from specialised centres.

Members of the Danish Fetal Medicine Research Group are: Hans Jakob Andersen, Jeanette Christensen, Vibeke Ersbak, Richard Farlie, Carsten Henriques, Annette Wind Olesen, Anni Holmskov, Lisa Neerup Jensen, Anette Kristiansen, Torben Larsen, Hedvig Poulsen, Jan Ramb, Lillian Skibsted, Peter Skovbo, Steffen Sommer, Lene Sperling, Susanne Vemmelund Juul, Helle Zingenberg. We thank Jan Hansen for retrieval of data from the Danish Central Cytogenetic Registry, Kirsten Rasmussen for retrieval of data from the Genetic Database of Funen, Steen Rasmussen (Danish National Board of Health) for providing gestational age at delivery in cases where no screening was done, and the departments of clinical biochemistry for analysing the biochemical samples used in the screening programme.

**Contributors:** See [bmj.com](http://bmj.com).

**Funding:** None.

**Competing interests:** None declared.

**Ethical approval:** This study was approved by the Danish Data Protection Agency and Danish National Board of Health.

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

- 1 Danish National Board of Health. Report from a working commission "Prenatal diagnosis and risk assessment." [www.sst.dk/upload/fosterdiagnostik1\\_001.pdf](http://www.sst.dk/upload/fosterdiagnostik1_001.pdf). 2003.
- 2 Morris JK, Mutton DE, Alberman E. Revised estimates of the maternal age specific live birth prevalence of Down's syndrome. *J Med Screen* 2002;9:2-6.
- 3 Snijders RJ, Sundberg K, Holzgreve W, Henry G, Nicolaides KH. Maternal age- and gestation-specific risk for trisomy 21. *Ultrasound Obstet Gynecol* 1999;13:167-70.
- 4 Alfirevic Z, Sundberg K, Brigham S. Amniocentesis and chorionic villus sampling for prenatal diagnosis. *Cochrane Database Syst Rev* 2003;(3):CD003252.
- 5 Nicolaides KH, Spencer K, Avgidou K, Faiola S, Falcon O. Multicenter study of first-trimester screening for trisomy 21 in 75 821 pregnancies: results and estimation of the potential impact of individual risk-orientated two-stage first-trimester screening. *Ultrasound Obstet Gynecol* 2005;25:221-6.

**Accepted:** 30 September 2008

# Comparison of different strategies in prenatal screening for Down's syndrome: cost effectiveness analysis of computer simulation

Jean Gekas,<sup>1</sup> Geneviève Gagné,<sup>2</sup> Emmanuel Bujold,<sup>3</sup> Daniel Douillard,<sup>2</sup> Jean-Claude Forest,<sup>4</sup> Daniel Reinharz,<sup>2</sup> François Rousseau<sup>4</sup>

**EDITORIAL** by Alfirevic  
**RESEARCH** p 449

<sup>1</sup>Centre de recherche du CHUQ, Service de Génétique Médicale, Unité de Diagnostic Prénatal, Faculté de Médecine, Université Laval, Québec city, Québec, Canada

<sup>2</sup>Centre de recherche du CHUQ, Département de médecine sociale et préventive, Faculté de Médecine, Université Laval, Cité universitaire, Québec city

<sup>3</sup>Centre de recherche du CHUQ, Unité de Diagnostic Prénatal, Département d'Obstétrique-Gynécologie, Faculté de Médecine, Université Laval, Québec city

<sup>4</sup>Centre de recherche du CHUQ, Département de biologie médicale, Faculté de Médecine, Université Laval, The CanGeneTest research consortium on genetic laboratory services, Québec city

Correspondence to: J Gekas, Centre Hospitalier de l'Université Laval (CHUL), 2705, boul. Laurier, bureau RC-9300, Sainte-Foy, Québec city, QC G1V 4G2, Québec, Canada  
jean.gekas@mail.chuq.qc.ca

Cite this as: *BMJ* 2009;338:b138  
doi:10.1136/bmj.b138

This article is an abridged version of a paper that was published on [bmj.com](http://bmj.com). Cite this article as: *BMJ* 2009;338:b138

## ABSTRACT

**Objectives** To assess and compare the cost effectiveness of three different strategies for prenatal screening for Down's syndrome (integrated test, sequential screening, and contingent screenings) and to determine the most useful cut-off values for risk.

**Design** Computer simulations to study integrated, sequential, and contingent screening strategies with various cut-offs leading to 19 potential screening algorithms.

**Data sources** The computer simulation was populated with data from the Serum Urine and Ultrasound Screening Study (SURUSS), real unit costs for healthcare interventions, and a population of 110 948 pregnancies from the province of Québec for the year 2001.

**Main outcome measures** Cost effectiveness ratios, incremental cost effectiveness ratios, and screening options' outcomes.

**Results** The contingent screening strategy dominated all other screening options: it had the best cost effectiveness ratio (\$C26 833 per case of Down's syndrome) with fewer procedure related euploid miscarriages and unnecessary terminations (respectively, 6 and 16 per 100 000 pregnancies). It also outperformed serum screening at the second trimester. In terms of the incremental cost effectiveness ratio, contingent screening was still dominant: compared with screening based on maternal age alone, the savings were \$C30 963 per additional birth with Down's syndrome averted. Contingent screening was the only screening strategy that offered early reassurance to the majority of women (77.81%) in first trimester and minimised costs by limiting retesting during the second trimester (21.05%). For the contingent and sequential screening strategies, the choice of cut-off value for risk in the first trimester test significantly affected the cost effectiveness ratios (respectively, from \$C26 833 to \$C37 260 and from \$C35 215 to \$C45 314 per case of Down's syndrome), the number of procedure related euploid miscarriages (from 6 to 46 and from 6 to 45 per 100 000 pregnancies), and the number of unnecessary terminations (from 16 to 26 and from 16 to 25 per 100 000 pregnancies).

**Conclusions** Contingent screening, with a first trimester cut-off value for high risk of 1 in 9, is the preferred option for prenatal screening of women for pregnancies affected by Down's syndrome.

## INTRODUCTION

Major advances have been made in prenatal screening for Down's syndrome, but there is still no consensus on

the most cost effective alternative to using maternal age alone as a screening strategy, and no national strategy exists in the United States or Canada.<sup>1,2</sup> Different approaches have been proposed<sup>2-4</sup> such as

- The integrated test; results from first trimester screening tests are not analysed until results from second trimester tests are evaluated, when both sets are assessed together
- Sequential screening; first trimester screening results determine whether second trimester testing is indicated: women with a positive first trimester result are offered invasive testing (chorionic villous sampling), whereas those with a negative result are offered another serum test
- Contingent screening; first trimester screening results are used to categorise women as high, intermediate, or low risk. High risk women are offered early diagnosis (chorionic villous sampling), low risk women are reassured and do not undergo second trimester testing, and intermediate risk women are offered second trimester testing. The cut-off values between the risk categories vary depending on how the groups are defined.

Cost effectiveness ratios and similar end points of screening strategies are needed for the development of optimal public health strategies,<sup>2</sup> but only contradictory and limited data exist on the three above methods of combining first and second trimester tests.<sup>2,4</sup> We therefore performed computer simulations to compare 19 different screening options based on the three screening approaches with various cut-off levels and using real data from the Serum Urine and Ultrasound Screening Study (SURUSS) trials.<sup>4,5</sup> Our aims were to

- Analyse the cost effectiveness ratio of the different screening options from a public health perspective
- Compare their performance estimates for an overall 90% detection rate by evaluating seven other relevant end points that cover the main outcomes in prenatal screening for Down's syndrome:
  - 1) False positive rate that defines the number of scheduled amniocentesis procedures
  - 2) Number of procedure related miscarriages of normal (euploid) fetuses
  - 3) Number of live births with Down's syndrome

- 4) Number of unnecessary terminations
- 5) Proportion of pregnancies affected by Down's syndrome that were screened by a first trimester test
- 6) Proportion of patients reassured early in gestation by first trimester testing
- 7) Proportion of continuing pregnancies that proceed to a second trimester test
- Compare the effects of eight different first trimester cut-off values in contingent and sequential screening strategies to determine the optimal value.

**MATERIAL AND METHODS**

**Screening strategies and tests**

The screening strategies considered were the integrated test, sequential screening, and contingent screening compared with maternal age alone ( $\geq 35$  years),<sup>6</sup> and the triple test, which has been widely available in Europe and Canada since 1991. See bmj.com for details.

**Diagnostic tests**

Our computer model accounted for the performance characteristics of the diagnostic tests: amniocentesis for prenatal diagnosis only after second trimester screenings results; and chorionic villous sampling for women with a positive result at the first trimester screening.

**Decision model**

We developed a computer program in C11 language that simulates all 19 screening scenarios. See bmj.com for a simplified version of the decision model, and also for tables showing the input variables and demographic characteristics of the virtual population used.

**Probabilities**

*Pregnancies affected by Down's syndrome and risk of miscarriage*—We estimated the risk of a pregnancy being affected by Down's syndrome by multiplying the specific odds by maternal age of having an affected live birth (corrected for the spontaneous miscarriage of fetuses with Down's syndrome) by the likelihood ratio (for a given set of marker values) obtained from the overlapping multivariate Gaussian distributions of marker levels in affected and unaffected pregnancies.<sup>4</sup>

*Test performance*—We used the distribution of marker values in pregnancies affected by Down's syndrome and in unaffected pregnancies reported in the SURUSS trials to determine the parametric values.<sup>4,5</sup> We included rates of false positive and false negative test results in the calculations.

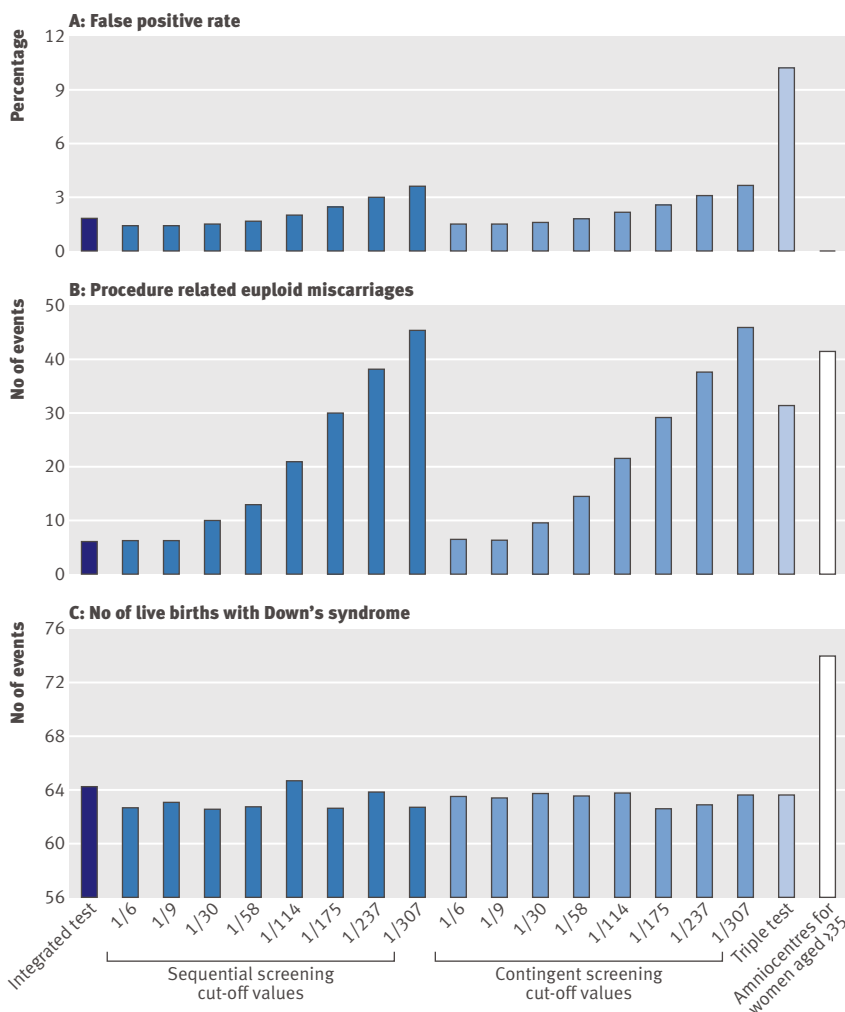
*Other probabilities*—Our simulation included the level of compliance previously estimated in the Québec population—that is, the proportion of women who consent to participate in prenatal screening for Down's syndrome and to a diagnostic test after a positive screening result. We also included the chance that a woman will choose elective abortion after a positive diagnostic test.

**Costs**

We considered only direct costs under the Ministry of Health and the Public Medical Assurance perspectives. Unit prices were averages for Quebec province calculated from government databases. See bmj.com.

**Cost effectiveness analysis**

We followed guidelines for cost effectiveness analysis in prenatal diagnosis.<sup>6</sup> The analyses were run to evaluate global costs, cost effectiveness ratios (costs per case of Down's syndrome diagnosed), and incremental cost effectiveness ratios comparing all 19 screening options. Incremental cost effectiveness ratio was calculated by dividing the difference in cost between each screening strategy and screening based on maternal age only (amniocentesis for women aged  $\geq 35$  years) by the difference in the number of cases of Down's syndrome detected by the two strategies. The incremental cost effectiveness ratio therefore



**Fig 1** | Effects of different prenatal screening strategies for Down's syndrome on (A) rate of false positive results, (B) number of procedure related euploid miscarriages, and (C) number of live births with Down's syndrome (all values are per 100 000 pregnancies)

represents the additional cost or savings per additional abnormality detected.

### Sensitivity analyses

In order to test the robustness of our findings, we performed a series of simulations in which we varied the major assumptions of the modelling with values proposed in the AETMIS report<sup>7</sup> and recalculated the cost effectiveness ratios.<sup>6</sup> See [bmj.com](#).

## RESULTS

### Cost effectiveness analysis

The most cost effective screening strategy seems to be contingent screening (cost effectiveness ratio \$C26 833 per case of Down's syndrome with a high risk cut-off value of 1 in 30). These results were robust in sensitivity analyses. Compared with the reference screening strategy (maternal age only), the

incremental cost effectiveness ratio of contingent screening (with cut-off value of 1 in 6) is  $-\$C30\,963$  per averted birth with Down's syndrome. The estimated costs per prevented birth with Down's syndrome were \$C35 215 for sequential screening, \$C38 944 for integrated screening, and \$C43 809 for the triple test. Consequently, the contingent screening strategy seems to be most cost effective. As expected, amniocentesis based on advanced maternal age only is the least cost effective option, with a cost effectiveness ratio of \$C74 037 per case of Down's syndrome.

### Screening strategies outcomes

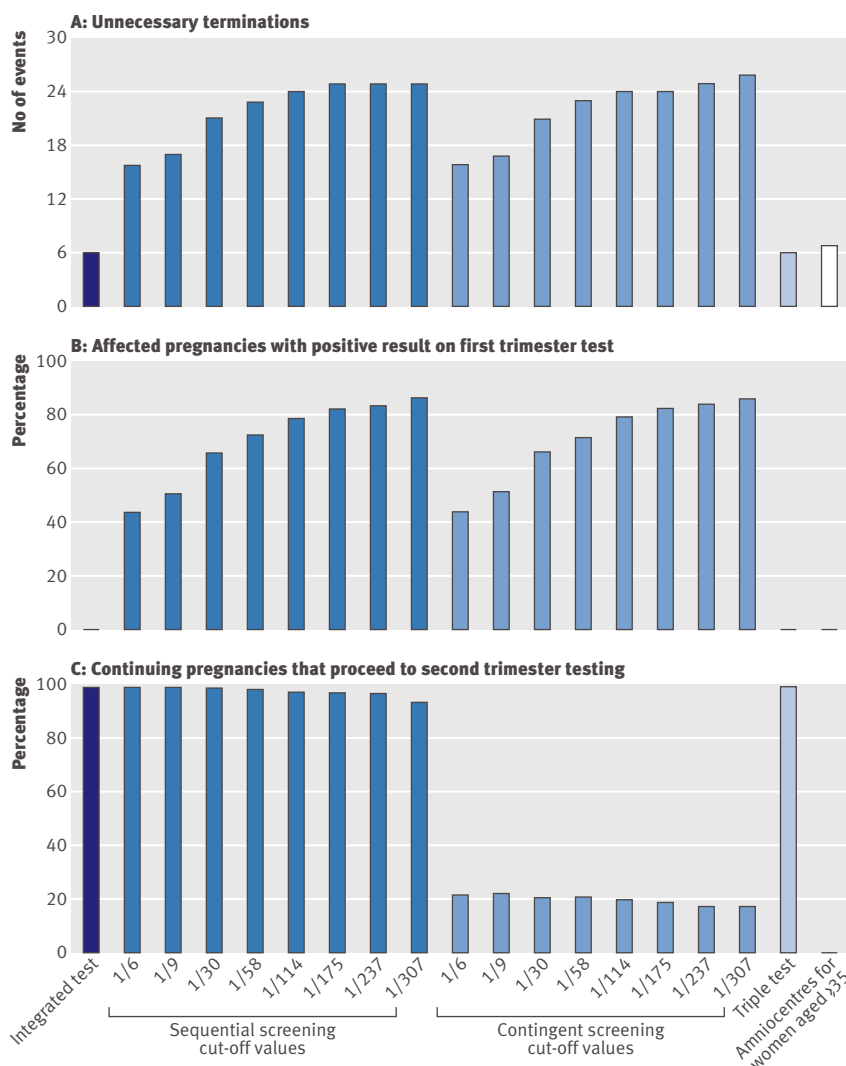
Depending on the specific end points that are considered, different screening approaches may seem more appropriate. The most appropriate screening test for Down's syndrome should have the lowest false positive rate, the highest detection rate, and the best cost effectiveness ratio.<sup>8</sup> The screening procedure (non-invasive screening and diagnostic technique) may also interfere with the number of live births with Down's syndrome. Figure 1 shows the effects of the different screening strategies on the major end points.

Early diagnosis of pregnancies affected by Down's syndrome may result in unnecessary terminations since some of the affected fetuses would be spontaneously aborted before term.<sup>4</sup> Some studies have suggest that, in order to avoid this problem, women prefer a lower false positive rate, but others suggest that women want early diagnosis. We evaluated the effect of the different screening strategies by calculating the number of unnecessary terminations, the proportion of pregnancies affected by Down's syndrome screened by a first trimester test, and the proportion of continuing pregnancies that proceed to a second trimester test (fig 2).

Contingent and sequential screening strategies produced similar results on the major outcomes—respectively, 1.43% and 1.55% for false positive results (fig 1), six procedure related euploid miscarriages each (fig 1), 16 unnecessary terminations each (fig 2), and 43.71% and 44.13% of affected pregnancies with a positive first trimester test (fig 2) (these results are those achieved with the most favourable cut-off value). However, contingent screening was unique in permitting 77.81% of patients to be reassured at the first trimester and 77.95% of women to avoid retesting in the second trimester (proportion of continuing pregnancies that proceed to a second trimester testing, fig 2).

### Impact of cut-off values in first trimester screening

We simulated the use of different cut-off values for the risk of a fetus having Down's syndrome in the first trimester tests of the contingent and sequential screening strategies. For both contingent and sequential screening, these had significant effects on



**Fig 2** | Effects of different prenatal screening strategies for Down's syndrome on (A) number of unnecessary terminations, (B) proportion of pregnancies affected by Down's syndrome screened by a first trimester test, and (C) proportion of continuing pregnancies that proceed to a second trimester test (all values are per 100 000 pregnancies)

- Cost effectiveness ratios (from \$C26 833 to \$C37 260 and from \$C35 215 to \$C45 314 per case of Down's syndrome respectively)
- False positive rate (1.55% to 3.69% and 1.43% to 3.66%, fig 1)
- Number of procedure related euploid miscarriages (6 to 46 and 6 to 45, fig 1)
- Number of unnecessary terminations (16 to 26 and 16 to 25, fig 2)
- Proportion of affected pregnancies with a positive result from the first trimester test (44.1% to 86.1% and 43.7% to 86.3%, fig 2).

## DISCUSSION

### Limitations of study

Our results are based on mathematical modelling and not on prospective observational data. However, we used empirical data and true healthcare costs as input parameters for the simulations and our reported cost effectiveness ratios were based on Québec province's healthcare system and costs. However, given the robustness of our findings observed in the sensitivity analyses, our conclusions are likely to be applicable to other jurisdictions. The demographic characteristics of the population we simulated are similar to those of other Western countries. With respect to healthcare costs, they could be estimated for other provinces by using other costs units in the simulation model.

Our study did not consider the logistical problems and costs of implementing a screening strategy that includes nuchal translucency measurement or chorionic villous sampling in North America. Neither technique is universally available throughout the US and Canada.

### Cut-off values for contingent and sequential screening options

There is no consensus on what are the most appropriate cut-off values for high risk in the first trimester tests for

contingent and sequential screening strategies,<sup>9</sup> even though they could have a major effect on screening performance.<sup>24</sup> Our data show the impact of modulating the cut-off values and confirm that, for risk assessment to be successful, a careful choice of cut-off is required. See [bmj.com](http://bmj.com) for discussion.

For contingent screening, the optimal cut-off value seems to be 1 in 9 cut-off with regard to cost effectiveness. The cut-off value of 1 in 30 showed the best cost effectiveness ratio, whereas the value of 1 in 6 produced the best incremental cost effectiveness ratio. An in-between value of 1 in 9 should achieve the best balance between these performance measures.

The cut-off value for the combined test (first trimester measurement of nuchal translucency, pregnancy associated plasma protein A, and  $\beta$  human chorionic gonadotrophin) which has become the de facto standard of care in the UK (and may spread in North America<sup>210</sup>), used a first trimester cut-off value of 1 in 670 for obtaining a 90% detection rate.<sup>5</sup> With strategies that combine first and second trimester tests,<sup>39</sup> healthcare workers would have to be prepared (and trained) for the use of the much higher level of 1 in 9, and appropriate information would be needed for patients.

**Contributors:** See [bmj.com](http://bmj.com)

**Funding:** This study was funded by the CanGèneTest research consortium on genetic laboratory services through a team grant from the Institute of Genetics of the Canadian Institutes for Health Research, the Heart and Stroke Foundation of Canada, the Canadian Agency for Drugs and Technologies in Health, and the Canadian Genetics Diseases Network of Centres of Excellence. EB holds a clinician-scientist award from the Canadian Institute for Health Research and the Jeanne et Jean-Louis Lévesque Perinatal Research Chair at Laval University.

**Competing interests:** None declared.

### WHAT IS ALREADY KNOWN ON THIS TOPIC

Prenatal screening for Down's syndrome is widely used, with many different screening strategies but still no consensus on the most cost effective approach

New strategies that provide a relatively high detection rate combine tests at both first and second trimesters

### WHAT THIS STUDY ADDS

For strategies that combine first and second trimester tests, the choice of cut-off value for risk in the first trimester test significantly influences the cost effectiveness ratios and outcomes

Contingent screening, with a first trimester cut-off value for high risk of 1 in 9, is the preferred option. This cut-off value is much higher than those generally used in screening programmes

- 1 Wyatt P. Integrated, sequential, and contingent screening for Down syndrome—local needs should drive methodology. *Prenat Diagn* 2007;27:186-7.
- 2 Ball RH, Caughey AB, Malone FD, Nyberg DA, Comstock CH, Saade GR, et al. First- and second-trimester evaluation of risk for Down syndrome. *Obstet Gynecol* 2007;110:10-7.
- 3 Reddy UM, Wapner RJ. Comparison of first and second trimester aneuploidy risk assessment. *Clin Obstet Gynecol* 2007;50:442-53.
- 4 Wald NJ, Rudnicka AR, Bestwick JP. Sequential and contingent prenatal screening for Down syndrome. *Prenat Diagn* 2006;26:769-77.
- 5 Wald NJ, Rodeck C, Hackshaw AK, Walters J, Chitty L, Mackinson AM. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *J Med Screen* 2003;10:56-104.
- 6 Caughey AB. Cost-effectiveness analysis of prenatal diagnosis: methodological issues and concerns. *Gynecol Obstet Invest* 2005;60:11-8.
- 7 AETMIS. Le dépistage prénatal du syndrome de Down et d'autres aneuploïdes au premier trimestre de la grossesse. 2003. (AETMIS 03-01) xxi:84. [www.aetmis.gouv.qc.ca/site/416.262.0.0.1.0.phtml](http://www.aetmis.gouv.qc.ca/site/416.262.0.0.1.0.phtml)
- 8 Summers AM, Langlois S, Wyatt P, Wilson RD. Prenatal screening for fetal aneuploidy. *J Obstet Gynaecol Can* 2007;29:146-79.
- 9 Reddy UM. The evolving prenatal screening scene. *Obstet Gynecol* 2007;110:2-4.
- 10 American College of Obstetricians and Gynecologists. First-trimester screening for fetal aneuploidy: opinion 296. *Obstet Gynecol* 2004;104:215-7.

**Accepted:** 12 December 2008

# Effect of topical alkane vapocoolant spray on pain with intravenous cannulation in patients in emergency departments: randomised double blind placebo controlled trial

Ramzi Hijazi,<sup>1</sup> David Taylor,<sup>1,2</sup> Joanna Richardson<sup>2</sup>

## EDITORIAL by Moore et al

<sup>1</sup>University of Melbourne, Parkville, Victoria, Australia 3052

<sup>2</sup>Austin Health, Melbourne, Victoria, Australia 3084

Correspondence to: D Taylor  
David.Taylor@austin.org.au

Cite this as: *BMJ* 2009;338:b215  
doi:10.1136/bmj.b215

## ABSTRACT

**Objective** To assess the efficacy, acceptability, and safety of a topical alkane vapocoolant in reducing pain during intravenous cannulation in adults.

**Design** Randomised double blind placebo controlled trial.  
**Setting** Emergency department of a metropolitan teaching hospital.

**Participants** 201 adult patients (54% male), mean (SD) age 58.2 (19.5) years, who required intravenous cannulation.

**Interventions** Less than 15 seconds before cannulation, the skin area was sprayed with either water (control, n=98) or vapocoolant (intervention, n=103), from a distance of 12 cm for 2 seconds. The intervention spray was a blend of propane, butane, and pentane.

**Main outcome measures** Pain with cannulation and discomfort with spray, measured with a 100 mm visual analogue scale.

**Results** Groups did not differ significantly in age, sex, indication for or site of cannulation, cannula size, or who cannulated the patient ( $P>0.05$ ). Median (interquartile range) pain scores for cannulation in the control and intervention groups were 36 (19-51) and 12 (5-40) mm, respectively ( $P<0.001$ ), and 59 (60%) and 33 (32%) reported pain scores  $\geq 30$  mm ( $P<0.001$ ). Scores for spray discomfort also differed significantly ( $P<0.001$ ) because of skewing to the right within the intervention group. The median discomfort scores, however, were 0 mm in both groups. Success rates for first cannulation attempt did not differ between groups ( $P=0.39$ ). Thirty four (39%) and 62 (62%) patients said they would choose the spray they received for analgesia in the future ( $P=0.002$ ). At follow-up at five days, two patients in the intervention group reported transient skin redness.

**Conclusions** Topical alkane vapocoolant spray is effective, acceptable, and safe in reducing pain with peripheral intravenous cannulation in adults in the emergency department.

**Trial registration** Australian Clinical Trials ACTRN12607000470493.

## INTRODUCTION

About half of patients report moderate to severe pain with cannulation and anxiety before the procedure.<sup>1</sup> On a 100 mm visual analogue scale pain scores in untreated adults ranged from 24 mm to 38 mm.<sup>1-3</sup>

Intradermal injection of lidocaine effectively reduces pain,<sup>2,3</sup> but the injection itself is painful,<sup>3</sup> and local tissue distortion, caused by the injection, might increase the rate of cannulation failure.<sup>1-3</sup>

Topical local anaesthetic agents must penetrate the stratum corneum barrier, which necessitates application 45 minutes in advance for Emla (lidocaine 2.5% and prilocaine 2.5%) and 30 minutes for Ametop (4% tetracaine), but in emergency departments immediate cannulation is often required.

Topical vapocoolant sprays can produce immediate skin anaesthesia. Rapid evaporation of volatile liquid spray from the skin surface causes a drop in temperature and results in temporary interruption of pain sensation and therefore offers a potentially convenient and effective anaesthetic for intravenous cannulation. Four randomised controlled trials have been reported. Two showed ethyl chloride to be effective,<sup>2,3</sup> while two others found ethyl chloride<sup>1</sup> and fluorohydrocarbon,<sup>4</sup> respectively, to be ineffective.

We assessed the efficacy, acceptability, and safety of a topical alkane vapocoolant spray for reducing pain with intravenous cannulation in adults by comparing its effects with a control (water) spray.

## METHODS

**Study design**—The trial took place from November 2007 to May 2008. It was a randomised double blind placebo controlled clinical trial set in a mixed (adult and paediatric) emergency department that treated about 55 000 patients a year. Patients were included if they were aged  $\geq 18$  and needed intravenous cannulation. Exclusion criteria were refusal to participate, inability to provide informed consent (non-English speaking, altered mental state, severe illness, urgent need for cannulation), moderate to severe discomfort or pain, skin disease associated with cold intolerance (such as Raynaud's phenomenon), known allergy to spray contents, peripheral neuropathy or numbness, parenteral analgesia within the previous four hours, and the use of other local anaesthesia.

**Recruitment**—We enrolled a convenience sample comprising consecutive patients who met inclusion criteria, usually between 9 am and 5 pm on weekdays, when the principal investigator was present in the emergency room. Participants received a verbal and

written explanation of the study and gave written consent.

**Randomisation**—Each enrolled patient was then assigned the next sequentially ordered study pack. Patients were block randomised (blocks of six) by an independent pharmacist using a computerised random number generator. Before informed consent had been obtained, only the pharmacist knew the randomisation status. After consent the principal investigator also knew status. All others were blinded to group allocation.

**Intervention and control sprays**—The vapocoolant (intervention) spray was COLD Spray, a propane, butane, and pentane blend, with an added fragrance, and is supplied in a handheld pressurised spray can, used for first aid treatment of muscular pain and other injuries. The control (placebo) spray was Evian Eau Minerale Naturelle, a pure water spray with hydrocarbon propellant, with similar packaging to the intervention spray.

**Spray application**—The intervention and control spray cans were masked, but because of the slight differences in the two sprays (variable transient skin blanching, jet force, and trajectory) the spray administrator could not be blinded. A blinded member of emergency department staff identified a suitable vein for cannulation. The overlying skin was wiped with an alcohol swab and allowed to dry. The principal investigator then administered the allocated spray from about 12 cm for two seconds to avoid “frosting” of vapocoolant on the skin and evaporation was allowed for up to 10 seconds. The area was again wiped with an alcohol swab and cannulation proceeded immediately. Cannulation had to be carried out within 15 seconds of administration of the spray.

**Outcome measures**—Our primary outcome measure was pain with cannulation. Secondary measures were discomfort with the spray on administration, success rate of cannulation, willingness of the patient to choose the allocated spray in the future, the patients’ guess at randomisation status, and unexpected events. About one minute after cannulation, the patient marked their perceived level of pain with cannulation, followed by level of initial discomfort with the spray on a visual analogue scale comprising a 100 mm horizontal line labelled “no pain” at the left end and “worst pain

imaginable” at the right. We then asked the patient about their willingness to choose the allocated spray in the future and to guess at their randomisation status. A blinded assistant (emergency department physician or nurse) not involved with the patient’s care collected all outcome data. The principal investigator recorded only patients’ demographics and cannulation details. The principal investigator attempted to follow-up all patients five days after cannulation, either by visiting the ward or by telephone at home. Patients were asked open and closed questions about any unexpected events experienced at the cannulation site. At least three attempts were made to contact each patient.

**Statistical analysis**—The mean pain score with cannulation has been reported as 30 mm (SD 25).<sup>1</sup> Our study was conservatively powered to detect a 10 mm difference between mean pain scores in the groups (30 mm *v* 20 mm, respectively). At least 98 patients were required in each group (power 0.8, level of significance 0.05). All data were analysed with the intention to treat principle and SPSS statistical software (SPSS, Chicago, IL) (level of significance 0.05).

## RESULTS

### Study population

Of 304 patients assessed for enrolment, 201 were randomised: 98 to the control group and 103 to the intervention group. The groups did not differ significantly ( $P>0.05$ ) in age, sex, reason for cannulation, cannulation site, cannula size, or who cannulated the patients. There were, however, five protocol violations. For one patient in the control group and two in the intervention group, the cannulation site was slightly away from the site sprayed. Also, for two patients in the intervention group, incomplete preparation resulted in a delay of more than 15 seconds between spraying and cannulation.

### Study outcomes

The table shows the main outcome measures. Patients in the intervention group reported significantly lower pain scores with cannulation; their median pain score was one third that of the control group. There were also significantly fewer patients in the intervention group who reported a pain score of  $\geq 30$  mm.

Discomfort from both sprays was generally slight, with median discomfort scores of zero in both groups. However, 23 (24%) and 50 (49%) patients in the control and intervention groups, respectively, reported a discomfort score of more than zero. Significantly more patients in the intervention group reported that they would choose the spray they received if they had a choice in the future. The nature of the spray did not affect success rates of cannulation.

Significantly more patients in the control group correctly guessed the nature of the spray they received. Despite this, almost a third and a half of patients in the

**Outcome measures in patients undergoing cannulation according to allocation to control (water spray) or intervention (vapocoolant spray). Figures are numbers (percentages) of patients unless specified otherwise**

	Control (n=98)	Intervention (n=103)	P value
Median (IQR) pain with cannulation	36 (19-51)	12 (5-40)	<0.001
Cannulation pain $\geq 30$ mm	59 (60)	33 (32)	<0.001
Median (IQR) discomfort with spray	0 (0-0)	0 (0-11)	<0.001
Successful cannulation	73 (75)	83 (81)	0.390
Future choice of same spray*	34 (39)	62 (62)	0.002
Correct guess at nature of spray	68 (69)	56 (54)	0.001

IQR=interquartile range.

\*Data missing for 10 in control group and three in intervention group.

control and intervention groups, respectively, did not correctly guess which spray they received.

At five days after cannulation, 73 (75%) and 83 (81%) patients in the control and intervention groups, respectively, were followed up. Of these, two patients in the intervention group reported transient redness at the site sprayed. No other unexpected events were reported.

## DISCUSSION

We have shown that, compared with control patients, those who received alkane vapocoolant had a 24 mm lower median pain score (18 mm lower mean) and significantly fewer had pain scores  $\geq 30$  mm.

There was a significant difference between the discomfort of placebo and vapocoolant application. The median discomfort score in both groups, however, was zero, and the absolute amount of discomfort from the vapocoolant was small. Furthermore, almost two thirds of patients who received the vapocoolant spray would choose this treatment in the future to reduce cannulation pain compared with about one third of patients who received the placebo.

Our unexpected event rate was low compared with those reported for tetracaine (erythema 34%, pruritus 6%),<sup>5</sup> Emla (erythema 6%),<sup>5</sup> and lidocaine (erythema 13%, swelling 53%).<sup>1</sup> Other risks associated with vapocoolant are likely to be minimal. Like many other vapocoolants, the one we examined is flammable and its use around heat or ignition sources is not recommended.<sup>6</sup> The manufacturer recommends a spray time of no longer than five seconds and a distance of at least 12 cm to avoid frostbite.<sup>6</sup>

Though vasoconstriction from cooling might increase the difficulty of cannulation, we found no significant difference in success rates of cannulation between the two groups.

### Strengths and limitations

Selection bias might have occurred if patients who refused or were excluded differed from those enrolled. Enrolment of consecutive patients probably minimised selection bias. Only 24 patients refused to take part, and there is no reason to believe that excluded patients differed substantially. Perception of cannulation pain is unaffected by the presence or absence of other painful conditions.<sup>7</sup> Hence, selection bias is unlikely to have affected the results. Although the baseline characteristics of the two groups were similar, we did not measure other potential confounders, such as pain threshold and needle anxiety. About two thirds of patients in the control group guessed their randomisation status, which might have resulted in measurement bias. All assessors were familiar in the use of the visual analogue scale, and any bias in variation between them is likely to have been balanced between the two groups.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

There have been conflicting reports from small unblinded studies on the efficacy of vapocoolant sprays to reduce pain with intravenous cannulation

Some clinicians use these agents for this indication

## WHAT THIS STUDY ADDS

Alkane vapocoolant spray results in significant reductions in pain with cannulation and is safe and acceptable to patients

### Comparison with other studies

Although methods differed, our findings are consistent with those of two studies that examined ethyl chloride for the same indication.<sup>2,3</sup> In contrast, other vapocoolant trials of ethyl chloride<sup>1</sup> and fluorohydrocarbon<sup>4</sup> did not show significant pain relief. The ethyl chloride study however, was not blinded, used a larger sized cannula (18 gauge rather than 20 gauge), sprayed the vapocoolant from 25 cm until a layer of frosting was seen,<sup>1</sup> and had a vapocoolant group comprising only 30 patients. The fluorohydrocarbon study was not blinded and was probably underpowered as the standard deviation used in the sample size calculation was considerably smaller than that observed in the data collected.<sup>4</sup> Hence, strengths of our study are the use of blinding and the considerably larger sample size.

### Recommendations

Notwithstanding these limitations, our findings indicate that vapocoolant spray might be useful for decreasing pain with cannulation and consideration should be given for its routine use.

**Contributors:** See [bmj.com](http://bmj.com).

**Funding:** This study was funded by an advanced medical science grant from the University of Melbourne, Victoria, Australia. There were no study sponsors. The authors were entirely independent from the funders. The funders had no role in the project.

**Competing interests:** None declared.

**Ethical approval:** The Austin Health human research and ethics committee approved the study. All patients gave written informed consent.

- 1 Biro P, Meier T, Cummins AS. Comparison of topical anaesthesia methods for venous cannulation in adults. *Eur J Pain* 1997;1:37-42.
- 2 Selby IR, Bowles BJ. Analgesia for venous cannulation: a comparison of EMLA (5 minutes application), lignocaine, ethyl chloride, and nothing. *J R Soc Med* 1995;88:264-7.
- 3 Armstrong P, Young C, McKeown D. Ethyl chloride and venepuncture pain: a comparison with intradermal lidocaine. *Can J Anaesth* 1990;37:656-8.
- 4 Hartstein BH, Barry JD. Mitigation of pain during intravenous catheter placement using a topical skin coolant in the emergency department. *Emerg Med J* 2008;25:257-61.
- 5 Browne J, Awad I, Plant R, McAdoo J, Shorten G. Topical amethocaine (Ametop (TM)) is superior to EMLA for intravenous cannulation. *Can J Anaesth* 1999;46:1014-8.
- 6 Reis EC, Jacobson RM, Tarbell S, Weniger BG. Taking the sting out of shots: control of vaccination-associated pain and adverse reactions. *Pediatr Ann* 1998;27:375-86.
- 7 Harris T, Cameron PA, Ugoni A. The use of pre-cannulation local anaesthetic and factors affecting pain perception in the emergency department setting. *Emerg Med J* 2001;18:175-7.

Accepted: 22 October 2008

# Combination antiretroviral therapy in population affected by conflict: outcomes from large cohort in northern Uganda

Andrew Kiboneka,<sup>1</sup> Ricky Jones Nyatia,<sup>2</sup> Christine Nabiryo,<sup>1</sup> Aranka Anema,<sup>3</sup> Curtis L Cooper,<sup>4</sup> Kimberly Ann Fernandes,<sup>3</sup> Julio S G Montaner,<sup>3</sup> Edward J Mills<sup>1,3</sup>

## EDITORIAL by Jacob and Abongomera

<sup>1</sup>The AIDS Support Organization (TASO), Kampala, Uganda

<sup>2</sup>TASO, Gulu, Uganda

<sup>3</sup>British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada

<sup>4</sup>Division of Infectious Diseases, Ottawa Hospital, Ottawa, Canada

Correspondence to: E Mills, St Paul's Hospital, Room 627, 1081 Burrard Street, Vancouver, BC, Canada V6Z 1Y6  
emills@cfcnet.ubc.ca

Cite this as: *BMJ* 2009;338:b201  
doi:10.1136/bmj.b201

## ABSTRACT

**Objective** To measure the clinical and immunological outcomes of HIV positive adult patients receiving combination antiretroviral therapy in conflict affected northern Uganda.

**Design** Prospective cohort study.

**Setting** Gulu District, northern Uganda.

**Participants** 1625 adults (aged over 14 years) receiving combination antiretroviral therapy.

**Main outcome measures** Primary outcome: all cause mortality. Secondary outcomes: impact of covariates (sex, age, CD4 count at start, adherence, tuberculosis at start, duration of treatment, and internally displaced person status) on mortality.

**Results** Sixty nine (4.2%) patients died during follow-up. The mortality incidence rate was 3.48 (95% confidence interval 2.66 to 4.31) per 100 person years. Patients started treatment with a median CD4 count of 157 (interquartile range 90-220) cells/ $\mu$ l; most (1009; 63%) had World Health Organization stage 2 defined illness. Sixty two patients had pulmonary tuberculosis at the start of treatment. Of the 1521 patients with adherence data, 118 (7.8%) had adherence of less than 95% and 1403 (92.2%) had adherence of 95% or above.

**Conclusion** Patients receiving combination antiretroviral therapy in conflict affected northern Uganda had a mortality comparable to that of patients in peaceful, low income settings and better adherence than patients in higher income settings. These favourable findings highlight the need to expand access to combination antiretroviral therapy in populations affected by armed conflict.

## INTRODUCTION

Sub-Saharan Africa is home to two thirds of the global HIV/AIDS epidemic.<sup>1</sup> The continent simultaneously has the world's highest rate of armed conflict, which affects more than a third of African countries.<sup>2</sup> International policy guidelines for humanitarian responses previously intimated that the provision of combination antiretroviral therapy was not feasible in complex emergencies,<sup>3,4</sup> owing to associated population movements and poor access to basic health services.<sup>5</sup> As a consequence, populations affected by conflict continue to have very limited access to combination antiretroviral therapy.<sup>1</sup>

A recent consensus statement from many international agencies, including the World Health

Organization and the United Nations High Commissioner for Refugees (UNHCR), indicated that provision of combination antiretroviral therapy to HIV infected people in emergency settings represents a public health and human rights imperative.<sup>6,7</sup> The UNHCR has since developed clinical and operational guidelines for the management of combination antiretroviral therapy in displaced populations.<sup>8</sup> A critical need exists for increased research in this area to inform evidence based humanitarian interventions.<sup>9</sup> We present clinical and immunological outcomes of a large cohort of adult patients receiving combination antiretroviral therapy in an ongoing complex humanitarian emergency.

## METHODS

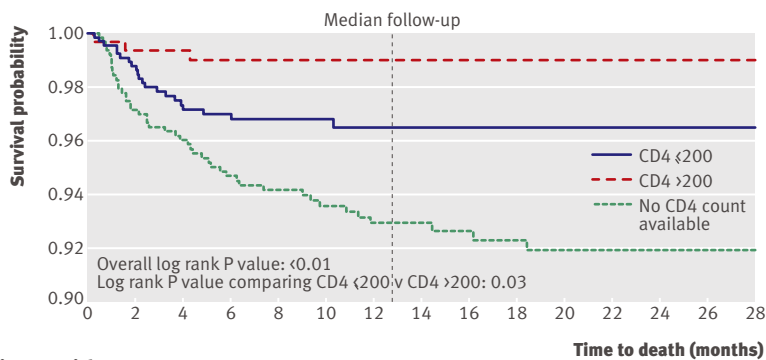
### Setting

Northern Uganda has been in a state of humanitarian emergency for more than 20 years. Of the estimated 1.6 million people displaced between 2002 and 2004, most continue to reside in protected camps.<sup>10</sup> As of June 2008, efforts towards peace were compromised and movement restrictions remain in place for those living in internally displaced persons' camps.

Between January and July 2005, Gulu district and municipality had a population of approximately 566 000 inhabitants. Most residents lived in internally displaced persons' camps. The crude mortality rate in the area was estimated to be high, reaching 1.22 (95% confidence interval 1.00 to 1.44) per 10 000 per day among camp residents and 1.29 (1.04 to 1.53) per 10 000 per day in the surrounding community. St Mary's Lacor Hospital, a large non-governmental hospital in Gulu, reports HIV/AIDS as the most common reason for death in 2005.<sup>11</sup> A sero-surveillance study among women attending antenatal clinics in Gulu District in 2006 found a prevalence of HIV infection of 10.3%.

### Programme

Before 2004, no combination antiretroviral therapy was provided publicly in Gulu and private provision of care was limited. Since 2004, The AIDS Support Organization (TASO) has started combination antiretroviral therapy in 1625 adults and 57 children. The primary initiation regimen is based on non-nucleoside reverse transcriptase inhibitors. At the time of data collection first line treatment typically comprised



No of patients at risk:		Time to death (months)														
		0	2	4	6	8	10	12	14	16	18	20	22	24	26	28
CD4 $\leq$ 200	664	570	448	281	109	99	53	32								
CD4 $>$ 200	318	285	251	188	133	131	104	64								
No CD4 count available	643	588	506	414	276	258	207	86								

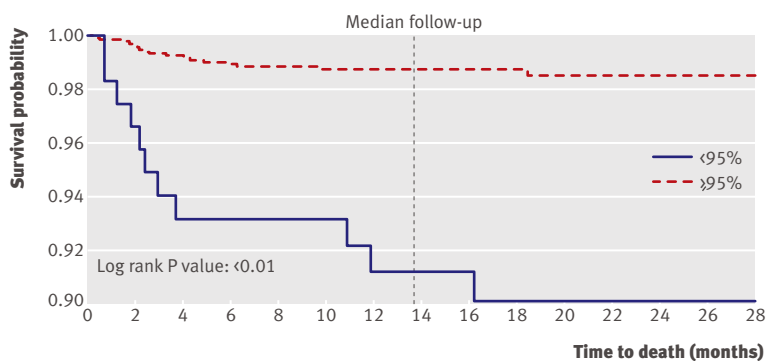
**Fig 1** | Kaplan-Meier plot for survival in patients with CD4 cell count at start of treatment  $>200$  cells/ $\mu$ l or  $\leq 200$  cells/ $\mu$ l and in those with no CD4 data available

nevirapine, lamivudine, and stavudine, and second line treatment comprised boosted lopinavir, didanosine, and zidovudine. Criteria used for starting treatment at this time included WHO stage 3 or 4 illness or a CD4 cell count below 200 cells/ $\mu$ l. Many patients will not have CD4 evaluations.

#### Data collection

Clinicians and field workers complete standardised forms detailing patients' demographics, as well as clinical, psychosocial, and drug use data at each visit. A field adherence monitoring team is responsible for active retention and follow-up of patients. The team visits patients who fail to attend any appointment and those who have home based care. Measurement of adherence involves a composite of pharmacy monitored drug possession ratio, pharmacy refill records, and a three day recall report by patients or care givers. TASO defines clinical adherence as  $\geq 95\%$ .

Patients began receiving combination antiretroviral therapy from 8 June 2005 and were followed until 29



No of patients at risk:		Time to death (months)														
		0	2	4	6	8	10	12	14	16	18	20	22	24	26	28
Adherence $<95\%$	118	106	100	91	84	82	65	34								
Adherence $\geq 95\%$	1403	1266	1056	764	415	389	290	144								

**Fig 2** | Kaplan-Meier plot for survival in patients defined as adherent or not adherent ( $<95\%$  or  $\geq 95\%$  adherence)

January 2008. The median number of patients who entered the programme was 47 (interquartile range 23-72) monthly.

#### Outcomes and analysis

The primary outcome was all cause mortality. Secondary outcomes included assessment of immunological status at start of combination antiretroviral therapy and patients' characteristics. Covariates included age, sex, CD4 count at start of treatment, presence of tuberculosis, internally displaced person status, and adherence defined as above.

We used Kaplan-Meier methods to assess survival over time. We used Cox proportional hazards to determine if covariates affected mortality.

#### RESULTS

A total of 1625 patients receiving combination antiretroviral therapy were included in our analysis—100% of all patients who ever received combination antiretroviral therapy from TASO in Gulu District. Most (72%) patients were women, and the median age was 39 (interquartile range 33-46) years. All patients were started on non-nucleoside reverse transcriptase inhibitor based regimens.

Patients started treatment with a median CD4 count of 157 (interquartile range 90-220) cells/ $\mu$ l, and most (63%) had WHO stage 2 disease. A total of 62 patients had been diagnosed as having pulmonary tuberculosis at baseline. The median follow-up time was 12.8 (7.7-23.1) months; 69 (4.2%) patients died during follow-up, giving a mortality incidence rate of 3.48 (95% confidence interval 2.66 to 4.31) per 100 person years.

Of the 982 patients with available baseline CD4 counts, 664 (68%) had CD4 counts of 200 cells/ $\mu$ l or less. Mortality varied by baseline CD4 cell count: 21/664 (3.2%, 3.22/100 patient years) patients with CD4 counts of 200 cells/ $\mu$ l or less died compared with 3/316 (0.9%, 0.68/100 patient years) patients with CD4 counts above 200 cells/ $\mu$ l. For the 643 patients with no initial CD4 count available, 45 patients died (7.0%, 50.8/100 person years). Figure 1 shows the Kaplan-Meier plot for survival in patients with CD4 counts of 200 cells/ $\mu$ l or lower and above 200 cells/ $\mu$ l and for those without CD4 cell evaluations.

Of the 1521 patients with complete data for adherence, 118 (7.8%) patients had adherence below 95% and 1403 (92.2%) had adherence of 95% or above. A total of 11/118 (9.3%, 5.24/100 patient years) patients with less than 95% adherence died compared with 17/1403 (1.2%, 1.00/100 patient years) patients with at least 95% adherence. Figure 2 shows the Kaplan-Meier plot for survival of patients defined as adherent or not adherent ( $<95\%$  or  $\geq 95\%$ ).

Lower mortality was associated with female sex (hazard ratio 0.70, 95% confidence interval 0.55 to 0.91,  $P=0.02$ ), higher baseline CD4 count (hazard ratio

per 100 cell increase 0.14, 0.06 to 0.34,  $P < 0.0001$ ), and at least 95% adherence (hazard ratio 0.14, 0.10 to 0.21,  $P < 0.0001$ ).

## DISCUSSION

Adults receiving treatment in Gulu, northern Uganda, showed clinical outcomes better than those found in the only other existing study of patients receiving combination antiretroviral therapy in a conflict setting.<sup>12</sup> The results compare favourably with those found in peaceful regions of Uganda and other low income countries,<sup>13-16</sup> and they are consistent with UNAIDS estimates of age specific HIV prevalence and access to combination antiretroviral therapy in Uganda.<sup>1</sup>

### Findings in context with other studies

The mortality incidence rate among adult patients receiving combination antiretroviral therapy in our cohort was about half of that found in Bukavu, Democratic Republic of Congo, a setting with ongoing active violence and noticeably fewer aid agencies available.<sup>12</sup> Interestingly, this is also lower than the mortality found in better resourced, politically stable areas of Uganda, such as Rakai, where the mortality has been 5.2 per 100 person years.<sup>13</sup> We found that residence in a camp was not associated with poorer outcomes.

### Strengths and limitations

Strengths of our study include the retention of patients in our cohort. Whereas many programmes in Africa are affected by considerable loss to follow-up,<sup>17</sup> our programme uses a counselling team on motorcycles to consistently track patients. Involvement in community groups assists patients in retention, countering stigma, and promoting health.

Regarding limitations, as with any cohort study, confounding variables of which we are unaware may be present. We have tried to control for these by using a priori explanations of bias. We did not have routine CD4 counts for all patients and do not do viral load evaluations. We did a sensitivity analysis in which the Cox proportional hazards survival analysis was re-done with adjustment for all variables except CD4 at start of treatment. We found the same magnitude of

association between adherence and mortality as in the previous analysis that excluded patients without CD4 counts.

### Possible explanations and implications

The adherence to treatment among our cohort was high. Whereas 92% of our patients had at least 95% adherence, pooled adherence rates across sub-Saharan Africa and North America show that 77% and 55% of patients adhered to combination antiretroviral therapy.<sup>18</sup> Duration of clinical follow-up may contribute to levels of adherence.

These impressive clinical outcomes may also be attributed to the stability of the population in Gulu after the influxes of 2002 and 2004.<sup>10</sup> Although violence continues, compared with other districts affected by conflict in northern Uganda Gulu has the largest number of health facilities.<sup>19</sup> Steps will need to be taken to ensure that combination antiretroviral therapy is not disrupted as a result of the relocation of people from the camp to their homes and vice versa.<sup>8,20</sup> This should include cooperation with HIV treatment facilities in neighbouring districts and delivery of additional emergency "relocation" stock of combination antiretroviral therapy and the creation of treatment information cards and duplicate medical records for patients.<sup>12</sup>

### Conclusions

Our findings show that the provision of combination antiretroviral therapy in contexts with armed conflict is both feasible and potentially highly successful. Our study validates recent policies set out by the WHO and UNHCR about HIV treatment in complex emergencies and shows the imperative of expanding access to combination antiretroviral therapy to populations affected by armed conflict.<sup>21</sup>

**Contributors:** See bmj.com.

**Funding:** None. The US President's Emergency Plan for AIDS Relief (PEPFAR) is the primary funding body associated with TASO; it provided no funding for this study and has not seen the findings. AK and CN receive support from the Canada-Africa Prevention Trials (CAPT) Network. EJM is supported by a Canada Research Chair.

**Competing interests:** None declared.

**Ethical approval:** The administrative headquarters of TASO Uganda, Kampala, and the Mbale Regional Referral Hospital Review Board approved the study.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Early guidelines suggested that antiretroviral therapy for populations affected by conflict was untenable

Access to antiretroviral therapy for conflict affected populations remains low

## WHAT THIS STUDY ADDS

Patients receiving antiretroviral therapy had favourable outcomes in mortality and adherence

Innovative strategies to retain patients in care may contribute to positive health outcomes and reduce mortality and loss to follow-up

Expansion of antiretroviral care to other conflict affected populations is warranted

- 1 Joint United Nations Programme on HIV/AIDS (UNAIDS). *Towards universal access: scaling up priority HIV/AIDS interventions in the health sector*. Geneva: WHO, 2007 (available at [www.who.int/hiv/mediacentre/universal\\_access\\_progress\\_report\\_en.pdf](http://www.who.int/hiv/mediacentre/universal_access_progress_report_en.pdf)).
- 2 Uppsala Conflict Data Program (UCDP). Customized report: active conflicts in Africa. 2006. [www.pcr.uu.se/database/CustomizedReport/subset.php](http://www.pcr.uu.se/database/CustomizedReport/subset.php).
- 3 Steering Committee for Humanitarian Response. *The Sphere project: humanitarian charter and minimum standards in disaster response*. Geneva: Sphere Project, 2004 (available at [www.humanitarianinfo.org/IMToolbox/10\\_Reference/Sphere\\_Guidelines\\_2004/2004\\_English\\_Sphere.pdf](http://www.humanitarianinfo.org/IMToolbox/10_Reference/Sphere_Guidelines_2004/2004_English_Sphere.pdf)).
- 4 Inter-Agency Standing Committee (IASC). *Guidelines for HIV/AIDS interventions in emergency settings*. 2003. <http://data.unaids.org/>

- Publications/External-Documents/IASC\_Guidelines-Emergency-Settings\_en.pdf.
- 5 Spiegel PB. HIV/AIDS among conflict-affected and displaced populations: dispelling myths and taking action. *Disasters* 2004;28:322-39.
  - 6 World Health Organization. Antiretroviral drugs in emergencies: neglected but feasible. Consensus statement. 2006. www.who.int/hac/techguidance/pht/HIV\_AIDS\_101106\_arvemergencies.pdf.
  - 7 United Nations High Commissioner for Refugees. *Antiretroviral medication policy for refugees*. Geneva: UNHCR, 2007 (available at [http://data.unaids.org/pub/Report/2007/20070326\\_unhcr\\_art\\_en.pdf](http://data.unaids.org/pub/Report/2007/20070326_unhcr_art_en.pdf)).
  - 8 United Nations High Commissioner for Refugees. *Clinical guidelines for antiretroviral therapy management for displaced populations: Southern Africa*. Geneva: UNHCR, 2007 (available at: [www.unhcr.org/protect/PROTECTION/4683b0522.pdf](http://www.unhcr.org/protect/PROTECTION/4683b0522.pdf)).
  - 9 Singh S, Orbinski JJ, Mills EJ. Conflict and health: a paradigm shift in global health and human rights. *Confl Health* 2007;1:1.
  - 10 United Nations High Commissioner for Refugees. Global appeal 2008-2009: Uganda. 2008. [www.unhcr.org/home/PUBL/474ac8d00.pdf](http://www.unhcr.org/home/PUBL/474ac8d00.pdf).
  - 11 Office for the Coordination of Humanitarian Affairs (OCHA). Northern Uganda briefing paper May 2006. [www.reliefweb.int/library/documents/2006/ocha-uga-31may.pdf](http://www.reliefweb.int/library/documents/2006/ocha-uga-31may.pdf).
  - 12 Culbert H, Tu D, O'Brien DP, Ellman T, Mills C, Ford N, et al. HIV treatment in a conflict setting: outcomes and experiences from Bukavu, Democratic Republic of the Congo. *PLoS Med* 2007;4:e129.
  - 13 Lutalo T, Gray RH, Wawer M, Sewankambo N, Serwadda D, Laeyendecker O, et al. Survival of HIV-infected treatment-naïve individuals with documented dates of seroconversion in Rakai, Uganda. *AIDS* 2007;21(suppl 6):S15-9.
  - 14 Spacek LA, Shihab HM, Kanya MR, Mwesigire D, Ronald A, Mayanja H, et al. Response to antiretroviral therapy in HIV-infected patients attending a public, urban clinic in Kampala, Uganda. *Clin Infect Dis* 2006;42:252-9.
  - 15 Oyugi JH, Byakika-Tusiime J, Ragland K, Laeyendecker O, Mugerwa R, Kityo C, et al. Treatment interruptions predict resistance in HIV-positive individuals purchasing fixed-dose combination antiretroviral therapy in Kampala, Uganda. *AIDS* 2007;21:965-71.
  - 16 Braitstein P, Brinkhof MW, Dabis F, Schechter M, Boule A, Miotti P, et al. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet* 2006;367:817-24.
  - 17 Rosen S, Fox MP, Gill CJ. Patient retention in antiretroviral therapy programs in sub-Saharan Africa: a systematic review. *PLoS Med* 2007;4:e298.
  - 18 Mills EJ, Nachega JB, Buchan I, Orbinski J, Attaran A, Singh S, et al. Adherence to antiretroviral therapy in sub-Saharan Africa and North America: a meta-analysis. *JAMA* 2006;296:679-90.
  - 19 Chamla DD, Olu O, Wanyana J, Natseri N, Mukooyo E, Okware S, et al. Geographical information system and access to HIV testing, treatment and prevention of mother-to-child transmission in conflict affected northern Uganda. *Confl Health* 2007;1:12.
  - 20 Mills EJ, Cooper C. Simple, effective interventions are key to improving adherence in marginalized populations. *Clin Infect Dis* 2007;45:916-7.
  - 21 Ford N. Treating AIDS in complex emergencies: the need for clear policy consensus. *Progress in Development Studies* 2009;9:55-61.

Accepted: 17 October 2008

## Outcome of depression in later life in primary care: longitudinal cohort study with three years' follow-up

E Licht-Strunk,<sup>1</sup> H W J Van Marwijk,<sup>1</sup> T Hoekstra,<sup>2</sup> J W R Twisk,<sup>3</sup> M De Haan,<sup>1</sup> A T F Beekman<sup>4</sup>

<sup>1</sup>Department of General Practice and the EMGO Institute for Health and Care Research of VU University Medical Centre, Van der Boechorststraat 7, 1081 BT, Amsterdam, Netherlands

<sup>2</sup>Department of Health Sciences, Faculty of Earth and Life Sciences, VU University, Netherlands

<sup>3</sup>Department of Methodology and Applied Biostatistics, Institute of Health Sciences, Faculty of Earth and Life Sciences, VU University

<sup>4</sup>Department of Psychiatry and the EMGO Institute for Health and Care Research of VU University Medical Centre

Correspondence to: E Licht-Strunk [e.licht@vumc.nl](mailto:e.licht@vumc.nl)

Cite this as: *BMJ* 2009;338:a3079 [doi:10.1136/bmj.a3079](https://doi.org/10.1136/bmj.a3079)

This article is an abridged version of a paper that was published on [bmj.com](http://bmj.com). Cite this article as: *BMJ* 2009;338:a3079

### ABSTRACT

**Objectives** To study the duration of depression, recovery over time, and predictors of prognosis in an older cohort ( $\geq 55$  years) in primary care.

**Design** Longitudinal cohort study, with three years' follow-up.

**Setting** 32 general practices in West Friesland, the Netherlands.

**Participants** 234 patients aged 55 years or more with a prevalent major depressive disorder.

**Main outcome measures** Depression at baseline and every six months using structured diagnostic interviews (primary care evaluation of mental disorders according to diagnoses in *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition) and a measure of severity of symptoms (Montgomery Åsberg depression rating scale). The main outcome measures were time to recovery and the likelihood of recovery at different time points.

Multivariable analyses were used to identify variables predicting prognosis.

**Results** The median duration of a major depressive episode was 18.0 months (95% confidence interval 12.8 to 23.1). 35% of depressed patients recovered within one year, 60% within two years, and 68% within three years. A poor outcome was associated with severity of depression at baseline, a family history of depression, and

poorer physical functioning. During follow-up functional status remained limited in patients with chronic depression but not in those who had recovered.

**Conclusion** Depression among patients aged 55 years or more in primary care has a poor prognosis. Using readily available prognostic factors (for example, severity of the index episode, a family history of depression, and functional decline) could help direct treatment to those at highest risk of a poor prognosis.

### INTRODUCTION

Although the importance of depression in later life is widely acknowledged and recent trials have shown convincingly that treatment can be effective<sup>1,2</sup> depression in most older patients remains undiagnosed.<sup>3-5</sup> We studied depression in patients aged 55 or more in primary care to determine the duration of depressive episodes, recovery over time, and predictors of prognosis.

### METHODS

We screened participants with major depressive disorder from 32 practices in West Friesland, the Netherlands. Briefly, consecutive patients aged 55 or more visiting their general practitioner were invited to fill in the geriatric depression scale-15 items.<sup>6,7</sup> Patients

scoring over 5 were invited for a diagnostic interview using the primary care evaluation of mental disorders.<sup>8</sup> Participants were interviewed every six months for three years.

A major depressive disorder was diagnosed using the primary care evaluation of mental disorders (range 0-9).<sup>8</sup> We assessed patients every six months using the primary care evaluation of mental disorders and the Montgomery Åsberg depression rating scale (range 0-60),<sup>9</sup> higher scores indicating more severe depression. A cut-off of 10 defines recovery.<sup>10</sup> We defined recovery as the patient no longer fulfilling the criteria for major depressive disorders and scoring less than 10 on the depression rating scale.

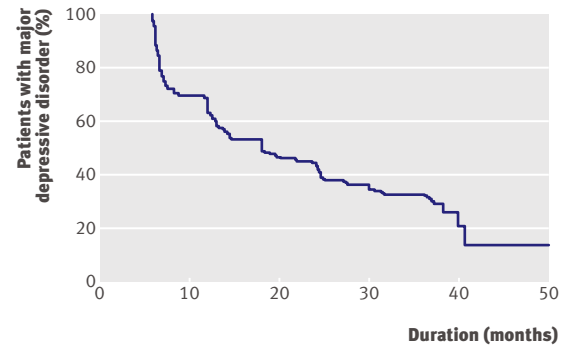
We used structured questionnaires at baseline and at one, two, and three years to collect information on personal characteristics. The questionnaire for chronic somatic comorbidity contains a list of diseases (see [bmj.com](http://bmj.com)).<sup>11</sup> We chose the cut-off of none versus one versus more than one disease to produce subgroups for comorbidity. We used the mini-mental status examination to measure cognitive decline.<sup>12</sup> A cut-off of 24 defines cognitive impairment.<sup>13</sup> Comorbid anxiety was assessed using the anxiety questionnaire of the primary care evaluation of mental disorders.<sup>8</sup> We used the diagnostic interview schedule<sup>14</sup> to assess the age at onset of the first depressive episode, family history of depression, and the number of previous episodes of depression. Early onset depression was defined as a first depressive episode before age 55. We categorised the number of previous episodes as none versus one versus more than one to produce subgroups for history of depression. Physical functioning was ascertained using the physical component scale of the medical outcome study 36-items, short form (range 0-100),<sup>15 16</sup> higher scores indicating better physical functioning.

#### Statistical analysis

We used independent *t* tests or  $\chi^2$  tests to compare the baseline characteristics of participants with only a baseline measurement with those who had measurements at follow-up. To explore the characteristics of dropouts we compared participants who had fewer than four measurements with those who had four or more.

We analysed the duration of the episodes of depression using time to recovery as the outcome measure. Time to recovery and the likelihood of recovery at different times were estimated using Kaplan Meier methods. For these analyses we included patients with at least one follow-up measurement.

We used Cox regression analyses to identify determinants at baseline predicting poor outcome (no recovery). Firstly, we investigated whether there was a linear relation between the potential predictor variables and the outcome. We divided those variables with non-linear relations into categories, using cut-off



Survival curve of 204 patients aged 55 years or more with major depressive disorder in primary care

scores from the literature or the median of the sum score. Secondly, we carried out univariable regression analyses for all potential predictors with the outcome measure. For the multivariable analyses we selected variables that might be associated with the outcome. Thirdly, we entered the predictors simultaneously in a multivariable regression model. We constructed the best predictive model using manual backward selection and deleted the variables with the lowest predictive value. The best fitting model was tested with the log likelihood ratio test. All analyses were carried out using SPSS version 15.0.

We also explored changes in predictors during follow-up. We compared the results of those who had recovered with those who still had depression.

#### RESULTS

Overall, 4222 of 5395 (78.3%) patients completed the depression questionnaire (see [bmj.com](http://bmj.com)). Non-responders were more often men and of similar age. After exclusions, 458 of 659 (70%) patients who completed the questionnaire scored over 5 and took part in a diagnostic interview. Responders and non-responders were not significantly different for age (mean difference 1.0 year, 95% confidence interval -0.5 to 2.6;  $P=0.175$ ), sex (odds ratio 1.0, 95% confidence interval 0.7 to 1.4;  $P=0.822$ ), and depression score (mean difference 0.13, -0.6 to 0.3;  $P=0.658$ ). Of the 458 patients interviewed, 244 (53%) had a major depressive disorder of whom 234 (96%) participated in the baseline interview (see [bmj.com](http://bmj.com) for characteristics of the cohort).

#### Dropouts

Two hundred and four participants (87%) had at least one follow-up measurement and were analysed. Thirty had only completed the baseline interview. They were older than those with follow-up measurements (mean difference 4.1 years, 0.8 to 7.4;  $P=0.04$ ), but were not significantly different for sex (odds ratio 1.1, 0.5 to 2.5;  $P=0.77$ ), level of education ( $P=0.87$ ), comorbidity ( $P=0.37$ ), living alone or with others (odds ratio 0.7, 0.3 to 1.7;  $P=0.52$ ), or baseline depression score (mean difference 0.4, -0.1 to 0.9;  $P=0.10$ ).

Overall, 175 respondents (75%) completed four or more assessments. Those with fewer than four were older (mean difference 4.6 years, 2.1 to 7.1;  $P<0.01$ ) and had a lower level of education ( $P=0.04$ ) but were not different for sex (odds ratio 0.6, 0.3 to 1.1;  $P=0.07$ ), comorbidity ( $P=0.25$ ), living situation (odds ratio 1.2, 0.6 to 2.2;  $P=0.60$ ), or baseline depression score (mean difference 0.2,  $-0.2$  to 0.5;  $P=0.44$ ).

#### Duration of depressive episodes and recovery over time

The mean time to recovery was 19.3 months (95% confidence interval 17.5 to 21.2), and the median time to recovery was 18.0 months (12.8 to 23.1; figure). Overall, 35.1% (95% confidence interval 28.3% to 42.0%) of participants had recovered at one year, 60.4% (53.0% to 67.7%) at two years, and 68.1% (60.9% to 75.3%) at three years.

#### Predictors of prognosis

Univariable Cox survival analysis showed that eight variables were associated with time to recovery (see [bmj.com](#)). In the best fitting multivariable model (see [bmj.com](#)) poor outcome was associated with more severe depression at baseline (hazard ratio 1.25, 95% confidence interval 1.08 to 1.45), a family history of depression (1.45, 0.97 to 2.17), and less physical functioning (0.98, 0.97 to 0.99).

Only 40% of the patients with depression were receiving treatment at baseline—31% were taking antidepressants and 9% were referred to specialised care (see [bmj.com](#)). No association was found between treatment for depression and recovery. Patients receiving treatment, however, had more severe depression at baseline (mean difference in score on depression rating scale 4.5;  $P<0.01$ ).

Cognitive decline did not change during follow-up, and no difference was found between the depressed and recovered groups (see [bmj.com](#)). The prevalence of chronic diseases increased over time and was more common in the depressed group than in the recovered group. The depressed group but not recovered group showed a decline in daily functioning. Finally, the number of patients receiving treatment hardly changed during follow-up. At three years, 37% of patients were being treated for depression, compared with 24% in the recovered group.

#### DISCUSSION

The prognosis for depression among older patients ( $\geq 55$  years) in primary care is poor. The median duration of a major depressive episode was 18 months. Thirty five per cent had recovered within one year, 60% after two years, and 68% at three years.

A systematic review on depression in later life reported a poorer outcome in inpatients; 13-18% had recovered after one year compared with 35% in our study.<sup>17</sup> A possible explanation for the difference is the higher prevalence of functional limitations among

inpatients, which was also a predictor of poor clinical outcome in our study. The review showed a better clinical outcome in hospital based studies,<sup>18</sup> possibly due to better adherence to treatment. In our cohort only 40% of patients were receiving treatment at baseline.

Poor outcome was associated with the severity of the index episode, a family history of depression, and functional decline. These clinical factors are available in routine practice and could be used to guide treatment. Our data showed that functional limitations increased over three years in the group with chronic depression but not in those who recovered.

Only a small proportion of patients with depression were being treated at baseline and during follow-up, similar to previous studies.<sup>19,20</sup> We found no association between treatment and prognosis. Patients being treated, however, had more severe depression at baseline, suggesting confounding by indication.

We provided data on factors that predict the prognosis of depression in later life, controlling for inter-relatedness. This is relevant for clinical practice, as it identifies those at highest risk of a poor prognosis.

#### Strengths and limitations

We used diagnostic and dimensional data ascertained at multiple times to study the course of depression in a large cohort of older patients in primary care. These strengths, together with follow-up for three years and access to several putative prognostic factors, allowed a thorough assessment of prognosis. As we included a sample of patients with depression in day to day clinical practice, those with existing depression might have led to over-representation of patients with chronic depression.

We screened patients regardless of the reason for consultation. As they were able to attend the practice and 98% were living independently, they were a selection of the more healthy older population. Furthermore, those lost to the analyses throughout the study were older. Given that functional limitations is a predictor for poor clinical outcome, our results may be an underestimation of the true course of depression

#### WHAT IS ALREADY KNOWN ON THIS TOPIC

Depression in later life is common in primary care and has an enormous impact on wellbeing and functioning

#### WHAT THIS STUDY ADDS

Depression in patients aged 55 or more in primary care has a poor prognosis

Identified prognostic factors (for example, severity of the index episode, a family history of depression, and functional decline) are readily available in clinical practice and could help direct treatment to those at highest risk of a poor prognosis

owing to attrition of the frailest patients during follow-up.

Using an interval of six months to assess patients may have led to an overestimation of the duration of depression if patients recovered earlier. We did not have access to data on the duration of the index episode. This systematically underestimates the duration and counteracts potential overestimation. The advantage of Cox survival analyses is that we used data on respondents with more than one measurement. However, in those respondents who missed interviews, either the follow-up was less than three years or the interval between measurements was more than six months, which may have led to bias. Those with fewer than four measurements were older and had a lower level of education than those with more measurements. Both variables were not associated with outcome.

### Conclusion

Our results support our hypothesis that the prognosis of depression in later life in primary care would be poor. One possible explanation for this is inadequate treatment in older patients, as most (60% at baseline, 71-77% during follow-up) did not receive treatment for depression. Analyses of data on patients with undetected depression during the first year of follow-up showed a poor prognosis; depression was still present in 67% after one year.<sup>5</sup> Identifying patients with depression at high risk of a poor outcome can help to improve treatment and prognosis. We found that a poor outcome was predicted by clinical factors that should be readily available in routine practice.

We thank the respondents for their participation.

**Contributors:** See [bmj.com](http://bmj.com).

**Funding:** The data reported on were collected in the context of the West Friesland Study, which was financed by The Netherlands Organisation for Health Research and Development (grant No 2200.0019) and the Dutch Organization for Scientific Research (MW grant 904.57.127).

**Competing interests:** None declared.

**Ethical approval:** This study was approved by the medical ethical committee of the VU University Medical Centre.

- 1 Chew-Graham CA, Lovell K, Roberts C, Baldwin R, Morley M, Burns A, et al. A randomised controlled trial to test the feasibility of a

- collaborative care model for the management of depression in older people. *Br J Gen Pract* 2007;57:364-70.
- 2 Unutzer J, Katon W, Callahan CM, Williams JW Jr, Hunkeler E, Harpole L, et al. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA* 2002;288:2836-45.
- 3 Young AS, Klap R, Sherbourne CD, Wells KB. The quality of care for depressive and anxiety disorders in the United States. *Arch Gen Psychiatry* 2001;58:55-61.
- 4 Crabb R, Hunsley J. Utilization of mental health care services among older adults with depression. *J Clin Psychol* 2006;62:299-312.
- 5 Licht-Strunk E, Beekman ATF, de Haan M, van Marwijk HWJ. The prognosis of undetected depression in older general practice patients. A one year follow-up study. *J Affect Disord* 2008; Jul 14 [Epub ahead of print].
- 6 Sheikh JJ. Geriatric depression scale: recent evidence and development of a shorter version. *Clin Gerontol* 1986;5:165-72.
- 7 Burke WJ, Roccaforte WH, Wengel SP. The short form of the geriatric depression scale: a comparison with the 30-item form. *J Geriatr Psychiatry Neurol* 1991;4:173-8.
- 8 Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV, III, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *JAMA* 1994;272:1749-56.
- 9 Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134:382-9.
- 10 Hawley CJ, Gale TM, Sivakumaran T. Defining remission by cut off score on the MADRS: selecting the optimal value. *J Affect Disord* 2002;72:177-84.
- 11 Kriegsman DM, Penninx BW, van Eijk JT, Boeke AJ, Deeg DJ. Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy. *J Clin Epidemiol* 1996;49:1407-17.
- 12 Folstein MF, Folstein SE, McHugh PR. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189-98.
- 13 Kafonek S, Ettlinger WH, Roca R, Kittner S, Taylor N, German PS. Instruments for screening for depression and dementia in a long-term care facility. *J Am Geriatr Soc* 1989;37:29-34.
- 14 Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health diagnostic interview schedule. Its history, characteristics, and validity. *Arch Gen Psychiatry* 1981;38:381-9.
- 15 Ware JE, Kosinski M, Keller SD. *SF-36 physical and mental summary scales: a users' manual*, 2nd ed. Boston, MA: Health Institute, 1994.
- 16 Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
- 17 Cole MG, McCusker J, Ciampi A, Windholz S, Latimer E, Belzile E. The prognosis of major and minor depression in older medical inpatients. *Am J Geriatr Psychiatry* 2006;14:966-75.
- 18 Cole MG, Bellavance F. The prognosis of depression in old age. *Am J Geriatr Psychiatry* 1997;5:4-14.
- 19 Lenze EJ, Schulz R, Martire LM, Zdaniuk B, Glass T, Kop WJ, et al. The course of functional decline in older people with persistently elevated depressive symptoms: longitudinal findings from the Cardiovascular Health Study. *J Am Geriatr Soc* 2005;53:575.
- 20 Verhaak PF. Analysis of referrals of mental health problems by general practitioners. *Br J Gen Pract* 1993;43:203-8.

**Accepted:** 20 October 2008

## What types of article does the *BMJ* consider?

We are delighted to receive articles for publication—from doctors and others—on the clinical, scientific, social, political, and economic factors affecting health. We give priority to articles that will help doctors to make better decisions. Please see our advice to authors at <http://resources.bmj.com/bmj/authors>, and if you would like to submit an article do so via our online editorial office at <http://submit.bmj.com>.

All original research articles are submitted, although we may invite submission (without promising acceptance) if we come across research being presented at conferences, if

we see it in abstract form, or if the authors make an inquiry about the suitability of their work before submission.

We are also pleased to consider submitted articles for sections which carry a mix of commissioned and submitted articles—editorials, analysis, clinical review, practice, fillers, and Career Focus. Please follow the specific advice on each of these article types (see <http://resources.bmj.com/bmj/authors/types-of-article>) before submitting your article. Some types of article—news, features, observations, head to head, views and reviews—are commissioned by the editors.