Appendix 1 (as supplied by authors): Power calculation

We assumed that there will be 1,000 patients per year for each hospital group. Then, the Infection and Pulmonary groups will each have 3000 patients (1000 patients during the baseline period and 2000 patients during the 2 year intervention period). We assumed an effect size of 30% decrease in incidence for both outcomes (nosocomial infection and bronchopulmonary dysplasia) from the baseline to the end of the study period.
For the nosocomial infection outcome, we assumed a baseline incidence rate of 25% (based on previous Canadian Neonatal Network data[1]); then for the Infection group, with a 30% decrease in incidence, the nosocomial infection incidence rate at the end of the intervention period should be 17.5%. For Pulmonary group, we assumed that the nosocomial infection incidence rate will not change (i.e. 25% at the beginning and end of the study).
For the bronchopulmonary dysplasia outcome, we assumed a baseline incidence rate of 30%. Then, for the Pulmonary group, with a 30% decrease in incidence, the bronchopulmonary dysplasia incidence rate should be 21% at the end of intervention period. For the Infection group, we assumed that the bronchopulmonary dysplasia incidence rate will not change (i.e. 30% at the beginning and end of the study). See the table below.

time point	Expected Sample Size		Expected Nosocomial Infection Rate		Expected Bronchopulmonary Dysplasia Rate	
	Infection Group	Pulmonary Group	Infection Group	pulmonary Group	Infection Group	Pulmonary group
baseline	1000	1000	25%	25%	30%	30%
end of intervention period	2000	2000	17.5%	25%	30%	21%

To account for the correlation effects of the outcome for patients within the same hospital, we used an intra-cluster correlation coefficient, ICC of 0.01,[2,3] based on the findings of Gulliford et al[4] and Adams et al,[5] to adjust the effective sample size based on the design effect formula: $D = 1 + (m - 1) \cdot ICC$, where m is the average number of patients for each hospital, and D is the design effect (adjustment factor for the effective sample size).
With type I error of 0.05, we calculated the powers of (1) between group comparison, and (2) within group comparison, based on the above assumptions.

Power of between group comparisons: For the between group comparison, power calculation is based on the following logistic regression model:

$$\log\left(\frac{p_{gkt}}{1-p_{gkt}}\right) = \alpha + \beta_g t$$

where p_{gkt} is the probability of *k* -th infant in group *g* admitted at time *t* (*t=0* if the infant was admitted at baseline year, and number of days after the study begins), and β_g is the time slope for group *g* (*g=0* or *1*, with *0* for control group and *1* for intervention group) , and α is the intercept. To account for the correlation of infants admitted to the same hospital, we use ICC=0.01 to adjust the nominal sample size to the effective sample size. The objective of the between group comparison is to detect a significant (p-value<0.05) test of $\beta_1 \neq \beta_0$. Although we have specified the baseline incidence and the target incidence rates, it is not easy to translate the target incidence rates to the values of the two slopes (β_0, and β_1), making it difficult to apply traditional power calculation methods for this situation. Instead, we follow the procedure of Stroup (2002):

(1) First, construct a data set based on the assumptions. Under the assumptions, both control and intervention groups have constant incidence rates (25% for nosocomial infection and 30% for bronchopulmonary dysplasia) throughout the baseline year; and during the 2-year study period, the logit of incidence rate in the intervention group decreases constantly (until 17.5% for nosocomial infection and 21% for bronchopulmonary dysplasia) while the incidence rate of the control group is the same as the incidence rate during the baseline year. For baseline year, we set time *t=0* and determine total number of patients (based on the effective baseline sample size) and total number of nosocomial infection/bronchopulmonary dysplasia patients according to the baseline incidence rate. For the 2-year study period, we separate it into 24 months (*t=1,2,…,24*) and assume that equal number of patients were admitted each month, and then determined for each month, the total numbers of patients (effective sample size for the study period) and total numbers of nosocomial infection/bronchopulmonary dysplasia patients according to the incidence rates for that month, for both intervention and control groups.

Appendix to: Lee SK, Aziz K, Singhal N, et al. Improving the quality of care for infants: a cluster randomized controlled trial. *CMAJ* 2009. DOI:10.1503/cmaj.081727.

(2) Fit the proposed model to the data to obtain the critical value for type I error $\alpha = 0.05$, and to determine the distribution under alternative hypothesis. The critical value $F_{critical}$ depends on type I error ($\alpha = 0.05$): $P_0(F > F_{critical}) = 0.05$, where P_0 denotes the distribution under null hypothesis. In this case, the null hypothesis is: $\beta_1 = \beta_0$, and the distribution under this null hypothesis is a central *F*-distribution). The distribution under alternative hypothesis is a non-central *F*-distribution, with the non-central parameter determined by the estimated effect size: $\hat{\beta}_1 - \hat{\beta}_0$, where $\hat{\beta}_0$ and $\hat{\beta}_0$ are estimated from the data set constructed in step (1).

(3) Finally, the power can be calculated using the critical value and the non-central parameter: $power = P_1(F > F_{critical})$, where P_1 denotes the distribution under alternative hypothesis, which is the non-central *F*-distribution with non-central parameter determined in step (2).

Following the above procedure and assumptions, we obtained the powers of between group comparison for nosocomial infection and bronchopulmonary dysplasia to be 0.89 and 0.83, respectively.

Power for within group comparison: For the within group comparisons, we calculated the power to detect the difference of incidence rates (25% vs 17.5% for nosocomial infection, and 30% vs 21% for bronchopulmonary dysplasia) using tests of two independent proportions. For nosocomial infection, the power is 0.91, and for bronchopulmonary dysplasia, the power of this test is 0.96.

References

1. Lee SK, McMillan DD, Ohlsson A, et al. Variations in outcomes and practices in the Canadian NICU Network: 1996-1997. *Pediatrics* 2000;106:1070-9.
2. Kerry SM, Bland JM. Statistics notes: Sample size in cluster randomisation. *BMJ* 1998;316:549.
3. Kerry SM, Bland JM. Statistical notes: the intra-cluster correlation coefficient in cluster randomization. *BMJ* 1998;316:1455-60.
4. Gulliford MC, Ukoumune OC, Chinn S. Components of variance and intraclass correlation for the design of community based surveys and intervention studies: data from the Health Survey for England 1994. *Am J Epidemiol* 1999;149:876-83.
5. Adams G, Gulliford MC, Ukoumunne OC, et al. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004;57:785-94.

Appendix to: Lee SK, Aziz K, Singhal N, et al. Improving the quality of care for infants: a cluster randomized controlled trial. *CMAJ* 2009. DOI:10.1503/cmaj.081727.