Open Access: free full text available online
For the full versions of these articles see bmj.com

# China's excess males, sex selective abortion, and one child policy: analysis of data from 2005 national intercensus survey

Wei Xing Zhu,[1] Li Lu,[2] Therese Hesketh[3]

EDITORIAL by Liu and Zhang

[1]College of Law, Political Science and Public Administration, Zhejiang Normal University, Jinhua, Zhejiang 310347, China
[2]Institute of Social and Family Medicine, Zhejiang University, Hangzhou 310016, China
[3]Centre for International Health and Development, University College London, London WC1N 1EH
Correspondence to: T Hesketh
t.hesketh@ich.ucl.ac.uk

Cite this as: *BMJ* 2009;338:b1211
doi:10.1136/bmj.b1211

## ABSTRACT

**Objectives** To elucidate current trends and geographical patterns in the sex ratio at birth and in the population aged under 20 in China and to determine the roles played by sex selective abortion and the one child policy.

**Design** Analysis of household based cross sectional population survey done in November 2005.

**Setting** All of China's 2861 counties.

**Population** 1% of the total population, selected to be broadly representative of the total.

**Main outcome measure** Sex ratio defined as males per 100 females.

**Results** 4 764 512 people under the age of 20 were included. Overall sex ratios were high across all age groups and residency types, but they were highest in the 1-4 years age group, peaking at 126 (95% confidence interval 125 to 126) in rural areas. Six provinces had sex ratios of over 130 in the 1-4 age group. The sex ratio at birth was close to normal for first order births but rose steeply for second order births, especially in rural areas, where it reached 146 (143 to 149). Nine provinces had ratios of over 160 for second order births. The highest sex ratios were seen in provinces that allow rural inhabitants a second child if the first is a girl. Sex selective abortion accounts for almost all the excess males. One particular variant of the one child policy, which allows a second child if the first is a girl, leads to the highest sex ratios.

**Conclusions** In 2005 males under the age of 20 exceeded females by more than 32 million in China, and more than 1.1 million excess births of boys occurred. China will see very high and steadily worsening sex ratios in the reproductive age group over the next two decades. Enforcing the existing ban on sex selective abortion could lead to normalisation of the ratios.

## INTRODUCTION

In the absence of intervention, the sex ratio at birth is consistent across populations at between 103 and 107 boys born for every 100 girls.[1 2] Higher early mortality among boys ensures a ratio of close to 100 in the reproductive years. However, in many countries, mainly in South and East Asia, the sex ratio deviates from this norm because of the tradition of preference for sons.[3] Historically, preference for sons has been manifest postnatally through female infanticide and the neglect and abandonment of girls.[4] However, since the early 1980s selection for males prenatally with ultrasonographic sex determination and sex selective abortion has been possible. The highest sex ratios are seen in countries with a combination of preference for sons, easy access to sex selective technology, and a low fertility rate.[5] In the era of the one child policy the fact that the problem of excess males in China seems to outstrip that of all other countries is perhaps no surprise.[6 7]

Some of the evidence for this sex imbalance in China has been challenged, because accurate population based figures have been difficult to obtain.[8 9] Births classified as "illegal," violating the one child policy, may be concealed to avoid penalties.[10 11] Under-reporting of births of girls may be more common in this context.[12 13] However, if girls are not reported at birth, they are likely to filter into the statistics later, as registration is necessary for immunisation or to start school.[14] Therefore, examining the sex ratio across different age bands provides a more accurate picture.

The objectives of this study were to elucidate current trends and geographical patterns in the sex ratio at birth and in the population under the age of 20 in China and to explore the role played by sex selective abortion and the one child policy in the sex imbalance.

## METHODS

We analysed data from the intercensus survey of 2005, which was carried out on a representative 1% of the total population in November 2005. The survey covered all of China's 2861 counties. Only data for the under 20 age group are reported here.

The major outcome variable is the sex ratio, defined as males per 100 females. We calculated the excess of males for all age groups by using an average of the mean sex ratios from 13 countries that have normal secondary sex ratios and little or no sex preference.[15] These were 105 for the 1-9 age group and 104 for the 10-19 age group.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;338:b1211

Table 1 | Sex ratio (95% confidence interval) by age and residence, under 20 year olds

| Residence | No (%) | Age (year of birth) | | | | |
|---|---|---|---|---|---|---|
| | | <1 year (2004-5) (n=182 393) | 1-4 years (2000-4) (n=724 709) | 5-9 years (1995-9) (n=1 060 664) | 10-14 years (1990-4) (n=1 353 263) | 15-19 years (1985-9) (n=1 443 483) |
| All | 4 764 512 | 119 (119 to 120) | 124 (123 to 124) | 119 (119 to 120) | 114 (113 to 114) | 108 (108 to 109) |
| Urban* | 1 073 229 (23) | 114 (112 to 115) | 116 (115 to 117) | 116 (115 to 117) | 112 (111 to 114) | 101 (100 to 103) |
| Town† | 813 386 (17) | 117 (115 to 119) | 122 (120 to 124) | 121 (120 to 122) | 116 (115 to 117) | 109 (107 to 111) |
| Rural‡ | 2 877 897 (60) | 122 (121 to 122) | 126 (125 to 126) | 120 (120 to 121) | 114 (113 to 114) | 111 (110 to 111) |

*Area with more than 100 000 non-agricultural population.
†Population of at least 20 000, where non-agricultural population exceeds 10%.
‡More than 90% agricultural workers.

## RESULTS

The survey counted 4 764 512 people under the age of 20: 1 073 229 (22%) urban residents, 813 386 (17%) town residents, and 2 877 897 (60%) rural inhabitants. In the 12 months before the study 161 109 births were reported: 23% in urban areas, 17% in towns, and 60% in rural areas. First order births accounted for 63% of the total, second order for 32%, third order for 4.3%, and fourth or higher order for 1%.

### Under 20 sex ratio

Table 1 shows the sex ratio by age group and type of residency. Sex ratios were consistently higher than normal across residency type and all age groups except for urban 15-19 year olds. Sex ratios peaked in the 1-4 age group; the highest was 126 (95% confidence interval 125 to 126) in rural areas. Table 2 shows the sex ratio by age group for all provinces. Only two provinces, Tibet and Xinjiang, had sex ratios within normal limits across the age range. Two provinces, Jiangxi and Henan, had ratios of over 140 in the 1-4 age group; four provinces—Anhui, Guangdong, Hunan, and Hainan—had ratios of over 130; and seven provinces had ratios between 120 and 129. The provinces with the highest sex ratios are clustered together in the central-southern region. The excess of males increased from 5.1% (n=142 634) in the cohort born between 1986 and 1995 to 9.4% (n=184 970) in the cohort born between 1996 and 2005 across the whole country.

### Sex ratio at birth

The total sex ratio at birth for the 12 months to October 2005 was 120 (119 to 121) for the whole sample (see bmj.com), with a gradient between urban (115, 113 to 117), town (120, 118 to 122), and rural (123, 121 to 124) areas. This equates to 11 320 excess boys born for the year for the whole sample. These overall figures conceal dramatic differences in sex ratio at birth by birth order. The sex ratio at birth for first order births was slightly high in cities and towns but was within normal limits in rural areas. However, the ratio rose very steeply for second and higher order births in cities 138 (132 to 144), towns 137 (131 to 143), and rural areas 146 (143 to 149), although the numbers of second order births in cities were low. For third births, the sex ratio rose to over 200 in four provinces, although third births accounted for only 4.3% of the total.

## DISCUSSION

The findings paint a discouraging picture of very high and increasing sex ratios in the reproductive age group in China for the next two decades. The sex ratio increased steadily from 108 in the cohort born between 1985 and 1989 to 124 in the 2000 to 2004 cohort. However, the ratio then declined to 119 for the 2005 cohort, perhaps indicating the beginning of a reduction in sex ratios for the future. Sex ratios were outside the normal range for almost all age groups in almost all provinces. The sex ratios rose dramatically between first and second order births, with very high sex ratios for the very few higher order births. The highest ratios were seen in the centre and south of the country. Extrapolating from this 1% sample to the whole country, we estimate that an excess of 1 132 000 boys were born in the 12 months to October 2005 and that an excess of 32 706 400 males under the age of 20 existed in the whole of China at that time, 18 497 000 of them under the age of 10.

This is the most recent nationwide demographic survey in China. A large survey aiming to represent 1% of the total population obviously has some limitations. Complete coverage of households and inhabitants is impossible on such a large scale. Furthermore, extrapolation to the whole population from a 1% sample should be done with caution. The small sample size at provincial level in some age bands and for second and higher order births leads to wide confidence intervals, illustrating the uncertainty around these figures. However, the overall credibility of the data is increased by the high sex ratios in older age groups, for which concealment and under-reporting of girls would be difficult, and by the number of births counted for the 12 months to October 2005 (161 109), which matches the estimate of 16 million births a year from other sources.[16]

**Table 2** | Sex ratios (95% confidence intervals) by age group and province, and excess males

| Region and province (population in millions) | No | Age group | | | | | Excess males <10 years (%) | Excess males 10-20 years (%) |
|---|---|---|---|---|---|---|---|---|
| | | <1 year | 1-4 years | 5-9 years | 10-14 years | 15-19 years | | |
| All | 4 764 512 | 119 (119 to 120) | 124 (123 to 124) | 119 (119 to 120) | 114 (114 to 115) | 108 (107 to 108) | 9.4 | 5.1 |
| North: | | | | | | | | |
| Beijing (14) | 35 657 | 114 (102 to 127) | 112 (105 to 118) | 110 (106 to 115) | 104 (100 to 109) | 111 (108 to 115) | 5.4 | 4.1 |
| Tianjin (10) | 29 170 | 114 (100 to 130) | 118 (111 to 126) | 115 (109 to 122) | 110 (106 to 115) | 99 (96 to 103) | 7.5 | 1.7 |
| Hebei (65) | 250 933 | 120 (116 to 125) | 122 (120 to 125) | 115 (113 to 117) | 111 (109 to 113) | 104 (102 to 105) | 8.4 | 3.2 |
| Shanxi (32) | 131 763 | 116 (110 to 123) | 112 (109 to 116) | 109 (107 to 111) | 109 (107 to 111) | 106 (104 to 108) | 5.1 | 3.6 |
| Inner Mongolia (23) | 76 693 | 114 (107 to 122) | 107 (103 to 111) | 109 (106 to 113) | 110 (107 to 113) | 107 (104 to 109) | 4.3 | 4.0 |
| Northeast: | | | | | | | | |
| Liaoning (42) | 118 018 | 113 (106 to 120) | 114 (111 to 117) | 111 (108 to 114) | 110 (108 to 113) | 106 (103 to 108) | 5.6 | 3.7 |
| Jilin (26) | 81 133 | 113 (105 to 121) | 112 (108 to 116) | 113 (110 to 117 | 110 (107 to 113) | 107 (104 to 109) | 5.9 | 3.8 |
| Heilongjiang (38) | 112 057 | 109 (103 to 116) | 111 (108 to 115) | 107 (104 to 110) | 108 (105 to 110) | 107 (105 to 109) | 4.1 | 3.4 |
| East: | | | | | | | | |
| Shanghai (17) | 37 406 | 117 (106 to 128) | 109 (103 to 115) | 111 (105 to 116) | 108 (103 to 113) | 98 (95 to 101) | 4.9 | 0.4 |
| Jiangsu (71) | 230 997 | 125 (120 to 131) | 123 (120 to 126) | 121 (119 to 124) | 118 (116 to 120) | 105 (103 to 106) | 10 | 5.0 |
| Zhejiang (43) | 150 125 | 114 (109 to 119) | 113 (111 to 116) | 113 (111 to 116) | 113 (111 to 116) | 108 (106 to 110) | 6.3 | 5.0 |
| Anhui (61) | 256 350 | 131 (126 to 137) | 138 (135 to 141) | 124 (122 to 127) | 115 (114 to 117) | 107 (106 to 109) | 12.9 | 5.5 |
| Fujian (33) | 129 146 | 122 (116 to 129) | 119 (116 to 122) | 124 (121 to 127) | 118 (116 to 121) | 101 (99 to 103) | 9.8 | 4.2 |
| Jiangxi (41) | 186 198 | 129 (121 to 137) | 143 (140 to 146) | 130 (128 to 133) | 118 (116 to 120) | 114 (112 to 116) | 14.8 | 7.6 |
| Shandong (91) | 303 287 | 114 (110 to 118) | 116 (114 to 118) | 116 (114 to 118) | 115 (114 to 117) | 106 (105 to 108) | 7.4 | 4.5 |
| Central: | | | | | | | | |
| Henan (95) | 386 594 | 122 (118 to 126) | 142 (140 to 144) | 131 (129 to 133) | 119 (118 to 121) | 110 (109 to 111) | 14.4 | 6.7 |
| Hubei (58) | 208 230 | 128 (122 to 135) | 129 (126 to 133) | 129 (126 to 132) | 120 (118 to 121) | 118 (117 to 120) | 12.6 | 8.7 |
| Hunan (64) | 232 938 | 122 (117 to 127) | 133 (130 to 136) | 122 (120 to 124) | 115 (113 to 117) | 112 (110 to 113) | 11.6 | 6.1 |
| Guangdong (70) | 384 845 | 119 (115 to 123) | 133 (131 to135) | 127 (126 to 129) | 115 (113 to 116) | 96 (95 to 97) | 12.3 | 2.2 |
| Guangxi (46) | 199 776 | 121 (116 to 126) | 122 (120 to 125) | 127 (125 to 130) | 122 (120 to 124) | 123 (121 to 125) | 11.0 | 10.2 |
| Hainan (7.5) | 36 427 | 123 (111 to 136) | 134 (127 to 141) | 135 (129 to141) | 120 (113 to 127) | 123 (118 to 128) | 6..2 | 9.3 |
| Southwest: | | | | | | | | |
| Chongqing (31) | 100 070 | 112 (104 to 120) | 119 (115 to 123) | 117 (114 to 120) | 113 (110 to 115) | 114 (111 to 117) | 7.9 | 6.2 |
| Sichuan (84) | 311 530 | 115 (110 to 119) | 116 (114 to 118) | 114 (112 to 116) | 111 (110 to 113) | 108 (106 to 109) | 6.8 | 4.6 |
| Guizhou (37) | 178 547 | 128 (112 to 134) | 127 (124 to 130) | 115 (113 to 117) | 112 (110 to 114) | 119 (117 to 122) | 9.0 | 7.2 |
| Yunnan (41) | 189 774 | 113 (108 to 118) | 115 (113 to117) | 112 (110 to 114) | 112 (110 to 114) | 112 (110 to 114) | 6.2 | 5.9 |
| Tibet (3) | 13 764 | 102 (88 to 120) | 104 (96 to113) | 105 (98 to 112) | 102 (95 to 108) | 105 (98 to 112) | 2.1 | 1.6 |
| Northwest: | | | | | | | | |
| Shaanxi (36) | 141 904 | 134 (126 to 143) | 125 (121 to 129) | 123 (120 to 126) | 117 (115 to 119) | 112 (110 to 114) | 10.9 | 6.6 |
| Gansu (25) | 112 399 | 116 (109 to 124) | 120 (117 to 124) | 116 (113 to 119) | 109 (107 to 111) | 106 (104 to 109) | 3.8 | 0.1 |
| Qinghai (5) | 23 483 | 117 (103 to 133) | 111 (105 to 118) | 109 (104 to 115) | 104 (98 to 110) | 100 (95 to 106) | 5.0 | 0.1 |
| Ningxia (5) | 27 373 | 107 (96 to 119) | 112 (106 to 119) | 108 (103 to 113) | 106 (101 to 110) | 104 (99 to 109) | 4.5 | 2.3 |
| Xinjiang (17) | 87 919 | 105 (99 to 112) | 106 (102 to 109) | 104 (101 to 107) | 106 (103 to 108) | 107 (104 to 110) | 2.2 | 3 |

### Role of sex selective abortion

The precise role of sex selective abortion in the sex imbalance has been unclear, not least because the practice is illegal in China and obtaining reliable figures is difficult. If under-registration of girls accounted for most of the excess births of boys, then sex ratios would fall from birth through early childhood, as girls are required to be registered for immunisation and school entry.[14] Our finding that the sex ratios for the 1-4 year old cohort are higher

## WHAT IS ALREADY KNOWN ON THIS TOPIC

The reported sex ratio (males per 100 females) in China is high, but accurate population based figures for actual sex ratios have been notoriously difficult to obtain

The role of sex selective abortion and the influence of the one child policy on the sex imbalance have been unclear

## WHAT THIS STUDY ADDS

China will see very high and steadily worsening sex ratios in the reproductive age group for the next two decades

Sex selective abortion accounts for almost all the excess males

One particular variant of the one child policy leads to the highest sex ratios

than those at birth and in infancy tends to refute this hypothesis. Comparison between each cohort specific sex ratio and the corresponding sex ratio at birth from previous census data also lends support to the assertion that under-registration of girls is not a major contributor to high sex ratios at birth.[17 18] Infanticide is another possible explanation for girls missing at birth, but is widely acknowledged to be very rare now.[19-22]

The dramatic increase in sex ratio with second births shows that couples are selecting to ensure a boy, the so called "at least one son practice."[14] In urban areas where few couples are allowed a second child, the high sex ratio for first order births suggests some sex selection occurring with the only child. Finally, the steady rise in sex ratios across the birth cohorts since 1986 mirrors the increasing availability of ultrasonography over that period.[23 24]

### Role of one child policy

The relation between the sex ratio and the one child policy is a complex one. The policy is implemented differently across the country (see bmj.com), and our data suggest that the sex ratio is related to the way in which the policy is implemented.[14] Whereas in most cities only one child is allowed, three main variants of the policy exist in rural areas. Our data show that the type 2 variant, which allows couples a second child after a girl, results in the highest sex ratios for second order births and the overall highest sex ratios.

Medium sex ratios were most common in the strict type 1 provinces where 40% of couples are allowed a second child but generally only if the first is a girl. However, these provinces are also wealthier, levels of education are higher, and traditional values of preference for sons are changing.[25] The lowest ratios are seen in the type 3, most permissive, provinces. However, these provinces are sparsely populated and poor, inhabited partly by ethnic groups who are generally less inclined to prefer sons and less accepting of abortion.[24]

The policy implications are clear. Changing the regulations in force in type 2 provinces, which permit most couples a second child after a female birth, could help to reduce the sex ratio.

**Contributors:** All the authors participated in the analysis and in preparing the tables and saw and approved the final version of the paper. ZWX is the guarantor.

**Funding:** This study was funded through a China-UK excellence fellowship for TH from the Department of Innovation, Universities and Skills.

**Competing interests:** None declared.

**Ethical approval:** Not needed.

1 James WH. The human sex ratio. Part 1: a review of the literature. *Human Biology* 1987;59:721-5.
2 Teitelbaum M. Factors affecting the sex ratio in large populations. *J Bio Sci* 1970:2:61-71.
3 Arnold F. The effect of son preference on fertility and family planning: empirical evidence. *Popul Bull UN* 1987;23:44-55.
4 Klasen S, Wink C. A turning point in gender bias in mortality? An update on the number of missing women. *Popul Dev Rev* 2002;28:285-312.
5 Park CB, Cho NH. Consequences of son preference in a low fertility society: imbalance of the sex ratio at birth in Korea. *Popul Dev Rev* 1995;21:59-84.
6 Hesketh T, Zhu WX. Abnormal sex ratios in human populations: causes and consequences. *Proc Natl Acad Sci USA* 2006;103:13271-5.
7 Gu B, Roy K. Sex ratio at birth in China, with reference to other areas in East Asia: what we know. *Asia Pac Popul J* 1995;10:17-42.
8 Short SE, Zhai FY. Looking locally at China's one-child policy. *Stud Fam Plan* 1998;29,4:373-87.
9 Merli MG, Raftery AE. Are births underreported in rural china? Manipulation of statistical records in response to China's population policies. *Demography* 2000;37:109-26.
10 Banister J. Shortage of girls in China today. *J Popul Res* 2004;21:19-45.
11 Johansson S, Nygren O. The missing girls of China: a new demographic account. *Popul Dev Rev* 1991;17:35-51.
12 Zeng Y, Tu P, Gu B, Xu L, Li B, Li Y. Causes and implications of the recent increase in the reported sex ratio at birth in China. *Popul Dev Rev* 1993;19:283-302.
13 Coale A. Excess female mortality and the balance of the sexes in the population: an estimate of the number of missing females. *Popul Dev Rev* 1991;17:518.
14 Attane I. China's family planning policy: an overview of its past and future. *Stud Fam Plan* 2002;33:103-13.
15 Ulizzi L, Astolfi P, Zonta LA. Sex ratio at reproductive age: changes over the last century in the Italian population. *Hum Biol* 2001;73:121-8.
16 China-Profile. Facts, figures, and analyses. www.china-profile.com.
17 Li S. Imbalanced sex ratio at birth and comprehensive intervention in China. Presentation at 4th Asia Pacific Conference on Reproductive and Sexual Health and Rights, Hyderabad, 29-31 October 2007. Available at www.unfpa.org/gender/docs/studies/china.pdf.
18 China Population Information and Research Centre. Basic population data of China: 1949-2000. www.cpirc.org.cn.
19 Hesketh T. Zhu WX. The one child family policy: the good, the bad and the ugly. *BMJ* 1997;314:1685-7.
20 Wu ZC, Viisainen K, Wang Y, Hemminki E. Perinatal mortality in rural China: retrospective cohort study. *BMJ* 2003;327:1319-22.
21 Hemminki E, Wu ZC, Cao GY. Illegal births and legal abortions—the case of China. *Reprod Health Matters* 2005;2:5.
22 Li S, Zhu C, Feldman M. Gender differences in child survival in contemporary rural China: a county study. *J Biosoc Sci* 2004;36:83-109.
23 Chu JH. Prenatal sex determination and sex-selective abortion in rural central China. *Popul Dev Rev* 2001;27:259-81.
24 Li R. An analysis of the sex ratio at birth in impoverished areas in China. *Chin J Popul Sci* 1998;10:65-73.
25 Winkler EA. Chinese reproductive policy at the turn of the millennium: dynamic stability. *Popul Dev Rev* 2002;28:379-418.

**Accepted:** 27 November 2008

# Frailty and type of death among older adults in China: prospective cohort study

Matthew E Dupre,[1] Danan Gu,[2] David F Warner,[3] Zeng Yi[4,5]

EDITORIAL by Feng

[1]Department of Sociology and Center for the Study of Aging and Human Development, Duke University Medical Center, Durham, NC 27710, USA
[2]Urban Studies and Planning, Portland State University, 506 SW Mill Street 570M, Portland, OR 97207, USA
[3]Department of Sociology, Case Western Reserve University, Cleveland, OH
[4]Center for the Study of Aging and Human Development, Duke University Medical Center, Durham, NC
[5]China Center for Economic Research, Peking University, Beijing, China
Correspondence to: M E Dupre med11@geri.duke.edu or D Gu gudanan@yahoo.com

Cite this as: *BMJ* 2009;338:b1175
doi:10.1136/bmj.b1175

## ABSTRACT

**Objective** To examine the association between frailty and type of death among the world's largest oldest-old population in China.

**Design** Prospective cohort study.

**Setting** 2002 and 2005 waves of the Chinese longitudinal healthy longevity survey carried out in 22 provinces throughout China.

**Participants** 13 717 older adults (aged ≥65).

**Main outcome measures** Type of death, categorised as being bedridden for fewer than 30 days with or without suffering and being bedridden for 30 or more days with or without suffering.

**Results** Multinomial analyses showed that higher levels of frailty significantly increased the relative risk ratios of mortality for all types of death. Of those with the highest levels of frailty, men were most likely to experience 30 or more bedridden days with suffering before death (relative risk ratio 8.70, 95% confidence interval 6.31 to 12.00) and women 30 or more bedridden days with no suffering (11.53, 17.84 to 16.96). Regardless of frailty, centenarians and nonagenarians were most likely to experience fewer than 30 bedridden days with no suffering, whereas those aged 65-79 and 80-89 were more likely to experience fewer than 30 bedridden days with suffering. Adjusting for compositional differences had little impact on the link between frailty and type of death for both sexes and age groups.

**Conclusions** The association between frailty and type of death differs by sex and age. Health scholars and clinical practitioners should consider age and sex differences in frailty to develop more effective measures to reduce preventable suffering before death.

## INTRODUCTION

Evidence suggests that frailty is a valid and reliable proxy of biological age, providing a robust measure of the balance between health assets and deficits across a variety of dimensions.[1-5] Quantifying frailty using a "frailty index" is increasingly being recognised as important for identifying population differences in ageing and as a tool for monitoring susceptibility to disease and death.[5-9] On the basis of research from cross sectional and non-representative studies, consensus is growing that the extent of suffering (pain and discomfort) and number of bedridden days are basic dimensions underlying the quality of death in late life.[10-12]

We carried out a prospective cohort analysis of the association between frailty and type of death among adults aged 65 to 109 in China and investigated whether differences in mortality varied by sex and age.

## METHODS

The Chinese longitudinal healthy longevity survey collected data on one of the largest samples of people in the oldest-old age group (≥80 years) in the world. Information was collected on a range of variables (see bmj.com). Older adults from 22 provinces were first interviewed in 1998, with follow-up interviews in 2000, 2002, and 2005. We utilised data from the 2002 and 2005 waves of the survey. In the 2002 wave 15 919 participants aged 65-109 were interviewed. Of these interviewees, 8090 (50.8%) were reinterviewed in 2005 and 5627 (35.3%) died before follow-up. After exclusion of losses to follow-up, the sample comprised 13 717 participants.

We used an objective indicator of physical failure with a subjective indicator of suffering before death to categorise types of death between the two surveys.[11 12] Firstly, we dichotomised bedridden days before death into fewer than 30 and 30 or more, then dichotomised the subjective painfulness of death reported by the next of kin (peaceful *v* non-peaceful). The four types of death were less than 30 bedridden days with or with no suffering and 30 or more bedridden days with or with no suffering. The reference category in the multinomial regression models was survival over the three years.

We constructed a frailty index using 39 variables. Each item was assigned a value of 1 in the presence of a deficit (otherwise 0), or 2 for people with two or more serious conditions that led to hospital stay or a period of confinement in bed.[6] The frailty index was a summary of all deficits, divided by the number of possible deficits. We split the index into fourths for each sex to account for non-linear relations between levels of frailty and type of death.

To obtain robust estimates we also adjusted analyses for several previously identified confounders.[13] All confounders except age were dichotomised. Demographic measures included age groups 65-79 (reference), 80-89, 90-99, and 100 and older, people from non-Han ethnic minorities, and urban dwellers. Measures of socioeconomic status included education, primary lifetime occupation as a white collar worker, economic independence, good economic standing, and being in receipt of adequate drugs for any illnesses. Social contact and support measures included marital

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;338:b1175

status, close proximity to children, and religious activity. Measures for health practices included exercising on a regular basis and having smoked in the past five years.

### Statistical analysis

We computed sample distributions of the study variables separately by sex and frailty. To test differences in the distributions of frailty for dichotomous variables we used Kendall's τ tests and for categorical variables Pearson's $\chi^2$ tests. Multinomial logistic regression models were used to estimate the relative risk ratios and 95% confidence intervals associated with frailty and types of death. We used two sets of nested regression models to adjust for the confounders. The first set of analyses tested the effects of frailty by sex while adjusting for basic personal information. In the second set of analyses we included socioeconomic status, social contact and support, and health practices. We then computed the predicted proportions of the types of death across age and frailty for both sexes.

## RESULTS

Among those in the lowest fourth (least frail) for frailty, 84.9% of men and 86.1% of women survived to 2005 compared with just over 25% of men and women in the highest fourth (see bmj.com). Among decedents, about 25% of men and women had fewer than 30 bedridden days before death, although women were less likely to suffer than men. A greater proportion of women than men were bedridden for 30 or more days with or without suffering before death. Those with the most frailty had the highest rates for all types of death. Men with the most frailty were more likely to suffer before death (17.9% <30 bedridden days, 15.7% ≥30 bedridden days) compared with women, who were more likely not to suffer before death (25.4% <30 bedridden days, 21.1% ≥30 bedridden days), especially after at least 30 bedridden days.

There was no difference between sex, frailty, and urban residence. At nearly all levels of frailty men had higher socioeconomic status, were more likely to be married, took regular exercise, and had smoked in the past five years compared with women. Women were more likely to live close to their children and engage in religious activities.

For both sexes the proportion of participants who did not suffer before death increased with age and the proportion who suffered decreased with age (see bmj.com). Across levels of frailty, the most pronounced patterns were the precipitous declines in the numbers of participants who experienced fewer than

**Relative risk ratios from multinomial logistic regression models of type of death by sex and frailty of participants in Chinese longitudinal healthy longevity survey**

| | Men | | Women | |
|---|---|---|---|---|
| Variables | Model 1*: relative risk ratio (95% CI) | Model 2†: relative risk ratio (95% CI) | Model 1*: relative risk ratio (95% CI) | Model 2†: relative risk ratio (95% CI) |
| **<30 bedridden days with no suffering *v* survival** | | | | |
| Frailty index fourth: | | | | |
| First (least frail) | 1.00 (reference) | 1.00 (reference) | 1.00 (reference) | 1.00 (reference) |
| Second | 1.21 (0.90 to 1.64) | 1.19 (0.88 to 1.62) | 1.83 (1.40 to 2.39) | 1.77 (1.35 to 2.32) |
| Third | 1.90 (1.48 to 2.44) | 1.87 (1.45 to 2.41) | 2.98 (2.26 to 3.94) | 2.81 (2.12 to 3.73) |
| Fourth (most frail) | 4.02 (3.44 to 6.43) | 4.16 (3.16 to 5.47) | 5.67 (4.29 to 7.49) | 5.28 (3.95 to 7.06) |
| **<30 bedridden days with suffering *v* survival** | | | | |
| Frailty index fourth: | | | | |
| First | 1.00 (reference) | 1.00 (reference) | 1.00 (reference) | 1.00 (reference) |
| Second | 1.14 (0.84 to 1.54) | 1.11 (0.82 to 1.50) | 1.85 (1.43 to 2.40) | 1.77 (1.37 to 2.30) |
| Third | 1.96 (1.53 to 2.51) | 1.82 (1.41 to 2.34) | 2.19 (1.65 to 2.91) | 2.02 (1.52 to 2.68) |
| Fourth | 4.42 (3.43 to 5.70) | 3.87 (1.96 to 3.32) | 4.27 (3.22 to 5.65) | 3.87 (2.90 to 5.16) |
| **≥30 bedridden days with no suffering *v* survival** | | | | |
| Frailty index fourth: | | | | |
| First | 1.00 (reference) | 1.00 (reference) | 1.00 (reference) | 1.00 (reference) |
| Second | 1.47 (0.93 to 2.36) | 1.45 (0.90 to 2.31) | 2.13 (1.44 to 3.17) | 2.06 (1.39 to 3.06) |
| Third | 2.32 (1.57 to 3.41) | 2.20 (1.49 to 3.24) | 4.29 (2.90 to 6.35) | 4.00 (2.68 to 5.96) |
| Fourth | 7.61 (5.21 to 11.13) | 6.67 (4.51 to 9.85) | 11.53 (7.84 to 16.96) | 10.53 (7.06 to 15.70) |
| **≥30 bedridden days with suffering *v* survival** | | | | |
| Frailty index fourth: | | | | |
| First | 1.00 (reference) | 1.00 (reference) | 1.00 (reference) | 1.00 (reference) |
| Second | 1.62 (1.10 to 2.39) | 1.58 (1.07 to 2.33) | 1.69 (1.22 to 2.35) | 1.61 (1.16 to 2.24) |
| Third | 2.36 (1.70 to 3.27) | 2.23 (1.60 to 3.09) | 2.97 (2.10 to 4.19) | 2.66 (1.87 to 3.76) |
| Fourth | 8.70 (6.31 to 12.00) | 7.75 (5.54 to 10.83) | 6.98 (5.00 to 9.75) | 5.96 (4.23 to 8.39) |

*Adjusted for age, ethnicity, and urban residence.
†Adjusted for age, ethnicity, urban residence, socioeconomic status, social contact and support, and health practices.

30 bedridden days with suffering and increases in the number of participants who experienced 30 or more bedridden days with no suffering, particularly among women.

The table presents the results from the multinomial analyses for the associations between frailty and type of death for both sexes. The relative risk ratios for model 1, adjusted for age, ethnicity, and urban residence, and for model 2 further adjusted for several confounders showed that increased frailty was associated with higher risks of death, regardless of type, and was especially pronounced for those with high levels of frailty. Adjusting for differences in socioeconomic status, social contact or support, and health practices in model 2 had little impact on the patterning or magnitude of the frailty effects and provided strong evidence that frailty is a robust predictor of type of death in the Chinese longitudinal healthy longevity survey.

The effect of frailty differed between the sexes. Overall, the relative risk ratios of dying across all categories of death (except for ≥30 bedridden days with suffering) related to increases in frailty were higher among women, presumably due to their older age and better physiological resilience compared with men. Frailty increased the risk of dying with some suffering for men and increased the risk of being bedridden for fewer than 30 days or for 30 or more days with no suffering for women.

The most common type of death for both sexes was 30 or fewer bedridden days with no suffering (see bmj.com). Given the same level of frailty and adjusting for confounders, the proportion of men who experienced 30 or more bedridden days with suffering was greater than the proportion of men bedridden with no suffering; however, the opposite was found among women. Analyses further indicated that the oldest participants (≥90 years) were more likely to experience fewer bedridden days with no suffering than younger participants. Fewer than 30 bedridden days with no suffering were more common than 30 or more bedridden days with no suffering, whereas fewer than 30 bedridden days with suffering were more common than 30 or more bedridden days with suffering for each sex and age group. Moreover, being bedridden for fewer than 30 days or for 30 or more days with suffering was less prevalent with increasing age compared with being bedridden for fewer than 30 days or 30 or more days with no suffering, which was especially apparent for the most frail participants.

Higher levels of frailty among men were associated with greater proportions of death with 30 or more bedridden days and with lesser proportions of death with fewer than 30 bedridden days for every age group. Among women, being bedridden for fewer than 30 days with suffering decreased and being bedridden for 30 or more days with no suffering increased across all levels of frailty within each age group; however, differences were modest for experiencing fewer than 30 bedridden days with no suffering and 30 or more bedridden days with suffering across levels of frailty.

## DISCUSSION

Results based on a prospective cohort of adults aged 65 to 109 in China showed that higher levels of frailty increased the risk of experiencing all four types of death that we examined; bedridden for fewer than 30 days with or with no suffering or bedridden for 30 or more days with or with no suffering. People with greater frailty were more likely to experience more bedridden days before death than those with less frailty. The increasing effects of frailty on suffering were primarily limited to those who experienced 30 or more bedridden days before death. The association between frailty and type of death differed by sex and age, and adjusting for differences in socioeconomic status, social support, and health practices had almost no impact on the relations. Overall, the results provided strong evidence that cumulative deficits, quantified with a frailty index, influenced not only the likelihood of dying but also the quality of life before death.[8 14]

Overall, compared with men, women exhibited higher risks of experiencing one of the four types of death as levels of frailty increased, owing to their older age and therefore increased physiological frailty. Evidence from Western nations and China concur that older women are often in poorer health yet live longer than men.[15 16] The implications of this finding are perhaps more pronounced in developing countries such as China. In many cases, Chinese women bear a disproportionate amount of the care giving to spouses and their respective parents. These women also have the fewest economic and familial resources and exhibit the most frailty as they age. Despite these disadvantages for men the risks increased most noticeably for deaths with suffering, whereas for women the risks increased much more for every level of frailty for deaths with no suffering.

The lack of suffering among extremely old people may be due to the precipitous withdrawal of life sustaining treatments,[17] that such people are psychologically or genetically robust and less likely to express pain or discomfort,[18 19] or that because participants and their families had sufficient time to prepare for death their perception of suffering might have been overlooked or expected.[20]

A major strength of this research is the application of a comprehensive measure of frailty in a large scale prospective sample. The frailty index had a significant impact on the type of death experienced by older adults and underscored a critical aspect of the quality of dying that is often overlooked in cohort studies. Our sensitivity analyses closely replicated age-sex distributions in frailty indices in other studies.[7 8 21] On the basis

## WHAT IS ALREADY KNOWN ON THIS TOPIC

An index for frailty is a valid construct for measuring mortality and the use of healthcare services

## WHAT THIS STUDY ADDS

The frailty index was strongly associated with suffering and the amount of time spent bedridden before death

Extremely old people (≥90 years) were most likely to experience 30 or fewer bedridden days with no suffering before death compared with other old adults

Women with the highest levels of frailty were most likely to experience 30 or more bedridden days with no suffering before death

of this evidence we are confident about the validity of our frailty measure.

One limitation of our study is that our longitudinal measures of the types of death include only two domains of the quality of dying previously identified in the literature.[22 23] Although the qualitative assessment of suffering before death is a key dimension of quality of death, the measure of suffering in the Chinese longitudinal healthy longevity survey was ascertained from next of kin and may be biased. This is because some research shows that proxy reports are not always consistent with reports by individuals before their death.[24] A final concern is that sample attrition from 2002 to 2005 was not random and may have introduced bias in our estimates. However, supplementary analyses indicated that the general patterns and conclusions of our analysis did not change regardless of whether we parameterised the lost sample as a categorical outcome or imputed the missing cases.

In medical settings, curative treatments are often eclipsed in moderate to extreme cases of frailty by palliative care in efforts to reduce discomfort and enhance the quality of life before death. The clinical application of a checklist (or index) for frailty can be a useful diagnostic tool that helps to characterise a patient's biological age compared with their chronological age, although unlike the phenotypic approach the frailty index may need translation in clinical practice for specific treatments.[21 25]

China is the world's largest developing country and is facing unique challenges to its healthcare system as an unbalanced population structure and a rapidly ageing population is straining the traditional family oriented system of care. We believe that the present study is an important step towards identifying frailty and its association with the quality of death in a rapidly developing nation.

**Contributors:** See bmj.com.
**Funding:** The data used in this study were from the 2002 and 2005 waves of the Chinese longitudinal healthy longevity survey, which was funded by the National Institute on Aging, the China Natural Science Foundation, the China Social Science Foundation, the United Nations Population Funds, and the Hong Kong Research Grant Council. DG's work was partly supported by a National Institute of Aging grant (R01 AG023627, PI: ZY) when he was at Duke University and was partly supported by a provost mini-grant for internationalisation at Portland State University. The work by MED and DFW was partially supported by the Carolina Population Center at the University of North Carolina at Chapel Hill postdoctoral training programme, funded by the National Institute of Child Health and Human Development grant NIH 5-T32-HD07168-28 (MED) and National Institute on Aging grant T32 AG00155 (DFW). ZY's work was supported by NIA grant R01 AG023627.
**Competing interests:** None declared.
**Ethical approval:** This study was approved by the institutional review board of Duke University health system institutional review board.

1 Markle-Reid M. Conceptualizations of frailty in relation to older adults. *J Adv Nurs* 2003;44:58-68.
2 Campbell AJ, Buchner DM. Unstable disability and the fluctuations of frailty. *Age Ageing* 1997;26:315-8.
3 Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci* 2001;56A:M146-57.
4 Rockwood K, Mogilner A, Mitnitski A. Changes with age in the distribution of a frailty index. *Mech Ageing Dev* 2004;125:517-9.
5 Morley JE, Haren NT, Rolland Y, Kim MJ. Frailty. *Med Clin North Am* 2006;90:837-47.
6 Gu D, Dupre ME, Sautter J, Zhu H, Liu Y, Zeng Y. Frailty and mortality among Chinese at advanced ages. *J Gerontol B Psychol Sci Soc Sci* 2009;64:279-89.
7 Mitnitski A, Graham JE, Mogilner AJ, Rockwood K. Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC Geriatr* 2002;2:1.
8 Mitnitski A, Song X, Skoog I, Broe GA, Cox JL, Grunfeld E, et al. Relative fitness and frailty of elderly men and women in developed countries and their relationship with mortality. *J Am Geriatr Soc* 2005;35:2184-9.
9 Puts MT, Lips P, Deeg DJ. Sex differences in the risk of frailty for mortality independent of disability and chronic diseases. *J Am Geriatr Soc* 2005;53:40-7.
10 Field MJ, Cassell CK, eds. *Approaching death: improving care at the end of life*. Washington DC: National Academic Press, 1997.
11 Curtis JR, Rubenfeld GD, eds. *Managing death in the intensive care unit*. New York: Oxford University Press, 2001.
12 Patrick DL, Engelberg RA, Curtis JR. Evaluating the quality of dying and death. *J Pain Symptom Manage* 2001;22:717-26.
13 Ferrucci L, Turchi A, Fumagallo S, Di Bari M, Silerstrini G, Zacchei S, et al. Sex-related differences in the length of disability prior to death in older persons. *Aging Clin Exp Res* 2003;15:310-4.
14 Pronovost P, Angus DG. Economics of managing death in the ICU. In: Curtis JR, Rubenfeld GD, eds. *Managing death in the intensive care unit*. New York: Oxford University Press, 2001:245-55.
15 Robine JM, Jagger C, Mathers CD, Crimmins EM, Suzman RM, eds. *Determining health expectancies*. New York: Wiley, 2003.
16 Dupre ME, Liu G, Gu D. Predictors of longevity: evidence from the oldest-old in China. *Am J Public Health* 2008;98:1203-8.
17 Hamel MB, Teno JM, Goldman L, Lynn J, Davis RB, Galanos AN, et al. Patient age and decisions to withhold life-sustaining treatments from seriously ill, hospitalized adults. *Ann Intern Med* 1999;130:116-25.
18 Willcox DC, Willcox BJ, Hsueh WC, Suzuki M. Genetic determinants of exceptional human longevity: insights from the Okinawa Centenarians Study. *Age* 2006;28:313-32.
19 Martin P, Poon LW, Kim E, Johnson MA. Social and psychological resources in the oldest old. *Exp Aging Res* 1996;22:121-39.
20 Pinquart M, Sorensen S. Preparation for death and preparation for care in older community-dwelling adults. *Omega* 2002;45:69-88.
21 Kulminski A, Ukraintseva S, Kulminskaya IV, Arbeev K, Land KC, Yashin A. Cumulative deficits better characterize susceptibility to death in elderly people than phenotypic frailty: lessons from the Cardiovascular Health Study. *J Am Geriatr Soc* 2008;56:898-903.
22 Patrick DL, Curtis JR, Engelberg RA, Nielsen E, McCown E. Measuring and improving the quality of dying and death. *Ann Intern Med* 2003;139:410-5.
23 Steinhauser KS, Clipp EC, Tulsky JA. Evolution in measuring the quality of dying. *J Palliat Med* 2002;5:407-14.
24 Hinton J. How reliable are relatives' retrospective reports of terminal illness? Patients and relatives' accounts compared. *Soc Sci Med* 1996;43:1229-36.
25 Rockwood K, Andrew M, Mtinitski A. A comparison of two approaches to measuring frailty in elderly people. *J Gerontol A Biol Sci Med Sci* 2007;62A:738-43.

**Accepted:** 11 December 2008

# Errors in the administration of parenteral drugs in intensive care units: multinational prospective study

Andreas Valentin,[1,2] Maurizia Capuzzo,[3] Bertrand Guidet,[4,5,6] Rui Moreno,[7] Barbara Metnitz,[8] Peter Bauer,[8] Philipp Metnitz,[9] on behalf of the Research Group on Quality Improvement of the European Society of Intensive Care Medicine (ESICM) and the Sentinel Events Evaluation (SEE) Study Investigators

[1]General and Medical Intensive Care Unit, Medical Department, Medical University of Vienna, Vienna, Austria
[2]Rudolfstiftung Hospital, Juchgasse 25, A-1030 Vienna
[3]Department of Anaesthesia and Intensive Care, University Hospital of Ferrara, Ferrara, Italy
[4]INSERM, Unité de Recherche en Epidemiologié, Systemes d'Information, et Modelisation, Paris, France
[5]Faculty of Medicine, University Pierre et Marie Curie, Paris
[6]Assistance Publique, Hôpitaux de Paris, Hôpital Saint-Antoine, Service de Réanimation Médicale, Paris
[7]Department of Intensive Care, Hospital de St António dos Capuchos, Centro Hospitalar de Lisboa (central, e.p.e), Lisbon, Portugal
[8]Section of Medical Statistics, Medical University of Vienna, Vienna
[9]Department of Anaesthesiology and General Intensive Care, Medical University of Vienna, Vienna

Correspondence to: A Valentin
andreas.valentin@meduniwien.ac.at

Cite this as: *BMJ* 2009;338:b814
doi:10.1136/bmj.b814

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;338:b814

## ABSTRACT

**Objective** To assess on a multinational level the frequency, characteristics, contributing factors, and preventive measures of administration errors in parenteral medication in intensive care units.

**Design** Observational, prospective, 24 hour cross sectional study with self reporting by staff.

**Setting** 113 intensive care units in 27 countries.

**Participants** 1328 adults in intensive care.

**Main outcome measures** Number of errors; impact of errors; distribution of error characteristics; distribution of contributing and preventive factors.

**Results** 861 errors affecting 441 patients were reported: 74.5 (95% confidence interval 69.5 to 79.4) events per 100 patient days. Three quarters of the errors were classified as errors of omission. Twelve patients (0.9% of the study population) experienced permanent harm or died because of medication errors at the administration stage. In a multiple logistic regression with patients as the unit of analysis, odds ratios for the occurrence of at least one parenteral medication error were raised for number of organ failures (odds ratio per increase of one organ failure: 1.19, 95% confidence interval 1.05 to 1.34); use of any intravenous medication (yes *v* no: 2.73, 1.39 to 5.36); number of parenteral administrations (per increase of one parenteral administration: 1.06, 1.04 to 1.08); typical interventions in patients in intensive care (yes *v* no: 1.50, 1.14 to 1.96); larger intensive care unit (per increase of one bed: 1.01, 1.00 to 1.02); number of patients per nurse (per increase of one patient: 1.30, 1.03 to 1.64); and occupancy rate (per 10% increase: 1.03, 1.00 to 1.05). Odds ratios for the occurrence of parenteral medication errors were decreased for presence of basic monitoring (yes *v* no: 0.19, 0.07 to 0.49); an existing critical incident reporting system (yes *v* no: 0.69, 0.53 to 0.90); an established routine of checks at nurses' shift change (yes *v* no: 0.68, 0.52 to 0.90); and an increased ratio of patient turnover to the size of the unit (per increase of one patient: 0.73, 0.57 to 0.93).

**Conclusions** Parenteral medication errors at the administration stage are common and a serious safety problem in intensive care units. With the increasing complexity of care in critically ill patients, organisational factors such as error reporting systems and routine checks can reduce the risk for such errors.

## INTRODUCTION

An investigation of 21 hospitals in the Netherlands showed that nearly 6% of 1.3 million hospital admissions in 2004 resulted in unintentional harm to the patient.[1] The complexity of processes and medical conditions dealt with in intensive care medicine makes this specialty particularly vulnerable to error.[2 3] In the first multinational sentinel events evaluation (SEE 1) study, 38.8 incidents per 100 patient days in five categories (drains and lines, artificial airway, equipment, handling of alarms, and medication) were observed in 205 intensive care units.[4] Medication errors at the prescription and administration stages were reported at a rate of 10.5 per 100 patient days. As such errors carry a particularly high potential for serious harm, this topic was chosen for the second study (SEE 2).

We conducted a prospective, observational, multinational study on the frequency, characteristics, contributing factors and impact of parenteral medication errors at the administration stage in intensive care units.

## METHODS

We used a cross sectional design with a 24 hour observation period. Ultimately, 113 units from 27 countries participated.

### Definition, assessment, and description of medication errors at the administration stage

A medication error at the administration stage was defined as an error of omission or commission in the context of parenteral drug administration that harmed or could have harmed a patient. We exclusively addressed five types of error: wrong dose, wrong drug, wrong route, wrong time, missed medication.

All nurses and physicians on duty were asked to fill in a structured questionnaire with a formalised and coded description of every medication error. We assessed contributing factors (communication-written, communication-oral, handover, workload/stress/fatigue, experience/knowledge/supervision, violation of protocol/standard, recently changed brand name of drugs, equipment failure, others); situational factors (admission/discharge, routine, emergency, movement with the hospital, intervention, urgent crisis of another patient, others); and grading of the impact of the error (a change registered or not, intervention necessary or not, no harm, temporary harm, permanent harm, death).

A coordinator provided information regarding characteristics of the unit, the staffing and patient flow, and the classification of the severity of illness and medical personnel workload for every patient.

### Patients and patient related factors

The study included all patients staying in the participating units. Each single dose of parenteral medication given to each patient was counted. For each patient we calculated the sequential organ failure assessment (SOFA) score[5] and the nursing manpower use score (NEMS)[6] to assess nursing workload. The total duration of each patient's stay, and the hours spent in the unit during the observation period were reported. The vital status was assessed at discharge or on day 28 after the study period.

### Intensive care unit related factors

Characteristics of hospital size (beds), type and size of intensive care unit, shift schedules, number of nurses and physicians in each shift, number of occupied and free beds, maximum number of patients, and number of admitted and discharged patients in each shift were recorded for every unit. Information about any system for formal critical incident reporting or computerised prescribing was recorded. Information about the process of parenteral medication administration was obtained.

**Table 1 | Odds ratios* (OR) for occurrence of at least one error in parenteral drug administration in intensive care unit (ICU). Univariate logistic regression**

| Variable | Variable measurement† | OR (95% CI) | P value |
|---|---|---|---|
| **All observed types of parenteral drug administration errors** | | | |
| ICU size (beds) | 1 | 1.02 (1.01 to 1.03) | <0.01 |
| ICU type: medical | 1 | 0.64 (0.42 to 0.96) | 0.03 |
| ICU type: mixed | 1 | 1.57 (1.04 to 2.38) | 0.03 |
| Patients per nurse | 1 | 1.26 (1.04 to 1.54) | 0.02 |
| CIRS in place | Yes/no | 0.67 (0.53 to 0.84) | <0.01 |
| Infusions previously prepared by pharmacist | Yes/no | 1.32 (1.02 to 1.70) | 0.03 |
| No of different types of infusion pumps | 1 | 0.89 (0.81 to 0.99) | 0.03 |
| Routine check at shift change | Yes/no | 0.63 (0.50 to 0.81) | <0.01 |
| Labelling of syringes | Yes/no | 0.61 (0.44 to 0.86) | <0.01 |
| Occupancy rate (%) | 10 | 1.02 (1.00 to 1.05) | 0.04 |
| Relative turnover | 1 | 0.75 (0.60 to 0.93) | 0.01 |
| **Errors of commission (wrong dose, wrong drug, wrong route)** | | | |
| Hospital size (beds) | 100 | 1.05 (1.02 to 1.08) | <0.01 |
| ICU size (beds) | 1 | 1.02 (1.01 to 1.03) | <0.01 |
| Patients per nurse | 1 | 1.51 (1.10 to 2.07) | 0.01 |
| Patients per physician | 1 | 1.10 (1.01 to 1.20) | 0.03 |
| ICU beds per nurse | 1 | 1.35 (1.02 to 1.77) | 0.03 |
| CIRS† in place | Yes/no | 0.36 (0.24 to 0.54) | <0.01 |
| Infusions previously prepared by pharmacist | Yes/no | 2.32 (1.57 to 3.41) | <0.01 |
| Electronic prescribing system in use | Yes/no | 0.62 (0.40 to 0.95) | 0.03 |
| **Errors requiring an intervention** | | | |
| CIRS in place | Yes/no | 0.44 (0.26 to 0.75) | <0.01 |
| Electronic prescribing system in use | Yes/no | 0.43 (0.23 to 0.82) | 0.01 |
| No of different types of perfusors | 1 | 1.16 (1.04 to 1.29) | 0.01 |
| Labelling of syringes | Yes/no | 0.31 (0.18 to 0.56) | <0.01 |

CIRS=critical incident reporting system.
*Odds ratios calculated by using dependent variable "occurrence of at least one medication error" with patients as unit of analysis. Table displays unit related variables reaching significance in different sets of error.
†Increment or binary.

### Data collection

Data collection started at the beginning of the nurses' day shift. The study period was designed to overlap a second day to include at least one day shift and one night shift.

### Statistical analysis

Odds ratios were calculated with a dichotomous outcome variable—no medication error versus occurrence of at least one medication error—with the patient as the unit of analysis. We evaluated univariate associations between patients' characteristics and characteristics of the intensive care unit and the outcome. We included variables that reached univariate significance in a multivariate logistic regression analysis. Clustering on the patient level or unit level was accounted for by including various patient and unit related variables to explain differences between patients and units. We performed the calculations separately for all types of errors.

## RESULTS

### ICU and patient sample

After exclusion of 57 patients under the age of 18 and 35 patients with conflicting information, the final sample consisted of 1328 patients from 113 units in 27 countries. Patients spent 0.6 to 24 hours in the unit during the 24 hour observation period, and only 23% spent less than 24 hours in the unit. A SOFA score of 5—the median value in the patients—reflects the dysfunction or failure of more than one organ system. A NEMS score of 27—the median value in the patients—reflects the equivalent of 59% of the workload that one unit nurse can perform in 24 hours.

Nearly half the units (48%) had a system for formal critical incident reporting.

### Process characteristics

In 37 (33%) units medication was prescribed by means of an electronic prescribing system. Further reporting showed that 26 (23%) used infusions previously prepared by a pharmacist, 76 (67%) used perfusors with a fixed standard preparation, 69 (61%) provided a dedicated area for preparation of medications, 78 (69%) routinely checked perfusors and infusion pumps at every shift change, and 97 (86%) labelled all syringes prepared with drugs before use.

### Occurrence and characteristics of errors

In the 1328 patients, 861 medication errors affecting 441 patients were reported. A total of 887 patients (67%) experienced no error, 250 (19%) experienced one error, and 191 patients (14%) experienced more than one. Of the 113 units, 21 (19%) reported no medication errors. There were 74.5 (95% confidence

**Table 2 | Odds ratios* (OR) for occurrence of at least one parenteral drug administration error in intensive care unit (ICU). Stepwise multiple logistic regression**

| Variable | Variable measurement† | OR (95% CI) | P value |
|---|---|---|---|
| **All observed types of parenteral drug administration errors** | | | |
| Patient-related variables: | | | |
| No of parenteral administrations | 1 | 1.06 (1.04 to 1.08) | <0.01 |
| No of organ failures | 1 | 1.19 (1.05 to 1.34) | <0.01 |
| NEMS item: | | | |
| Basic monitoring | Yes/no | 0.19 (0.07 to 0.49) | <0.01 |
| IV medication | Yes/no | 2.73 (1.39 to 5.36) | <0.01 |
| Specific interventions in ICU | Yes/no | 1.50 (1.14 to 1.96) | <0.01 |
| ICU related variables: | | | |
| ICU size (beds) | 1 | 1.01 (1.00 to 1.02) | 0.04 |
| Patients per nurse | 1 | 1.30 (1.03 to 1.64) | 0.03 |
| CIRS in place | Yes/no | 0.69 (0.53 to 0.90) | <0.01 |
| Routine check at shift change | Yes/no | 0.68 (0.52 to 0.90) | <0.01 |
| Occupancy rate (%) | 10 | 1.03 (1.00 to 1.05) | 0.03 |
| Relative turnover | 1 | 0.73 (0.57 to 0.93) | 0.01 |
| **Errors of commission (wrong dose, wrong drug, wrong route)** | | | |
| Patient related variables: | | | |
| No of parenteral administrations | 1 | 1.05 (1.02 to 1.07) | <0.01 |
| NEMS item: | | | |
| Multiple vasoactive medication | Yes/no | 2.43 (1.41 to 4.18) | <0.01 |
| ICU related variables: | | | |
| Patients per physician | 1 | 1.12 (1.02 to 1.23) | 0.01 |
| CIRS in place | Yes/no | 0.34 (0.22 to 0.52) | <0.01 |
| Infusions previously prepared by pharmacist | Yes/no | 2.36 (1.55 to 3.60) | <0.01 |
| **Errors requiring an intervention** | | | |
| Patient related variables: | | | |
| No of parenteral administrations | 1 | 1.08 (1.05 to 1.12) | <0.01 |
| NEMS item: | | | |
| Multiple vasoactive medication | Yes/no | 2.63 (1.37 to 5.07) | <0.01 |
| Specific interventions outside ICU | Yes/no | 2.25 (1.16 to 4.39) | 0.02 |
| ICU related variables: | | | |
| Electronic prescribing system in use | Yes/no | 0.32 (0.16 to 0.64) | <0.01 |

NEMS=nine equivalents of nursing manpower use score; CIRS=critical incident reporting system.
*Odds ratios calculated by using dependent variable "occurrence of at least one medication error" with patients as unit of analysis. Model accounts for patient and ICU characteristics and different sets of error.
†Increment or binary.

interval 69.5 to 79.4) errors per 100 patient days. The most frequent errors were related to wrong time of administration (n=386) and missed medication (n=259), followed by wrong dose (n=118), wrong drug (n=61), and wrong route (n=37).

Medication errors occurred most frequently during routine situations (n=595; 69%) and least frequently during admission and discharge procedures (n=73; 8%), movement with the hospital (n=40; 5%), undefined situations (n=41; 5%), emergencies (n=38; 4%), interventions (n=36; 4%), and urgent crisis with another patient in the unit (n=29; 3%).

Regarding the type of administration, 505 errors occurred during 5622 intravenous bolus administrations (9%), 279 during 5034 continuous intravenous administrations (6%), and 69 during 1069 subcutaneous administrations (6%). Aside from unclassified drugs, most errors occurred in antimicrobial drugs and sedation or analgesia.

Unit staff reported workload/stress/fatigue as a contributing factor in 32% (n=272) of all errors. Other contributing factors were recently changed drug name (n=155; 18%), communication-written (n=124; 14%), communication-oral (n=83; 10%), experience/knowledge/supervision (n=81; 9%), violation of protocol/standard (n=76; 9%), handover (n=53; 6%), equipment failure (n=0), and others (n=0).

### Reported impact of errors

Participating units reported that 71% of parenteral medication errors resulted in no change in the status of the patient. In contrast, according to the review of reporting units, 12 patients (0.9% of the total study population) experienced permanent harm (n=7) or death (n=5) in relation to a total of 15 errors. In eight cases, trainees were involved. Unit mortality (reported up to day 28) was 14.5% and 22.8% in patients without and with a parenteral medication error, respectively. Mortality was not predicted by the occurrence of a parenteral medication error.

### Predictors of parenteral medication errors

Univariate analysis showed that a higher severity of illness, a higher level of care, and a higher rate of parenteral drug administrations were associated with increased odds for the occurrence of at least one medication error. With respect to unit related variables, the odds for the occurrence of at least one medication error were increased at a higher patient to nurse ratio and were decreased when a critical incident reporting system was in place. Table 1 gives details of all unit related variables reaching significance in different sets of type of error.

In a multiple logistic regression analysis, six unit related and five patient related variables remained in the final model when we included all types of error (table 2). A higher severity of illness, a higher level of care, and a higher rate of parenteral drug administrations were associated with increased odds for the occurrence of at least one medication error. Results were robust for the existence of a critical incident reporting system both in the univariate and multivariate analysis for all three different types of error. There was one exception in the multivariate analysis, where "electronic prescribing system" was the only significant unit related variable. In this case, however, "electronic prescribing system" might cover the influence of "critical incident reporting," both being positively correlated ($\chi^2$ test, P=0.03) over units.

## DISCUSSION

In five categories of parenteral medication errors at the administration stage we found a total prevalence of 74.5 errors per 100 patient days. In 71% of errors there was no change in the patient's status, but 12 patients (0.9%) experienced permanent harm or died. The administration of parenteral medication is a weak

point in intensive care. This problem is not attributable to suboptimal care in a few individual units but represents a common pattern. Only 19% of participating units reported no parenteral medication errors during the 24 hour observation period.

Reported error rates in medication administration range from 3.3% to 6.2%, 6.5%,[7-9] and up to 56%,[10] depending on factors such as the inclusion or exclusion of different routes of administration and timing errors. Although 75% of all medication errors in our study were related to wrong timing or missed medication, the potential impact of such occurrences should not be underestimated. More than half of the errors with reported subsequent serious harm were attributable to errors of omission.

### How and why errors occur

We identified several contributing factors for errors in the administration of parenteral medication, though our observational design means we cannot confirm any causal relation. Analyses showed that more severely ill patients, who receive a higher level of care and increased use of parenteral medication, are more likely to experience a medication error. A coupling of an increase in complexity of care and the risk for error is consistent with existing literature.[4 11-13]

Unit related factors are the most important starting point for changes in the management of care. Most medication errors occurred during routine care of patients. In accordance with several other studies,[14-16] we found that increased workload—as measured by the patient to nurse ratio, the occupancy rate, and the ratio of beds per nurse—is associated with a higher risk for adverse events. A more complex effect was related to the size of the participating units: the complexity of organisation and communication in a given unit increases with the number of beds and makes the system more prone to error. However, this ratio showed a decreased risk for adverse events in units with more beds in relation to the turnover of patients. Interestingly, the provision of infusions prepared by a pharmacist increased the risk for a medication error. The risks were lower when nurses labelled syringes they themselves had prepared. This is an example of reducing complexity and avoiding gaps in information and communication in the process of care. Staff reported the use of drugs with recently changed brand names as a contributing factor in 18% of errors.

We found that an existing critical incident reporting system was an independent predictor for a decreased risk of medication errors with respect to all types of error. In addition, the process of routine checking at nurses' shift changes significantly reduced the risk for medication errors.

As trainees were involved in more than half of the errors with subsequent serious harm, the supervision of trainees should be a focus of concern. Given the frequency and impact of errors of omission, preventive measures for this type of error should be investigated in further studies.

### Limitations

Components that have been shown to influence the occurrence of errors—such as variations in unit organisation[10] and communication[2 17]—are difficult to measure and were outside the scope of our observation. With self reporting, there are several problems in assessing errors. Different formats of data collection will lead to different findings.[18] Furthermore, a self reporting method carries the risk of underreporting.[2 19 20] Moreover, the possibility of volunteer bias needs to be considered because the units studied were self selecting.

We thank Mary McKenney for editorial advice, Gerhard Krenn for programming the study website, and the staff of all the units that contributed to the study. A list of people in charge of the study at the unit level can be found at www.hsro-esicm.org.

**Contributors:** See bmj.com.

**Funding:** The SEE study was supported and funded by the Austrian Center for Documentation and Quality Assurance in Intensive Care Medicine (ASDI) and endorsed by the European Society of Intensive Care Medicine (ESICM).

**Competing interests:** None declared.

**Ethical approval:** Because the study was observational and no additional interventions were performed, the need for informed consent was waived by the institutional review board. Each unit, however, was made responsible for obtaining local permissions, as necessary, according to local regulations.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Serious concerns about medication safety in intensive care units have been raised, mostly in single centre studies

The extent to which medication safety represents a common problem in units is unknown

Cause of medication errors is related to human factors and, more important, system failures, but preventive factors are only partly explored

## WHAT THIS STUDY ADDS

The administration of parenteral medication seems to represent a common pattern of weakness in patients' safety in intensive care units

Organisational factors such as error reporting systems and routine checks can reduce the risk of parenteral administration errors at the administration stage

1 Sheldon T. Dutch study shows that 40% of adverse incidents in hospital are avoidable. *BMJ* 2007;334:925.
2 Donchin Y, Gopher D, Olin M, Badihi Y, Biesky M, Sprung CL, et al. A look into the nature and causes of human errors in the intensive care unit. *Crit Care Med* 1995;23:294-300.
3 Rothschild JM, Landrigan CP, Cronin JW, Kaushal R, Lockley SW, Burdick E, et al. The critical care safety study: the incidence and nature of adverse events and serious medical errors in intensive care. *Crit Care Med* 2005;33:1694-700.
4 Valentin A, Capuzzo M, Guidet B, Moreno RP, Dolanski L, Bauer P, et al. Patient safety in intensive care: results from the multinational sentinel events evaluation (SEE) study. *Intensive Care Med* 2006;32:1591-8.
5 Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996;22:707-10.
6 Reis Miranda D, Moreno R, Iapichino G. Nine equivalents of nursing manpower use score (NEMS). *Intensive Care Med* 1997;23:760-5.

7 Calabrese AD, Erstad BL, Brandl K, Barletta JF, Kane SL, Sherman DS. Medication administration errors in adult patients in the ICU. *Intensive Care Med* 2001;27:1592-8.
8 Fahimi F, Ariapanah P, Faizi M, Shafaghi B, Namdar R, Ardakani MT. Errors in preparation and administration of intravenous medications in the intensive care unit of a teaching hospital: an observational study. *Aust Crit Care* 2008;21:110-6.
9 Kopp BJ, Erstad BL, Allen ME, Theodorou AA, Priestley G. Medication errors and adverse drug events in an intensive care unit: direct observation approach for detection. *Crit Care Med* 2006;34:415-25.
10 Van den Bemt PM, Fijn R, van der Voort PH, Gossen AA, Egberts TC, Brouwers JR. Frequency and determinants of drug administration errors in the intensive care unit. *Crit Care Med* 2002;30:846-50.
11 Bracco D, Favre JB, Bissonnette B, Wasserfallen JB, Revelly JP, Ravussin P, et al. Human errors in a multidisciplinary intensive care unit: a 1-year prospective study. *Intensive Care Med* 2001;27:137-45.
12 Cullen DJ, Sweitzer BJ, Bates DW, Burdick E, Edmondson A, Leape LL. Preventable adverse drug events in hospitalized patients: a comparative study of intensive care and general care units. *Crit Care Med* 1997;25:1289-97.
13 Webster CS. The nuclear power industry as an alternative analogy for safety in anaesthesia and a novel approach for the conceptualisation of safety goals. *Anaesthesia* 2005;60:1115-22.
14 Tarnow-Mordi WO, Hau C, Warden A, Shearer AJ. Hospital mortality in relation to staff workload: a 4-year study in an adult intensive-care unit. *Lancet* 2000;356:185-9.
15 Ream RS, Mackey K, Leet T, Green MC, Andreone TL, Loftis LL, et al. Association of nursing workload and unplanned extubations in a pediatric intensive care unit. *Pediatr Crit Care Med* 2007;8:366-71.
16 Hugonnet S, Chevrolet JC, Pittet D. The effect of workload on infection risk in critically ill patients. *Crit Care Med* 2007;35:76-81.
17 Wasserfallen JB, Butschi AJ, Muff P, Biollaz J, Schaller MD, Pannatier A, et al. Format of medical order sheet improves security of antibiotics prescription: the experience of an intensive care unit. *Crit Care Med* 2004;32:655-9.
18 Beckmann U, Bohringer C, Carless R, Gillies DM, Runciman WB, Wu AW, et al. Evaluation of two methods for quality improvement in intensive care: facilitated incident monitoring and retrospective medical chart review. *Crit Care Med* 2003;31:1006-11.
19 Vincent C, Stanhope N, Crowley-Murphy M. Reasons for not reporting adverse incidents: an empirical study. *J Eval Clin Pract* 1999;5:13-21.
20 Stanhope N, Crowley-Murphy M, Vincent C, O'Connor AM, Taylor-Adams SE. An evaluation of adverse incident reporting. *J Eval Clin Pract* 1999;5:5-12.


**Accepted:** 5 December 2008

# Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews

Fujian Song,[1] Yoon K Loke,[1] Tanya Walsh,[2] Anne-Marie Glenny,[2] Alison J Eastwood,[3] Douglas G Altman[4]

[1]Faculty of Health, University of East Anglia, Norwich NR4 7TJ
[2]School of Dentistry, University of Manchester, Manchester
[3]Centre for Reviews and Dissemination, University of York, York
[4]Centre for Statistics in Medicine, Oxford

**Correspondence to:** F Song
fujian.song@uea.ac.uk

Cite this as: *BMJ* 2009;338:b1147
doi:10.1136/bmj.b1147

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;338:b1147


## ABSTRACT

**Objective** To investigate basic assumptions and other methodological problems in the application of indirect comparison in systematic reviews of competing healthcare interventions.

**Design** Survey of published systematic reviews.

**Inclusion criteria** Systematic reviews published between 2000 and 2007 in which an indirect approach had been explicitly used. Identified reviews were assessed for comprehensiveness of the literature search, method for indirect comparison, and whether assumptions about similarity and consistency were explicitly mentioned.

**Results** The survey included 88 review reports. In 13 reviews, indirect comparison was informal. Results from different trials were naively compared without using a common control in six reviews. Adjusted indirect comparison was usually done using classic frequentist methods (n=49) or more complex methods (n=18). The key assumption of trial similarity was explicitly mentioned in only 40 of the 88 reviews. The consistency assumption was not explicit in most cases where direct and indirect evidence were compared or combined (18/30). Evidence from head to head comparison trials was not systematically searched for or not included in nine cases.

**Conclusions** Identified methodological problems were an unclear understanding of underlying assumptions, inappropriate search and selection of relevant trials, use of inappropriate or flawed methods, lack of objective and validated methods to assess or improve trial similarity, and inadequate comparison or inappropriate combination of direct and indirect evidence. Adequate understanding of basic assumptions underlying indirect and mixed treatment comparison is crucial to resolve these methodological problems.

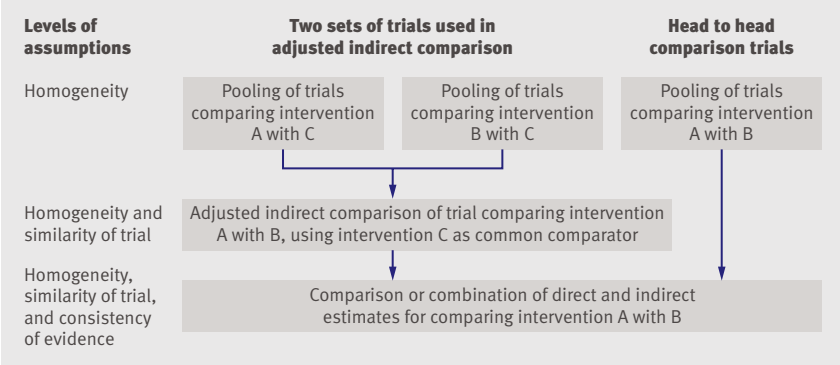

## INTRODUCTION

Well designed head to head randomised controlled trials are generally considered to provide the most rigorous research evidence on the relative effects of interventions.[1] Evidence from such trials is often limited or unavailable, however, and indirect comparison may be necessary.[2 3]

A simple but inappropriate method is to compare the results of individual arms from different trials as if they were from the same trial. This unadjusted indirect comparison has been criticised for discarding the within trial comparison, increasing liability to bias and over- precise estimates.[2] The adjusted indirect comparison can, however, take advantage of the strength of randomised controlled trials in making unbiased comparisons (see box).[4 5] Here the indirect comparison of different interventions is adjusted by comparing the results of their direct comparisons with a common control group.

### *Indirect comparison*

The results of placebo controlled trials suggested that both active treatments are more effective than placebo for smoking cessation. The results of the two sets of placebo controlled trials can also be used to indirectly compare the active treatments. Compared with placebo, the magnitude of treatment effect of bupropion (odds ratio 0.51, 95% confidence interval 0.36 to 0.73) was similar to that of nicotine replacement therapy




**Assumptions underlying adjusted indirect and mixed treatment comparison**

(0.57, 0.48 to 0.67). Therefore it could be indirectly concluded that the treatments were equally effective. The adjusted indirect comparison can also be formally done, using one of several sound methods. The result of adjusted indirect comparison suggests that bupropion was as effective as nicotine replacement therapy for smoking cessation (0.90, 0.61 to 1.34), although the confidence interval is wide. The validity of the adjusted indirect comparison depends on a similarity assumption, assuming that the two sets of placebo controlled trials are sufficiently similar for moderators of relative treatment effect.

*Comparison of direct and indirect estimates*

The result of the head to head comparison trial suggested that bupropion was more effective than nicotine replacement therapy for smoking cessation (0.48, 0.28 to 0.82). The discrepancy between the direct and indirect estimate was marginally significant ($I^2$=71%, P=0.06). Statistical methods are available to combine the results of direct and indirect evidence (combined odds ratio 0.68, 95% confidence interval 0.37 to 1.25). A consistency assumption is, however, required to combine the estimates. The combination of inconsistent evidence from different sources may provide invalid and misleading results.

To improve statistical power, evidence generated by indirect comparison can be combined with evidence from head to head trials,[9-11] facilitated by the development of network meta-analysis[12] and Bayesian hierarchical models for mixed treatment comparisons.[13]

Empirical evidence indicates that the results of an adjusted indirect comparison usually but not always agree with the results of direct comparison trials.[4] Recently, conflicting evidence has emerged about the validity of indirect comparison,[6] therefore the potential usefulness of adjusted indirect comparison is still overshadowed by concern about bias resulting from its misuse.

Existing statistical methods for adjusted indirect comparison and mixed treatment comparison are unbiased, but only if some assumptions are fulfilled.[2] The description of important assumptions underlying indirect comparison may not be clear in some methodological studies. For mixed treatment comparison it was noted that "the only additional assumption is that the similarity of the relative effects of treatment holds across the entire set of trials, irrespective of which treatments were actually evaluated."[9] However, the additional assumption may hold to a subset of trials but not across the entire set of trials. We suggest a framework to delineate the main assumptions related to indirect and mixed treatment comparison (figure).

Assumptions concerning adjusted indirect comparison and mixed treatment comparison are similar to but more complex than the underlying assumption for standard meta-analysis. At least three issues of comparability need consideration: a homogeneity assumption for each meta-analysis, where different trials are sufficiently homogeneous and estimate the same treatment effect (fixed effect model) or different treatment effects distributed around a typical value (random effects model); a similarity assumption for individual adjusted indirect comparison, where trials are similar for moderators of relative treatment effect; and a consistency assumption for the combination of evidence from different sources (figure).

We report findings from a survey of methodological problems in the application of indirect and mixed treatment comparison.

## METHODS

We searched PubMed for systematic reviews or meta-analyses published between 2000 and 2007 in which indirect comparison had been explicitly used (see bmj.com). The titles and abstracts of retrieved references were independently assessed by two reviewers to identify relevant reviews.

We extracted data on clinical indications, interventions compared, comprehensiveness of the literature search for trials used in indirect comparison, methods for indirect comparison, and whether direct evidence from head to head comparison trials was also available. We examined whether the assumption of similarity was explicitly mentioned and whether any efforts were made to investigate or improve the similarity of trials for indirect comparison. One reviewer extracted data and another checked each study.

## RESULTS

Overall, 88 review reports (91 publications) were included: 59 were reviews of effectiveness of interventions, 19 were reports of health technology assessment or cost effectiveness analysis, six were Cochrane reviews, and four were reviews used to illustrate methods for indirect comparisons.

Indirect comparison was used to evaluate drug interventions in 72 of the reviews: 43 compared drugs of different classes, 17 drugs of the same class, and 10 different formats or modes of delivery of the same drug. Two reviews compared the relative efficacy of an active drug with placebo. Non-drug interventions were indirectly compared in 16 reviews.

The most commonly used approach (49/88) was the adjusted indirect comparison using classic frequentist methods (see bmj.com). More complex methods

(network or Bayesian hierarchical meta-analysis) were used in 18 reviews. In 13 reviews, indirect comparison was informal, without calculation of relative effects or testing for statistical significance. In six reviews results from different trials were naively compared without using a common treatment control.

Direct evidence from head to head comparison trials was available in 40 reviews (see bmj.com), including 15 that used simple adjusted methods, 16 that used complex methods, and six that used informal methods. Compared with simple adjusted methods, complex methods were more likely to be used to combine the direct and indirect evidence. Where direct comparison was available, direct and indirect evidence were combined in 15 of the reviews that used complex methods and in only two of the reviews that used simple methods (see bmj.com). Furthermore, direct and indirect evidence were less likely to be explicitly compared in reviews that used complex rather than simple methods (9 *v* 11).

The assumption of trial similarity was explicitly mentioned or discussed in only 40 reviews (see bmj.com). Explicit mention of the similarity assumption was associated with efforts to examine or improve the similarity between trials for indirect comparisons (30/40 *v* 19/48). Methods to investigate or improve trial similarity included subjective judgment by a comparison of study characteristics (n=26) and subgroup and metaregression analysis to identify or adjust for possible moderators of treatment effects (n=23). The assumption of consistency was not explicit in most cases where direct and indirect evidence were compared or combined (18/30; see bmj.com).

In eight reviews, indirect comparison was based on data from other published systematic reviews or meta-analyses (see bmj.com). Evidence from head to head comparison trials was not systematically searched for or not included in nine cases (see bmj.com).

**A simple example of indirect comparison**

The case study compared bupropion with nicotine replacement therapy patch for smoking cessation.[6] The outcome was the number of smokers who failed to quit at 12 months (table). Indirect comparison can be made using two sets of randomised controlled trials: nine that compared bupropion with placebo and 19 that compared nicotine replacement therapy with placebo. One trial also compared bupropion with nicotine replacement therapy.[7]

**Number of smokers failing to quit at 12 months, according to treatment group**

| Comparison | No of trials | Odds ratio (95% CI) | $I^2$ (%) |
|---|---|---|---|
| Bupropion *v* placebo | 9 | 0.51 (0.36 to 0.73) | 54 |
| NRT patch *v* placebo | 19 | 0.57 (0.48 to 0.67) | 12 |
| Bupropion *v* NRT patch: | | | |
| Direct comparison | 1 | 0.48 (0.28 to 0.82) | — |
| Adjusted indirect comparison | 9+19 | 0.90 (0.61 to 1.34) | — |
| Combined (direct +indirect) | 1+(9+19) | 0.68 (0.37 to 1.25) | 71 |

NRT=nicotine replacement therapy.
See Bucher et al[5] and Song et al[4] for indirect comparison methods. Random effects model was used in meta-analyses of trials and for combination of direct and indirect estimates.[8]

## DISCUSSION

Indirect comparison is being increasingly used for the evaluation of a wide range of healthcare interventions. In this study, 16 of the 88 included reviews were health technology assessment reports. In many such reports, indirect comparison had not been done for clinical effectiveness but was used in the economic evaluation.

In the literature, several related but different assumptions underlying adjusted indirect comparison (figure) have not been clearly distinguished, resulting in methodological and practical problems in the interpretation of indirect or mixed treatment comparison. The problems include unclear understanding of underlying assumptions, inappropriate search and selection of relevant trials, use of inappropriate or flawed methods, lack of objective and validated methods to assess or improve trial similarity, and inadequate comparison or inappropriate combination of direct and indirect evidence.

Indirect comparison was explicit but informal in 13 reviews—neither relative effects nor statistical significance were calculated. Since the use of indirect comparison is often inevitable, a more explicit and formal approach is preferable. In six reviews, the results from individual arms of different trials were compared naively as if they were from one controlled trial. This approach is flawed because the strength of randomisation is disregarded.[2]

The strength of randomisation could be preserved in adjusted indirect comparison. The most common scenario was the indirect comparison of two competing interventions adjusted by common comparators using classic frequentist methods (including simple metaregression). The advantages of the simple methods include ease of use and transparency. However, when there are several alternative interventions to be compared, the simple adjusted indirect comparison may become inconvenient. More complex methods, such as network meta-analysis, are being increasingly used to make simultaneous comparisons of multiple interventions.[10 12 13] These methods treat all included interventions equally rather than focusing on one particular comparison of two interventions.

Subgroup analysis and metaregression are commonly used to assess or improve trial similarity for adjusted indirect comparison (see bmj.com). Their usefulness may be limited because the number of trials involved in adjusted indirect comparison was usually small and it was uncertain whether the important study level variables were reported in all relevant trials.

Trial similarity was often assessed by examining heterogeneity across trials and by a narrative comparison of trial characteristics for the different treatment comparisons being included, which may be deemed informal and subjective.

When data from head to head comparison trials are available, consideration needs to be given to whether an indirect comparison is justified when direct comparison trials are available; any discrepancies between direct and indirect evidence need to be sensibly

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Indirect comparisons can be valid if some basic assumptions are fulfilled

The related but different methodological assumptions have not been clearly distinguished

## WHAT THIS STUDY ADDS

Certain methodological problems may invalidate the results of evaluations using indirect comparison approaches

Understanding basic assumptions underlying indirect and mixed treatment comparison is crucial to resolve these problems

A framework can help clarify homogeneity, similarity, and consistency assumptions underlying adjusted indirect comparisons

interpreted; and could direct evidence be combined with the results of indirect comparison.

It is controversial whether indirect evidence needs to be considered when there is evidence from direct comparison trials.[5 9] Indirect comparison was considered helpful by authors of the 40 reviews in which both direct and indirect evidence were available. Such evidence was less likely to be explicitly compared and more likely to be combined in reviews that used complex rather than simple methods (see bmj.com). Since the evidence consistency is usually assessed informally and subjectively,[9] transparency is important to allow others to make their own judgment.

Reviews may include trials with three or more arms. Some reviews separately compared two active treatments with placebo within the same trial, and then the results of two separate comparisons were used in adjusted indirect comparison. This downgrades direct evidence to indirect evidence, reduces precision, and uses data from the same placebo arm twice.

In nine reviews, direct comparison trials were excluded or not searched for systematically. In reviews that included only placebo controlled trials, it was often unclear whether there were other active treatment controlled trials that could also be used for adjusted indirect comparison. Some indirect comparisons seemed to be done on an ad hoc basis, using data from existing systematic reviews and meta-analyses.

Reviews were included in this survey only if the indirect comparison was explicit in their titles and abstracts, and if they were indexed in PubMed. Thus we may have missed reports with indirect comparisons. Missed reviews may have been less explicit and less formal than included ones, therefore not mentioned in the abstract.

Empirical evidence on the validity of indirect and mixed treatment comparison is still limited and many questions remain unanswered. In addition, there is only limited empirical evidence to show that improved trial similarity is associated with improved validity of indirect and mixed treatment comparison.

**Contributors:** See bmj.com.
**Funding:** No specific funding was received for this study.
**Competing interests:** None declared.
**Ethical approval:** Not required.

1 Pocok SJ. *Clinical trials: a practical approach*. New York: Wiley, 1996.
2 Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, et al. Indirect comparisons of competing interventions. *Health Technol Assess* 2005;9:1-134.
3 Ioannidis JP. Indirect comparisons: the mesh and mess of clinical trials. *Lancet* 2006;368:1470-2.
4 Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326:472-5.
5 Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50:683-91.
6 Song F, Harvey I, Lilford R. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. *J Clin Epidemiol* 2008;61:455-63.
7 Jorenby DE, Leischow SJ, Nides MA, Rennard SI, Johnston JA, Hughes AR, et al. A controlled trial of sustained-release bupropion, a nicotine patch, or both for smoking cessation. *N Engl J Med* 1999;340:685-91.
8 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
9 Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;331:897-900.
10 Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;15:2733-49.
11 Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. *Control Clin Trials* 2000;21:488-97.
12 Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;21:2313-24.
13 Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;23:3105-24.

**Accepted:** 10 November 2008

pico

# Total mortality after changes in leisure time physical activity in 50 year old men: 35 year follow-up of population based cohort

Liisa Byberg,[1] Håkan Melhus,[2] Rolf Gedeborg,[3] Johan Sundström,[4] Anders Ahlbom,[5] Björn Zethelius,[6] Lars G Berglund,[7] Alicja Wolk,[8] Karl Michaëlsson[1]

[1]Department of Surgical Sciences, Section of Orthopaedics, and Uppsala Clinical Research Centre, Uppsala University, SE-75185 Uppsala, Sweden
[2]Department of Medical Sciences, Section of Clinical Pharmacology, Uppsala University
[3]Department of Surgical Sciences, Section of Anaesthesiology and Intensive Care, and Uppsala Clinical Research Centre, Uppsala University
[4]Department of Medical Sciences, Section of Acute and Internal Medicine, Uppsala University
[5]Department of Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, SE-17177 Stockholm, Sweden
[6]Department of Public Health and Caring Sciences, Section of Geriatrics, Uppsala University
[7]Uppsala Clinical Research Centre, Uppsala University
[8]Department of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet

**Correspondence to:** L Byberg
liisa.byberg@surgsci.uu.se

Cite this as: *BMJ* 2009;338:b688
doi: 10.1136/bmj.b688

**STUDY QUESTION** How does increased physical activity after middle age influence mortality and what is the size of the effect in comparison with smoking cessation?

**SUMMARY ANSWER** Increased physical activity in middle age increases longevity after an induction period of up to 10 years of no benefit. After 10 years of follow-up, however, increased physical activity between the ages 50 and 60 halved mortality compared with continued inactivity, and the effect was similar to that seen after smoking cessation (compared with continued smoking).

### Participants and setting

Participants in our study were 50 year old men in Uppsala, Sweden, who were examined in 1970-3 and re-examined at ages 60, 70, 77, and 82 years.

### Design, size and duration

We individually linked the 2205 men in this population based cohort with the population register. At the end of follow-up in 2006, 1329 of the men had died. Information on physical activity was obtained by questionnaire at each examination and was categorised as low, medium, and high. We used time updated variables in our analyses, taking changes over time into account. We studied changed physical activity between 50 and 60 years in 1759 men who participated in both examinations, of whom 998 died.

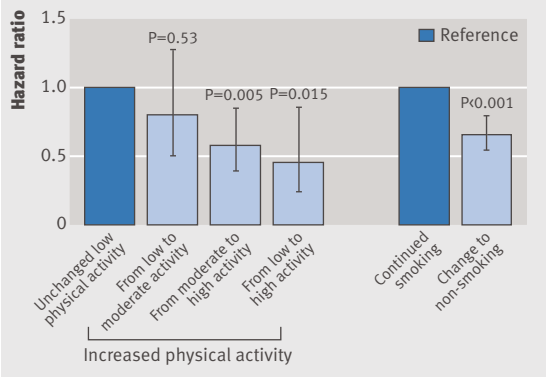
### Main results and the role of chance

Mortality (per 1000 person years) was 27.1, 23.6, and 18.4 in the groups with low, moderate, and high physical activity, respectively. Men who increased their physical activity level between the ages of 50 and 60 continued to have higher mortality during the first five years of follow-up compared with unchanged high physical activity (adjusted hazard ratio 2.64, 95% confidence interval 1.32 to 5.27). Given the small numbers of deaths, we are reluctant to place a strong emphasis on this higher risk, especially as mortality was not higher than that in men who continued to be sedentary. After 10 years of follow-up, increased physical activity was associated with reduced mortality to the level of men with unchanged high physical activity (adjusted hazard ratio 1.10, 0.87 to 1.38). The impact of increased physical activity on mortality was on a par with the effect of smoking cessation and independent of potential confounders (see figure).

### Bias, confounding, and other reasons for caution

We took account of changes over time in potential confounders including smoking, obesity, self perceived health, and morbidity, and in classic risk factors for mortality including hypertension and total cholesterol concentration. We also adjusted our estimates for socioeconomic group and educational level. Potential sources of bias that might conservatively influence our results include assessment of physical activity by questionnaire and adjustment for variables that can be regarded not only as confounders but also as intermediates on the causal pathway—such as perceived health, body weight, and diabetes.

**EFFECT OF CHANGED PHYSICAL ACTIVITY AND SMOKING STATUS ON MORTALITY RISK 10 YEARS LATER**




### Generalisability to other populations

The generalisability to women is yet to be determined. Mechanisms and reasons for an active choice to increase physical activity are not fully understood and may be different in other populations.

### Study funding/potential competing interests

All researchers are independent of the study funders, the Swedish Research Council.

## BMJ pico: advice to authors

For Research articles, we routinely post the full version only on bmj.com, and prepare an abridged version for the print journal.

To increase readership of research articles in the print *BMJ* and to give authors more control over the abridging, we are piloting a new way of abridging research articles for the print *BMJ*—publishing what is essentially an evidence abstract called BMJ pico. We hope that you will want to take part in this pilot if your research article is accepted. There is no need to prepare a BMJ pico in advance, however—please wait until we have offered to publish your article.

This is a summary of a paper that was published on bmj.com as *BMJ* 2009;338:b688