eMethods

1. *Study population (additional information)*

During the study period prophylactic corticosteroids were not routinely administered to cardiac surgery patients in the LUMC. Data on administration of prophylactic corticosteroids were extracted from the automated registration system for the operating room in which all administered medications were registered. Of the 476 patients in this study, 115 received prophylactic corticosteroids. Of the 361 patients who did not receive prophylactic corticosteroids, 73 did receive corticosteroids during the surgical procedure or as treatment of protamine-allergy at the end of surgery. After surgical intervention all patients were admitted to the cardio-thoracic intensive care unit (ICU).

2. *Data extraction and study end-points*

Data were extracted from electronic patient record databases, routinely used in the operating room and in the ICU (Metavision®, Mirador®) in which clinical parameters are collected automatically. In case data was missing in these electronic records, data was extracted from paper patient charts. These were kept simultaneously during the conversion phase from paper patient charts to electronic patient records. We extracted data on demographic features and type of surgical intervention. Furthermore, the logistic EuroSCORE, routinely computed and registered by the thoracic surgery department, was obtained for all patients. This a validated prognostic score of in-hospital mortality related to cardiac surgery, based on patient-related factors (age, sex, chronic pulmonary disease, extra cardiac arteriopathy, neurological dysfunction, previous cardiac surgery, serum creatinine, active endocarditis, critical pre-operative state), cardiac-related factors (unstable angina, left ventricular dysfunction, recent myocardial infarct, pulmonary hypertension) and operation-related factors (emergency, other than isolated CABG, surgery on thoracic aorta, post infarct septal rupture).[1;2] The following clinical study end-points were recorded: 30-day mortality, ventilation time, duration of ICU and hospital stay. These study end-points are collected and checked systematically on a weekly basis by a quality manager and by the hospital billing department. The following clinical parameters were extracted from the electronic patient records and recorded for the study: highest necessary dose of norepinephrine, highest glucose value and highest leukocyte count in the first 24 hours after intervention. The

occurrence of atrial fibrillation, infections, heart failure or delirium during hospital stay was also extracted. Infection was defined as clinical symptoms requiring new antibiotic treatment; heart failure was defined as a clinical diagnosis requiring additive diuretic, or invasive supportive (intra-aortic balloon pump, assist device) treatment; delirium was defined as the need for haloperidol.

3. Main assumptions for instrumental variable analyses

In order to be valid, an instrumental variable should fulfill three main assumptions, which we will discuss specifically applied to our study (see also eFigure 1). The first assumption is that anesthesiologist's preference affects the probability that a patient receives corticosteroids. The second assumption is that the anesthesiologist's preference for corticosteroids does not affect the outcome in other ways than through the decision of whether to administer corticosteroids (exclusion restriction); the third is that the anesthesiologist's preference for corticosteroids is not related to characteristics of his patient population (independence assumption).[3;4] The difference in outcomes can then be attributed entirely to the difference in the probability of receiving corticosteroids (based on the anesthesiologist's preference).

We explored whether there was variation in corticosteroid administration amongst anesthesiologists and whether this seemed independent of their patient population. The lower part of eFigure 2 shows the proportion of patients to whom the anesthesiologists administered prophylactic corticosteroids; the upper part shows box plots of the EuroSCORE of these patients. The percentage of patients to whom the anesthesiologists administered corticosteroids showed considerable variation, ranging from 0% to 63%. In our data, there is no consistent pattern in the EuroSCORE with increasing prescription of corticosteroids (in accordance with the independence assumption), giving general reassurance that we could use anesthesiologist's preference as an instrumental variable

4. Instrumental variable selection

In our study population there was large variation among anesthesiologists in frequency of administration of prophylactic corticosteroids, ranging from 0% to 63%. This indicated that

anesthesiologist's preference regarding administration of prophylactic corticosteroids was a potentially suitable instrument. We considered several estimates of anesthesiologist's preference for use as an instrument, based on one, two, five, ten or all previous patients. For a given patient the proportions of these preceding patients who received prophylactic corticosteroids were calculated, to provide estimates of the anesthesiologist's relative preference for prophylactic corticosteroids at the time of the treatment decision for this specific patient.

To identify which of our candidate instruments was most strongly related to treatment, we carried out the first stage of the two-stage least squares instrumental variable regression only, by means of linear regression of the treatment on the candidate instrument.[7] We selected the strongest instrumental variable based on the F-statistic and partial r^2 of the first stage of the two-stage least squares regression and on the range of predicted probabilities of treatment. An F-statistic greater than 10 suggests that small sample bias is negligible and that the instrument is therefore sufficiently strong.[8] The partial r^2 indicates which proportion of the variance of the treatment is explained by the instrumental variable.[9]

eTable 1 displays the regression coefficients, the F-statistic and the partial r^2 for the first stage regression using each of the candidate instruments. The regression coefficient can be interpreted as follows for the instrument based on the last patient only: for a patient treated by an anesthesiologist who administered corticosteroids to the previous patient the probability of receiving corticosteroids was 0.28 higher than for a patient treated by an anesthesiologist who did not administer corticosteroids to the previous patient. Analogously, for a patient treated by an anesthesiologist who administered corticosteroids to all previous patients the expected probability of receiving corticosteroids would be 0.82 higher than for a patient treated by an anesthesiologist who administered corticosteroids to none of their previous patients. The strengths of instrumental variables based on 10 previous prescriptions or all previous prescriptions were very similar, with a partial r^2 of 0.21 and 0.22 and F-statistics of 131 and 126 respectively. These instruments were considerably stronger than the instruments based on just one or two previous prescriptions. Although the partial r^2 and F-statistic were slightly higher for the instrument based on 10 previous prescriptions than for the instrument based on all previous

prescriptions, the range of predicted probabilities of treatment was slightly larger for the latter instrument. We therefore selected the proportion of all previous patients who received prophylactic corticosteroids for use as an instrument in subsequent analyses.

5. Conventional statistical analyses

Crude analysis

For continuous outcomes we calculated a mean difference (MD) with 95% confidence interval (CI) between treatment groups. For binary outcomes we calculated a risk difference (RD) with 95% CI, because this effect measure can be compared directly to two-stage least squares instrumental variable results. Ventilation time in hours and duration of ICU and hospital stay in days were dichotomized (as shorter or longer than the median). Robust standard errors were used for dichotomous outcomes.

Multivariable model and propensity score adjusted analyses

The above analyses were repeated using multivariable adjustment and propensity score adjustment. The multivariable model was adjusted for age, sex, diabetes mellitus, EuroSCORE and type of surgical procedure, for the 470 patients with information on all included covariates. Operating surgeon was not included in the multivariate regression models because for many outcomes there were few events, limiting the number of covariates that can be included in the regression model. The propensity score was calculated by first performing a logistic regression model with receipt of prophylactic corticosteroids as the dependent variable and all variables used in the multivariable model plus the operating surgeon as covariates and then predicting the probabilities of treatment for each patient based on this model. This was done for the 464 patients with information on all variables.

6. Sensitivity analyses

(1) an instrumental variable analysis adjusted for age, sex, EuroSCORE, type of intervention and diabetes, to explore the effect of additional adjustments.

(2) an instrumental variable analysis using an alternative instrument based on treatment of the previous 5 patients only, which might accommodates preference changes over time better than an instrument based on all previous patients.

(3) an analysis in which we replaced the second stage of the instrumental variable regression with a generalized linear model with a log-link, which gives relative risk estimates, because two-stage least squares regression is based on linear models and may pose problems if exposures and outcomes are dichotomous, including predicted values below 0 or above 1.

7. Timeline of analyses relative to the DECS randomized trial[10]

All major decisions about patient selection, choice of instrument, outcomes and types of analysis were made before we knew the DECS trial results, to which we compared our results. After the trial was published we performed additional analyses with cut-off points similar to those in the trial for a better comparison.

Reference List

(1) Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999;16:9-13.

(2) Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. *Eur Heart J* 2003;24:881-882.

(3) Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006;17:268-275.

(4) Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol* 2009;62:1226-1232.

(5) Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006;17:360-372.

(6) Swanson SA, Hernan MA. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013;24:370-374.

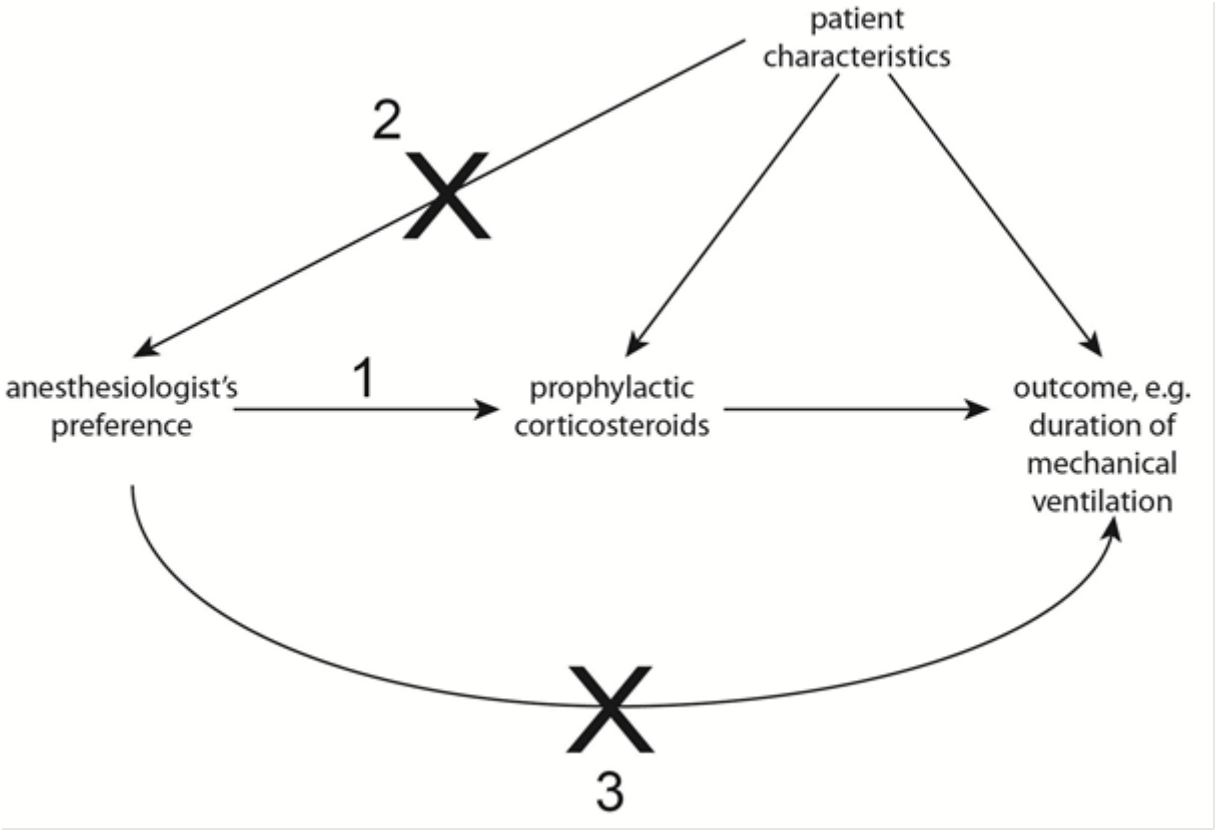
(7) Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010;19:537-554.

(8) Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology* 2006;17:260-267.

(9) Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009;62:1233-1241.

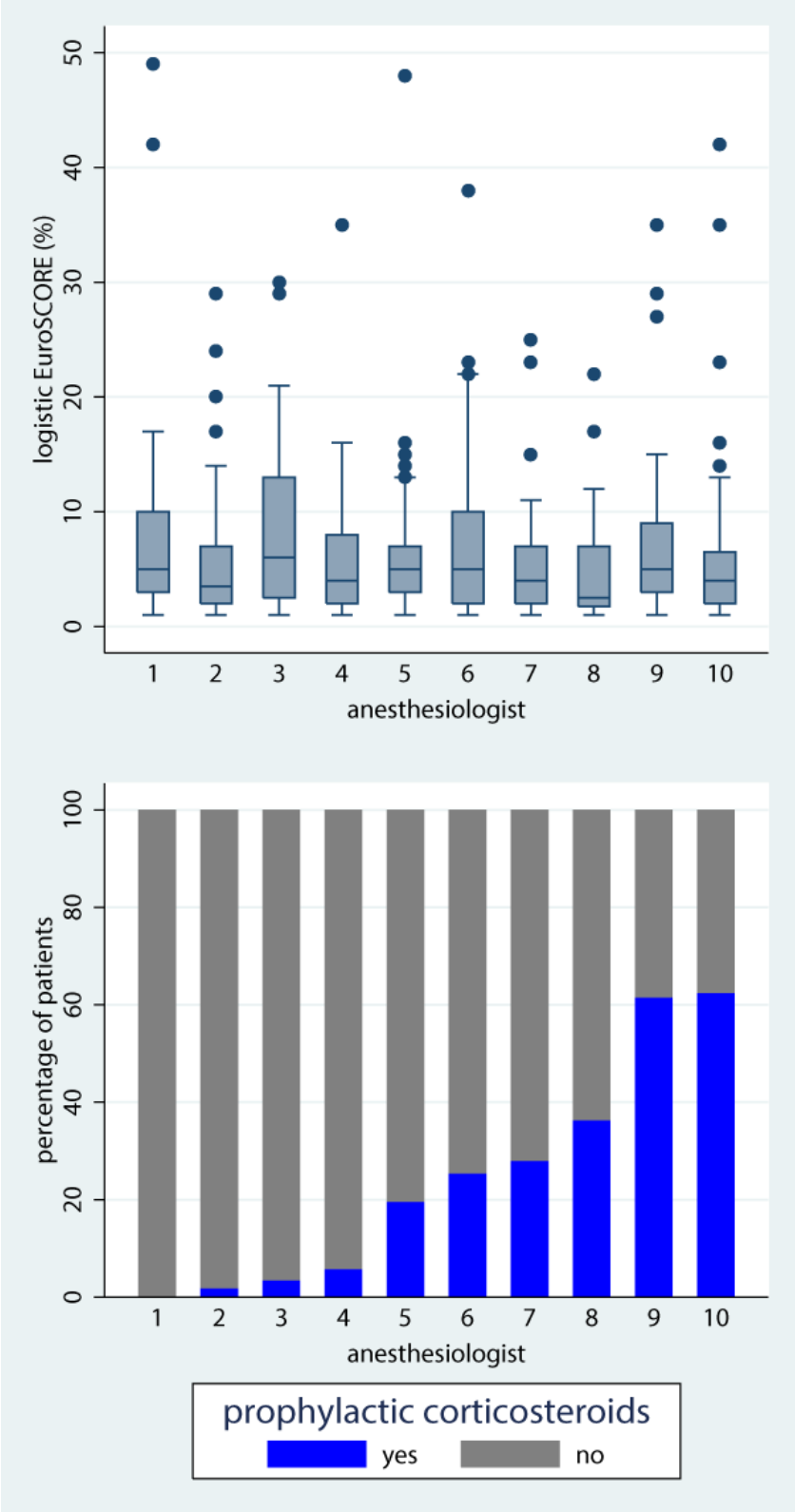
(10) Dieleman JM, Nierich AP, Rosseel PM et al. Intraoperative high-dose dexamethasone for cardiac surgery: a randomized controlled trial. *JAMA* 2012;308:1761-1767.

eFigure 1. Instrumental variable assumptions in this study.



Causal diagram depicting instrumental variable assumptions if anesthesiologist's preference is used as an instrumental variable in a study investigating the effect of prophylactic corticosteroids in cardiac surgery patients.

eFigure 2.



Proportion of patients to whom an anesthesiologist administered prophylactic corticosteroids (lower part) and distribution of the EuroSCORE of these patients (upper part).

eTable 1. Strength of instruments based on 5 different preference assignments.

Instrument	Difference in probability of treatment (95% CI)*	*F*-statistic	Partial r^2
Previous patient†	0.28 (0.19-0.37)	39	0.08
Last 2 patients‡	0.47 (0.36-0.58)	74	0.14
Last 5 patients‡	0.68 (0.56-0.80)	117	0.20
Last 10 patients‡	0.77 (0.64-0.91)	131	0.22
All previous patients‡	0.82 (0.67-0.96)	126	0.22

†Administration of prophylactic corticosteroids in the previous patient of the same anesthesiologist.

‡Proportion of the last 2/ last 5/ last 10/ all previous patients of the same anesthesiologist who received prophylactic corticosteroids.

*All differences stated are regression coefficients and represent the difference in the probability of receiving prophylactic corticosteroid between patients with values of the instrument of 1 and 0. For the instrument based on the previous patient 1 indicates that the previous patient received the treatment and 0 denotes that the previous patient did not receive the treatment. For the instrument based on all previous patient 1 would denote that all previous patients received the treatment, and 0 would denote that none of the previous patients received the treatment.

eTable 2. Outcomes by treatment status and estimates of the treatment effect from four different analyses.

	No prophylactic corticosteroids (n=361)	Prophylactic corticosteroids (n=115)	Crude	Multivariable model	Propensity score adjusted	Instrumental variable
Primary outcome†						
Mortality (30 days)	10 (2.8)	4 (3.5)	0.7 (-3.1,4.5)	-0.7 (-4.3,2.9)	-1.0 (-4.5,2.6)	-2.6 (-8.4,3.2)
Ventilation time >11 hrs	158 (46)	52 (49)	3.1 (-7.8,14.1)	-1.8 (-11.8,8.1)	-2.0 (-12.8,8.8)	-28.1 (-52.4,-3.9)
ICU stay >1 day	169 (47)	63 (55)	7.9 (-2.6,18.5)	2.5 (-7.6,12.6)	3.0 (-7.7,13.8)	-18.4 (-41.8,5.0)
Hospital stay >8 days	139 (39)	50 (44)	5.0 (-5.4,15.5)	0.4 (-9.7,10.5)	0.4 (-10.2,11.0)	-22.7 (-44.9,-0.5)
Ventilation time >24 hrs*	60 (17)	21 (20)	2.4 (-6.3,11.0)	-1.1 (-8.9,6.6)	-1.9 (-9.8,6.0)	-16.3 (-33.2,0.5)
ICU stay>2 days*	115 (32)	43 (38)	5.5 (-4.7,15.7)	1.9 (-8.1,11.9)	1.7 (-8.7,12.0)	-16.2 (-37.2,4.8)
Clinical parameters						
Highest norepinephrine dose > 0.1µg/kg/min †	112 (33)	35 (32)	-0.8 (-11.0,9.4)	-5.6 (-15.2,4.0)	-4.4 (-14.5,5.7)	-27.1 (-47.9,-6.2)
Highest glucose (mmol/l) ‡	10.4 (0.14)	11.4 (0.25)	0.96 (0.40,1.52)	0.82 (0.27,1.37)	0.83 (0.27,1.40)	0.94 (-0.20,2.08)
Highest leukocyte count (10^9/L) ‡	13.4 (0.23)	15.6 (0.51)	2.29 (1.30,3.27)	2.30 (1.29,3.30)	2.33 (1.32,3.35)	3.01 (1.01,5.00)
Complications †						
Atrial fibrillation	173 (48)	50 (44)	-4.5 (-15.0,6.1)	-4.2 (-14.7,6.4)	-4.4 (-15.3,6.5)	5.4 (-17.4,28.1)
Infection	52 (15)	15 (13)	-1.6 (-8.8,5.6)	-4.1 (-11.3,3.1)	-4.9 (-12.2,2.4)	-14.1 (-30.1,1.8)
Heart failure	48 (13)	22 (19)	5.8 (-2.3,13.8)	1.9 (-5.7,9.5)	2.6 (-5.1,10.5)	-9.5 (-24.8,5.9)
Delirium	54 (15)	20 (18)	2.6 (-5.4,10.6)	1.9 (-5.9,9.6)	1.2 (-6.9,9.3)	-6.9 (-23.4,9.6)

† n (%), risk difference in % (95% CI)
‡ mean (SE), mean difference (95% CI)
*Additional analyses for comparison to RCT results.

eTable 3. Sensitivity analyses.

	Instrumental variable: unadjusted	Instrumental variable: adjusted[a]	Instrumental variable (last 5)[b]
Primary outcome†			
Mortality (30 days)	-2.6 (-8.4,3.2)	-1.9 (-7.7,4.0)	-5.1 (-10.5,0.3)
Ventilation time >11 hrs	-28.1 (-52.4,-3.9)	-26.1 (-48.3,-4.0)	-28.7 (-54.1,-3.4)
ICU stay >1 day	-18.4 (-41.8,5.0)	-16.6 (-38.4,5.1)	-17.2 (-41.0,6.6)
Hospital stay >8 days	-22.7 (-44.9,-0.5)	-22.7 (-44.6,-0.8)	-27.5 (-50.3,-4.6)
Ventilation time >24 hrs	-16.3 (-33.2,0.5)	-14.7 (-30.3,1.0)	-18.7 (-35.5,-2.0)
ICU stay>2 days	-16.2 (-37.2,4.8)	-13.2 (-33.1,6.7)	-24.0 (-45.7,-2.4)
Clinical parameters			
Highest norepinephrine dose † > 0.1µg/kg/min	-27.1 (-47.9,-6.2)	-26.5 (-45.9,-7.2)	-28.1 (-49.9,-6.4)
Highest glucose (mmol/l)‡	0.94 (-0.20,2.08)	0.82 (-0.28,1.92)	0.40 (-0.78,1.57)
Highest leukocyte count (10^9/L)‡	3.01 (1.01,5.00)	3.15 (1.15,5.15)	3.09 (1.01, 5.16)
Complications †			
Atrial fibrillation	5.4 (-17.4,28.1)	8.4 (-13.9,30.7)	11.7 (-12.1,35.5)
Infection	-14.1 (-30.1,1.8)	-16.1 (-32.3,0.0)	-10.4 (-26.5,5.7)
Heart failure	-9.5 (-24.8,5.9)	-8.6 (-23.1,5.8)	-17.5 (-32.8,-2.2)
Delirium	-6.9 (-23.4,9.6)	-8.2 (-24.6,8.2)	-6.2 (-23.3,10.8)

† risk difference in % (95% CI)
‡ mean difference (95% CI)

a. Instrumental variable analysis adjusted for age, sex, EuroSCORE, type of intervention and diabetes.
b. Instrumental variable analysis using the proportion of the last 5 patients treated with corticosteroids as an instrument.

eTable 4. Relative risk estimates.

	RR (95% CI*)
Primary outcome	
Mortality (30 days)	0.38 (0.03,3.16)
Ventilation time >11 hrs	0.50 (0.25,0.89)
ICU stay >1 day	0.68 (0.42,1.08)
Hospital stay >8 days	0.56 (0.26,0.93)
Ventilation time >24 hrs	0.36 (0.10,1.03)
ICU stay>2 days	0.61 (0.27,1.16)
Clinical parameters	
Highest norepinephrine dose > 0.1μg/kg/min	0.42 (0.18,0.80)
Complications	
Atrial fibrillation	1.12 (0.69,1.71)
Infection	0.33 (0.06,1.21)
Heart failure	0.50 (0.12,1.31)
Delirium	0.62 (0.14,1.84)

Relative risk estimates obtained using a two-stage model with a linear first stage and a generalised linear model with log-link second stage. The instrumental variable used was the proportion of all previous patients treated with corticosteroids. Confidence intervals were obtained using a bootstrap procedure with 1000 samples, bias corrected.

eResults. Comparison to DECS trial results.

The RCT found 3.4% of patients in the dexamethasone group and 4.9% in the placebo group had a ventilation time >24 hours, a difference of -1.5% (95% CI -2.7%,-0.3%). The percentage of patients with an ICU stay >48 hours was 10.2% in the dexamethasone group and 14.0% in the placebo group, a difference of -3.8% (95%CI -5.7%,-1.9%). For atrial fibrillation the percentages were 33.1% and 35.2% respectively, a difference of -2.1% (95% CI -4.9%, 0.7%); for infections 9.8% and 14.8%, a difference of -5.3% (95% CI -7.2%, -3.4%); for delirium 9.2% and 11.7%, a difference of -2.5% (95% CI -4.3%, -0.7%).[1] In general the effects on these outcomes were similar in direction to the results of our instrumental variable analyses, but with considerably smaller effect sizes.

References

(1) Dieleman JM, Nierich AP, Rosseel PM et al. Intraoperative high-dose dexamethasone for cardiac surgery: a randomized controlled trial. *JAMA* 2012;308:1761-1767.

eAppendix. Simulation study to investigate the influence of finite sample bias.

Monte Carlo simulations for a series of study population sizes of 100, 200, 300, 400 and 500. The instrument *P* was generated from the standard uniform distribution U(0,1). An unmeasured confounder C_u was generated from the uniform distribution U(0,1).

Treatment *X* was generated from a binomial distribution with individual patients' probabilities of treatment dependent on *P* and C_u according to the following equation:

$P(X=1|P,C_u)= 0.7P + 0.2C_u$

Binary outcome *Y* was generated from a binomial distribution with individual patients' probabilities of the outcome dependent on treatment *X* and on C_u as follows:

$P(Y=1| X,C_u)=0.2 - 0.1X + 0.7C_u$.

Next, the treatment effect was estimated in each sample using ordinary least squares regression and two-stage least squares regression. The mean estimates and their standard deviation for each sample size across 2000 simulations are displayed in the table below. The mean partial r^2 in the simulations was 0.17, slightly lower than in our study data. Even at sample size 100 the mean 2-SLS is very close to the true effect of -0.10, indicating small sample bias is not a concern. However, the 2-SLS estimates are very variable, as indicated by their large standard deviations.

Sample size	OLS estimates, mean (SD)	2-SLS estimates, mean (SD)
100	-0.054 (0.100)	-0.101 (0.270)
200	-0.051 (0.069)	-0.107 (0.182)
300	-0.053 (0.056)	-0.107 (0.145)
400	-0.052 (0.051)	-0.100 (0.128)
500	-0.054 (0.045)	-0.101 (0.111)

Stata code for simulations

```
*create a file in which to store results
drop _all
clear all
postfile simres ssize b1 b2 pr2 F using "filename", replace

*programme for creating one dataset (called "finite")
drop _all
capture program drop finite
program finite, rclass
drop _all
// ssize = sample size, as a macro
args ssize
//generate patients
set obs `ssize'
gen n=_n
//generate the instrument P
gen P=runiform()
//generation of an unmeasured confounder U
gen U=runiform()
//generation of treatment X
gen PrX= 0.7*P+0.2*U
gen X1 = runiform()
gen X=recode(X1,PrX,1)
recode X (1=0) (else=1)
drop X1
//generation of outcome Y
gen PrY= 0.2-0.1*X+0.7*U
gen Y1 = runiform()
gen Y=recode(Y1,PrY,1)
recode Y (1=0) (else=1)
drop Y1
//ordinary least squares regression
quietly regress Y X
scalar b1 = _b[X]
//two-stage least squares regression
quietly ivreg2 Y (X=P), first
scalar b2 = _b[X]
*also save first stage partial r2 and F-statistic
matrix tmp2 = e(first)
scalar pr2 = tmp2[2,1]
scalar F = tmp2[3,1]
post simres (`ssize') (b1) (b2) (pr2) (F)
end



*run the simulations
foreach ssize in 100 200 300 400 500{
simulate, reps(2000) seed(312): finite `ssize'
                  }

postclose simres

*analyse the results
use "filename", clear
sort ssize
//calculation of mean and standard deviation of the OLS and 2-SLS estimates
//per sample size across 2000 simulations
by ssize: summarize b1 b2 pr2, detail
```