



The Results of Direct and Indirect Treatment Comparisons in Meta-Analysis of Randomized Controlled Trials

Heiner C. Bucher,* Gordon H. Guyatt, Lauren E. Griffith, and Stephen D. Walter

DEPARTMENT OF CLINICAL EPIDEMIOLOGY AND BIOSTATISTICS,
McMASTER UNIVERSITY, HAMILTON, ONTARIO,
CANADA, L8N 3Z5

ABSTRACT. When little or no data directly comparing two treatments are available, investigators often rely on indirect comparisons from studies testing the treatments against a control or placebo. One approach to indirect comparison is to pool findings from the active treatment arms of the original controlled trials. This approach offers no advantage over a comparison of observational study data and is prone to bias.

We present an alternative model that evaluates the differences between treatment and placebo in two sets of clinical trials, and preserves the randomization of the originally assigned patient groups. We apply the method to data on sulphamethoxazole–trimethoprim or dapsone/pyrimethamine as prophylaxis against *Pneumocystis carinii* in HIV infected patients. The indirect comparison showed substantial increased benefit from the former (odds ratio 0.37, 95% CI 0.21 to 0.65), while direct comparisons from randomized trials suggests a much smaller difference (risk ratio 0.64, 95% CI 0.45 to 0.90; *p*-value for difference of effect = 0.11).

Direct comparisons of treatments should be sought. When direct comparisons are unavailable, indirect comparison meta-analysis should evaluate the magnitude of treatment effects across studies, recognizing the limited strength of inference. J CLIN EPIDEMIOL 50;6:683–691, 1997. © 1997 Elsevier Science Inc.

KEY WORDS: Meta-analysis, methodology, clinical epidemiology, outcome, HIV infection, *Pneumocystis carinii* pneumonia

INTRODUCTION

Physicians in almost every field of medicine face an increasing number of treatment options. Deciding between these treatments is often difficult. For instance, two drugs may have demonstrated effectiveness in comparison to placebo or standard treatment in well-executed randomized trials, but direct comparisons of effectiveness between these agents may be unavailable.

In recent years meta-analysis has become a widespread method to improve clinical decision-making by simultaneously considering the results of multiple randomized trials. Meta-analysis provides a potential solution to the problem of deciding between treatments that have not been directly compared. For example, Felson *et al.* [1] conducted a meta-analysis examining the relative effectiveness and toxicity of second-line drugs in rheumatoid arthritis. Because few of the drugs of interest had been subject to direct comparisons, the investigators chose to compare patients in different trials given competing second-line agents, rather

than comparing them with the placebo-treated control subjects in their own trials. Although all data in this meta-analysis came from randomized controlled trials, the power of randomization is lost and the data are reduced to the equivalent of those derived from contemporaneous or historical cohort studies. This comparison is no stronger than one obtained from an observational study in which patients are drawn from separate populations, with the likelihood that differences in prognosis unrelated to treatment will bias the comparison of effectiveness [2].

Indirect comparisons can also form the basis of models used by health economists for their cost-effectiveness analyses. In the absence of direct comparison of enoxaparin or warfarin for the prophylaxis of post-surgical deep vein thrombosis, O'Brien *et al.* [3] combined data from the treatment arms of ten randomized trials that compared either enoxaparin or warfarin versus placebo. Because, once again, the power of randomization is lost, and the groups are likely to differ in their prognosis, it is unlikely that an unbiased comparison will result from this strategy.

When direct comparisons are unavailable, is there any alternative strategy for pooling data that is less susceptible to bias? Conceptually, we could examine the magnitude of

*Address for correspondence: Heiner C. Bucher, Medizinische Universitäts-Poliklinik Kantonsspital Basel, CH-4031 Basel, Switzerland
Accepted for publication on 3 March 1997.

the treatment effect in studies of different treatments and compare those treatment effects. With this approach it is possible that even if both control and treatment groups differ in their baseline characteristics, we may obtain an unbiased estimate of treatment effect. The only requirement is that the magnitude of the treatment effect is constant across differences in the populations' baseline characteristics.

In this paper, we present a model for making indirect comparisons of the magnitude of treatment effects that preserves the randomization of the originally assigned patient groups. We illustrate this model with an example from a meta-analysis of randomized controlled trials that compares two experimental prophylactic regimens against the standard prophylaxis for the prevention of *Pneumocystis carinii* pneumonia in patients with HIV infection. We contrast the results derived from analyses that compare results of treatment groups from different trials, the results from randomized controlled trials investigating the two prophylactic regimens of interest by direct comparisons, and our model of indirect comparisons.

A Model for the Indirect Comparison of Proportional Data

In the discussion that follows, we will focus on the odds ratio as our measure of treatment effect. The method could be equally applied to estimates of relative risk. The appendix presents more detailed derivations of the formulae while the following text presents the ideas conceptually.

In meta-analysis the overall measure of association is typically taken as the weighted average of the individual studies with weights being the inverse of the variance for each study. Let us consider a situation where g studies have compared treatment A to standard treatment B and h studies have compared treatment C to standard treatment B. Our interest is in obtaining an indirect estimate of the association of A versus C. To obtain an indirect estimate of the association A versus C from the paired comparisons of A versus B and of C versus B, first an overall measure of variation can be taken as the sum of the total chi-squared values from the paired comparisons with $g + h$ degrees of freedom. A summary odds ratio of the indirect comparison can then be computed by taking the ratio of the log odds ratio from studies comparing A versus B and from studies comparing C versus B.

The estimate we derive from this approach will be unbiased in large samples if there is no interaction between covariates defining subgroups of patients (reflected, for instance, in different inclusion criteria in different studies) and the magnitude of the treatment effect. The appendix provides a proof of unbiasedness of our estimate if no such interaction is present. If an interaction is present, then the pooling of data across studies is questionable, even for direct comparisons.

This model then allows application of standard methods

whereby we express the association of a treatment effect from a number of studies that each estimate a level of association between an intervention and an outcome by a chi-squared distribution. If the value of chi-squared is large we may conclude that there is association between treatment and outcome somewhere within the studies. However, we would not know whether the association was consistent across all studies or whether it varied from one group of studies to another. We can elucidate this issue by partitioning the chi-squared into two components, a chi-squared which assesses the degree of heterogeneity and a chi-squared that assesses the significance of the average degree of association.

Application of the Model and Comparison with Other Approaches

We tested our model in a meta-analysis of randomized controlled trials that compare two experimental (trimethoprim-sulphamethoxazole and dapsone/pyrimethamine) and one standard regimen (aerosolized pentamidine) for primary and secondary prevention of *Pneumocystis carinii* pneumonia in HIV infection. In a systematic review of the literature we identified all randomized controlled trials comparing either trimethoprim-sulphamethoxazole dapsone/pyrimethamine with the standard treatment with aerosolized pentamidine. In addition, we identified all trials providing a direct estimate of the relative effectiveness by comparing trimethoprim-sulphamethoxazole with dapsone or dapsone/pyrimethamine. Details of search strategies, inclusion criteria and the description of included trials are provided in an other publication [4]. For the purpose of this study we report only on the single outcome of *Pneumocystis carinii* pneumonia.

We identified 22 trials. Nine trials compared trimethoprim-sulphamethoxazole with aerosolized pentamidine [5–13]. Two trials compared dapsone/pyrimethamine [14,15] and three trials dapsone [16–18] with pentamidine. Four trials compared trimethoprim-sulphamethoxazole with dapsone/pyrimethamine [19–22]. Four trials included three treatment arms e.g., one trial compared dapsone [23] and three trials compared dapsone/pyrimethamine [24–26] with trimethoprim-sulphamethoxazole and aerosolized pentamidine. These four trials were only used for the direct comparison of trimethoprim-sulphamethoxazole versus dapsone/pyrimethamine. In these 22 trials a total of 1484 patients had been treated with trimethoprim-sulphamethoxazole, 1547 patients with dapsone or dapsone/pyrimethamine, and 1837 patients with aerosolized pentamidine. We combined nine trials for the comparison of trimethoprim-sulphamethoxazole and aerosolized pentamidine, five trials for the comparison of dapsone/pyrimethamine and aerosolized pentamidine, and eight trials for the direct comparison of dapsone/pyrimethamine and trimethoprim-sulphamethoxazole. In Table 1 we show important baseline characteris-

TABLE 1. Odds ratios and 95% confidence interval in randomized controlled trials of indirect comparison with either trimetho-prim-sulfamethoxazole or dapsone/pyrimethamine compared to aerosolized pentamidine for prevention of *Pneumocystis carinii* pneumonia in HIV-infected subjects

Trials: author, year of publication [Reference]	Odds ratio (95% CI)	No. of Pcp and subjects	Mean follow-up (person-years)	Mean CD4 cells (mm ³)	Patients with		Study quality score
					AIDS <i>n</i> (%)	Antiretroviral therapy <i>n</i> (%)	
Comparing TMP-SMX with AP							
Rozenbaum 1991 [12]							
TMP-SMX	0.30 (0.01–7.67)	0/29	21.7	124	NA	29 (100)	0.47
AP		1/27	20.3	117		27 (100)	
Hardy 1992 [5]							
TMP-SMX	0.34 (0.18–0.66)	14/154	202.4	62	154 (100)	139 (90)	0.68
AP		36/156	191.4	51	156 (100)	138 (89)	
Schneider 1992 [6]							
TMP-SMX	0.04 (0.00–0.64)	0/142	56.0	90	53 (75)	39 (55)	0.71
AP		6/71	53.8	95	46 (65)	31 (44)	
Smith 1992 [11]							
TMP-SMX	0.45 (0.11–1.87)	3/27	49.5	39	27 (100)	22 (76)	0.48
AP		6/26	47.6	69	26 (100)	21 (81)	
Michelet 1993 [9]							
TMP-SMX	0.33 (0.05–2.16)	1/53	105.6	159	NA	51 (96)	0.35
AP		4/55	105.6	140		52 (94)	
May 1994 [7]							
TMP-SMX	0.43 (0.09–1.98)	2/108	112.7	200	24 (22)	73 (68)	0.61
AP		5/106	118.0	200	22 (21)	70 (66)	
Stellini 1994 [10]							
TMP-SMX	0.16 (0.01–3.56)	0/26	35.9	137	2 (8)	18 (69)	0.44
AP		2/23	25.5	137	3 (13)	16 (69)	
Nielsen 1995 [8]							
TMP-SMX	0.15 (0.03–0.90)	1/47	66.8	59	47 (100)	45 (96)	0.59
AP		8/48	68.6	74	47 (100)	44 (94)	
Rizzardi 1995 [13]							
TMP-SMX	0.89 (0.28–2.88)	5/95	175.5	128	10 (11)	90 (95)	0.70
AP		6/101	162.5	127	21 (21)	85 (84)	
Trials comparing D/P or D with AP							
Slavin 1992 [16]							
D	1.04 (0.37–2.89)	9/50	49.7	128	16 (32)	N.A.	0.56
AP		8/46	56.0	126	18 (39)		
Girard 1993 [14]							
D/P	1.02 (0.42–2.46)	10/173	238.3	112	115 (67)	162 (94)	0.73
AP		10/176	243.4	117	128 (73)	152 (86)	
Torres 1993 [17]							
D	1.23 (0.58–2.60)	15/126	102.0	192	71 (56)	109 (72)	0.73
AP		15/152	129.0	163	93 (61)	88 (70)	
Opravil 1995 [15]							
D/P	0.76 (0.35–1.67)	12/291	342.8	116	52 (18)	262 (90)	0.78
AP		13/242	304.9	105	51 (21)	220 (91)	
Salmon 1995 [18]							
D	0.46 (0.16–1.29)	5/92	96.0	57	93 (100)	46 (50)	0.71
AP		12/102	116.0	59	103 (100)	47 (46)	
Comparing TMP-SMX with D/P or D and AP ^a							
Antinori 1992 [24]							
TMP-SMX	0.13 (0.02–0.76)	1/66	49.7	107	17 (26)	55 (83)	0.45
D/P		9/63	28.4	99	27 (43)	40 (64)	
Mallolas 1992 [25]							
TMP-SMX	0.43 (0.12–1.53)	3/107	100.0	160	45 (42)	104 (97)	0.61
D/P		8/116	97.0	188	70 (60)	105 (91)	
Tocchetti 1994 [26]							
TMP-SMX	0.31 (0.01–8.29)	0/15	12.7	91	8 (53)	15 (100)	0.39
D/P		1/15	13.8	94	14 (93)	14 (93)	
Bozzette 1995 [23]							
TMP-SMX	1.08 (0.68–1.72)	42/276	672.0	152	NA	276 (100)	0.62
D		41/288	696.0	148		288 (100)	

(continued)

TABLE 1. (Continued)

Trials: author, year of publication [Reference]	Odds ratio (95% CI)	No. of Pcp and subjects	Mean follow-up (person-years)	Mean CD4 cells (mm ³)	Patients with		Study quality score
					AIDS <i>n</i> (%)	Antiretroviral therapy <i>n</i> (%)	
Comparing TMP-SMX with D/P or D							
Blum 1992 [20]							
TMP-SMX	1.21 (0.12–12.00)	1/39	64.7	118	7 (18)	24 (62)	0.58
D		1/47	71.8	111	9 (19)	33 (70)	
Podzameczer 1993 [19]							
TMP-SMX	0.24 (0.07–0.81)	3/81	95.0	109	33 (41)	40 (49)	0.56
D/P		13/85	78.0	128	38 (45)	44 (52)	
Podzameczer 1995 [21]							
TMP-SMX	0.07 (0.004–1.20)	0/104	122.5	148	29 (28)	92 (89)	0.72
D/P		6/96	113.1	132	25 (26)	83 (87)	
Sirera 1995 [22]							
TMP-SMX	0.60 (0.21–1.70)	6/115	212.0	113	NA	115 (100)	0.34
D/P		9/105	225.0	118		105 (100)	

Abbreviations: AIDS = acquired immune deficiency syndrome; AP = aerosolized pentamidine; CI = confidence interval; D = dapsone; D/P = dapsone/pyrimethamine; HIV = human immunodeficiency virus; NA = not applicable; Pcp = *Pneumocystis carinii* pneumonia; TMP-SMX = trimethoprim-sulphamethoxazole.

*The number of events in the aerosolized pentamidine groups of trials with 3 treatment arms are not presented.

tics, the event rates for *Pneumocystis carinii* pneumonia and single odds ratio of each trial.

Table 2 shows the different estimates for the direct and indirect comparison of two drug regimens for the prophylaxis of *Pneumocystis carinii* pneumonia. While trimethoprim-sulphamethoxazole showed a clear benefit when compared with the standard treatment aerosolized pentamidine, dapsone/pyrimethamine showed only a relatively weak trend. The odds ratio of trimethoprim-sulphamethoxazole versus aerosolized pentamidine for *Pneumocystis carinii* pneumonia was 0.48 (95% confidence interval 0.36 to 0.65) and for dapsone/pyrimethamine versus aerosolized pentamidine 0.89 (95% confidence interval 0.68 to 1.17). For both sets of paired comparisons the test for homogeneity was not statistically significant.

For the indirect comparisons of trimethoprim-sulphamethoxazole versus dapsone/pyrimethamine the odds ratio was considerably lower than the point estimate derived from trials that allowed for the direct comparisons of trimethoprim-sulphamethoxazole and dapsone/pyrimethamine. The odds ratio from indirect comparison was 0.37 (95% confidence interval 0.21 to 0.65) favoring trimethoprim-sulphamethoxazole over dapsone/pyrimethamine for the prevention of *Pneumocystis carinii* pneumonia. The odds ratio for *Pneumocystis carinii* pneumonia from the eight trials that compared the drugs directly was 0.64 (95% confidence interval 0.45 to 0.90, test for heterogeneity $p = 0.41$).

This direct meta-analysis provides support for the superiority of trimethoprim-sulphamethoxazole over dapsone/pyrimethamine for the prevention of *Pneumocystis carinii*

TABLE 2. Direct and indirect comparison in meta-analysis on the efficacy of two drug regimens compared to standard prophylaxis in randomized controlled trials of primary and secondary prevention of *Pneumocystis carinii* pneumonia in patients with HIV infection

Type of comparison	No. of trials in comparison	Odds ratio (95% CI)	Test for heterogeneity
Indirect comparison			
Trimethoprim-sulfamethoxazole vs. dapsone/pyrimethamine	14	0.37 ^a (0.21–0.65)	$p = 0.03$
Direct comparison			
Trimethoprim-sulfamethoxazole vs. dapsone/pyrimethamine	8	0.64 ^a (0.45–0.90)	$p = 0.41$
Dapsone/pyrimethamine vs. aerosolized pentamidine	9	0.89 (0.68–1.17)	$p = 0.27$
Trimethoprim-sulfamethoxazole vs. aerosolized pentamidine	13	0.48 (0.36–0.65)	$p = 0.98$

Abbreviations: CI = confidence interval; HIV = human immunodeficiency virus.

^a p Value for the difference of summary ratio = 0.11.

TABLE 3. Differences in baseline characteristics of randomized controlled trials selected for the indirect comparison of either trimethoprim-sulfamethoxazole or dapsone/pyrimethamine with aerosolized pentamidine for prophylaxis of *Pneumocystis carinii* pneumonia in patients with HIV infection

Trials	No. of trials in comparison	Baseline characteristics ^a				
		Mean person years follow-up	Mean CD4 cells (mm ³)	Patients with		Individual study quality score ^b
				AIDS (%)	Antiretroviral therapy (%)	
Trimethoprim-sulfamethoxazole vs. aerosolized pentamidine	9	1045.9 (256.6)	111.7 (19.1)	59.9 (16.0)	82.7 (5.0)	0.55 (0.04)
Dapsone/pyrimethamine vs. aerosolized pentamidine	5	2013.0 (596.9)	117.3 (19.1)	45.8 (11.3)	74.8 (10.5)	0.61 (0.09)
<i>p</i> Value		0.08	0.77	0.49	0.61	0.61

Abbreviations: AIDS = acquired immune deficiency syndrome; HIV = human immunodeficiency virus; SE = standard error.

^aSE in parentheses.

^bQuality scores according to Chalmers *et al.* [27]: minimum value 0, maximum value 1.

pneumonia, but the magnitude of difference is less than indicated by the indirect comparison. The difference in point estimates that were derived from the direct or indirect comparisons was, however, not statistically significant [$p = 0.11$].

To explore the possible reasons for the differences in the magnitude of the effect in the indirect and direct comparisons, we first examined whether there were differences in important prognostic cofactors in trials comparing either trimethoprim-sulphamethoxazole or dapsone/pyrimethamine with pentamidine. We looked at the following variables CD4 cell count, time of follow-up, the number of patients with antiretroviral therapy at the start of the trial, and the methodologic quality of the study [27] and found no statistically significant differences (Table 3). We then explored whether such differences existed in trials that were included in either direct or indirect comparison (Table 4).

There was a trend showing that patients in trials with direct comparisons were somewhat less immunosuppressed with a higher mean CD4 cells at study entry (group mean 126.2 versus 113.7, $p = 0.48$) and were less likely to have a diagnosis of AIDS (41.9% versus 54.7%, $p = 0.43$). Trials with direct comparisons had a longer follow-up than the ones with indirect comparison (mean person follow-up 1990.7 versus 1391.3, $p = 0.59$). The numbers of patients with antiretroviral therapy at study onset were similar in trials with direct or indirect comparisons. Trials included with direct comparisons had comparable methodologic quality scores as trials included with indirect comparisons.

DISCUSSION

We have presented an alternative to simple comparisons of treatment arms for dealing with situations when randomized

TABLE 4. Differences in baseline characteristics of randomized controlled trials included in either direct or indirect meta-analytic comparison on the efficacy of two drug regimens compared to standard prophylaxis of *Pneumocystis carinii* pneumonia in patients with HIV infection

Type of comparison	No. of trials in comparison	Baseline characteristics ^a				
		Mean person years follow-up	Mean CD4 cells (mm ³)	Patients with		Individual study quality score ^b
				AIDS (%)	Antiretroviral therapy (%)	
Direct comparison						
Trials comparing trimethoprim-sulfamethosazole with dapsone/pyrimethamine	8	1990.7 (928.7)	126.2 (9.3)	41.9 (8.1)	83.6 (6.4)	0.53 (0.05)
Indirect comparison						
Trials comparing trimethoprim-sulfamethoxazol or dapsone/pyrimethamine with aerosolized pentamidine	14	1391.3 (285.9)	113.7 (12.0)	54.7 (10.8)	80.3 (4.6)	0.57 (0.04)
<i>p</i> Value		0.59	0.48	0.43	0.67	0.54

Abbreviations: AIDS = acquired immune deficiency syndrome; HIV = human immunodeficiency virus; SE = standard error.

^aSE in parentheses.

^bQuality scores according to Chalmers *et al.* [27]: minimum value 0, maximum value 1.

trials comparing two innovative treatments with placebo or standard treatment, but not comparing the two treatments directly with one another, are available. Our model is also based on meta-analysis, but instead of direct comparison of treatment arms, it compares the magnitude of treatment effects in randomized trials of the interventions in comparison to placebo or control groups.

Our example demonstrates that while this new model protects against some biases, it may still lead to an inaccurate estimate of treatment effect. We had access not only to indirect, but to direct comparison of treatment effects, and found that the indirect comparison led to a substantial overestimation of the relative effectiveness of trimethoprim-sulphamethoxazole relative to dapsone/pyrimethamine when compared with the estimate derived from trials with direct comparisons. Considering why this might be involves reviewing the reasons a meta-analysis utilizing indirect comparisons might yield an inaccurate estimate of relative effectiveness.

One reason for discrepancies in effect sizes between direct and indirect comparison could be related to the different efficiency of the two methods to assess underlying treatment effects. We have tried to assess the efficiency of the two methods.

From equation 5 (see appendix), we see that $\text{Var}[\ln \text{OR}_{\text{ind}}]$ is made up of two components, the variances associated with the AB and CB studies. To consider the efficiency of the indirect method, suppose for simplicity that all the studies have the same weight w . Then the variance of the indirect estimate is equal to $(g + h)/w$.

In contrast, consider the estimation of OR_{AC} if it were possible to carry out direct comparison studies, with the same number of patients as in the indirect studies. Then there would be twice as many patients per group available (assuming equal group sample sizes), and hence the variance would be $(g + h)/2w$. In other words, if one had the choice of allocating subjects into direct or indirect comparison studies, the direct method would be twice as efficient as the indirect method. Note, however, that in practice meta-analyses are carried out on existing studies, so the precision of the estimated AC effect is not under the control of the analyst; questions of efficiency are therefore somewhat moot.

Studies of one treatment may be methodologically inferior with, for example, randomization open to manipulation, or inadequate blinding. Typically, these deficiencies will lead to overestimates of treatment effect [28]. While we were unable to detect methodologic differences, this does not exclude their presence.

A second source of bias may be in the measurement of outcomes. Criteria for, or the thoroughness with which the diagnosis of *Pneumocystis carinii* pneumonia is sought may differ in trials of one drug against standard care versus the other against standard care. The length of follow-up may also differ. For example we found that trials used for direct comparison followed study subjects for an appreciably

longer time. The likelihood that patients included in these trials experience an event as a consequence of the continuing deterioration of their immune system, or switch to less powerful drugs because of side effects, is considerably higher than in trials with a shorter follow-up. This difference in length of follow-up may therefore explain the smaller difference in treatment effect we observed in the direct comparisons.

Finally, the efficacy of treatment may differ in subpopulations of patients. Treatment may have a larger impact in less severely affected patients, in older or young patients, or in more compliant groups. We found that patients in the direct comparison trials had less advanced HIV infection with higher CD4 cell counts and fewer patients with AIDS. These important differences in baseline characteristics and study design may illustrate to some extent why the indirect comparison lead to a bigger effect size than the one derived from direct comparison. Considering these issues, one could easily imagine a treatment of equal effectiveness showing greater impact if tested in more responsive populations.

Is it possible that direct and indirect comparisons yield different results but that both are valid? This would occur if the difference between treatments A and C differed in different populations (say, populations X and Y). Assume the direct comparisons of A versus C all included only population X, and the indirect comparisons (involving trials of A versus B and C versus B) only included population Y. Were this the case, then the direct comparisons would yield valid estimates of the effect of A versus C in population X, and the indirect comparisons would yield valid estimates A versus C in population Y. In the current example, we found no such clear differences in the populations included in the direct and indirect comparisons.

Is the primacy of the direct comparisons threatened if the direct comparison trials are methodologically weaker than the indirect comparison trials? In this situation, the indirect comparisons may indeed yield a more accurate estimate of treatment effect. Unfortunately, if direct and indirect estimates differ under these circumstances, either or both of methodological superiority of the indirect comparison trial or underlying differences in the populations could explain the differences. There would be no way of distinguishing between these possibilities. In the current example, the similarity of the methodological quality of the direct and indirect comparison trials reduces this particular concern.

In the example we presented, we can only speculate on whether the differences between the trials explain the different estimates of treatment effect. Thus, our findings suggest that the strength of inference associated with indirect comparisons is inevitably limited, even in the absence of demonstrable sources of bias. Given the inevitable limitations of indirect comparisons, clinicians may put greater weight on factors other than apparent relative effectiveness in choosing between treatments. These factors may include toxicity, convenience, and cost.

This is not to say that indirect comparisons should be

discarded altogether. Better quality trials, more precise estimates of effectiveness, greater similarity between the populations, and larger differences in effectiveness estimates between treatments all add weight to a finding from indirect comparison of differences between two treatments' effect.

In conclusion, summary estimates in meta-analysis should be based on direct comparisons whenever possible. Indirect comparisons, when undertaken, should involve comparisons of the magnitude of effect in the two sets of controlled trials, rather than focusing exclusively on the treatment arms. Investigators making indirect comparisons should look carefully for differences in methodology, outcome measurement, or the populations included in studies of the two drugs. Even in the absence of evident differences, the strength of inference from indirect comparisons is limited, and if the questions is important then direct comparisons should be undertaken. In the absence of such comparisons, considerations other than relative drug effectiveness gain greater importance in making treatment decisions.

Dr. Bucher was supported by a grant from the Swiss National Research Foundation (grant number 32-38793.93). Dr. Walter holds a National Health Scientist award from Health Canada. We are indebted to Robin Roberts for suggestions and development of our statistical model.

References

1. Felson DT, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis. *Arthritis Rheum* 1990; 33: 1449-1461.
2. Boers M, Guyatt GH, Oxman AD. Combined effect size: Comment on the meta-analysis of second-line drugs in rheumatoid arthritis. *Arthritis Rheum* 1991; 34: 1342.
3. O'Brien BJ, Anderson DT, Goeree R. Cost-effectiveness of enoxaparin versus warfarin prophylaxis against deep-vein thrombosis after total hip replacement. *Can Med Assoc J* 1994; 150: 1083-1090.
4. Bucher HC, Griffith L, Guyatt GH, Opravil M. Meta-analysis of prophylactic treatments against *Pneumocystis carinii* pneumonia and toxoplasmic encephalitis. *Acquir Immune Defic Syndr* 1997. (In press)
5. Hardy WD, Feinberg J, Finkelstein DM, Power ME, He W, Kaczka C. A controlled trial of trimethoprim-sulphamethoxazole or aerosolized pentamidine for secondary prophylaxis of *Pneumocystis carinii* pneumonia in patients with the acquired immunodeficiency syndrome. *N Engl J Med* 1992; 327: 1842-1848.
6. Schneider MME, Hoepelman AIM, Karrel J, Schattenkerk E, Nielsen TL, Van der Graff Y. A controlled trial of aerosolized pentamidine or trimethoprim-sulphamethoxazole as primary prophylaxis against *Pneumocystis carinii* pneumonia in patients with human immunodeficiency virus infection. *N Engl J Med* 1992; 327: 1836-1841.
7. May T, Beuscart C, Reynes J, Marchou B, Leclercq P, Lebas B, et al. Trimethoprim-sulphamethoxazole versus aerosolized pentamidine for primary prophylaxis of *Pneumocystis carinii* pneumonia: A prospective, randomized, controlled clinical trial. *J Acquir Immune Defic Syndr* 1994; 7: 457-462.
8. Nielsen TL, Jensen BN, Nelsing S, Mathiesen LR, Skinhøj P, Nielson JO. Randomized study of sulphamethoxazole-trimethoprim versus aerosolized pentamidine for secondary prophylaxis of *Pneumocystis carinii* pneumonia in patients with the acquired immunodeficiency syndrome. *Scand J Infectious Dis* 1995; 27: 217-220.
9. Michelet C, Raffi F, Besnier JM, Chennebault JM, Moulichon F, Milpied B, et al. Cotrimoxazole versus aerosolized pentamidine for primary prophylaxis of *Pneumocystis carinii* pneumonia (abstract 1071). In: **Program and Abstracts of the 33rd Inter-science Conference on Antimicrobial Agents and Chemotherapy, Washington, DC.** American Society for Microbiology; 1993: 315.
10. Stellini R, Castelnuovo F, Parainfo G, Scalzini A. Trimethoprim-sulphamethoxazole versus pyrimethamine in the primary prophylaxis of toxoplasmic encephalitis in HIV positive patients: Preliminary results of a randomized prospective study (Abstract PO-B10-1505). In: **Program and Abstracts of the 9th International Conference on AIDS Berlin, Germany;** 1993.
11. Smith DE, Hills DA, Gazzard BG. Patient tolerance of nebulized pentamidine or cotrimoxazole as secondary prophylaxis for *Pneumocystis carinii* pneumonia. *AIDS* 1992; 6: 1044-1045.
12. Rozenbaum W, Tavares M, Navarette MS, Gharakhanian S, Cardon B, Lebrette MG. Evaluation of trimethoprim sulphamethoxazole vs aerosolized pentamidine with concomitant onset of zidovudine in primary *Pneumocystis carinii* pneumonia prophylaxis (abstract no. W.B. 2247). In: **Program and Abstracts of 7th International Conference on AIDS.** Amsterdam 1991 June 16-21; 7: 243.
13. Rizzardi GP, Lazzarin A, Musicco M, Frigerio D, Maillard M, Lucchini M, et al. Risks and benefits of aerosolized pentamidine and cotrimoxazole in primary prophylaxis of *Pneumocystis carinii* pneumonia in HIV-1-infected patients: A two-year Italian multicentric randomized controlled trial. *J Infection* 1996; 32: 123-131.
14. Girard PM, Landman R, Gaudebout C, Oliveras R, Saimot AG, Jelazko P, et al. Dapsone-pyrimethamine compared with aerosolized pentamidine as primary prophylaxis against *Pneumocystis carinii* pneumonia and toxoplasmosis. *N Engl J Med* 1993; 328: 1514-1520.
15. Opravil M, Hirschel B, Lazzarin A, Heald A, Pechere M, Rüttimann S, et al. Once weekly administration of dapsone/pyrimethamine vs. aerosolized pentamidine as combined prophylaxis for *Pneumocystis carinii* pneumonia and toxoplasmic encephalitis in human immunodeficiency virus-infected patients. *Clin Infect Dis* 1995; 20: 531-541.
16. Slavin MA, Hoy JF, Stewart K, Pettinger MB, Lucas CR, Kent SJ. Oral dapsone versus nebulized pentamidine for *Pneumocystis carinii* pneumonia prophylaxis: An open randomized prospective trial to assess efficacy and hematological toxicity. *AIDS* 1992; 6: 1160-1174.
17. Torres R, Barr M, Thorn M, Gregory G, Kiely S, Chanin E, et al. Randomized trial of dapsone and aerosolized pentamidine for the prophylaxis of *Pneumocystis carinii* pneumonia and toxoplasmic encephalitis. *Am J Med* 1993; 95: 573-582.
18. Salmon D, Saba J, Fontbonne A, Aboulker JP, Schwartz D, Vilde JL. Dapsone vs. pentamidine aerosols for secondary prophylaxis of *Pneumocystis carinii* pneumonia in AIDS patients. (abstract 1474) In: **Program and Abstracts of the 32nd Inter-science Conference on Antimicrobial Agents and Chemotherapy, Washington, DC.** American Society for Microbiology; 1992: 356.
19. Podzamczar D, Santin M, Jimenez J, Casanova A, Bolao F, Rifiand G, et al. Thrice-weekly cotrimoxazole is better than weekly dapsone-pyrimethamine for the primary prevention of *Pneumocystis carinii* pneumonia in HIV-infected patients. *AIDS* 1993; 7: 501-506.
20. Blum RN, Miller LA, Gaggini LC, Cohn DL. Comparative trial of dapsone versus trimethoprim-sulphamethoxazole for

- primary prophylaxis of *Pneumocystis carinii* pneumonia. *J Acquir Immune Defic Syndr* 1992; 5: 341–347.
21. Podzamczar D, Salazar A, Jiménez J, Consiglio E, Santín M, et al. Intermittent trimethoprim-sulphamethoxazole compared with dapsone-pyrimethamine for the simultaneous primary prophylaxis of *Pneumocystis pneumonia* and toxoplasmosis in patients infected with HIV. *Ann Intern Med* 1995; 122: 755–761.
 22. Sirera G, Clotet B, Romeu J, Raventos A, Gomez M, Paredes R. A randomized study of primary prophylaxis of *Pneumocystis carinii* pneumonia and cerebral toxoplasmosis with twice weekly dapsone-pyrimethamine versus thrice weekly cotrimoxazole (abstract 295). In: **Program and Abstracts of the 2nd National Conference on Human Retroviruses and Related Infections, Washington, DC; January 29–February 2, 1995.**
 23. Bozzette SA, Finkelstein DM, Spector SA, Frame P, Powderly WG, He W, et al. A randomized clinical trial of three anti-pneumocystis agents in patients with advanced HIV infection. *N Engl J Med* 1995; 332: 639–699.
 24. Antinori A, Murri R, Tamburrini E, De Luca A, Ortona L. Failure of low dose dapsone-pyrimethamine in primary prophylaxis of *Pneumocystis carinii* pneumonia. *Lancet* 1992; 340: 788.
 25. Mallolas J, Zamora L, Gatell JM, Miró JM, Vernet E, Valls ME et al. Primary prophylaxis for *Pneumocystis carinii* pneumonia: A randomized trial comparing cotrimoxazole, aerosolized pentamidine and dapsone plus pyrimethamine. *AIDS* 1993; 7: 59–64.
 26. Tocchetti A, Tambini R, Allegro A, Longoni E, Rinaldi E. Four different regimens for prevention of *Pneumocystis carinii* pneumonia and toxoplasma encephalitis in HIV-infected patients. *AIDS* 1994; 8: 272–274.
 27. Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, Reitman D, et al. A method for assessing the quality of a randomized control trial. *Contr Clin Trials* 1981; 2: 31–49.
 28. Schulz KF, Chalmers I, Hayes RJ, Altman G. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408–412.
 29. Fleiss JF. **Statistical Methods for Rates and Proportions.** New York: John Wiley & Sons; 1981: 160–187.

APPENDIX

Effect Size Estimate

Suppose that the outcome rate for patients on treatment A in subgroup i is P_{Ai} , and similarly for the other treatments. Subgroups might be defined in terms of age, severity of disease, or other relevant covariates. From studies comparing treatments A and B in subgroup i the treatment effect might be assessed through the odds ratio

$$OR_{AB}(i) = \frac{P_{Ai}/(1 - P_{Ai})}{P_{Bi}/(1 - P_{Bi})}$$

An overall estimate of treatment effect can be obtained from the Mantel-Haenszel method, for instance, as a weighted average of the subgroup estimates, e.g.,

$$OR_{AB} = \sum w_i OR_{AB}(i) / \sum w_i \quad (1)$$

where w_i is a suitable weight reflecting the variance of the estimate $OR_{AB}(i)$ [29]. A similar comparison from the C versus B studies

gives

$$OR_{CB} = \sum w'_i OR_{CB}(i) / \sum w'_i \quad (2)$$

where w'_i denotes a weight appropriate to the CB studies, in general not equal to w_i .

The ratio of (1) and (2) provides the indirect estimate of the difference in effect between treatments A and C. Note that if (I) $OR_{AB}(i) = OR_{AB}$ and (II) $OR_{CB}(i) = OR_{CB}$, then the ratio becomes

$$\frac{OR_{AB}}{OR_{CB}} = OR_{ind} \quad (3)$$

where OR_{ind} is, conceptually, the A versus C odds ratio for any subgroup that is in common between the AB and CB studies. The special conditions (I) and (II) are equivalent to no treatment $\times$ subgroup interaction in either set of studies. Result (3) indicates that if no such interaction is present, the indirect estimate is consistent for the true AC odds ratio. A log transformation of (3) yields

$$\ln(OR_{ind}) = \ln(OR_{AB}) - \ln(OR_{CB}). \quad (4)$$

Because OR_{AB} and OR_{CB} are estimated from different studies, they are statistically independent, and hence the variance of OR_{ind} can be obtained from

$$\text{Var}(\ln OR_{ind}) = \text{Var}(\ln OR_{AB}) + \text{Var}(\ln OR_{CB}) \quad (5)$$

Derivation of the Test Statistics

In meta-analysis the overall measure of association, $\bar{y}$ (such as the odds ratio) is usually taken as a weighted average of the measure from the k individual studies, with weights being the inverse of the variance for each study, as in equation (1). Following Fleiss [29], under a null hypothesis of no association in any of the k studies ($y_i = 0$), then $\chi^2_{total} = \sum w y^2$ is distributed approximately as χ^2 with k degrees of freedom, i.e., χ^2_k . The statistic

$$\chi^2_{association} = (\sum w y)^2 / (\sum w) \quad (6)$$

is distributed as χ^2_1 , and measures the average association across the studies. To assess the variation between studies, the term

$$\chi^2_{heterogeneity} = \chi^2_{total} - \chi^2_{association}$$

is used, distributed as χ^2_{k-1} . Note also that

$$\chi^2_{heterogeneity} = \sum w (y - \bar{y})^2. \quad (7)$$

Let us now consider the situation where g studies have compared treatment A to standard treatment B or placebo and h studies have compared treatment C to standard treatment B or placebo. Our interest is in obtaining an indirect estimate of the association of A versus C. For the comparison of A versus B the overall association is given by

$$\chi^2_{ABtotal} = \sum_{AB} w y^2$$

(with the sum being over the AB studies) and its components are

$$\chi^2_{ABassociation} = (\sum_{AB} w y)^2 / (\sum_{AB} w)$$

and

$$\chi_{AB \text{ heterogeneity}}^2 = \sum_{AB} w(y - \bar{y}_{AB})^2.$$

Similarly for the comparison for C versus B, the overall association is given by

$$\chi_{CB \text{ total}}^2 = \sum_{CB} w\bar{y}^2$$

and its components are

$$\chi_{CB \text{ association}}^2 = (\sum_{CB} w\bar{y})^2 / (\sum_{CB} w)$$

and

$$\chi_{CB \text{ heterogeneity}}^2 = \sum_{CB} w(y - \bar{y}_{CB})^2.$$

To obtain an indirect test of the A versus C effect from the paired comparisons of A versus B and of C versus B, first note that the overall measure of variation can be taken as the sum of the total χ^2 values from the paired comparisons:

$$\chi_{\text{total}}^2 = \sum_{AB+CB} w\bar{y}^2 = \chi_{AB \text{ total}}^2 + \chi_{CB \text{ total}}^2$$

which is distributed as χ_{g+h}^2 .

From (6), note that the association over all studies is measured by

$$\chi_{\text{association}}^2 = (\sum_{AB+CB} w\bar{y})^2 / (\sum_{AB+CB} w)$$

which is distributed as χ_1^2 . The heterogeneity among all studies is, from (7), measured by

$$\chi_{\text{heterogeneity}}^2 = \sum_{AB+CB} w(y - \bar{y})^2 \quad (8)$$

which is distributed as χ_{g+h-1}^2 .

For the indirect comparison of treatment A versus treatment C, we propose the statistic $\chi_{AC \text{ indirect}}^2$ being the difference of the over-

all heterogeneity (equation 7) and the component AB and BC measures of heterogeneity, i.e.,

$$\chi_{AC \text{ indirect}}^2 = \chi_{\text{heterogeneity}}^2 - \chi_{AB \text{ heterogeneity}}^2 - \chi_{CB \text{ heterogeneity}}^2 \quad (9)$$

We now show that χ^2 is distributed approximately as χ_1^2 , and can be used to test $H_0: \bar{y}_{AB} = \bar{y}_{BC}$, i.e., to provide the indirect test of treatment A versus treatment C. From the definitions of the terms of the right-hand side of (9), we have

$$\begin{aligned} \chi_{AC \text{ indirect}}^2 &= \sum_{AB+CB} w(y - \bar{y})^2 - \sum_{AB} w(y - \bar{y})^2 \\ &\quad - \sum_{CB} w(y - \bar{y})^2 \end{aligned}$$

Expanding each of the squares, terms in y^2 cancel, leaving

$$\begin{aligned} \chi_{AC \text{ indirect}}^2 &= -2 \sum_{AB+CB} w\bar{y}y + 2 \sum_{AB} w\bar{y}_{AB} y \\ &\quad + 2 \sum_{CB} w\bar{y}_{CB} y + \sum_{AB+CB} w\bar{y}^2 - \sum_{AB} w\bar{y}_{AB}^2 \\ &\quad - \sum_{CB} w\bar{y}_{CB}^2 = \sum_{AB} w\bar{y}_{AB}^2 + \sum_{CB} w\bar{y}_{CB}^2 \\ &\quad - \sum_{AB+CB} w\bar{y}^2 \end{aligned} \quad (10)$$

Now $\bar{y}$ is a weighted average of the mean effect sizes in the AB and CB studies, so

$$\bar{y} = \frac{\bar{y}_{AB} \sum_{AB} w + \bar{y}_{CB} \sum_{CB} w}{\sum_{AB} w + \sum_{CB} w} \quad (11)$$

Substituting from (11) and (10) and then simplifying, we have that

$$\chi_{AC \text{ indirect}}^2 = \frac{(\bar{y}_{AB} - \bar{y}_{CB})^2 (\sum_{AB} w)(\sum_{CB} w)}{\sum_{AB} w + \sum_{CB} w} \quad (12)$$

Thus testing $H_0: \bar{y}_{AB} = \bar{y}_{BC}$, and hence the indirect AC comparison of interest.