ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 1311−1316

Journal of Clinical Epidemiology

# GRADE guidelines: 9. Rating up the quality of evidence

Gordon H. Guyatt[a,b,*], Andrew D. Oxman[c], Shahnaz Sultan[d], Paul Glasziou[e], Elie A. Akl[f], Pablo Alonso-Coello[g], David Atkins[h], Regina Kunz[i,j], Jan Brozek[a], Victor Montori[k], Roman Jaeschke[b], David Rind[l,m], Philipp Dahm[n], Joerg Meerpohl[o,p], Gunn Vist[c], Elise Berliner[q], Susan Norris[r], Yngve Falck-Ytter[s], M. Hassan Murad[k], Holger J. Schünemann[a,b], The GRADE Working Group[1]

[a]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Room 2C12, 1200 Main Street, West Hamilton, Ontario L8N 3Z5, Canada*
[b]*Department of Medicine, McMaster University, Room 2C12, 1200 Main Street, West Hamilton, Ontario L8N 3Z5, Canada*
[c]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs plass, 0130 Oslo, Norway*
[d]*Division of Gastroenterology, Hepatology, and Nutrition, Department of Medicine, University of Florida, Gainesville, FL, USA*
[e]*Centre for Research in Evidence-Based Practice, Faculty of Health Sciences, Bond University, Gold Coast, Queensland, 4229, Australia*
[f]*Department of Medicine, State University of New York at Buffalo, NY, USA*
[g]*Iberoamerican Cochrane Center-Servicio de Epidemiología Clínica y Salud Pública and CIBER de Epidemiología y Salud Pública (CIBERESP), Hospital de Sant Pau, Universidad Autónoma de Barcelona, Barcelona 08041, Spain*
[h]*QUERI Program, Office of Research and Development, Department of Veterans Affairs, Washington, DC, USA*
[i]*Academy of Swiss Insurance Medicine (asim), University Hospital Basel Petergraben 4, CH-4031 Basel, Switzerland*
[j]*The Basel Institute of Clinical Epidemiology, University Hospital Basel Hebelstrasse 10, 4031 Basel, Switzerland*
[k]*Knowledge and Encounter Research Unit, Mayo Clinic, Rochester, MN, USA*
[l]*Harvard Medical School, Boston, MA, USA*
[m]*UpToDate, Boston, MA, USA*
[n]*Department of Urology, University of Florida, Gainesville, FL, USA*
[o]*German Cochrane Center, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, 79104 Freiburg, Germany*
[p]*Division of Pediatric Hematology and Oncology, Department of Pediatric and Adolescent Medicine, University Medical Center Freiburg, 79106 Freiburg, Germany*
[q]*Technology Assessment Program, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, USA*
[r]*Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR 97239-3098, USA*
[s]*Division of Gastroenterology, Case and VA Medical Center, Case Western Reserve University, Cleveland, OH 44106, USA*

Accepted 5 June 2011; Published online 30 July 2011

**Abstract**

The most common reason for rating up the quality of evidence is a large effect. GRADE suggests considering rating up quality of evidence one level when methodologically rigorous observational studies show at least a two-fold reduction or increase in risk, and rating up two levels for at least a five-fold reduction or increase in risk. Systematic review authors and guideline developers may also consider rating up quality of evidence when a dose−response gradient is present, and when all plausible confounders or biases would decrease an apparent treatment effect, or would create a spurious effect when results suggest no effect. Other considerations include the rapidity of the response, the underlying trajectory of the condition, and indirect evidence. 

*Keywords:* GRADE; Guidelines; Level of evidence; Observational studies; Large effects; Risk of bias

## 1. Introduction

In prior papers in this series devoted to exploring GRADE's approach to rating the quality of evidence and grading strength of recommendations, we have dealt with issues of framing the question; introduced GRADE's conceptual approach to rating the quality of a body of evidence; and presented five reasons for rating down the quality

[1] The GRADE system has been developed by the GRADE Working Group. The named authors drafted and revised this article. A complete list of contributors to this series can be found on the journal's Web site at www.elsevier.com.

* Corresponding author. CLARITY Research Group, Department of Clinical Epidemiology and Biostatistics, McMaster University, Room 2C12, 1200 Main Street, West Hamilton, Ontario L8N 3Z5, Canada. Tel.: 905-525-9140; fax: 905-523-8781.

*E-mail address*: guyatt@mcmaster.ca (G.H. Guyatt).

0895-4356/$ - see front matter 
doi: 10.1016/j.jclinepi.2011.06.004

**Key points**

- GRADE includes three criteria for rating up quality of evidence particularly applicable to observational studies.
- Rating up one or even two levels is possible when effects in observational studies are sufficiently large, particularly if they occur over short periods of time.
- A dose–response gradient, or a conclusion that plausible residual confounding would further support inferences regarding treatment effect, may also raise the quality of the evidence.

of evidence: risk of bias, imprecision, inconsistency, indirectness, and publication bias. This ninth article in the series examines the criteria for rating up the quality of evidence.

The three primary reasons for rating up the quality of evidence are (Table 1) as follows:

1. When a large magnitude of effect exists,
2. When there is a dose–response gradient, and
3. When all plausible confounders or other biases increase our confidence in the estimated effect.

We have noted previously that GRADE is relevant to rating evidence regarding the impact of interventions on patient-important outcomes—rather than, for instance, prognostic studies that identify patient characteristics associated with desirable or adverse outcomes. Using the GRADE framework, evidence from observational studies is generally classified as low. Unsystematic clinical observations are usually at a high risk of bias and therefore generally receive a rating of very low quality evidence. There are times, however, when we have high confidence in the estimate of effect from such studies. GRADE has therefore suggested mechanisms for rating up the quality of evidence from observational studies.

The circumstances under which we may wish to rate up the quality of evidence for intervention studies will likely occur infrequently and are primarily relevant to observational studies (including cohort, case–control, before–after, and time series studies) and to nonrandomized experimental or interventional studies (e.g., providing treatment to one of the two matched groups). Indeed, although it is theoretically possible to rate up results from randomized control trials (RCTs), we have yet to find a compelling example of such an instance.

**Table 1.** Factors that may increase the quality of evidence

- Large magnitude of effect (direct evidence, relative risk [RR] = 2–5 or RR = 0.5–0.2 with no plausible confounders); very large with RR > 5 or RR < 0.2 and no serious problems with risk of bias or precision (sufficiently narrow confidence intervals); more likely to rate up if effect rapid and out of keeping with prior trajectory; usually supported by indirect evidence.
- Dose-response gradient.
- All plausible residual confounders or biases would reduce a demonstrated effect, or suggest a spurious effect when results show no effect.

## 2. Large magnitude of effect

For some clinical interventions (e.g., hip replacement to reduce pain and functional limitations in severe osteoarthritis, epinephrine to prevent mortality in anaphylaxis, and insulin to prevent mortality in diabetic ketoacidosis), clinicians are, correctly, extremely confident of their effectiveness. Moreover, in each of these situations we are also extremely confident that the impact of the intervention is substantial. Thus, using GRADE's definition of quality of evidence, the underlying quality of evidence to support these clinical interventions would be considered high although the evidence comes from observational studies or from unsystematic clinical observations.

Moderate- or high-quality evidence can also come from epidemiological studies of public health interventions. For example, a systematic review of observational studies examining the relationship between infant sleeping position and sudden infant death syndrome (SIDS) found an odds ratio (OR) of 4.1 (95% confidence interval [CI]: 3.1, 5.5) of SIDS occurring with front vs. back sleeping positions [1]. Furthermore, ''back to sleep'' campaigns that were started in the 1980s to encourage back sleeping position were associated with a relative decline in the incidence of SIDS by 50–70% in numerous countries [1].

The most striking aspect of these examples is the large magnitude of effect. Although the treatment or intervention effect comes from observational studies or time series studies of public health interventions, the large effect and additional population-based epidemiological evidence merit rating up the quality of evidence at least one level.

However, rating up for a large effect raises conceptual challenges. If methodologically rigorous observational studies (those that comprehensively and accurately measure prognostic factors associated with the outcome of interest; minimize loss to follow-up; accurately measure outcome; and conduct an adjusted analysis that accounts for differences in the distribution of prognostic factors between intervention and control groups) show a sufficiently large effect, one can reasonably deduce that effect is real (that is, nonzero, and causally attributable to the intervention). It is considerably more problematic to deduce that the very large estimate of effect is accurate, and not a biased overestimate of effect. Given GRADE's definition of quality of evidence as confidence in the estimate (magnitude) of effect, this is the inference we are making in the hip replacement, epinephrine, and insulin examples.

Modeling studies addressing the degree of associations between causal factors and confounders, and between

confounders and outcomes, needed to induce different effect sizes of confounding support the decision to rate up for a large magnitude of effect. This modeling suggests that confounding (from nonrandom allocation) alone is unlikely to explain associations with a relative risk (RR) greater than 2 (or less than 0.5), and very unlikely to explain associations with an RR greater than 5 (or less than 0.2) [2].

These conclusions are supported by some empirical work. A Cochrane methods review of 35 comparisons (from 15 studies) of randomized vs. nonrandomized trials of the same intervention found 22 comparisons in which the nonrandomized trials had larger estimates of effect, 8 with similar results, and 4 in which nonrandomized trials found smaller effects [3]. The difference between randomized and nonrandomized trials ranged from a 76% smaller (1/4) to a 400% larger (4/1) effect. Although further research is warranted, both modeling and empirical work suggest the size of bias from confounding is unpredictable in direction but bounded in size. Hence, the GRADE group has previously suggested guidelines for rating quality of evidence up by one category (typically from low to moderate) for associations greater than 2, and up by two categories for associations greater than 5 [4]. Other simulations have suggested that a threshold RR of 10 may be more appropriate for rating up by two levels [5,6].

Note that in both cases these thresholds refer to risk ratios. When baseline risk is low—say, below 20%—odds and risk ratios are very similar and one can comfortably apply the risk ratio criteria. The OR is always farther from 1.0 than the risk ratio, and when baseline risk is high (say, over 40%), and the effect size is substantial in terms of odds, OR can be far larger in magnitude than risk ratios. Under such circumstances, a higher threshold for ORs may be appropriate.

Furthermore, when considering rating up the quality of evidence for magnitude of effect, factors relating to the magnitude include rapidity of treatment response, and the previous underlying trajectory of the condition [6]. For example, we feel confident that hip replacement has a large effect not only because of the size of the treatment response, but because the natural history of hip osteoarthritis is a progressive deterioration that surgery rapidly and uniformly reverses. The rapidity of response compared with the known trajectory of the condition can also be considered (and calculated [6]) as a large effect size.

An additional factor mitigating the problem of rating up the quality because of a large effect is that indirect evidence usually provides further support for large treatment effects. For example, oral anticoagulation in mechanical heart valves has not been compared with placebo in an RCT, but evidence from observational studies suggests a large effect of oral anticoagulation in decreasing thromboembolic events [7,8]. Supplementary indirect evidence from randomized trials that have demonstrated large reductions in the RR of thrombosis with anticoagulation in analogous conditions such as atrial fibrillation further increases our confidence in the beneficial effect of anticoagulation [9].

Similarly, the effectiveness of antibiotic prophylaxis in a variety of other situations supports observational studies that suggest that antibiotic prophylaxis results in an 89% RR reduction in meningococcal disease in contacts of patients who have suffered the illness [10].

Another situation allows an inference of a strong association without a formal comparative study. Consider the question of the impact of routine colonoscopy vs. no screening for colon cancer on the rate of perforation associated with colonoscopy. Here, a large series of representative patients undergoing colonoscopy will provide high-quality evidence on the risk of perforation associated with colonoscopy. When control rates are near 0 (i.e., we are certain that the incidence of spontaneous colon perforation in patients not undergoing colonoscopy is very low), case series of representative patients (one might call these cohort studies of affected patients if they include large numbers of patients) can provide high-quality evidence of adverse effects associated with an intervention, thereby allowing us to infer a strong association from even a limited number of events. One should not confuse the situation highlighted in the previous example with isolated case reports of associations between exposures and rare adverse outcomes (as have, for instance, been reported with vaccine exposure).

We complete our discussion of rating up with a note of caution. Systematic review and guideline authors may hesitate to attribute causation to even large effects when outcomes are subjective. For example, three small unblinded trials of pacemaker insertion showed significant reductions in the RR of recurrent syncope of over 80%, with corresponding absolute reductions of over 20% [11−13]. The quality of evidence might be considered moderate, or even low as a result of risk of bias and imprecision; the large effect provides a rationale for rating up, potentially to high quality. Despite this impressive result, however, a subsequent blinded trial failed to establish a reduction in syncope and suggested that the effect, if present at all, is more modest (RR of syncope 0.68, 95% CI: 0.44, 1.02) [14].

Furthermore, residual confounding may be substantial even when good prognostic data are available, and adjusting for case mix is not always successful [15]. Thus, a conservative approach to rating up for a large effect unlikely to be explained by confounding is likely wise. In general, we should not rate up for a large effect size if we have major concerns about other issues, including risk of bias, precision, and publication bias.

## 3. Dose−response gradient

The presence of a dose−response gradient has long been recognized as an important criterion for believing a putative cause−effect relationship [16]. Such a gradient may

increase our confidence in the findings of observational studies and thus enhance the assigned quality of evidence (Table 1).

For example, our confidence in the results of observational studies that show an increased risk of bleeding in patients who have supra-therapeutic anticoagulation levels is increased by the finding that there is a dose–response gradient between higher levels of the international normalized ratio and the increased risk of bleeding [17]. Similarly, infant growth is slowest in infants fed exclusively with breast milk, accelerated to some extent in infants fed in part with breast milk and part formula, and further accelerated in infants fed exclusively with formula [18]. A systematic review of observational studies investigating the effect of cyclooxygenase-2 inhibitors on cardiovascular events found an RR with rofecoxib of 1.33 (95% CI: 1.00, 1.79) with doses less than 25 mg/d and an RR of 2.19 (95% CI: 1.64, 2.91) with doses more than 25 mg/d [19].

A final example is the striking dose–response gradient associated with the rapidity of antibiotic administration in patients presenting with sepsis and hypotension (Fig. 1) [20]. This dose–response relationship increases our confidence that the effect on mortality (large absolute increases in mortality with each hour's delay) is real and substantial.

## 4. Plausible confounding can increase confidence in estimated effects

Rigorous observational studies will accurately measure prognostic factors associated with the outcome of interest and will conduct an adjusted analysis that accounts for differences in the distribution of these factors between intervention and control groups. The reason that in most instances we consider observational studies as providing only low-quality evidence is that unmeasured or unknown determinants of outcome unaccounted for in the adjusted analysis are likely to be distributed unequally between intervention and control groups. The technical language of observational epidemiology characterizes this phenomenon as ''residual confounding'' or ''residual biases.''

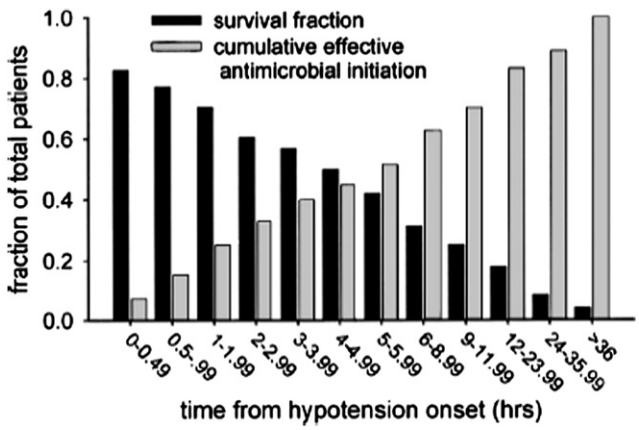


**Fig. 1.** Cumulative effective antimicrobial initiation following onset of septic shock-associated hypotension and associated survival. The *x*-axis represents time (h) following first documentation of septic shock-associated hypotension. *Black bars* represent the fraction of patients surviving to hospital discharge for effective therapy initiated within the given time interval. The *gray bars* represent the cumulative fraction of patients having received effective antimicrobials at any given time point.

On occasion, all plausible confounders and biases from observational studies unaccounted for in the adjusted analysis (i.e., all residual confounders) of a rigorous observational study would result in an underestimate of an apparent treatment effect. If, for instance, only sicker patients receive an experimental intervention or exposure, yet they still fare better, it is likely that the actual intervention or exposure effect is even larger than the data suggest.

For example, a rigorous systematic review of observational studies including a total of 38 million patients demonstrated higher death rates in private for-profit vs. private not-for-profit hospitals [21]. One possible source of bias relates to different disease severity between patients in the two hospital types. It is likely, however, that patients in the not-for-profit hospitals were sicker than those in the for-profit hospitals. Thus, to the extent that residual confounding existed, it would bias results against the not-for-profit hospitals.

A second possible bias was that higher numbers of patients with excellent private insurance coverage could lead to a hospital having more resources and a spillover effect that would benefit those without such coverage. Because for-profit hospitals are likely to admit a larger proportion of such well-insured patients than not-for-profit hospitals, the bias is once again against the not-for-profit hospitals. Because all the plausible biases would diminish the observed effect, one might consider the evidence from these observational studies as moderate rather than low quality.

In another example of this phenomenon, an unpublished systematic review addressed the effect of condom use on HIV infection among men who have sex with men. The pooled effect estimate of RR from the five eligible observational studies was 0.34 [0.21, 0.54] in favor of condom use compared with no condom use. Two of these studies [22,23] that examined the number of partners in those using condoms and not using condoms found that condom users were more likely to have more partners (but did not adjust for this confounding factor in their analyses). Considering the number of partners would, if anything, strengthen the effect estimate in favor of condom use.

A parallel situation exists when observational studies have failed to demonstrate an association. An example comes from the now discredited putative association between vaccination and autism [24]. Subsequent observational studies failed to confirm the association [25,26]. This lack of association occurred despite the empirically confirmed bias that parents of autistic children diagnosed after the publicity associated with the original article would be more likely to remember their vaccine experience than

parents of children diagnosed before the publicity [27]—and presumably, than parents of nonautistic children. The negative findings despite this form of recall bias suggest rating up the quality of evidence.

This situation also arises in the exploration of other apparent harmful effects. For example, because the hypoglycaemic drug phenformin causes lactic acidosis, the related agent metformin is under suspicion for the same toxicity. Nevertheless, large observational studies have failed to demonstrate an association. Given the likelihood that clinicians would be more alert to lactic acidosis in the presence of the agent, one might consider this moderate- or even high-quality evidence refuting a causal relationship between typical therapeutic doses of metformin and lactic acidosis.

## 5. Other considerations

Particular design features of extremely rigorous well-conducted observational studies may warrant consideration for rating up quality of evidence. For instance, a case-control study found that sigmoidoscopy was associated with a reduction in colon cancer mortality for lesions in range of the sigmoidoscope (OR 0.30, 95% CI: 0.19, 0.48), but not beyond the range of the sigmoidoscope (OR 0.96, 95% CI: 0.61, 1.50) [28]. Possible bias because of unmeasured confounders should have been very similar if not identical in the two situations, considerably raising confidence in the causal effect of the sigmoidoscopy.

## 6. A final note of caution

Consideration of all our previously presented criteria for rating down quality of evidence (risk of bias, imprecision, inconsistency, indirectness, and publication bias) must precede consideration of reasons for rating up quality. The decision to rate up should only rarely be made if serious limitations are present in any of these areas. In particular, decisions to rate up because of large or very large effects should consider not only the point estimate but also the width of the CI around that effect: one should rarely rate up for large effects if the CI overlaps substantially with effects smaller than the chosen threshold.

## 7. Conclusions

In summary, there are three factors that might increase the quality of evidence. In general, these three factors, mostly applicable to observational studies, are encountered infrequently. Although most observational studies, even if well done, yield low-quality evidence, one can consider rating up the quality of evidence when there is a large or a very large magnitude of effect, when consideration of all plausible residual confounders and biases would reduce a demonstrated effect, or suggest a spurious effect when results show no effect, or when there is an evidence of a dose−response gradient. Rarely, other considerations that do not easily fit into the above categories may constitute reasons for rating up quality of evidence.

## References

[1] Gilbert R, Salanti G, Harden M, See S. Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. Int J Epidemiol 2005;34:874−87.
[2] Bross ID. Pertinency of an extraneous variable. J Chronic Dis 1967;20:487−95.
[3] Kunz R, Vist G, Oxman AD. Randomization to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2007;2. MR000012.
[4] Weiner SJ. Contextualizing medical decisions to individualize care: lessons from the qualitative sciences. J Gen Intern Med 2004;19:281−5.
[5] Rothman K, Greenland S. Modern epidemiology. 2nd ed. Lippincott Raven; 1998.
[6] Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. BMJ 2007;334:349−51.
[7] Cannegieter SC, Rosendaal FR, Briet E. Thromboembolic and bleeding complications in patients with mechanical heart valve prostheses. Circulation 1994;89:635−41.
[8] Baudet EM, Puel V, McBride JT, Grimaud J, Rogues F, Clerc F, et al. Long-term results of valve replacement with the St. Jude Medical prosthesis. J Thorac Cardiovasc Surg 1995;109:858−70.
[9] Singer DE, Albers GW, Dalen JE, Fang M, Go A, Halperin J, et al. Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest 2008;133:546S−92S.
[10] Purcell B, Samuelsson S, Hahne SJ, Ehrhard I, Heuberger S, Camaroni I, et al. Effectiveness of antibiotics in preventing meningococcal disease after a case: systematic review. BMJ 2004;328:1339.
[11] Connolly SJ, Sheldon R, Roberts RS, Gent M. The North American Vasovagal Pacemaker Study (VPS). A randomized trial of permanent cardiac pacing for the prevention of vasovagal syncope. J Am Coll Cardiol 1999;33:16−20.
[12] Ammirati F, Colivicchi F, Santini M. Permanent cardiac pacing versus medical treatment for the prevention of recurrent vasovagal syncope: a multicenter, randomized, controlled trial. Circulation 2001;104:52−7.
[13] Sutton R, Brignole M, Menozzi C, Ravieli A, Alboni P, Giani P, et al. Dual-chamber pacing in the treatment of neurally mediated tilt-positive cardioinhibitory syncope:pacemaker versus no therapy: a multicenter randomized study. The Vasovagal Syncope International Study (VASIS) Investigators. Circulation 2000;102:294−9.
[14] Connolly SJ, Sheldon R, Thorpe KE, Roberts R, Ellenbogen K, Wilkoff B, et al. Pacemaker therapy for prevention of syncope in patients with recurrent severe vasovagal syncope: Second Vasovagal Pacemaker Study (VPS II): a randomized trial. JAMA 2003;289:2224−9.
[15] Deeks JJ, Dinnes J, D'Amico R, Sowden A, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7:iii−x. 1−173.
[16] Hill A. The environment and disease: association or causation? Proc R Soc Med 1965;58:295−300.
[17] Schulman S, Beyth RJ, Kearon C, Levine M. Hemorrhagic complications of anticoagulant and thrombolytic treatment: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest 2008;133:257S−98S.
[18] Kramer MS, Guo T, Platt RW, Shapiro S, Collet J, Chalmers B, et al. Feeding effects on growth during infancy. J Pediatr 2004;145:600−5.

[19] McGettigan P, Henry D. Cardiovascular risk and inhibition of cyclooxygenase: a systematic review of the observational studies of selective and nonselective inhibitors of cyclooxygenase 2. JAMA 2006;296: 1633−44.

[20] Kumar A, Roberts D, Wood KE, Light B, Parillo J, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med 2006;34:1589−96.

[21] Devereaux PJ, Choi PT, Lacchetti C, Weaver B, Schunemann H, Haines T, et al. A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals. CMAJ 2002;166:1399−406.

[22] Detels R, English P, Visscher BR, Jacobson L, Kingsley L, Shmiel J, et al. Seroconversion, sexual activity, and condom use among 2915 HIV seronegative men followed for up to 2 years. J Acquir Immune Defic Syndr 1989;2:77−83.

[23] Difranceisco W, Ostrow DG, Chmiel JS. Sexual adventurism, high-risk behavior, and human immunodeficiency virus-1 seroconversion among the Chicago MACS-CCS cohort, 1984 to 1992. A case-control study. Sex Transm Dis 1996;23:453−60.

[24] Wakefield AJ, Murch SH, Anthony A, Linnell J, Casson D, Malik M, et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet 1998;351: 637−41.

[25] Elliman DA, Bedford HE. MMR vaccine—worries are not justified. Arch Dis Child 2001;85:271−4.

[26] Taylor B, Miller E, Farrington CP, Petropoulos M, Favot-Mayaud I, Waight P. Autism and measles, mumps, and rubella vaccine: no epidemiological evidence for a causal association. Lancet 1999;353: 2026−9.

[27] Andrews N, Miller E, Taylor B, Lingam R, Simmons A, Stowe J, et al. Recall bias, MMR, and autism. Arch Dis Child 2002;87: 493−4.

[28] Selby JV, Friedman GD, Quesenberry CP Jr, Weiss N. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. N Engl J Med 1992;326:653−7.