DDT Vol. 9, No. 21 November 2004

Beyond classical meta-analysis: can inadequately reported studies be included?

Chris Robertson, Nik Ruzni Nik Idris and Peter Boyle

Classical meta-analysis requires the same data from each clinical trial, thus data-reporting must be of a high-quality. Imputation methods are used to include studies that provide incomplete information on variability and the fixed and random effects of a drug. Regression models can be used to include studies other than randomized placebo-controlled studies. In the example outlined here, the use of non-randomized single-arm studies and studies against comparator treatments has little influence on the estimation of the treatment effect in comparison with placebo, an effect that is based on the randomized placebo-controlled studies. The inclusion of other studies serves to increase the precision of the effect of the treatment compared with baseline. Although multiple imputation techniques enable a larger number of studies to be included, which will typically increase the precision of the estimated effect, a careful sensitivity analysis is also required.

Chris Robertson*
Nik Ruzni Nik Idris
Department of Statistics and Modelling Science
University of Strathclyde
26 Richmond Street
Glasgow
Scotland G1 1XH
*e-mail:
chris@stams.strath.ac.uk
Chris Robertson
Scottish Centre for Infection and Environmental Health
Clifton House
Clifton Place
Glasgow
Scotland G3 7LN
Peter Boyle
International Agency for Research on Cancer
150 Cours Albert Thomas
69372 Lyon Cedex 08
France

▼ Meta-analysis is increasingly used in drug development [1] and studies are often designed with the consideration that a meta-analysis could be performed [2]. Good meta-analyses successfully combine information from different studies to provide a better understanding of the effect of a treatment [3]; however, there are examples of meta-analyses that have generated controversial results [4,5].

The meta-analysis of Finasteride [3] in benign prostatic hyperplasia (BPH) was based on all randomized placebo-controlled trials that were performed over a period of at least 12 months: three of the studies had been published, one was reported as an abstract and two were complete but unpublished. To avoid selection bias, all placebo-controlled trials were included. Furthermore, the authors had access to individual patient data [6] from all studies. In the comparison of Finasteride with placebo among the six studies, the results revealed heterogeneity that was related to prostate volume. The effect of Finasteride was greater in studies in which men averaged larger prostate volumes, an effect that was not investigated in the individual studies.

In the meta-analysis of all eight breast cancer screening trials [4,5], the published data was studied and prespecified inclusion and exclusion criteria were used. One of the conclusions reached was that there was no mortality benefit associated with screening. However, this deduction was based on only two of the studies, which were deemed to have no major defects in their design and execution, and resulted in controversy [7–11]. The meta-analysis of breast cancer screening trials proved more difficult because the study protocols varied.

In the meta-analysis of Finasteride, although the entry criteria varied, a similar protocol was used in each study. The two meta-analyses started from the same point of using data on all randomized studies. In one study, data from all studies are included and a meta-regression model is used to investigate heterogeneity; in the second study [4,5], studies are included only if they satisfy the entry criterion. Although there are strong reasons for omitting studies in a meta-analysis that are concerned with known bias, it is believed that it is reasonable to start from the premise of using as much data as possible.

This review highlights and evaluates the methods that can be used to include studies in a meta-analysis when: (i) data on the variability of the treatment effect cannot be obtained from the study publication; and (ii) when there is data available from non-randomized studies. Such methods might be used in a meta-analysis that incorporates published reports. Methods used to deal with missing vari-

1359-6446/04/$ – see front matter PII: S1359-6446(04)03274-X

Table 1. Data on duration of study and peak urinary flow rate extracted from published papers and reports

Study type	Group	Duration of study (days)	Peak flow (ml s^{-1})				N	Refs
			Baseline		End of study			
			Mean	SD	Mean	SD		
OL	Permixon	90	11.74	8.82	14.67	15.93	592	[24]
RPC	Permixon	30	11.84	7.49	15.26	11.89	82	[15]
	Placebo	30	12.42	8.25	13.48	8.59	94	
RPC	Permixon	30	10.70	10.24	16.10	16.75	46	[16]
	Placebo	30	10.08	10.24	10.58	13.12	39	
RPC	Permixon	30	10.33	3.42	13.70	3.56	15	[17]
	Placebo	30	9.23	2.64	9.43	2.72	15	
RPC	Permixon	58	12.90	NA	16.20	NA	14	[18]
	Placebo	69	11.20	NA	11.80	NA	13	
RPC	Permixon	60	9.59	NA	13.72	NA	11	[19]
	Placebo	60	10.22	NA	12.18	NA	11	
RPC	Permixon	84	6.15	NA	8.50	NA	33	[20]
	Placebo	84	6.30	NA	8.60	NA	37	
RC	Permixon	84	9.75	7.29	11.25	8.77	20	[23]
	Prazosin	84	10.36	7.86	10.83	11.07	22	
RC	Permixon	180	10.62	2.78	13.30	6.72	467	[21][a]
	Finasteride	180	10.76	3.09	14.02	7.38	484	
RC	Permixon	21	10.40	2.70	13.20	4.20	31	[22]
	Alfuzosin	21	9.20	2.70	13.90	7.90	32	

[a]In the study performed by Carraro *et al.* [21], changes in the mean of 2.68 ml s^{-1} (SD of 6.36) and 3.26 ml s^{-1} (SD of 6.84) were reported for Permixon and Finasteride, respectively. Imputed SDs were 6.32 and 11.19 ml s^{-1}.
Abbreviations: N, number of subjects for which data is available; NA, not available; OL, open-label observational studies with no randomization; RC, randomized studies with comparator drug therapies; RPC, randomized placebo-controlled studies; SD, standard deviation.

ability estimates in study publications are unnecessary when individual patient data is available [6,12,13]. Data are presented that illustrate the problems and then demonstrate the use of multiple imputation methods to overcome the lack of data on treatment effect variability and meta-regression methods that are employed when data from non-randomized studies can be accessed.

Missing study-level variability estimates

Meta-analysis of Permixon

A meta-analysis of all available clinical trial data was undertaken to estimate the effects of Permixon [14], a drug therapy for BPH. Data from all published trials involving Permixon, where peak urinary flow (Q_{max}) was recorded, were available. There were six randomized placebo-controlled studies [15–20], and data from a further three randomized clinical trials of Permixon against other drugs were available. The largest study compared Permixon to Finasteride [21]; other studies compared Permixon with Alfuzosin [22] and with Prazosin [23]. Data from one large open-label study of Permixon [24] were also included.

The raw data (Table 1) were extracted from published papers and technical reports. Data quality is not uniform and some early studies publish mean values without standard deviations (SDs). Most studies quote the mean for the baseline and end of study without the SD of the difference. This is a serious limitation and one which the CONSORT guidelines [25] seek to address. Table 1 was constructed using the techniques for obtaining SDs, such as using *p*-values, *F*- and *t*-statistics and calculations from histograms [13,26]. A meta-analysis involves the calculation of a weighted average of study estimates of treatment difference, where the weight is typically the inverse variance of the study estimate. The estimates of treatment difference can be calculated from the trial publication (Table 1) on the assumption that the same subjects contribute to baseline and end of study, but it is the variance that must be imputed.

If the criteria of only using studies that publish appropriate data are adopted, which are normally taken to be an estimated effect and its standard error (SE), then only one study [21] (not one comparing Permixon with placebo) can be used, which is the largest and most recent randomized study. Thus, data from eight out of nine randomized studies are discarded, which clearly represents a huge loss of information. In terms of patients, this study represents 954 out of 1499 patients in randomized studies. In terms of understanding heterogeneity, it is the number of studies included that is more important than the number of patients [27].

Statistical issues

The effect for treatment group k in study j is given by Equation 1:

$$x_{jk} = \left(\overline{y}_{e.jk} - \overline{y}_{b.jk}\right) \qquad \text{[Eqn 1]}$$

where $\overline{y}_{e.jk}$ denotes the mean end of study value and $\overline{y}_{b.jk}$ is the corresponding baseline mean. The SE can be calculated using Equation 2:

$$s_{jk} = \sqrt{\frac{s^2_{ejk}}{n_{ejk}} + \frac{s^2_{bjk}}{n_{bjk}} - 2r_{jk}\frac{s_{ejk}}{\sqrt{n_{ejk}}}\frac{s_{bjk}}{\sqrt{n_{bjk}}}} \qquad \text{[Eqn 2]}$$

where, s^2_{ejk} is the end of study variance and n_{ejk} the sample size, with s^2_{bjk} and n_{bjk} the corresponding values at baseline. To calculate the mean difference, each subject needs a baseline and end of study value, therefore $n_{bjk} = n_{ejk}$. The repeat observations for each individual imply that the baseline and end of study values have a positive correlation (denoted by r_{jk}).

If the SDs and the correlation are known, then the SE of the difference can be calculated and this forms the basis of the imputation [26]; from the individual patient data for one study [21], the correlations were calculated to be 0.34 for Permixon and 0.38 for Finasteride. For those studies that published means and SDs at baseline and end of study (seven studies), the effect was calculated as the difference in the means and the SD was calculated using a common value of 0.36 for the correlation. With more complete data, it would be appropriate to check the validity of this approach and if necessary use different correlations.

For the three studies that published only mean values and not SDs, the SDs at baseline and end of study were separately imputed from the weighted average of the SDs from the other studies. At baseline this value can be calculated from Equation 3:

$$s_{b,imputed} = \sqrt{\frac{\sum_{jk} n_{bjk} s^2_{bjk}}{\sum_{jk} n_{bjk}}} \qquad \text{[Eqn 3]}$$

If some of the studies are small, then $n_{bjk} - 1$ could be used in place of n_{bjk}. Because there is little evidence to suggest that the SD is dependent on treatment, we advocate the pooling of treatment arms in the imputation of the SD. Indeed, the pooled two-sample *t*-test, a common test of efficacy, assumes common SDs. If there is evidence of gross differences in SDs over a treatment group, then it would not be appropriate to pool the treatment arms and individual imputed SDs should be calculated.

The general strategy is one of using the available information, that is, imputing values from the other studies when information is unavailable. This is one of the unsatisfactory aspects of meta-analyses of published data. To investigate the consequences of these imputations, for the example outlined here, sensitivity analyses and a multiple imputation analysis were performed.

Fixed effect meta-analysis

An estimate of the treatment difference between Permixon and placebo and the SE of this treatment difference are required from each study. The estimate of the treatment difference is given by Equation 4 and the SE of the treatment difference is calculated from Equation 5:

$$\hat{\theta}_j = x_{j,Permixon} - x_{j,Placebo} \qquad \text{[Eqn 4]}$$

$$s(\hat{\theta}_j) = \sqrt{s^2_{j,Permixon} + s^2_{j,Placebo}} \qquad \text{[Eqn 5]}$$

The fixed effect estimate is a weighted average where the weights are the inverse of the SEs (Equation 6 gives rise to Equations 7 and 8) [13].

$$w_j = \frac{1}{[s(\hat{\theta}_j)]^2} \qquad \text{[Eqn 6]}$$

$$\hat{\theta} = \frac{\sum_j w_j \hat{\theta}_j}{\sum_j w_j} \qquad \text{[Eqn 7]}$$

$$s(\hat{\theta}) = \sqrt{\frac{1}{\sum_j w_j}} \qquad \text{[Eqn 8]}$$

The imputation method

Three randomized studies [15–17] for Permixon can be used after imputation of the correlation between the baseline and end of study values. Analysis of the data from these three studies indicates that the effects of different correlations are minimal on the estimated effect (Table 2). The larger the correlation, the smaller the SE leading to a smaller SE for the estimate.

To include the three randomized placebo-controlled studies without SD information, it is necessary to perform

Table 2. Effects of imputation on the randomized placebo-controlled studies

Study	Baseline SD (ml s^{-1})	End of study SD (ml s^{-1})	Correlation	Estimate	Standard error	Lower 95% CI	Upper 95% CI
Three studies with SD information	NA	NA	0.20	3.04	1.07	0.94	5.13
	NA	NA	0.36	3.04	0.96	1.16	4.92
	NA	NA	0.50	3.04	0.85	1.37	4.71
Six studies with variable correlation	6.32	11.19	0.20	2.64	0.96	0.77	4.52
	6.32	11.19	0.36	2.65	0.86	0.96	4.34
	6.32	11.19	0.50	2.67	0.77	1.16	4.17
Six studies with fixed correlation	2.64	2.72	0.36	1.55	0.48	0.60	2.49
	2.64	11.19	0.36	2.64	0.86	0.96	4.33
	2.64	16.75	0.36	2.84	0.91	1.06	4.63
	6.32	2.72	0.36	2.16	0.72	0.75	3.57
	6.32	11.19	0.36	2.65	0.86	0.96	4.34
	6.32	16.75	0.36	2.84	0.91	1.05	4.62
	10.24	2.72	0.36	2.58	0.84	0.93	4.23
	10.24	11.19	0.36	2.73	0.88	1.00	4.45
	10.24	16.75	0.36	2.85	0.91	1.06	4.64

The columns detailing values for end of study, correlation and estimate give the values used in the imputation process.
The columns detailing values for estimate, standard error and lower and upper 95% CI give the estimated treatment difference between Permixon and placebo.
For the three studies for which SDs were available, there is no imputation of SDs.
Abbreviations: CI, confidence interval; NA, not applicable; SD, standard deviation.

additional imputation of the SDs at baseline and end of study. The imputed SDs were calculated by pooling the arms of all the studies (including the open-label and comparative studies) with SDs at baseline and end of study to give imputed SDs of 6.32 ml s^{-1} and 11.19 ml s^{-1} for the baseline and end of study, respectively; the two values are derived from data from seven studies comprising 13 treatment arms and 1939 patients. To give a more precise estimate [26], all studies were included in the analysis. However, there is an argument for including only the randomized placebo-controlled trials, in this scenario the estimates would have been 8.33 ml s^{-1} and 11.49 ml s^{-1} for the baseline and end of study, respectively. Two of the three studies that did not provide SDs were small, comprising only 22 and 27 patients, and the treatment difference is smaller than in the three randomized studies with SD information. Consequently, direct comparison of the estimates obtained for all six studies with a variable correlation (Table 2) with the estimates obtained for the three randomized trials shows a lower estimate, 2.6 ml s^{-1} compared with 3.0 ml s^{-1}. The SEs are smaller for the six studies because more studies, and consequently more patients, are included. Changing the correlation primarily influences the SE.

To investigate the sensitivity of the estimates to the baseline and end of study SDs, the correlation is maintained at a fixed value; all six randomized studies were used. Here, the SDs vary from the minimum observed to the maximum. The end of study SD is larger, more variable from study to study and has a greater influence on the meta-analysis estimate compared with the baseline value, which has little effect.

After imputation, the overall effect of Permixon is to increase peak flow by 2.65 ml s^{-1} (SE of 0.86 ml s^{-1}) compared with placebo. This SE does not take into account uncertainty that is associated with the imputation of the unknown values. However, multiple imputation is used to overcome this issue. Values for the missing SDs and correlations are repeatedly imputed from probability distributions. The baseline and end of study variances are sampled from gamma distributions that have the same expected value and variance as the observed baseline and end of study variances. The imputed SDs are the square roots of these values. The correlations are sampled from a normal distribution with mean 0.36 and SD 0.07, which was chosen to ensure that ~95% of correlations would lie between 0.22 and 0.50.

Multiple imputation is repeated 500 times and the mean is 2.649 ml s^{-1}, with a mean for the SEs of 0.860 ml s^{-1}, which correspond to the original estimates. The variability in the estimates is the variability that is induced by the imputation. The SD of these 500 estimates is 0.0577. This is added to the original SE to take into account the imputa-

tion. This gives a value for the SE of 0.862 (calculated as shown in Equation 9).

$$\sqrt{0.8600^2 + 0.0577^2} = 0.8620 \qquad \text{[Eqn 9]}$$

There is a small increase in the SE because the overall estimate is not sensitive to different imputed values (Table 2). This is a typical occurrence because important information about the SE of the treatment difference is contained in the known sample sizes.

The heterogeneity statistic takes a value of Q = 1.73 on five degrees of freedom and there is no need for random effects. Typically, a random effects estimate would be used and the multiple imputation procedure can easily be adapted.

Inclusion of studies other than randomized placebo-controlled studies

Classical meta-analysis is based on the analysis of randomized trials only. There is merit in including non-randomized single-arm studies and studies in which the active treatment is compared with a comparative treatment rather than placebo. This gives no advantage for the estimation of the difference between Permixon and placebo. However, there is a benefit to the estimation of the effect of Permixon

Box 1. Statistical model

Each arm of all the studies is summarized by the effect and its standard error (SE). The meta-analysis model is given by Equation i:

$$x_{jk} = \mu + \tau_k + u_j + \varepsilon_{jk} \qquad \text{[Eqn i]}$$

where μ represents the overall mean effect (this is the average change in the Permixon group), τ_k is the effect of treatment k relative to Permixon, u_j represents the random effect of study j and ε_{jk} represents the sampling variability of the effect, which is assumed to be known and equal to s^2_{jk}.

This is a multilevel or hierarchical model [36] and the estimates were obtained using the MLwiN software (Centre for Multilevel Modelling; http://multilevel.ioe.ac.uk) [37]. Initially, a random effects model was used because it was anticipated that the inclusion of more studies would give rise to a greater heterogeneity. The first level is given by Equation ii and the second level is represented by Equation iii:

$$x_{jk} = \mu_j + \tau_k + \varepsilon_{jk} \qquad \text{[Eqn ii]}$$

$$\mu_j = \mu + u_{oj} \qquad \text{[Eqn iii]}$$

The second level model can be extended to take into account study variables that might influence the average study effect in a systematic way. Such variables include the length of the study, dummy variables for the individual studies and study type; this extension gives Equation iv:

$$\mu_j = \mu + \beta c_j + u_{oj} \qquad \text{[Eqn iv]}$$

where β measures the effect of the study level covariate c_j on the study effect, μ_j.

The first level of the model could also include covariates that might explain imbalance in the treatment arms. In this analysis, we permitted the treatment effects μ and τ_k to vary randomly over the studies (Equation v).

$$\tau_{kj} = \tau_k + u_{1j} \qquad \text{[Eqn v]}$$

However, no study level variation was observed. This meta-regression model is one approach for investigating heterogeneity [29,38,39].

In the analysis of randomized placebo-controlled studies with dummy variables representing the studies for comparison with the imputation method analysis, the model fitted is given by Equations vi and vii:

$$x_{jk} = \mu + \delta_j + \tau_{j1} + \varepsilon_{jk} \qquad \text{[Eqn vi]}$$

$$\tau_{j1} = \tau_1 + u_j \qquad \text{[Eqn vii]}$$

where μ represents the effect of Permixon in the reference study, δ_j is the fixed effect of study j and τ_{j1} is the random treatment difference effect.

Apart from a random effect, this is identical to the fixed effect model used when illustrating the variance imputations. In the analysis of randomized placebo-controlled studies excluding study dummy variables (based on randomized studies only), the model fitted is given by Equations viii and ix:

$$x_{jk} = \mu + \tau_{j1} + \varepsilon_{jk} \qquad \text{[Eqn viii]}$$

$$\tau_{j1} = \tau_1 + u_j \qquad \text{[Eqn ix]}$$

where μ represents the effect of Permixon and τ_{j1} is the random treatment difference effect.

In the randomized placebo-controlled model with dummy variables, the estimated treatment difference is a pooled within-study estimate. This is not the case in this model, where it is a difference of effects averaged over studies.

The model fitted in analysis of all studies with terms for different study types and comparative drugs is given by Equations x and xi:

$$x_{jkl} = \mu + \tau_{jk} + \rho_{j1} + \varepsilon_{jk} \qquad \text{[Eqn x]}$$

$$\tau_{j1} = \tau_1 + u_j \qquad \text{[Eqn xi]}$$

where μ represents the effect of Permixon, τ_{jk} the treatment effect where $k = 1$ corresponds to the comparison of placebo with Permixon and $k > 1$ corresponds to the other drugs, ρ_{j1} represents the effect of the other study types relative to randomized placebo-controlled trials and l indexes the type of study.

The model used for the analysis of all studies that ignore differences among study types is given by Equations xii and xiii:

$$x_{jk} = \mu + \tau_{jk} + \varepsilon_{jk} \qquad \text{[Eqn xii]}$$

$$\tau_{j1} = \tau_1 + u_j \qquad \text{[Eqn xiii]}$$

from baseline to end of study. Such studies can be included within a statistical model [28–31] (Box 1).

Results

Although patients on placebo have a slight increase in mean Q_{max}, the 95% confidence interval (CI) contains zero (i.e. no effect) in all studies (Figure 1). Patients receiving Permixon display a clear increase in peak flow and only in four small studies does the 95% CI contain zero.

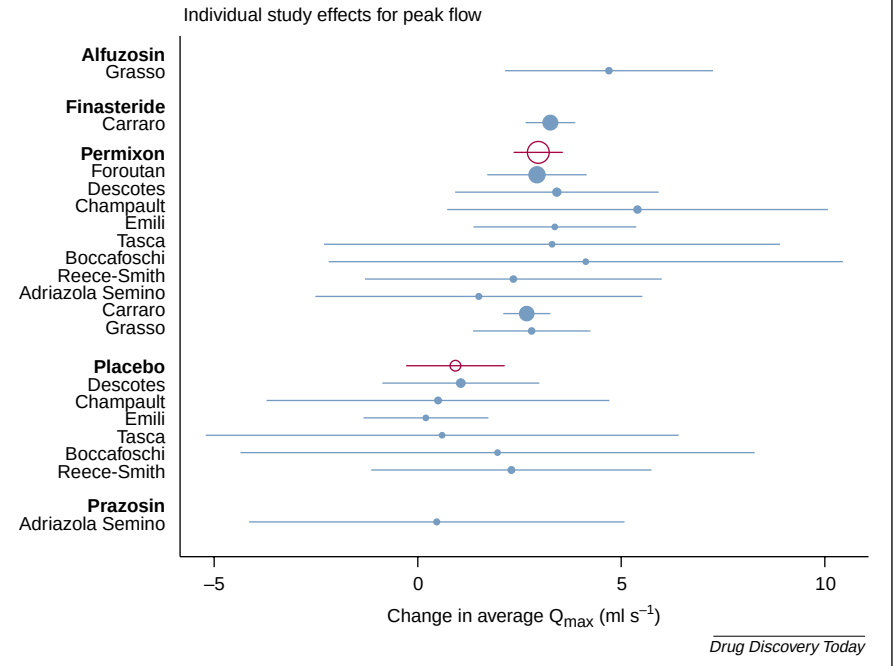

Figure 1. Changes in mean peak urinary flow by study and treatment group. The size of the points is proportional to the square root of the number of subjects. The summaries for the Permixon and placebo arms of the studies are plotted with open circles. The horizontal lines indicate the 95% confidence intervals.

Estimates from the meta-regression models with covariates representing the different study types are presented in Table 3. Four analyses are presented: (i) randomized placebo-controlled studies with dummy variables representing the studies for comparison with the imputation method analysis; (ii) randomized placebo-controlled studies excluding study dummy variables; (iii) all studies with terms for different study types and comparative drugs; and (iv) all studies but ignoring differences among study types. There was no evidence of significant heterogeneity and in all models the variance of the random effect was estimated as zero.

The parameter estimate for placebo gives the estimated change in the mean Q_{max} from baseline to end of study in the placebo arm compared with the Permixon arm; this is the treatment difference effect. It has a negative value because the increase in Q_{max} on placebo is less than the increase on Permixon. The parameter estimate for Permixon gives the change in Q_{max} from baseline to end of study for Permixon. Analysis of the randomized placebo-controlled study with dummy variables gives exactly the same estimated treatment difference of 2.65 (SE of 0.86) ml s^{-1} as the fixed effect analysis. Removing the dummy variables, yields a mean increase on Permixon of 2.70 (SE of 0.86) ml s^{-1} compared with placebo. The difference from the study that incorporated the dummy variables is a result of the later model no longer yielding a pure within-study estimate. The estimated total increase in peak urinary flow using Permixon from randomized placebo-controlled studies is 3.44 ml s^{-1}.

In the model that accounts for different study types and comparative drugs, the Permixon effect corresponds to the randomized studies only and the placebo estimate is the treatment difference, which is also only present in randomized studies; identical values to the model of randomized placebo-controlled studies that exclude dummy variables are obtained. The terms for open and comparative studies give the difference between the effect of Permixon in these studies compared with the randomized trials. The differences in effect between the open and comparative studies and the randomized study are negative, indicating that the change in Q_{max} from baseline is less in these studies compared with the randomized placebo-controlled studies. However, the SEs in these cases are large, which indicates that there is no significant difference. We conclude that Permixon has the same general effect on Q_{max} in all studies and therefore exclude the terms that represent differences among the study types. This gives an increase in Q_{max} associated with Permixon of 2.81 (SE of 0.23) ml s^{-1}. The SEs are now significantly smaller because more studies, including a large open-label study, have been used to estimate the effect of Permixon. The estimated treatment difference is now 2.06 (SE of 0.59) ml s^{-1}. Although this has a smaller SE than in the model of a randomized placebo-controlled trial that includes dummy variables, it is no longer a pure within-study estimate because it compares the average of the placebo arms with the average of the Permixon arms. It is easy to include other study level variables and the estimated effect of the use of

Table 3. Parameter estimates

	Randomized placebo-controlled[a]		Randomized placebo-controlled ignoring study differences		All study type with terms for different study types		All studies ignoring study type differences	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Permixon	3.60	0.95	3.44	0.67	3.44	0.67	2.81	0.23
Placebo	-2.65	0.86	-2.70	0.86	-2.70	0.86	-2.06	0.59
Open	-	-	-	-	-0.51	0.91	-	-
Comparative	-	-	-	-	-0.76	0.72	-	-
Prazosin	-	-	-	-	-2.21	2.37	-2.34	2.36
Finasteride	-	-	-	-	0.58	0.41	0.46	0.39
Alfuzosin	-	-	-	-	2.02	1.33	1.90	1.32

[a]Estimates for individual studies were (SE are given in parentheses and Descotes *et al.* [15] was used as the reference study): Champault *et al.* [16], estimate of 0.56 (1.78); Emili *et al.* [17], estimate of –0.56 (1.00); Tasca *et al.* [18], estimate of –0.33 (2.20); Boccafoschi *et al.* [19], estimate of 0.77 (2.41); and Reece-Smith *et al.* [20], estimate of 0.12 (1.50).
Abbreviation: SE, standard error.

Permixon for one month is to reduce peak urinary flow by 0.086 (SE of 0.110) ml s^{-1}.

Discussion

Treatment with Permixon is associated with an increase in peak urinary flow of approximately 2.0 to 2.5 ml s^{-1}. Over the sensitivity ranges considered here, there was not a substantial change to the overall estimate. Imputation has a greater affect on the precision of the estimate. There is no imputation of the primary response from each study, only imputation of the variability.

The inclusion of non-randomized single-arm studies and the use of data from randomized studies that are not placebo-controlled [32] have an effect on the estimated treatment difference, but, in the example used, this effect is only slight. Using only the randomized placebo-controlled studies gives an estimate of treatment difference that is an average of within-study comparisons. However, the inclusion of other single-arm studies signifies that the treatment difference is no longer completely a within-study comparison. A marked difference between the estimate from all the studies and the estimates from the randomized studies implies considerable heterogeneity and it is unwise to include the non-placebo-controlled studies. If there are no large discrepancies between the estimates, then the estimate based on all studies is preferable because there is a wider range of applicability and the precision will be smaller.

It is important to perform a full sensitivity analysis. In the example used in this review, similar estimates were obtained, which implies that the results are not sensitive to the studies included in the analysis. The precisions are sensitive to the studies included and the inclusion of large single-arm studies will increase the precision of the effect of the treatment from baseline to end of study. The issues here are similar to those involved in synthesizing information, and similar statistical methods are used [33].

If the CONSORT guidelines for reporting clinical trials are followed, then there should be no need for the imputation methods outlined in this review. Although these guidelines, and the corresponding procedures for meta-analysis [34], should improve meta-analyses in the future, imputation could still be essential when using studies already published. In terms of the methodology used and the reporting of procedures, earlier trials might not compare as favourably with later trials. However, this is a feature of the evolving nature of clinical studies.

Imputation of the correlation between the baseline and end of study measurements is another problem. In publications, there is limited information for the estimation of this correlation and access to individual patient data is required. This is most likely to happen when the meta-analysis of all clinical trials of a drug are carried out by a statistician who is working for the company producing the drug. The correlation affects the significance of the results. If the correlation coefficient is equal to zero, then all the estimated effects in this review are too precise. The larger the correlation, the more precise the individual study effects and the greater the heterogeneity in effects observed throughout the studies. In the absence of reliable information, it is better to err on the side of using a low positive

correlation because this will lead to conservative conclusions (wider CIs).

Although there was little numerical effect in the Permixon example used here, multiple imputation is a useful tool for correcting variability estimates. By comparison with the available data on the sample sizes and baseline and end of study means, there is only a small amount of missing information that needs to be imputed to calculate the SEs. The greatest benefit comes from the initial performance of the imputation because this process increases the number of studies that can be included. The example here is extreme because no single randomized placebo-controlled study provided the SE of the treatment difference. Without imputation, no meta-analysis could be attempted. The multiple imputation model is similar to Bayesian models [35].

This leads to the issue of whether or not the pooling of data should be attempted if some SEs are missing. Surely missing variability estimates mean poor studies and so only a qualitative summary should be carried out. We believe that omitting studies without SE information is analogous to missing out unpublished studies – a bias could arise. Imputation is a means of including more studies and provides a method for performing a sensitivity analysis of studies providing complete information. We do not advocate an uncritical use of this technique, but suggest it as a strategy for widening the range of applicability of the analysis.

References

1 Anello, C. (1999) Emerging and recurrent issues in drug development. *Stat. Med.* 18, 2301–2309
2 Yusuf, S. (1997) Meta-analysis of randomized trials: looking back and looking ahead. *Control. Clin. Trials* 18, 594–601
3 Boyle, P. *et al.* (1996) Prostate volume predicts outcome of treatment of benign prostatic hyperplasia with Finasteride: meta-analysis of randomised clinical trials. *Urology* 48, 398–405
4 Gotzsche, P.C. and Olsen, O. (2000) Is screening for breast cancer with mammography justifiable? *Lancet* 355, 129–134
5 Olsen, O. and Gotzsche, P.C. (2001) Cochrane review on screening for breast cancer with mammography. *Lancet* 358, 1340–1342
6 Higgins, J.P. *et al.* (2001) Meta-analysis of continuous outcome data from individual patients. *Stat. Med.* 20, 2219–2241
7 Dixon-Woods, M. *et al.* (2001) Screening for breast cancer with mammography. *Lancet* 358, 2166–2167
8 Lee, J.H. and Zukerman, D. (2001) Screening for breast cancer with mammography. *Lancet* 358, 2164–2165
9 Miller, A.B. (2001) Screening for breast cancer with mammography. *Lancet* 358, 2164–2168
10 Senn, S. (2001) Screening for breast cancer with mammography. *Lancet* 358, 2165–2168
11 Thornton, H. (2001) Screening for breast cancer with mammography. *Lancet* 358, 2165–2168
12 Stewart, L.A. and Clarke, M.J. (1995) Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. *Stat. Med.* 14, 2057–2079
13 Whitehead, A. (2002) *Meta-Analysis of Controlled Clinical Trials*, Wiley
14 Boyle, P. *et al.* (2000) Meta-analysis of clinical trials of Permixon in the treatment of symptomatic benign prostatic hyperplasia. *Urology* 55, 533–539
15 Descotes, J.L. *et al.* (1995) Placebo-controlled evaluation of the efficacy and tolerability of Permixon® in benign prostatic hyperplasia after exclusion of placebo responders. *Clin. Drug Invest.* 9, 291–297
16 Champault, G. *et al.* (1984) A double-blind trial of an extract of the plant *Serenoa repens* in benign prostatic hyperplasia. *Br. J. Clin. Pharmacol.* 18, 461–462
17 Emili, E. *et al.* (1983) Risulti clinici su un nuovo farmaco nella terapia dell'ipertrofia della prostata (Permixon®). *Urologia* 50, 1042–1049
18 Tasca, A. *et al.* (1985) Treatment of obstruction in prostatic adenoma using extract of *Serenoa Repens*. Double-blind clinical test v. placebo. *Minerva Urol. Nefrol.* 37, 87–91
19 Boccafoschi, C. and Annoscia, S. (1983) Comparison of *Serenoa Repens* extract and placebo in controlled clinical trial in patients with prostatic adenomatosis. *Urologia* 50, 1–14
20 Reece, S.H. *et al.* (1986) The value of Permixon in benign prostatic hypertrophy. *Br. J. Urol.* 58, 36–40
21 Carraro, J.C. *et al.* (1996) Comparison of phytotherapy (Permixon) with finasteride in the treatment of benign prostate hyperplasia: a randomized international study of 1,098 patients. *Prostate* 29, 231–240
22 Grasso, M. *et al.* (1995) Comparative effects of alfuzosin versus *Serenoa repens* in the treatment of symptomatic benign prostatic hyperplasia. *Arch. Esp. Urol.* 48, 97–103
23 Adriazola Semino, M. *et al.* (1992) Symptomatic treatment of benign hypertrophy of the prostate. Comparative study of prazosin and *Serenoa repens*. *Arch. Esp. Urol.* 45, 211–213
24 Foroutan, F. (1997) Wirksamkeit und Vertraglichkeit von Permixon® bei einem größeren Patientenkollektiv (592 Patienten) unter Praxisbedingungen. *Journal für Urologie und Urogynäkologie* 2, 17–21
25 Moher, D. *et al.* (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 357, 1191–1194
26 Follmann, D. *et al.* (1992) Variance imputation for overviews of clinical trials with continuous response. *J. Clin. Epidemiol.* 45, 769–773
27 Flather, M.D. *et al.* (1997) Strengths and limitations of meta-analysis: larger studies may be more reliable. *Control. Clin. Trials* 18, 568–579
28 Stram, D.O. (1996) Meta-analysis of published data using a linear mixed effects model. *Biometrics* 52, 536–544
29 van Houwelingen, H.C. *et al.* (2002) Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat. Med.* 21, 589–624
30 Berkey, C.S. *et al.* (1996) Multiple-outcome meta-analysis of clinical trials. *Stat. Med.* 15, 537–557
31 Berkey, C.S. *et al.* (1998) Meta-analysis of multiple outcomes by regression with random effects. *Stat. Med.* 17, 2537–2550
32 McAuley, L. *et al.* (2000) Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 356, 1228–1231
33 Ades, A.E. (2003) A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Stat. Med.* 22, 2995–3016
34 Moher, D. *et al.* (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. *Lancet* 354, 1896–1900
35 Sutton, A.J. and Abrams, K.R. (2001) Bayesian methods in meta-analysis and evidence synthesis. *Stat. Methods Med. Res.* 10, 277–303
36 Bryk, A.S. and Raudenbush, S.W. (1992) *Hierarchical Linear Models*. Sage Publications
37 Woodhouse, G. *et al.* (1995) *MLn Command Reference*. Multilevel Models Project, Institute of Education, London
38 Higgins, J.P. and Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539–1558
39 Thompson, S.G. and Higgins, J.P. (2002) How should meta-regression analyses be undertaken and interpreted? *Stat. Med.* 21, 1559–1573