Papers

The effectiveness of glucocorticoids in treating croup: meta-analysis

Monica Ausejo, Antonio Saenz, Ba' Pham, James D Kellner, David W Johnson, David Moher, Terry P Klassen

Abstract

Objective To determine the effectiveness of glucocorticoid treatment in children with croup.

Design Meta-analysis of randomised controlled trials that examine the effectiveness of glucocorticoid treatment in children with croup.

Main outcome measures Score on scale measuring severity of croup, use of cointerventions (adrenaline (epinephrine), antibiotics, or supplemental glucocorticoids), length of stay in accident and emergency or in hospital, and rate of hospitalisation.

Results Twenty four studies met the inclusion criteria. Glucocorticoid treatment was associated with an improvement in the croup severity score at 6 hours with an effect size of -1.0 (95% confidence interval -1.5 to -0.6) and at 12 hours -1.0 (-1.6 to -0.4); at 24 hours this improvement was no longer significant (-1.0, -2.0 to 0.1). There was a decrease in the number of adrenaline treatments needed in children treated with glucocorticoids: a decrease of 9% (95% confidence interval 2% to 16%) among those treated with budesonide and of 12% (4% to 20%) among those treated with dexamethasone. There was also a decrease in the length of time spent in accident and emergency (-11 hours, 95% confidence interval -18 to 4 hours), and for inpatients hospital stay was reduced by 16 hours (-31 to 1 hour). Publication bias seems to play a part in these results.

Conclusions Dexamethasone and budesonide are effective in relieving the symptoms of croup as early as 6 hours after treatment. Fewer cointerventions are used and the length of time spent in hospital is decreased in patients treated with glucocorticoids.

Introduction

Croup (laryngotracheobronchitis) is a common cause of upper airway obstruction in children and is characterised by hoarseness, a barking cough, and inspiratory stridor. These symptoms are thought to occur as a result of oedema of the larynx and trachea which has been triggered by a recent viral infection. Parainfluenza virus type 1 is the agent most commonly identified in cases of croup.[1]

Although croup is a self limiting illness, it is a large burden on healthcare systems because of the frequent visits made to doctors and accident and emergency departments and, when necessary, hospitalisations. The annual incidence of croup in children younger than 6 years ranges from 1.5% to 6%.[2] Admission rates for croup in children seen in outpatient settings range from 1.5% to 31% of cases seen; these figures vary widely, depending on hospital admission practices and the severity of the disease in the population being assessed.[3][4]

The standard management of croup includes mist treatment (that is, treatment with humidified air), although there is little evidence that this is effective.[5] Racemic adrenaline (epinephrine), or L-adrenaline, has been shown to provide temporary relief to patients with croup but is not thought to have longer term benefits.[6] Since the late 1980s it has been recognised that glucocorticoids provide some clinical benefit in children with croup. In 1989, Kairys et al published a meta-analysis of clinical trials examining the benefit of glucocorticoids.[7] However, since then a number of randomised trials have been published, and there has been increasing interest in the use of glucocorticoids to treat outpatients with croup. The objective of this meta-analysis was to provide evidence to guide clinicians in their treatment of patients with croup, to examine the effectiveness of glucocorticoids in these patients, and to identify areas of uncertainty for future research.

Methods

Study identification

We searched Medline from January 1966 to August 1997, exploding glucocorticoid treatment (and each of the terms for corticosteroids) and croup; we restricted the search to randomised controlled trials using a previously validated strategy (see appendix 1 on the *BMJ*'s website). We searched Excerpta Medica and Embase from January 1974 to August 1997 (appendix 1). The Controlled Trials Register of the *Cochrane Library* was also searched; it includes studies identified by the Acute Respiratory Infection Review Group through the hand searching of key journals. We also sent letters to the authors of trials published in the past five years to enquire whether they knew of any other published or unpublished trials. Two researchers (TPK, MA) then selected studies as being potentially relevant based on a review of the titles and abstracts, if available. The complete text of these studies was then retrieved.

Thomas C Chalmer's Center for Systematic Reviews, Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, Ontario K1H 8LI, Canada
Monica Ausejo
visiting fellow
Antonio Saenz
visiting fellow
Ba' Pham
statistician
David Moher
director
Terry P Klassen
director, children's hospital

Department of Pediatrics, University of Calgary, Alberta Children's Hospital, 1820 Richmond Road, Calgary, Alberta T2T 5C7, Canada
James D Kellner
assistant professor
David W Johnson
assistant professor

Correspondence to: TP Klassen Department of Pediatrics, University of Alberta, 2C3.67 Walter C Mackenzie Health Sciences Center, Edmonton, Alberta T6G 2R7 terry.klassen@ sympatico.ca

website extra

All studies that had been retrieved were reviewed independently by two reviewers (AS, TPK). To be eligible for inclusion in this review a study had to meet all of the following criteria: it had to have studied patients with croup; an intervention with glucocorticoid had to have been compared with either placebo or any other active treatment; clinically relevant outcome measures had to have been used, such as the clinical score, hospitalisation rate (in outpatient studies only), length of time in hospital, or additional interventions used; and patients had to have been randomly assigned to treatment groups. Studies written in any language were eligible for inclusion. The weighted κ score was used to measure interrater agreement. Differences over which studies should be included were resolved by consensus reached after discussion.

Data extraction

Once we identified studies as being relevant for review, they were masked by obscuring the authors' names and institutions, the location of the study, reference lists, and any other potential identifiers; this was done by an independent research assistant who was not involved with the abstraction of data. Data were extracted using a structured form that captured patient status (inpatient or outpatient), the intervention and its control, the name of drug, the route of administration, and the dose. Additionally, data were collected on the primary outcome measure; clinical croup score at baseline and at any subsequent assessment times; length of stay in hospital or accident and emergency in hours; whether the patient had improved (coded yes or no); and the use of additional interventions such as adrenaline, supplemental glucocorticoids, mist treatment, intubation, or antibiotic treatment. Data were extracted by one reviewer (MA) and checked for accuracy by a second reviewer (TPK).

Quality assessment of trials

We assessed quality using empirically derived items. We used the previously validated Jadad 5 point scale to assess randomisation (0-2 points), double blinding (0-2 points), and withdrawals and dropouts (0-1 point).[8] For component assessment, concealment of allocation was described either as adequate, inadequate, or unclear.[9] Sponsorship of studies was coded as either pharmaceutical company, other sources, or not mentioned.[10] Two observers independently assessed quality (MA, JK), and interrater agreement was measured by the intraclass correlation.[11] Differences were resolved by consensus.

Data analysis

All comparisons were performed between treatment and control groups thus preserving randomisation. The main outcome measure was the difference between treatment groups in the mean change from croup score at baseline. We derived the outcome measures from cross sectional summaries (for example, at baseline, 6, and 12 hours) in cases in which outcome measures were not reported directly. The variance of an effect size was derived from the common variance of a single croup score assuming a correlation of 0.5 between pre-treatment and post-treatment scores. Other variance imputations were performed according to the work of Follman et al.[12] Variances of a single score were derived from the P values of the Mann-Whitney test[w8 w9 w23] and from the measurement of confidence intervals.[w4]

The croup score was reported inconsistently because of the different scales used in each study, hence trial effect sizes were used in the pooled estimates.[13] A treatment effect divided by its measurement variation (for example, a pooled standard deviation) gives an effect size. To aid in the interpretation of pooled results reported by standardised effect size, we converted the effect size scale back to the croup score using a subset of trials in which such scores were available. Another way to express the croup score is by determining a clinically important change in the score in the individual patient and then calculating the proportion of patients who had significant improvement among the patients treated with glucocorticoids or placebo.

In addition to funnel plots, we used the rank correlation test[14] and a graphical method[15] for the detection of publication bias.[16] Adjustment for publication bias in the pooled estimates was performed using the graphical method,[15] a selection model approach,[17] and the trim and fill method.[18] We used more than one method since the relative merits of the methods are not well established. Tests of homogeneity were performed with the χ^2 statistic for between study variation.[13] For the analyses of croup scores and secondary outcomes, fixed effect models were used to combine treatment effects if there was no evidence of heterogeneity across studies; otherwise, the more conservative estimates from random effect models were reported. For binary data (such as improvement in signs and symptoms and the presence of various additional interventions), rate differences and the number needed to treat were derived. For the number needed to treat, we inverted the differences in the proportions improved and their 95% confidence intervals.

Heterogeneity between studies was explored using sensitivity and subgroup analyses performed on the primary outcome of the change in croup scores from baseline at 6 hours. Westley scores were the scores most commonly used in the trials.[19] Westley scores use a 17 point scale to assess air entry (2 points), stridor (2 points), intercostal retractions (indrawing of the chest wall between the ribs on inspiration) (3 points), cyanosis (5 points), and level of consciousness (5 points). Treatment differences in Westley scores were calculated in place of effect sizes to provide an approximate conversion between the two scales. Differences between estimates derived from Westley and other scores were assessed.

A trial effect size was defined as the difference between the two treatments in the mean change from croup score at baseline. We derived effect sizes from cross sectional summaries (for example, baseline, 6, 12 hours) for trials not reporting effect sizes directly. The standardised effect size (that is, an effect size divided by the common standard deviation of the change from baseline) was used to combine trials reporting different versions of the croup score. Sensitivity analyses were based on the type and dose of glucocorticoid administered. The quality score of the included trials was incorporated into the pooled estimates using the method proposed by Moher.[20] In addition, the impact

of the concealment of treatment allocation on the pooled estimates was assessed.[9]

Results

Study identification and characteristics

Ninety seven studies were identified as being potentially relevant and retrieved. Two of these studies were in press at the time of data extraction and have since been published.[w11] [w13] Forty four studies were excluded because they were reviews or commentaries, 12 did not study croup, nine had inadequate randomisation strategies, four were retrospective studies, two had no control group, one had no outcome of interest, and one was a duplication. Therefore, 24 studies were included (references and full details of these studies can be found in table A on the *BMJ*'s website). The weighted κ score between two reviewers was 0.89, indicating substantial agreement.

Twenty two of the included studies had been published in English, one in French, and one in Spanish. Dexamethasone was evaluated in 17 trials, budesonide in nine, and methylprednisolone in three; some studies examined more than one drug. Five of the trials compared active treatments; 19 were placebo controlled. The mean age of the children in the different studies ranged from 13 months to 45 months; the minimum age was 4 months and the maximum was 12 years. Fourteen trials were conducted on inpatients, and 10 were conducted on outpatients. However, studies tended to be small with a median of 40 (interquartile range 36 to 60) participants. The pooled baseline rates using fixed effect models were reported.

Quality assessment of trials

The intraclass correlation between two reviewers was 0.63 for the Jadad scale, 0.98 for allocation concealment, and 1.0 for sponsorship, indicating at least substantial agreement in all cases. The median Jadad score was 3 (interquartile range 2.75 to 4) or 60% (55% to 80%) for the best quality of reporting. Allocation concealment was adequate in 11 (46%) of the studies, inadequate in one (4%), and unclear in 12 (50%). Pharmaceutical sponsorship was identified in three (13%) studies, support was from other sources in three (13%), and not mentioned in 18 (75%). Overall, the quality of studies was better than has been observed for other diseases.[9 20 21]

Croup score

The most frequent outcome utilised in 13 studies was the clinical croup score based on a 17 point ordinal scale developed by Westley.[19] Other scoring systems, none of which have been validated, were utilised in five studies; in six studies no clinical score was reported.

The improvement in the Westley croup score at 6 hours was 2.8 (95% confidence interval 2.2 to 3.5) for dexamethasone or budesonide versus 1.0 (0.3 to 1.7) for placebo. The difference in improvement in the Westley score between treatment arms at 6 hours was 1.6 (1.1 to 2.2). The pooled standardised effect size was 1 (0.6 to 1.5) at 6 hours and 1 (0.4 to 1.6) at 12 hours. From our data, a standard effect size of 1.2 (0.7 to 1.7) corresponded with an improvement of 1.6 (1.1 to 2.2) in a Westley score (fig 1) (see appendix 2 on the *BMJ*'s
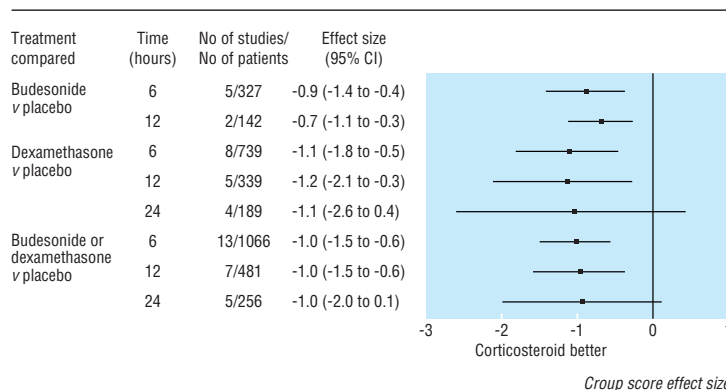

Treatment compared	Time (hours)	No of studies/ No of patients	Effect size (95% CI)
Budesonide v placebo	6	5/327	-0.9 (-1.4 to -0.4)
	12	2/142	-0.7 (-1.1 to -0.3)
Dexamethasone v placebo	6	8/739	-1.1 (-1.8 to -0.5)
	12	5/339	-1.2 (-2.1 to -0.3)
	24	4/189	-1.1 (-2.6 to 0.4)
Budesonide or dexamethasone v placebo	6	13/1066	-1.0 (-1.5 to -0.6)
	12	7/481	-1.0 (-1.5 to -0.6)
	24	5/256	-1.0 (-2.0 to 0.1)

Croup score effect size

Fig 1 Pooled effect sizes (95% confidence intervals) of glucocorticoid treatment for croup versus placebo. All estimates had significant heterogeneity among trials

website for a list of included trials). This change was not significant at 24 hours; however, fewer patients were evaluated at 24 hours and hence the lack of significance may be a reflection of a lack of statistical power. The magnitude of change of −1 is similar to that seen at earlier evaluation points but the 95% confidence interval crosses 0. A decrease in effect size of 1 from baseline is thought to be a clinically important change.

At 6 hours, the difference in risk was 15% (95% confidence interval 2% to 28%) with a number needed to treat of 7 (4 to 50). The baseline rate of clinical improvement was 41% (32% to 50%). At 12 hours the risk difference was 21% (9% to 33%) with a number needed to treat of 5 (1 to 11). The baseline rate of clinical improvement was 68% (58% to 77%). At 24 hours, the risk difference was 12% (3% to 22%) and the number needed to treat was 8 (5 to 33). The baseline rate of clinical improvement was 83% (75% to 91%). Although not all studies contributing to the effect size expressed their results as improved versus not improved, the degree of benefit of a number needed to treat of 5 to 7 patients (at different assessment times) would be sufficient to support the use of glucocorticoids over placebo.

Additional interventions

There was no significant increase in the use of antibiotics among those treated with glucocorticoids as compared with those treated with placebo when expressed as the difference in risk. This was consistent for the dexamethasone group (4%, −20% to 27%) and the budesonide group (−2%, −17% to 13%). There was a significant decrease noted in the use of adrenaline in the glucocorticoid groups with a difference in risk of −9% (−16% to −2%) in the budesonide group (number needed to treat 10; baseline rate 16%) and −12% (−20% to −4%) in the dexamethasone group (number needed to treat 8; baseline rate 23%). There was no significant impact on the use of supplemental glucocorticoids among either those treated with dexamethasone (4%, −4% to 13%) or those treated with budesonide (−15%, −32% to 2%).

When any glucocorticoid was compared with placebo (11 studies, 1150 patients) there was no significant change in the rate of difference of intubation or
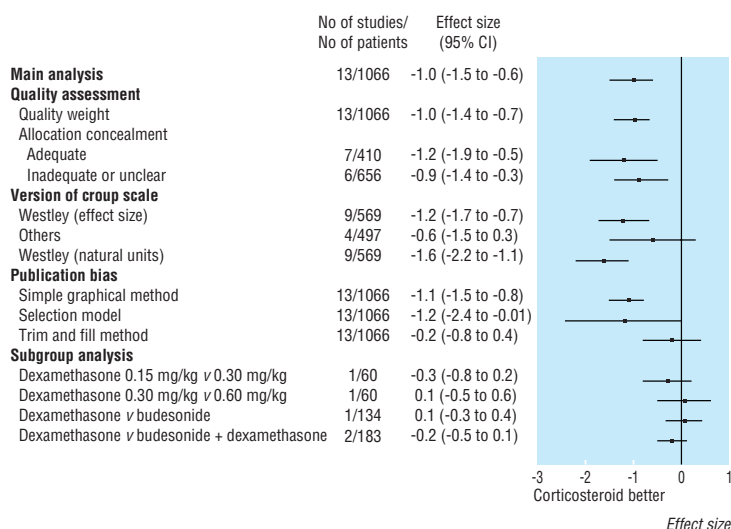
	No of studies/ No of patients	Effect size (95% CI)	
Main analysis	13/1066	-1.0 (-1.5 to -0.6)	
Quality assessment			
Quality weight	13/1066	-1.0 (-1.4 to -0.7)	
Allocation concealment			
Adequate	7/410	-1.2 (-1.9 to -0.5)	
Inadequate or unclear	6/656	-0.9 (-1.4 to -0.3)	
Version of croup scale			
Westley (effect size)	9/569	-1.2 (-1.7 to -0.7)	
Others	4/497	-0.6 (-1.5 to 0.3)	
Westley (natural units)	9/569	-1.6 (-2.2 to -1.1)	
Publication bias			
Simple graphical method	13/1066	-1.1 (-1.5 to -0.8)	
Selection model	13/1066	-1.2 (-2.4 to -0.01)	
Trim and fill method	13/1066	-0.2 (-0.8 to 0.4)	
Subgroup analysis			
Dexamethasone 0.15 mg/kg v 0.30 mg/kg	1/60	-0.3 (-0.8 to 0.2)	
Dexamethasone 0.30 mg/kg v 0.60 mg/kg	1/60	0.1 (-0.5 to 0.6)	
Dexamethasone v budesonide	1/134	0.1 (-0.3 to 0.4)	
Dexamethasone v budesonide + dexamethasone	2/183	-0.2 (-0.5 to 0.1)	

-3   -2   -1   0   1
Corticosteroid better

Effect size

Fig 2 Sensitivity and subgroup analyses of change in croup score from baseline at 6 hours. Negative effect sizes indicate relative improvement with glucocorticoid treatment

tracheotomy -2% (-14% to 10%; baseline rate 3.2%, 2.9% to 3.5%).

Hospitalisation

Overall, a significantly shorter time was spent in accident and emergency when children were treated with a glucocorticoid as compared with placebo (5 studies, 596 patients); the weighted mean difference was -11 (-18 to 4) hours. For inpatients, the difference was -16 (-31 to 1) hours.

There was a non-significant decrease of -16% (-39% to 6%) in the rate of hospitalisation for patients treated with budesonide versus patients treated with placebo (baseline rate 32%, 24% to 39%). This was also true for patients treated with dexamethasone as compared with patients treated with placebo (-2%, -31% to 5%) or if any glucocorticoid was compared with placebo (-14%, -12% to 5%). The more conservative random effects model was used to derive the overall estimate of the difference in hospitalisation rates because there was significant heterogeneity between studies. If the fixed effects model estimate was used there was a significant decrease in hospital admissions between patients treated with budesonide and those treated with placebo (-15%, -20% to -10%).

Sensitivity and subgroup analyses

The sensitivity analysis showed that the method of scoring the severity of croup was important (fig 2). An effect size of -1.2 (-1.7 to -0.7) was identified when the Westley croup score was used (9 studies, 569 patients) as compared with an effect size that was 50% smaller when other croup scores were used (4 studies, 497 patients; -0.6, -1.5 to 0.3), a size that was no longer significant. The Westley score is the only method that has undergone validation and reliability testing and been shown to be sensitive to important changes in a patient's clinical status. The smaller treatment effect noted with non-Westley scores could be the result of sensitivity to change or perhaps a greater degree of variability caused by low reliability.

We were unable to compare the route of administration of glucocorticoids in a meaningful way because of the lack of standardisation of scores between studies. The quality weighting of the effect size did not change the estimate or the width of the 95% confidence interval; this is in part explained by the high methodological quality of the studies. The estimate derived from studies in which allocation was adequately concealed was -1.2 (-1.9 to -0.5) and for the studies in which it was inadequately concealed or in which it was unclear was -0.9 (-1.4 to -0.3). These differences are probably not clinically or statistically significant.

Publication bias

We identified a marked publication bias, and there is also the possibility that small studies that showed that glucocorticoids had no effect were suppressed from publication. There was a significant correlation between treatment effect and sample size (for example, rank correlation test P=0.013; graphical method P=0.004). The Dear-Begg estimate of this correlation was 0.29. Pooled effect size at 6 hours calculated using the simple graphical method was -1.1 (-1.5 to -0.8); with the selection model it was -1.2 (-2.4 to -0.01); and with the trim and fill method it was -0.2 (-0.8 to 0.4). The trim and fill method suggested that seven small trials were suppressed because their results were not significant.

Discussion

Efficacy of steroids

This meta-analysis has shown that treatment with glucocorticoids is effective in improving symptoms of croup in children by as early as 6 hours and for up to at least 12 hours after treatment. This is shown by the significant improvement in scores of croup severity, shorter hospital stays, and the fact that adrenaline was used less often as an additional intervention. Although the decrease in the rate of hospitalisation was not significant, this outcome criterion varies from hospital to hospital and the direction of the change was towards effectiveness. The degree of benefit identified would merit the use of glucocorticoids, since 5 to 7 patients would need to be treated with glucocorticoids for one patient to experience a significant improvement in symptoms.

This finding did not change when the quality of the studies included was incorporated into our pooled estimate. We found a significant improvement even though almost half of the patients included were assessed using scoring tools that have not been validated and may be less sensitive to important changes in the patient's clinical status.

Publication bias

Of more importance is the fact that publication bias seems to be a modifier of this result, and it is likely that our analysis did not include smaller studies that had statistically negative results. Publication bias is an important threat to the validity of systematic reviews and is difficult to combat except through the registration of all randomised controlled trials on human participants. The existence of this bias suggests

that this meta-analysis overestimates the effectiveness of treatment with glucocorticoids. The results indicate that the number needed to treat at 12 hours is 5 patients for one patient to experience improvement. If publication bias exists and has exaggerated the benefit of treatment, then the number needed to treat would be greater. Thus, clinicians will have to decide whether it is still worth treating patients for croup. Considering the comparative safety and low cost of dexamethasone, it probably makes sense to continue using glucocorticoids. In cases in which the effect of adopting treatment with glucocorticoids has been examined, there has been evidence for a decline in hospital admission rates, fewer admissions to the intensive care unit, and shorter lengths of stay.[22 23]

The small numbers of patients in each study and confounding variables make it difficult to make definitive recommendations regarding the superiority of any glucocorticoid, dose, or route of administration. In the absence of further evidence, an oral dose of dexamethasone, probably 0.6 mg/kg, should be preferred because of its safety and efficacy. In a child who is vomiting, nebulised budesonide or intramuscular dexamethasone might be preferable.

Our results are mostly consistent with those of the meta-analysis by Kairys et al which found that glucocorticoids are beneficial in patients with croup[7] but there are some important differences. Because of the lower probability of bias in such studies we included only randomised controlled trials; hence some studies that were included by Kairys et al were not included in our meta-analysis. These excluded studies tended to be older and used techniques of quasirandomisation, such as alternate allocation.[24–26] Additionally, 15 randomised controlled trials on this topic have been published since 1989, many of them outpatients trials examining the effectiveness of budesonide or dexamethasone. The differences between inclusion criteria in our meta-analysis and that by Kairys et al may account for Kairys et al's finding that glucocorticoids significantly decrease the risk of intubation, whereas we did not observe a significantly decreased risk.

Study quality
The quality of the studies included was good. The median Jadad score in our study was 3; in other studies the median is often 2 or less.[20 21] In 46% of the trials allocation was adequately concealed whereas in most other studies about 10% to 15% of the trials being assessed have adequate concealment.[9 20] Although quality assessment and methods of its incorporation into systematic reviews are controversial, we have recently shown the importance of such assessments in detecting bias and have proposed a method of quality weighting.[20]

Outcome measures
Outcome measures are important in detecting significant change in a patient's clinical status. It is important that this measure is valid (that it measures what it ought to be measuring) and responsive (that it is sensitive to change).[27] This meta-analysis supports the importance of using valid and responsive outcome measures since the magnitude of the effectiveness of the treatment in this study was dependent on which scoring method

---

Key messages

- Most trials evaluating the treatment of croup are of high methodological quality and hence have a low risk of bias

- Publication bias, however, seems to be a problem, making the results of this meta-analysis somewhat less certain

- Glucocorticoids seem to bring about clinical improvement in children with croup within 6 hours

- Nebulised budesonide or dexamethasone, given either orally or intramuscularly, is effective in treating croup

- The use of glucocorticoids is associated with a lower rate of use of cointerventions and shortens the time spent in hospital

---

was used. We have shown that the Westley score is valid, responsive, and a reliable measure.[28] Although further validation and modification could be made to the Westley score, it should remain the primary outcome measure in trials currently being conducted.

Future trials may want to explore which dose of dexamethasone is most effective: is 0.15 mg/kg really as effective as 0.6 mg/kg? This meta-analysis supports the use of glucocorticoids to treat any patient with croup who has any signs of respiratory distress.

1 Henrickson KJ, Kuhn SM, Savatski LL. Epidemiology and cost of infection with human parainfluenza virus types 1 and 2 in young children. *Clin Infect Dis* 1994;18:770-9.
2 Denny FW, Murphy TF, Clyde WA, Collier AM, Henderson FW. Croup: an 11 year study in pediatric practice. *Pediatrics* 1983;71:871-6.
3 To T, Young W. Hospitalizations for croup in Ontario [abstract]. *Clin Invest Med* 1994;17:A25.
4 Marx A, Torok TJ, Holman RC, Clarke MJ. Pediatric hospitalizations for croup (laryngotracheitis): biennal increases associated with human parainfluenza virus 1 epidemics. *J Infect Dis* 1997;176:1423-7.
5 Skolnik NS. Treatment of croup: a critical review. *Am J Dis Child* 1989;143:1045-9.
6 Waiisman Y, Klein BL, Boenning DA, Young GM, Chamberlain JM, O'Donnell R, et al. Prospective randomized double-blind study comparing L-epinephrine and racemic epinephrine aerosols in the treatment of laryngotracheitis (croup). *Pediatrics* 1992;89:302-6.
7 Kairys SW, Olmstead EM, O'Connor GT. Steroid treatment of laryngotracheitis: a meta-analysis of the evidence from randomized trials. *Pediatrics* 1989;83:683-93.

8 Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12.
9 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
10 Cho MK, Bero LA. The quality of drug studies published in symposium procedings. *Ann Intern Med* 1996;124:485-9.
11 Bartko JJ. General methodology. II. Measures of agreement: a single procedure. *Stat Med* 1994;13:737-45.
12 Follman D, Elliott P, Sul I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. *J Clin Epidemiol* 1992;45:769-73.
13 Hedges LV, Olkin I. *Statistical methods for meta-analysis*. Orlando, FL: Academic Press, 1995.
14 Begg CB, Mazundar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:1088-1101.
15 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
16 Light RS, Pillemar DB. *Summing up: the science of reviewing research*. Cambridge, MA: Havard University Press, 1984.
17 Hedges LV. Modelling publication selection effects in meta-analysis. *Stat Sci* 1992;7:246-55.
18 Taylor S, Tweedie R. *Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analyses*. Fort Collins, CO: Colorado State University, 1998.
19 Westley CR, Cotton EK, Brook JG. Nebulized racemic epinephrine by IPPB for the treatment of croup: a double-blind study. *Am J Dis Child* 1978;132:484-7.
20 Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does the poor quality of reports of randomized trials exaggerate estimates of intervention effectiveness reported in meta-analysis? *Lancet* 1998;352:609-13.
21 Moher D, Fortin P, Jadad AR, Juni P, Klassen T, LeLorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;347:363-6.
22 Klassen TP. Glucocorticoids in the treatment of croup: barking up the right tree. *Can Med Assoc J* 1998;159:1121-2.
23 Geelhoed GC. Sixteen years of croup in a Western Australian teaching hospital: effects of routine steroid treatment. *Ann Emerg Med* 1996;28:621-6.
24 Martensson B, Nilson G, Torbjar J. The effect of corticosteroids in the treatment of pseudo-croup. *Acta Otolaryngol* 1960;158(suppl):62-71.
25 Novik A. Corticosteroid treatment of non-diptheritic croup. *Acta Otolaryngol* 1960;158(suppl):20-3.
26 Sussman S, Grossman M, Magoffin R, Schieble J. Dexamethasone (16 alpha-methyl, 9 alpha-fluoroprednisolone) in obstructive respiratory tract infections in children. *Pediatrics* 1964;34:851-5.
27 Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992;45:1341-5.
28 Klassen TP, Rowe PC. The croup score as an evaluative instrument in clinical trials [abstract]. *Arch Pediatr Adolesc Med* 1995;149:60.

(Accepted 2 June 1999)

Predictors and consequences of unemployment among construction workers: prospective cohort study

Päivi Leino-Arjas, Juha Liira, Pertti Mutanen, Antti Malmivaara, Esko Matikainen

Department of Epidemiology and Biostatistics, Finnish Institute of Occupational Health, FIN-00250 Helsinki
Päivi Leino-Arjas
senior researcher
Pertti Mutanen
statistician

Work Ability Centre, Finnish Institute of Occupational Health
Juha Liira
senior consultant

Department of Occupational Medicine, Finnish Institute of Occupational Health
Antti Malmivaara
assistant chief medical officer

Local Government Pensions Institution Finland, Albertinkatu 34, FIN-00100, Helsinki
Esko Matikainen
chief medical officer

Correspondence to: P Leino-Arjas plei@occuphealth.fi

Abstract

Objective To study predictors and consequences of unemployment.
Design Prospective cohort study.
Setting 11 construction companies in southern Finland.
Participants 586 male employees, aged 40-59 years at baseline in 1991 and not retired during a 4 year follow up.
Main outcome measures Long term unemployment, stress symptoms, disease, alcohol consumption, exercise activity, and body mass index.
Results In a multiple logistic regression model, long term unemployment (> 24 months $v \leq 24$ months) was predicted by age 50-54 years v 40-44 years (odds ratio 2.0, 95% confidence interval 1.1 to 3.7), ≤ 3 years' employment in the present job (3.1, 1.9 to 5.1), previous unemployment (2.1, 1.2 to 3.8), being single (1.8, 1.1 to 3.1), current smoking (2.6, 1.4 to 4.7), high alcohol consumption (2.1, 1.1 to 4.3), body mass index < 23 kg/m^2 v 23-29 kg/m^2 (2.4, 1.3 to 4.4), frequent stress symptoms (2.0, 1.2 to 3.2), mental disorders (7.8, 1.5 to 40.0), and skin diseases (2.0, 1.0 to 3.9). Workers who were unemployed long term reported increased stress (2.1, 1.2 to 3.5) but fewer incident diseases (0.6, 0.4 to 0.9), decreased alcohol consumption (2.9, 1.6 to 5.2), increased exercise (1.9, 1.2 to 3.0), and increased body mass index (2.3, 1.3 to 4.0) compared with the rest of the cohort.
Conclusions The workers' perceptions of work did not predict unemployment. Health based selection to long term unemployment was shown. Smoking and high alcohol consumption predated unemployment, but favourable lifestyle changes were observed among the unemployed workers. Stress predicted unemployment, which further increased the stress.

Introduction

A relation between unemployment and ill health seems established: several studies reported mental distress,[1-3] frequency of use of health care,[4-6] and mortality[5-8] to be high among unemployed individuals. The mechanisms of this relation are not, however, obvious. It is possible that *(a)* poor health leads to a low status on the labour market and increases the risk of unemployment, *(b)* job loss and unemployment have adverse effects on health, or *(c)* both mechanisms are active.[9]

The ideal way to investigate predictors and consequences of unemployment would be to conduct a prospective study of a cohort of employed people at baseline, with a subgroup becoming unemployed, and who were followed up over a long period. Studies on the effects of factory closure provide an option, but a suitable comparison group is difficult to find and potential participants are usually alerted to the possibility of closure around the same time as the researchers.[1 10] Socioeconomically disadvantaged people are at a high risk of both health problems and unemployment; it has not always been possible to control for this source of confounding. It seems plausible that characteristics apart from health—for example, lifestyle or attitudes towards work and family—would play a role in unemployment or re-employment.

We studied some predictors of unemployment in a cohort of construction workers during Finland's recession in the early 1990s. We also studied changes in health and lifestyle with duration of unemployment