Network Meta-Analysis: Development of a Three-Level Hierarchical Modeling Approach Incorporating Dose-Related Constraints

Rhiannon K. Owen, BSc, MSc[1],*, Douglas G. Tincello, BSc, MBChB, MD, FRCOG[2], Keith R. Abrams PhD, CStat[1]

[1]Department of Health Sciences, University of Leicester, Leicester, UK; [2]Reproductive Science Section, Department of Cancer Studies and Molecular Medicine, University of Leicester, Leicester, UK

A B S T R A C T

Background: Network meta-analysis (NMA) is commonly used in evidence synthesis; however, in situations in which there are a large number of treatment options, which may be subdivided into classes, and relatively few trials, NMAs produce considerable uncertainty in the estimated treatment effects, and consequently, identification of the most beneficial intervention remains inconclusive. **Objective:** To develop and demonstrate the use of evidence synthesis methods to evaluate extensive treatment networks with a limited number of trials, making use of classes. **Methods:** Using Bayesian Markov chain Monte Carlo methods, we build on the existing work of a random effects NMA to develop a three-level hierarchical NMA model that accounts for the exchangeability between treatments within the same class as well as for the residual between-study heterogeneity. We demonstrate the application of these methods to a continuous and binary outcome, using a motivating example of overactive bladder. We illustrate methods for incorporating ordering constraints in increasing doses, model selection, and assessing inconsistency

between the direct and indirect evidence. **Results:** The methods were applied to a data set obtained from a systematic literature review of trials for overactive bladder, evaluating the mean reduction in incontinence episodes from baseline and the number of patients reporting one or more adverse events. The data set involved 72 trials comparing 34 interventions that were categorized into nine classes of interventions, including placebo. **Conclusions:** Bayesian three-level hierarchical NMAs have the potential to increase the precision in the effect estimates while maintaining the interpretability of the individual interventions for decision making.

Keywords: network meta-analysis, statistical methods, mixed treatment comparisons, overactive bladder.

Introduction

Network meta-analyses (NMA) are widely used in an evidence synthesis setting due to the attractive nature of utilizing all relevant information from both direct and indirect evidence [1–4]. Nevertheless, in situations in which there are a large number of interventions of interest and relatively few trials, there is a potential issue with the sparsity of data in the treatment networks, which can lead to parameter uncertainty. Collapsing the intervention arms into their respective treatment classes increases the evidence base and precision in the effect estimates, but with such a class-based approach, the direct interpretation of individual intervention effects is lost, which makes decision making difficult. To overcome this issue, a three-level

hierarchical NMA can be applied [5–7]. This approach incorporates the exchangeability between interventions of the same class to predict an effect estimate for each of the interventions individually [8]. Thus, this approach allows strength to be borrowed within the classes of interventions, strengthening inferences and potentially reducing the uncertainty around the individual intervention effects, and consequently increasing the ability to rank these and inform decision-making frameworks. To further increase the precision in the effect estimates, constraints can be applied on increasing doses of the same intervention, making the assumption that higher doses have an effect greater or equal to that of lower doses [9,10].

To illustrate the use of the hierarchical framework, we applied the proposed methods to a real clinical question in overactive

bladder (OAB) syndrome. To manage the OAB syndrome, the National Institute for Health and Care Excellence in the United Kingdom [11] currently recommends a course of supervised pelvic floor muscle training, behavioral therapy, anticholinergic medication, sacral nerve stimulation, and more recently, botulinum toxin type A (BoNTA). Given the availability of numerous interventions and emerging alternative treatments such as BoNTA, there is an increasing need to identify the most beneficial intervention. However, given the large number of interventions and the limited evidence base, in terms of both the number of trials and the number of direct comparisons between active interventions, the estimated intervention effects from a standard NMA will have a considerable level of uncertainty associated with them. In situations in which there are a limited number of trials in a meta-analysis, estimating the heterogeneity between trials may also be problematic. One approach to overcome this issue, and increase precision in the treatment effects, involves incorporating external information from similar studies relevant to the treatment of interest [12]. In an NMA that includes all available trials in a specific field, however, such external information may be limited. The aim of this article was to develop and apply hierarchical NMAs to evaluate the clinical effectiveness of interventions for the OAB syndrome by borrowing strength between interventions of the same class and applying ordering constraints on increasing doses of BoNTA, thus increasing the precision that we have in our effect estimates but maintaining the interpretability of results at the individual intervention level. For illustration purposes, we focus on two outcomes associated with intervention effectiveness (mean change in incontinence episodes from baseline) and treatment tolerability (number of patients reporting one or more adverse events).

In this article, we demonstrate the individual treatment, class-based, and three-level hierarchical random effects model

approaches, and where applicable we demonstrate the use of extending hierarchical NMAs to incorporate ordering constraints. We apply these models to a motivating clinical example in the OAB syndrome. Furthermore, we demonstrate a comprehensive technique to assess inconsistency between the direct and indirect estimates of an extensive network using the method of node-splitting [13] and assess model fit using residual deviance [14] and the deviance information criterion (DIC) [15].

Methods

Illustrative Data Set

Almost all published articles reporting data on interventions for the OAB syndrome compare the intervention against placebo, which makes comparison across active interventions difficult without using indirect comparisons or NMA. This is particularly evident for trials evaluating anticholinergic drugs. Only three meta-analyses have been undertaken in the field of the OAB syndrome [16–18]. The interventions were assessed on a head-to-head basis, where studies comparing the interventions directly were pooled in a pairwise meta-analysis. Chapple et al. [16] focused on the evaluation of the clinical effectiveness of anticholinergic drugs compared with placebo, while Novara et al. [17] compared the efficacy of increased doses of each anticholinergic drug with that of their respective lower dose. Anger et al. [18] evaluated the effect of BoNTA against that of a placebo intervention. In the current literature, there is no coherent comparison between all the available interventions, and consequently, there is little information of a superior treatment for the OAB syndrome.
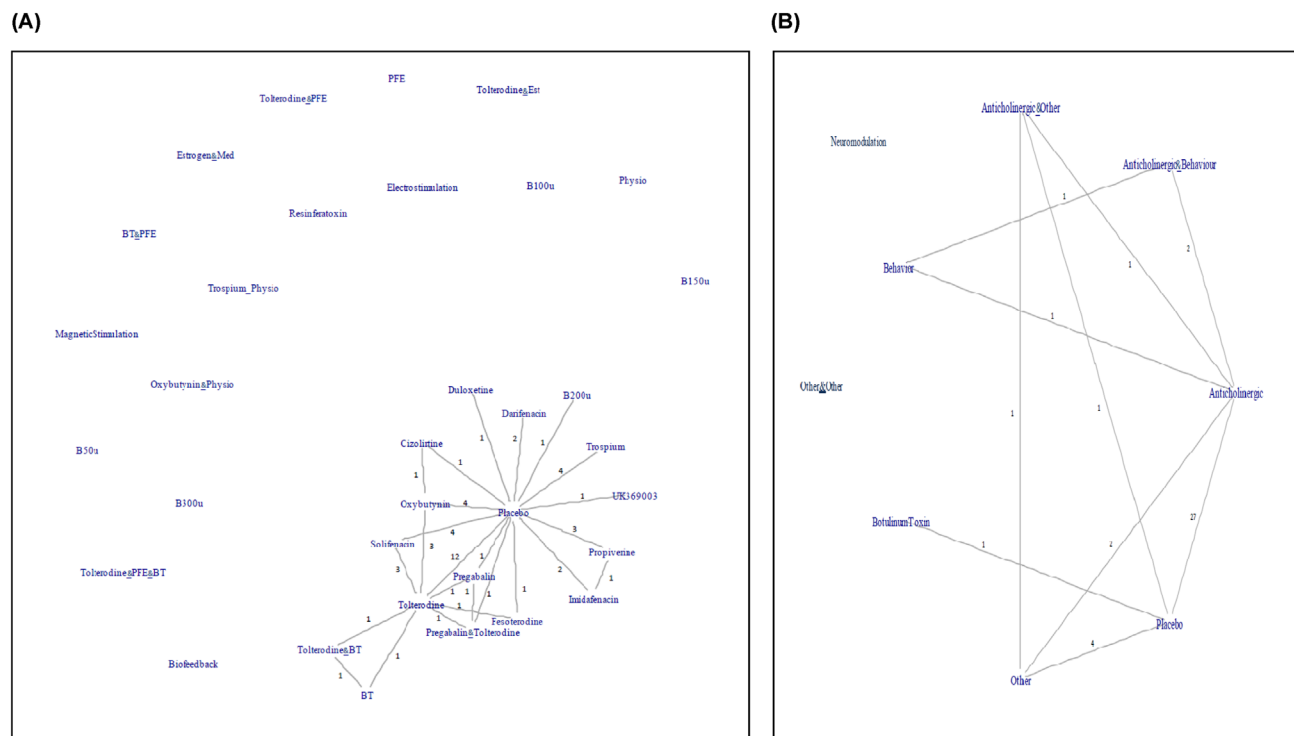
Figure 1A,B illustrates the network diagrams of direct comparisons for the individual intervention and classes of

(A)   **(B)**

Fig. 1 – Network diagrams for urinary incontinence. (A) Individual and hierarchical network diagram. (B) Classified network diagram. B50u, Botulinum toxin type A 50 units; B100u, Botulinum toxin type A 100 units; B150u, Botulinum toxin type A 150 units; B200u, Botulinum toxin type A 200 units; B300u, Botulinum toxin type A 300 units; BT, Bladder training; Est, Estrogen; Med, Medroxyprogesterone; PFE, Pelvic floor exercises; Physio, Physiotherapy.

(A)

(B)

Fig. 2 – Network diagrams for adverse events. (A) Individual and hierarchical network diagram. (B) Class-based network diagram. B50u, Botulinum toxin type A 50 units; B100u, Botulinum toxin type A 100 units; B150u, Botulinum toxin type A 150 units; B200u, Botulinum toxin type A 200 units; B300u, Botulinum toxin type A 300 units; BT, Bladder training; Est, Estrogen; Med, Medroxyprogesterone; PFE, Pelvic floor exercises; Physio, Physiotherapy.

interventions that evaluate the mean reduction in incontinence episodes, respectively. Similarly, Figure 2A,B demonstrates the network diagrams of the individual intervention and classes of interventions that evaluate the number of patients reporting one or more adverse events, respectively. The nodes represent either the individual intervention or classes of interventions. The interconnecting lines demonstrate a direct comparison, and the corresponding values represent the number of trials that directly compare those interventions. For interventions identified in the systematic literature review but disconnected from the network (i.e., fail to report outcome of interest), we were unable to obtain effect estimates and thus these were excluded from the analysis.

For analyses associated with the class of interventions, treatments were grouped, according to clinical opinion (D.G.T.), into anticholinergic drug therapy, botulinum toxin, neuromodulation, behavior therapy, other drugs, anticholinergic drugs in combination with behavior therapy, anticholinergic drugs in combination with other drugs, and a combination of other drugs. Figure 3 demonstrates the classification of each of the individual interventions, where the central node represents the class of treatments and the linked arms represent each of the individual interventions within that class. In this example, anticholinergic drug therapy consisted of all the members of the anticholinergic class of drugs. The botulinum toxin group contained all BoNTA interventions regardless of the site of administration or dose. The neuromodulation classification included all interventions involved in nerve stimulation or electrostimulation. *Behavior therapy* was defined as interventions that focused on attaining change in behavioral factors relevant to symptoms of the OAB syndrome, including physiotherapy, bladder retraining, and biofeedback. *Other drugs* were defined as all other pharmacotherapy interventions that were not classified as anticholinergic or BoNTA therapies.

Individual Treatment and Class-Based NMA Models

Equations 1 and 2 illustrate the general model described by Welton et al. [19] for the continuous and binary outcome case, respectively. It is these models that form the foundation for both the individual treatment and class-based NMAs.

For an intervention j, in study i, the continuous outcome can be interpreted as the mean change in 3-day diary data for the number of incontinence episodes from baseline y_{ij} and assumed to follow a normal distribution with mean equal to the underlying intervention effect $\theta_{i,j}$ and observed standard error $SE_{i,j}$. Let μ_i represent the baseline mean change in the number of incontinence episodes corresponding to the b_i intervention arm in the ith study, and let $\delta_{i,j}$ represent the mean difference in change in the number of incontinent episodes of intervention j relative to the b_i intervention arm. $\delta_{i,j}$ is obtained from a normal distribution with the mean equal to the mean differences $(d_j - d_{b_i})$ and between-study variance τ^2, where d_j and d_{b_i} represent the effect estimate of intervention j and study-specific baseline intervention b_i, respectively. Notably, when the between-study variance is zero, that is, $\tau^2 = 0$, we obtain a fixed effects model. Thus, the overall model is based on a linear regression model on a natural additive scale:

$$y_{i,j} \sim Normal(\theta_{i,j}, SE_{i,j}{}^2)$$

where

$$\theta_{i,j} = \begin{cases} \mu_i & \text{Intervention } b_i \\ \mu_i + \delta_{i,j} & \text{Intervention } j \end{cases}$$

and

$$\delta(i,j) \sim Normal((d_j - d_{b_i}), \tau^2) \qquad (1)$$

The number of reported adverse events is considered a binomial count $r_{i,j}$ from a sample number at risk $n_{i,j}$ for an intervention j
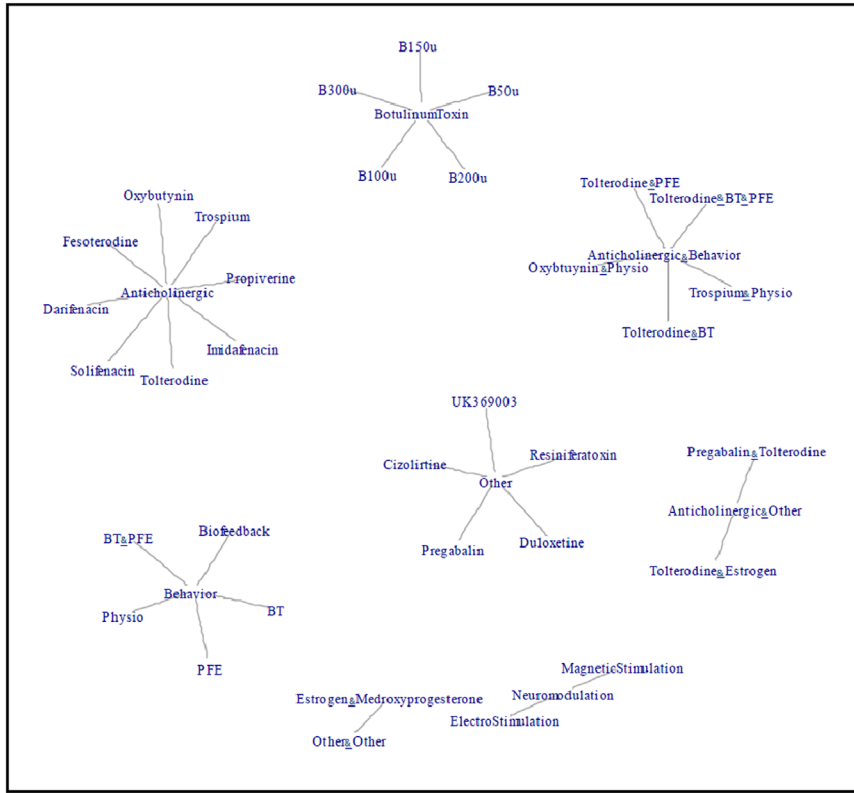
Fig. 3 – Treatment classification. B50u, Botulinum toxin type A 50 units; B100u, Botulinum toxin type A 100 units; B150u, Botulinum toxin type A 150 units; B200u, Botulinum toxin type A 200 units; B300u, Botulinum toxin type A 300 units; BT, Bladder training; PFE, Pelvic floor exercises; Physio, Physiotherapy.

within the ith study. This information allows estimation of the probability $p_{i,j}$, which is associated with the risk of adverse events. We assume a logistic regression model for the binary outcome. Thus, $\varphi_{i,j}$ represents the log-odds of treatment j relative to a baseline treatment b_i with between-study variance τ^2. In this case, d_j and d_{b_i} represent the estimated log-odds of intervention j and baseline intervention b_i, respectively. Therefore, the overall model is given by

$$r_{i,j} \sim Binomial(p_{i,j}, n_{i,j})$$

where

$$logit(p_{i,j}) = \begin{cases} \mu_i & \text{Intervention } b_i \\ \mu_i + \delta_{i,j} & \text{Intervention } j \end{cases}$$

and

$$\varphi(i,j) \sim Normal((d_j - d_{b_i}), \tau^2) \tag{2}$$

The baseline intervention means μ_i are assumed to have a normal (0, 1000) prior distribution. Therefore, for the continuous case, the mean reduction in incontinence episodes from baseline for the reference intervention could plausibly be in the range of 0 ± 62 incontinence episodes. For the binary case, the log-odds of an adverse event could plausibly be in the range of 0 ± 62. The between-study standard deviation values of τ are assumed to have a uniform (0, 5) prior distribution, suggesting that the between-study SD can take any value between, but not including, 0 and 5, and small values of τ are equally likely as large values [1]. A value of 5, for example, would indicate that for a random pair of studies, the difference in the mean reduction in incontinence episodes from baseline could be as large as 5.5 while the ratio of odds ratios could be as large as 232.8. Sensitivity analyses considering two other variance-component priors were considered: 1) gamma (0.001, 0.001)

on the precision scale, that is, 1/variance, and 2) half-normal (0, 1) on the SD scale [1].

Hierarchical NMA

A random effects model was used to estimate the effect of each individual intervention for both continuous (Equation 3) and binary (Equation 4) outcomes. To account for the exchangeability between the treatments within each class, the treatments within each class were assumed to follow a normal distribution with a class-specific mean and variance (Equation 5).

$$y_{i,j} \sim Normal(\theta_{i,j}, SE_{i,j}{}^2)$$

$$\theta_{i,j} = \begin{cases} \mu_i & \text{Intervention } b_i \\ \mu_i + \delta^*_{i,j,k} & \text{Intervention } j \end{cases} \tag{3}$$

$$r_{i,j} \sim Binomial(p_{i,j}, n_{i,j})$$

$$logit(p_{i,j}) = \begin{cases} \mu_i & \text{Intervention } b_i \\ \mu_i + \varphi^*_{i,j,k} & \text{Intervention } j \end{cases} \tag{4}$$

Following Dakin et al. [5] and Warren et al. [6], the effect estimate for a specific intervention class combination $d_{j,k}$ is described as

$$d_{j,k} \sim Normal(\mu_k, \sigma^2) \tag{5}$$

where μ_k denotes the pooled effect estimate for the kth class of interventions, with a common between-intervention variance σ^2. Class-specific between-intervention variances, σ_k^2, were also considered in exploratory analyses and assessed through model fit statistics.

At the individual intervention level, the effect estimate compared with that of a baseline treatment $\delta_{i,j,k}^{*}$ and $\varphi_{i,j,k}^{*}$ for the continuous case and the binary case, respectively, is expressed in terms of the effect estimate for a specific individual intervention class $d_{j,k}$, compared with a class-specific baseline treatment $d_{b_{i,k}}$, given by

$$\delta_{i,j,k}^{*}\sim\text{Normal}((d_{j,k}-d_{b_{i,k}}),\tau^2)\quad\text{for the continuous case}$$

and

$$\varphi_{i,j,k}^{*}\sim\text{Normal}((d_{j,k}-d_{b_{i,k}}),\tau^2)\quad\text{for the binary case}\qquad(6)$$

where $\delta_{i,j,k}^{*}$ represents the estimated mean difference of treatment j compared with a baseline treatment b_i and $\varphi_{i,j,k}^{*}$ represents the estimated log-odds of treatment j relative to a baseline treatment b_i. At level 3 in the hierarchical model, the between-study variance τ^2 was given a uniform (0, 5) prior distribution as in the random effect NMA model (see Equations 1 and 2). The pooled class-effect estimates μ_k were assumed to have a vague normal (0, 1000) prior distribution, with a common between-intervention variance σ^2 assumed to follow a uniform (0, 5) distribution on the SD scale. Thus, the pooled class-effect estimates could plausibly take values in the range of 0 ± 62 change in incontinent episodes from baseline for the continuous case and 0 ± 62 log-odds of an adverse event for the binary case. The common between-intervention variance could plausibly take values between, but not including, 0 and 5. Two other prior distributions for the variance for σ^2 were considered in the form of a sensitivity analysis: 1) gamma (0.001, 0.001) on the precision scale, and 2) half-normal (0, 1) on the SD scale [1].

For the continuous outcome, treatments were ranked on the basis of posterior distributions of the relative effect estimate δ and δ^*, where treatments with the largest relative reduction in mean incontinence episodes were ranked first for each Markov chain Monte Carlo (MCMC) iteration from the individual treatment and the hierarchical model, respectively. The estimated rankings overall were then calculated from a summary of these individual ranks at each iteration. Therefore, a higher rank indicates a more efficacious intervention overall. Similarly, for the binary outcome, the treatments were individually ranked on their posterior summaries φ and φ^*, where treatments with the highest log-odds of an adverse event were ranked in the first place for the individual treatment and the hierarchical model, respectively, where a higher rank indicates a larger prevalence. Thus, interventions ranked first are regarded as the "worst" treatments associated with adverse events. The corresponding probabilities were calculated by monitoring the number of MCMC iterations for which each of the treatments was ranked in the first place.

Incorporating Constraints on Increasing Doses

Ordering constraints were placed on multiple doses of BoNTA interventions, with the assumption that larger doses would have a greater or equal treatment effect compared with its respective lower dose (e.g., $d_1 \le d_2 \le \cdots \le d_m$). We applied these constraints by assigning an indicator function γ, equal to 1, given by

$$\gamma=\prod_{l=1}^{m-1}I(d_{l+1}-d_l)\qquad(7)$$

where $I(x) = 1$, if $x \ge 0$, and $I(x) = 0$, otherwise. This forces $(d_{l+1}-d_l) \ge 0$ and consequently imposes ordering constraints on the treatment effects of increasing doses (i.e., $d_l \le d_{l+1}$) [7]. Ordering constraints can be placed in either direction depending on the outcome of interest [6].

Assessment of Inconsistency

Consistencies between direct and indirect comparisons were evaluated using the method of "node-splitting" [13]. This approach allows the calculation of two posterior distributions, one of which is derived from trials that directly compare the interventions (e.g., interventions X and Y), $d_{\text{XY}}^{\text{Dir}}$, whereas the other is indirectly derived from the remaining trials $d_{\text{XY}}^{\text{Ind}}$. The fundamental model described in Equations (1) and (2) remains the same; however, for direct comparison, the effect estimates $\delta_{i\text{XY}}$ obtained from splitting the (X, Y) node are selected from a normal distribution with mean $d_{\text{XY}}^{\text{Dir}}$ and variance sd^2, that is,

$$\delta i\text{XY}\sim N(d_{\text{XY}}^{\text{Dir}},sd^2)\qquad(8)$$

Simultaneously, indirect comparisons are obtained using the consistency assumption, which states that for treatment effects d_{XY}, d_{XZ}, and d_{YZ}, relative to treatments X, Y, and Z,

$$\begin{aligned}d_{\text{XY}}&=d_{\text{YZ}}-d_{\text{XZ}}\\d_{\text{YZ}}&=d_{\text{XZ}}-d_{\text{XY}}\\d_{\text{XZ}}&=d_{\text{YZ}}-d_{\text{XY}}\end{aligned}\qquad(9)$$

To test for consistency between direct and indirect estimates, we simply calculated the difference for each pair of interventions, together with the probability that the direct estimate surpasses that of the indirect estimate. Thus, a Bayesian P value can be calculated using the derived test statistic and comparing it with a standard normal distribution. This method, however, can only be applied to interventions within a closed loop, meaning that there is both direct and indirect evidence available for all pairs of interventions under consideration [13].

Model Fit and Selection

The DIC [15] was used to compare models. It is a measure of the deviance, estimated by the posterior mean of minus twice the log-likelihood plus the effective number of parameters in the model. Thus, it is considered as a Bayesian measure of goodness of fit that can be used as a relative measure of model suitability and easily applied to hierarchical modeling [20]. In parallel, the total residual deviances for each of the models were also compared with the respective number of data points. To illustrate and assess the goodness of fit for each of the models, we plotted the residual deviances for each of the included studies against the respective number of data points for that study [14], that is, 2 for two arm studies, 3 for three arm studies, and so forth.

Model Estimation

All models were estimated using WinBUGS 1.4.3 [21]. The results are based on 60,000 samples, where the first 10,000 samples were discarded from the analyses as a "burn-in." Three individual chains with disparate starting values were analyzed and convergence was assessed using Brooks-Gelman-Rubin plots [22]. Sensitivity analyses were also undertaken to assess the impact of the choice of prior distributions especially for the variance parameters [1,23]. Full codes for continuous and binary models are given in Appendices A and B in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.10.006, respectively. Both node-splitting analyses and network diagrams were implemented in R [24] using the "R2WinBUGs" software package [25] and the GeMTC package [26], respectively.

Results

Model Fit and Selection

Table 1 contains the goodness-of-fit statistics for each of the models individually. Notably, analyses for class-based models were calculated on different data sets—a consequence of the treatment clustering into endonodal treatment classes [27], which resulted in the omission of several studies that compared

Table 1 – Goodness-of-fit statistics for fixed and random effects models.

Outcome	Model		Between-study SD (95% CrI)	Residual deviance (no. of data points)	DIC
Number of incontinence episodes	Individual	FE	–	158.1 (112)	51.68
		RE	0.20 (0.12–0.31)	116.9 (112)	29.05
	Class*	FE	–	118.3 (87)	1.89
		RE	0.15 (0.06–0.25)	90.66 (87)	–11.53
	Hierarchical	FE	–	158.6 (112)	41.33
		RE	0.20 (0.12–0.30)	114.9 (112)	17.02
	Hierarchical with constraints	FE	–	158.8 (112)	41.37
		RE	0.20 (0.12–0.30)	114.5 (112)	15.81
Number of patients reporting one or more adverse events	Individual	FE	–	134.5 (79)	591.97
		RE	0.32 (0.21–0.49)	74.42 (79)	547.74
	Class*	FE	–	155.6 (68)	554.81
		RE	0.36 (0.25–0.52)	65.74 (68)	484.85
	Hierarchical	FE	–	135 (79)	592.66
		RE	0.30 (0.20–0.46)	75.43 (79)	548.68

CrI, credible interval; DIC, deviance information criterion; FE, fixed effects; RE, random effects.
* Class-based analyses conducted on a different population and are not directly comparable.

two or more interventions from the same class. Thus, model fit statistics for class-based NMAs are solely presented for completeness and cannot be directly compared with the remaining models. In relation to individual analyses, hierarchical models appeared to have a slightly better fit to the data for both outcome measures as illustrated through the reduced residual deviance for continuous (Fig. 4A) and binary (Fig. 4B) outcomes. For the continuous outcome, incorporating ordering constraints slightly improved model fit further with respect to both the DIC and the total residual deviance. The hierarchical random effects model with ordering constraints had a lower DIC (15.81) than did the individual random effects model (29.05). Similarly, for the hierarchical random effects model with constraints, the total residual deviance of 114.5 was closer to the number of data points (112)

compared with the individual random effects model, which had a total residual deviance of 116.9. The random effects models were of a better fit to the data in comparison to the fixed effects models for all sets of analyses and thus, for illustration purposes, the results presented are based on estimates derived from the random effects NMAs.

Number of Incontinence Episodes

Table 2 illustrates the interventions ranked in order of their estimated efficacy for reducing incontinence episodes. Effect estimates derived from hierarchical models correspond with estimates derived from the individual analysis; however, there was a substantial increase in the precision surrounding effect
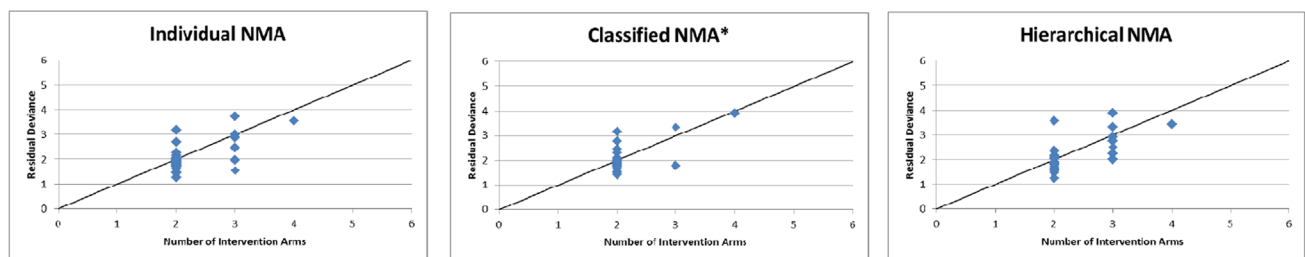
(A)
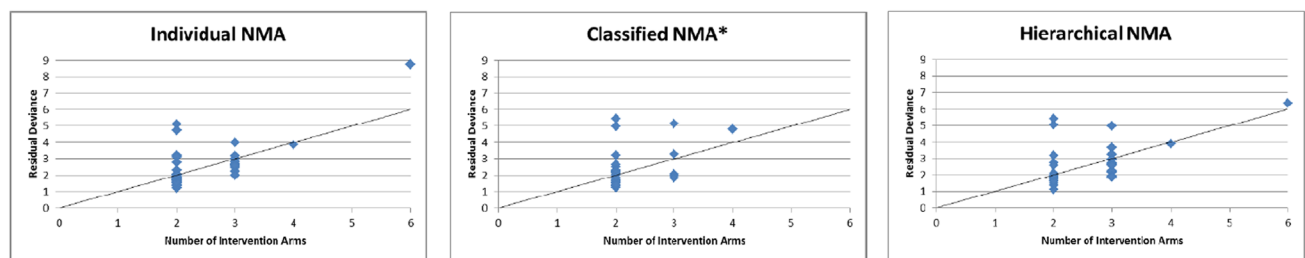
(B)

Fig. 4 – Residual deviance plots. (A) Incontinence episodes. (B) Number of patients reporting adverse events. NMA, network meta-analysis. *Classified analyses conducted on a different population and are not directly comparable.

Table 2 – Treatments placed in ranked order of their relative difference compared with placebo (δ and δ*) and corresponding probabilities for the mean reduction in incontinence episodes from baseline.

S. no.	Individual NMA				Hierarchical NMA				Hierarchical NMA with constraints			
	Treatment	Rank (95% CrI)	p (best) (%)	δ (95% CrI)	Treatment	Rank (95% CrI)	p (best) (%)	δ* (95% CrI)	Treatment	Rank (95% CrI)	p (best) (%)	δ* (95% CrI)
1	Trospium and physiotherapy	2 (1–10)	47.66	−1.89 (−2.97 to 0.82)	BoNTA 200 U	3 (1–13)	20.58	−1.33 (−2.13 to −0.61)	BoNTA 300 U	1 (1–7)	83.18	−1.44 (−2.09 to −0.78)
2	BoNTA 150 U	4 (1–22)	17.67	−1.45 (−2.73 to −0.17)	BoNTA 150 U	3 (1–16)	19.32	−1.31 (−2.15 to −0.55)	BoNTA 200 U	2 (2–10)	0	−1.38 (−2 to −0.73)
3	BoNTA 200 U	4 (1–12)	9.46	−1.46 (−2.28 to −0.65)	BoNTA 50 U	4 (1–16)	14.97	−1.29 (−2.13 to −0.54)	BoNTA 150 U	3 (3–12)	0	−1.32 (−1.94 to −0.68)
4	BoNTA 50 U	6 (1–23)	4.35	−1.17 (−2.26 to −0.09)	BoNTA 300 U	4 (1–17)	12.66	−1.27 (−2.12 to −0.51)	BoNTA 100 U	4 (4–14)	0	−1.26 (−1.89 to −0.61)
5	Physiotherapy	6 (2–19)	1.08	−1.15 (−1.93 to −0.38)	BoNTA 100 U	4 (1–18)	11.75	−1.27 (−2.11 to −0.49)	BoNTA 50 U	5 (5–18)	0	−1.19 (−1.84 to −0.51)
6	BoNTA 300 U	8 (2–25)	2.06	−0.97 (−2.10 to 0.17)	BT and PFE	9 (1–25)	5.29	−0.80 (−1.59 to −0.05)	BT and PFE	9 (1–24)	4.54	−0.82 (−1.53 to −0.14)
7	BT and PFE	9 (1–26)	11.16	−0.86 (−2.94 to 1.17)	Trospium and physio-therapy	9 (1–22)	3.648	−0.84 (−1.36 to −0.38)	Trospium and physio-therapy	9 (1–20)	2.88	−0.85 (−1.31 to −0.43)
8	Oxybutynin and physiotherapy	10 (2–25)	1.22	−0.81 (−1.77 to 0.11)	PFE	10 (1–25)	3.67	−0.77 (−1.55 to −0.03)	PFE	10 (1–24)	2.85	−0.79 (−1.47 to −0.13)
9	BoNTA 100 U	10 (2–26)	1.17	−0.85 (−1.99 to 0.31)	Oxybutynin and physio-therapy	10 (2–23)	1.748	−0.78 (−1.26 to −0.33)	Tolterodine, PFE, and BT	10 (2–21)	1.22	−0.79 (−1.19 to −0.39)
10	PFE	11 (1–26)	3.07	−0.72 (−2.61 to 1.15)	Physio-therapy	11 (2–24)	1.724	−0.75 (−1.46 to −0.08)	Oxybutynin and physio-therapy	10 (3–22)	1.14	−0.78 (−1.2 to −0.37)
11	Tolterodine, PFE, and BT	11 (3–24)	0.57	−0.78 (−1.55 to −0.02)	Duloxetine	11 (2–23)	1.216	−0.76 (−1.21 to −0.33)	Physio-therapy	11 (2–24)	1.43	−0.77 (−1.37 to −0.16)
12	Tolterodine and estrogen	11 (4–21)	0.1	−0.78 (−1.27 to −0.29)	Tolterodine and PFE	11 (2–23)	1.11	−0.75 (−1.21 to −0.30)	Duloxetine	11 (3–22)	0.68	−0.76 (−1.15 to −0.37)
13	Solifenacin	11 (6–16)	0	−0.77 (−0.96 to −0.57)	Bladder training	13 (2–26)	0.994	−0.67 (−1.46 to 0.12)	Tolterodine and PFE	11 (3–22)	0.63	−0.75 (−1.14 to −0.34)
14	Duloxetine	14 (5–24)	0.04	−0.60 (−1.22 to 0.02)	Tolterodine and estrogen	13 (5–23)	0.426	−0.66 (−1.02 to −0.28)	Bladder training	13 (3–26)	0.75	−0.68 (−1.35 to 0.04)
15	Trospium	14 (7–22)	0	−0.61 (−0.97 to −0.27)	Solifenacin	13 (6–21)	0.068	−0.66 (−0.86 to −0.46)	Solifenacin	13 (6–21)	0.06	−0.66 (−0.86 to −0.48)
16	Cizolirtine	17 (5–26)	0.04	−0.49 (−1.15 to 0.19)	Tolterodine, PFE, and BT	14 (4–24)	0.536	−0.64 (−1.05 to −0.23)	Tolterodine and estrogen	15 (5–24)	0.36	−0.63 (−1.02 to −0.24)
17	Tolterodine and PFE	17 (5–26)	0.03	−0.49 (−1.17 to 0.20)	Pregabalin and tolterodine	16 (6–25)	0.144	−0.57 (−0.93 to −0.17)	Pregabalin and tolterodine	17 (6–25)	0.15	−0.54 (−0.94 to −0.12)
18	Tolterodine	17 (12–21)	0	−0.48 (−0.62 to −0.35)	Trospium	17 (9–23)	0.004	−0.54 (−0.75 to −0.33)	Trospium	18 (10–24)	0	−0.54 (−0.75 to −0.33)
19	Oxybutynin	17 (11–22)	0	−0.48 (−0.69 to −0.28)	Propiverine	18 (9–24)	0.004	−0.52 (−0.72 to −0.30)	Propiverine	18 (10–24)	0	−0.52 (−0.72 to −0.3)
20	Propiverine	17 (9–23)	0	−0.49 (−0.82 to −0.16)	Oxybutynin	18 (11–24)	0	−0.51 (−0.66 to −0.33)	Tolterodine	19 (13–23)	0	−0.5 (−0.63 to −0.38)
21	Fesoterodine	18 (11–23)	0	−0.46 (−0.70 to −0.23)	Fesoterodine	18 (10–24)	0	−0.50 (−0.67 to −0.32)	Oxybutynin	19 (11–24)	0	−0.51 (−0.67 to −0.34)
22	Bladder training	21 (3–28)	0.14	−0.30 (−2.31 to 1.74)	Tolterodine	19 (12–23)	0	−0.50 (−0.63 to −0.37)	Fesoterodine	19 (11–24)	0	−0.5 (−0.67 to −0.32)
23	Pregabalin and tolterodine	21 (9–26)	0	−0.28 (−0.80 to 0.24)	Resinifer-atoxin	24 (12–27)	0.032	−0.19 (−0.70 to 0.39)	Cizolirtine	24 (12–27)	0.01	−0.24 (−0.71 to 0.22)
24	Estrogen and medroxy-progesterone	23 (9–27)	0	−0.10 (−0.83 to 0.63)	Cizolirtine	24 (11–27)	0.014	−0.24 (−0.71 to 0.25)	Resinifer-atoxin	24 (13–27)	0.01	−0.19 (−0.7 to 0.36)

	Placebo	Pregabalin	Estrogen and medroxy-progesterone	Placebo	Electro-stimulation
25 Placebo	24 (21–27)	NA	0.002	24 (14–27)	–0.19 (–0.65 to 0.29)
26 Pregabalin	24 (13–27)	0.01 (–0.63 to 0.66)	0.084	26 (9–28)	–0.1 (–0.81 to 0.62)
27 Resiniferatoxin	27 (7–28)	0.60 (–1.09 to 2.28)	0 / 0.006	26 (23–27) / 26 (23–28)	NA
28 Electros-stimulation	28 (23–28)	1.27 (–0.19 to 2.76)		28 (25–28) / 28 (23–28)	1.21 (–0.3 to 2.66)
	0 / 0	0.19 / 0	–0.18 (–0.64 to 0.31) / –0.10 (–0.79 to 0.58) / NA / 1.19 (–0.13 to 2.69)	0 / 0.12 / 0 / 0	

BoNTA, botulinum toxin type A; BT, bladder training; NA, not available/not applicable; NMA, network meta-analysis; PFE, pelvic floor exercises.

estimates produced by hierarchical analyses. For example, in comparison to placebo, BoNTA 150 U had a similar reduction of –1.45 (95% credible interval [CrI] –2.73 to –0.17) and –1.31 (95% CrI –2.15 to –0.55) leakage episodes per 24 hours for individual and hierarchical NMAs, respectively, though the precision increased by approximately 150% for the effect estimate derived from the hierarchical analysis. Imposing ordering constraints on increasing doses of BoNTA further increased the precision of the effect estimates, given that the estimated mean reduction in incontinence episodes for BoNTA 150 U became –1.32 (95% CrI –1.94 to –0.68), with the corresponding precision increasing by approximately 310% compared with the individual analysis. The reduction in posterior uncertainty is particularly apparent through the narrower CrIs. The synthesis of all available data in a hierarchical analysis suggested that BoNTA 200 U is the most efficacious intervention for reducing incontinent episodes with a corresponding probability of 20.58% of it being the "best" intervention overall. Imposing additional ordering constraints identified BoNTA 300 U as the most efficacious intervention with an increased probability of 83.18% of being the best intervention compared with an estimated 12.66% from the unconstrained model.

In addition, the posterior summaries for the class effects obtained from both class-based and hierarchical analyses also agree with one another (Table 3). These suggest that botulinum toxins, as a class of interventions, are the most efficacious at reducing the number of incontinent episodes per 24 hours, with an estimated mean reduction of –1.40 (95% CrI –2.14 to –0.66), –1.29 (95% CrI –2.12 to –0.56), and –1.32 (95% CrI –1.95 to –0.66) for class-based, hierarchical, and hierarchical with constraints models, respectively. Although categorizing the interventions into their respective classes for the class-based NMA increased the precision of the effect estimates compared with the individual NMA, it restricts the interpretability of the result at an individual intervention level.

Number of Patients Reporting Adverse Events

Table 4 presents the interventions ranked in order of the estimated odds of a patient reporting an adverse event. The estimated odds are comparable in both individual and hierarchical models; however, precisions in estimates derived from hierarchical analyses have substantially increased. In both sets of analyses, pregabalin is identified as the "worst" intervention for causing patients to report adverse events. The estimated odds ratio relative to placebo is 3.25 (95% CrI 1.52–7.03) for the individual NMA and 2.77 (95% CrI 1.55– 5.05) for the hierarchical NMA. Although the effect estimates derived from both models are broadly similar, there is a 67% increase in the posterior precision of the estimate obtained from the hierarchical model. This increase in precision is demonstrated through the consistently narrower CrIs. There is still considerable uncertainty in the estimated odds and consequently in the estimated ranks of the interventions, which is further highlighted by the associated 95% CrIs. Although pregabalin is ranked "worst" overall, it is ranked worst only 30% of the time in both individual and hierarchical NMAs, and thus this result should be interpreted cautiously [13].

The median of posterior distributions for class effects are comparable in both class-based and hierarchical NMAs (Table 5). In addition, both sets of analyses suggest that other drugs have the highest prevalence of patients reporting one or more adverse events, with 70.73% and 60.52% probability of these being the highest ranking intervention for causing one or more adverse events for class-based and hierarchical NMA, respectively.

Sensitivity Analysis

Sensitivity analyses suggested that changing the prior distributions of the variance parameters had very little impact on the estimated treatment effects for both the outcome measures (see Appendices C and D in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.10.006). Sensitivity to both the prior distribution for the between-study variance and the between-intervention class-specific variances showed little evidence of an impact on the overall treatment effect estimates and precision for individual and hierarchical models, respectively, thus suggesting that all sets of analyses are insensitive to the choice of the prior distributions.

Node-Splitting

There was little evidence of an inconsistency between direct and indirect estimates obtained from hierarchical NMAs as assessed by methods of node-splitting (see Appendices E and F in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.10.006). For the individual and hierarchical analysis, tolterodine, oxybutynin, BoNTA 100 U, solifenacin, and trospium demonstrate an inconsistent direct and indirect estimate for the mean reduction in incontinent episodes from baseline, relative to placebo (see Appendix E in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.10.006), although between the active interventions, there was little evidence of an inconsistency. Further investigation of the individual classes of treatments (e.g., anticholinergic drugs alone) showed little evidence of an inconsistency between direct and indirect estimates when compared with placebo. This would suggest that the potential pooling of the placebo interventions *between* classes might not be an appropriate assumption because placebo for one class of interventions could be very different from that for another class of interventions. For the adverse event outcome, however, there was very little evidence of any inconsistency between direct and indirect estimates for any set of pairwise interventions (see Appendix F in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.10.006).

Discussion

In this article, we have described and demonstrated the use of hierarchical modeling for mixed treatment NMAs–a useful methodology that can be used in clinical areas in which the available interventions are particularly extensive and the evidence base is somewhat limited both in terms of the number of trials and the number of direct comparisons [6]. With the development of MCMC methods for fitting these models implemented in WinBUGS, hierarchical NMAs are not only computationally feasible but also widely applicable to other clinical settings.

Characteristically, NMAs performed on large networks with a relatively small evidence base frequently evaluate the interventions using an individual treatment NMA, thereby presenting extremely uncertain effect estimates. Alternatively, and in the case of the OAB syndrome example, the NMAs will focus on analyzing a specific set, or class of interventions. Both methodological approaches, however, can often make it difficult to infer the most efficacious intervention, making health policy decision making difficult. Conducting an individual treatment NMA, with a limited evidence base, produces considerable uncertainty in the effect estimates and thus any inferences regarding treatment effectiveness will remain cautious. Reducing the network, by collapsing arms to compare classes of interventions, will severely hinder the ability to specifically identify the most efficacious treatment overall. For example, the class-based NMA identified botulinum toxin to be the most efficacious class of interventions

Table 3 – Classification of treatments placed in ranked order of their relative difference compared with placebo (δ and δ*) and corresponding probabilities for the mean reduction in incontinence episodes from baseline.

S. no.	Class-based NMA			Hierarchical NMA				Hierarchical NMA with constraints			
	Treatment class	Rank (95% CrI)	p (best) %	δ (95% CrI)	Treatment class	Rank (95% CrI)	p (best) %	δ* (95% CrI)	Treatment class	Rank (95% CrI)	p (best) %	δ* (95% CrI)
1	Botulinum toxin	1 (1–3)	89.38	−1.40 (−2.14 to −0.66)	Botulinum toxin	1 (1–4)	11.75	−1.29 (−2.12 to −0.56)	Botulinum toxin	1 (1–3)	82.6	−1.32 (−1.95 to −0.66)
2	Anticholinergic drugs and behavior therapy	2 (1–5)	7.216	−0.83 (−1.21 to −0.44)	Anticholinergic drugs and behavior therapy	3 (1–6)	3.67	−0.78 (−1.23 to −0.36)	Behavior therapy	3 (1–7)	9.93	−0.76 (−1.41 to −0.12)
3	Anticholinergic and other drugs	3 (2–6)	1.096	−0.65 (−0.95 to −0.35)	Behavior therapy	3 (1–7)	0	−0.74 (−1.5 to −0.03)	Anticholinergic drugs and behavior therapy	3 (1–5)	5.71	−0.79 (−1.15 to −0.43)
4	Behavior therapy	4 (2–7)	2.122	−0.60 (−1.11 to −0.08)	Anticholinergic and other drugs	4 (2–6)	0.536	−0.62 (−0.98 to −0.24)	Anticholinergic and other drugs	4 (2–6)	1.25	−0.59 (−1 to −0.17)
5	Anticholinergic drugs	5 (3–6)	0.004	−0.54 (−0.63 to −0.45)	Anticholinergic drugs	5 (3–6)	0	−0.53 (−0.69 to −0.38)	Anticholinergic drugs	5 (3–6)	0.05	−0.54 (−0.69 to −0.38)
6	Other drugs	6 (4–7)	0.016	−0.39 (−0.64 to −0.15)	Other drugs	6 (3–7)	0.006	−0.20 (−0.66 to 0.30)	Other drugs	6 (4–7)	0.03	−0.21 (−0.67 to 0.26)
7	Other and other drug	7 (3–9)	0.158	−0.09 (−0.77 to 0.59)	Other and other drug	7 (2–8)	0.068	−0.11 (−0.84 to 0.64)	Other and other drug	7 (2–8)	0.43	−0.1 (−0.87 to 0.67)
8	Placebo	8 (7–8)	0	NA	Neuromodulation	8 (7–8)	0	1.19 (−0.16 to 2.72)	Neuromodulation	8 (6–8)	0	1.21 (−0.31 to 2.69)
9	Neuromodulation	9 (8–9)	0.01	1.29 (−0.02 to 2.61)	Placebo	NA	NA	NA	Placebo	NA	NA	NA

NA, not available/not applicable; NMA, network meta-analysis.

Table 4 – Treatments placed in ranked order of their relative odds compared with placebo (exp(φ) and exp(φ^*)) and corresponding probabilities for the number of patients reporting one or more adverse events.

		Individual NMA			Hierarchical NMA			
	Treatment	Rank worst (95% CrI)	p (worst) (%)	exp(φ) (95% CrI)	Treatment	Rank worst (95% CrI)	p (worst) (%)	exp(φ^*) (95% CrI)
1	Pregabalin	2 (1–9)	30.41	3.25 (1.52–7.03)	Pregabalin	3 (1–9)	27.66	2.77 (1.55–5.05)
2	Duloxetine	3 (1–11)	26.06	3.03 (1.32–7.03)	Duloxetine	3 (1–10)	23.27	2.66 (1.45–4.95)
3	Oxybutynin	3 (1–7)	14.09	3.06 (1.97–4.85)	Oxybutynin	3 (1–8)	16.09	2.61 (1.77–3.94)
4	Propiverine	4 (1–9)	6.14	2.61 (1.62–4.22)	Cizolirtine	4 (1–13)	14.29	2.36 (1.16–4.81)
5	Cizolirtine	5 (1–15)	15.88	2.39 (–0.81 to 7.17)	Propiverine	5 (1–10)	7.08	2.29 (1.54–3.53)
6	Darifenacin	5 (1–10)	5.21	2.41 (1.36–4.31)	Darifenacin	6 (1–11)	4.19	2.12 (1.36–3.39)
7	Pregabalin and tolterodine	7 (2–14)	1.08	1.86 (0.90–3.82)	UK-369,003	7 (1–14)	3.51	1.92 (0.96–3.60)
8	Solifenacin	8 (4–11)	0.05	1.77 (1.26–2.53)	Pregabalin and tolterodine	8 (1–15)	3.15	1.77 (0.9–3.49)
9	Imidafenacin	9 (4–14)	0.14	1.60 (0.93–2.75)	Solifenacin	8 (4–12)	0.2	1.77 (1.31–2.44)
10	UK-369,003	11 (3–16)	0.83	1.34 (0.56–3.22)	Imidafenacin	9 (4–14)	0.2	1.62 (1.06–2.48)
11	Fesoterodine	11 (5–15)	0.1	1.31 (0.69–2.46)	Fesoterodine	10 (4–15)	0.2	1.49 (0.92–2.39)
12	Trospium	11 (7–15)	0.001	1.27 (0.88–1.82)	Trospium	11 (7–14)	0.01	1.36 (0.99–1.88)
13	Tolterodine	12 (9–14)	0	1.17 (0.92–1.49)	Tolterodine	13 (10–14)	0	1.21 (0.97–1.52)
14	Tolterodine and BT	14 (7–16)	0.02	1.0 (0.52–2.03)	Tolterodine and BT	14 (6–16)	0.15	1.04 (0.54–2.03)
15	Placebo	14 (12–15)	0	NA	Placebo	14 (13–15)	0	NA
16	BoNTA 200 U	16 (12–16)	0.002	0.50 (0.21–1.19)	BoNTA 200 U	16 (13–16)	0.004	0.50 (0.21–1.16)
17	BT	17 (17–17)	0	0.001 (0.00–0.1)	BT	17 (17–17)	0	0.02 (0.00–0.21)

BT, bladder training; NMA, network meta-analysis.

with a probability of 89.38%, though it is unclear which specific BoNTA intervention, that is, dose, is the most efficacious overall.

In the current literature, use of the term "hierarchical NMA" is intermittently used to describe what is commonly known as a "random effects NMA" with variance components at two levels in the model, one at the within-study level and one at the between-study (within intervention) level. In this article, we demonstrate a third level in the model, accounting for an additional variance component between interventions within a class. Adding an additional level to the model changes the assumption of exchangeability and, consequently, the degree of shrinkage [1]. For this reason, there is a notable change in the estimated mean treatment effects and precision of the hierarchical NMA compared with that of the individual treatment NMA.

In comparison to the above methods, use of a hierarchical model as described in this article has several advantages. Principally, there is a substantial increase in the precision surrounding the effect estimates, and this was particularly apparent for the interventions for which there are few trials and a limited number of direct comparisons between other active interventions [6]. In addition, the hierarchical model maintains the interpretability of the effect estimates at an individual intervention level.

Nevertheless, the hierarchical models make a fundamental assumption that the intervention effects, within treatment classes, are exchangeable, and a judgment of appropriateness of such an assumption has to be made [1]. If this assumption does not hold for every class of interventions, use of a hierarchical model will introduce inappropriate results; thus, it is important for researchers to classify treatments into clinically plausible classes. Of course, the treatments do not have to be classified if there is no reason to do so. A further limitation of the hierarchical model is the subjective classification of the interventions when

there is potential treatment overlap. For example, in the OAB syndrome case, trospium and physiotherapy as a combination intervention will overlap with both anticholinergic and behavior therapy classes. Moreover, the combination of interventions individually estimated in the NMA could be modeled as the sum of individual components, with the potential to incorporate a synergistic or subadditive interaction between the interventions [19]. Furthermore, the number of interventions and trials within each class can vary substantially. Therefore, in particular classes in which there are few interventions and a small evidence base, the estimates will remain fairly uncertain. In situations such as these, the impact of the prior distributions on the variance parameters could be substantial, and use of extensive sensitivity analyses would be crucial [1,23].

Extending the hierarchical model to incorporate ordering constraints [9,10] on the BoNTA interventions for the OAB syndrome example resulted in the highest dose, BoNTA 300 U, to consistently be the most effective dose and therefore all other BoNTA interventions have a 0% probability of being the "best" intervention overall. In other examples, including ordering constraints in this way, that is, allowing the treatment effects of higher doses to be greater than or *equal* to those of lower doses, allows lower doses of interventions to have an equal probability of being the best intervention. Introducing these constraints for the mean reduction in incontinence episodes resulted in an estimate of –1.44 (95% CrI –2.09 to –0.78) for BoNTA 300 U compared with an estimate of –1.27 (95% CrI –2.12 to –0.51) for the unconstrained hierarchical model. Thus, assuming that larger doses of an intervention have a greater or equal effect to that of lower doses does not alter the effect estimates to any noticeable extent. This approach does, however, reduce the uncertainty in the effect estimates, the estimated ranking of

the interventions, and the probability that each of the interventions is the most effective, and consequently aids decision making.

The hierarchical model can be further extended to fit a multivariate hierarchical NMA [28] in which all outcomes are measured simultaneously [29]. This method of analysis can help ameliorate the potential effect of outcome reporting bias in trials that fail to report all the outcome measures of interest. This approach estimates a correlation between the outcomes to predict a value for the missing data points conditional on the outcome measures already reported and the model [30].

To further investigate the inconsistency detected between the direct and indirect evidence for the continuous outcome, exploratory analyses could investigate the association of baseline risk and treatment effect. Baseline risk represents the average response of a patient under the control group (e.g., placebo). If inconsistency is a result of pooling placebo interventions, incorporating baseline risk in to the model will explain some of the heterogeneity between studies [31].

In summary, we have shown that the use of hierarchical modeling, in NMAs, can increase the precision of intervention estimates, without hindering the interpretability of individual treatments. As demonstrated by the OAB syndrome example, borrowing strength within the classes of treatments reduced the uncertainty in individual estimates, yet estimated relative effects were still comparable with results obtained from individual analyses.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version at http://dx.doi.org/10.1016/j.jval.2014.10.006 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

R E F E R E N C E S

[1] Spiegelhalter D, Abrams KR, Myles J. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. England: John Wiley, 2003.
[2] Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health 2008;11:956–64.
[3] Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331:897–900.
[4] Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. Stat Med 2013;32:2935–49.
[5] Dakin HA, Welton NJ, Ades AE, et al. Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. Stat Med 2011;30:2511–35.
[6] Warren FC, Abrams KR, Sutton AJ. Hierarchical Bayesian network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. Stat Med 2014;33:2449–66.
[7] Haas DM, Caldwell DM, Kirkpatrick P, et al. Tocolytic therapy for preterm delivery: systematic review and network meta-analysis. BMJ 2012;345:e6226.
[8] Soares MO, Dumville JC, Ades AE, Welton NJ. Treatment comparisons for decision making: facing the problems of sparse and few data. J R Stat Soc A Stat Soc 2014;177:259–79.
[9] Demiris N, Lunn D, Sharples LD. Survival extrapolation using the poly-Weibull model. Stat Methods Med Res 2011 Nov 21. http://dx.doi.org/10.1177/0962280211419645: [Epub ahead of print].
[10] Prevost TC, Abrams KR, Jones DR. Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. Stat Med 2000;19:3359–76.
[11] National Institute for Health and Care Excellence. Urinary incontinence: the management of urinary incontinence in women. 2013. Available from: http://publications.nice.org.uk/urinary-incontinence-cg40. [Accessed January 5, 2013].
[12] Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996;15:2733–49.
[13] Dias S, Welton NJ, Caldwell D. Evidence synthesis for decision making in healthcare. Stat Med 2010;29:932–44.
[14] Welton NJ, Sutton AJ, Cooper NJ, et al. Evidence Synthesis for Decision Making in Healthcare. Chichester, UK: John Wiley & Sons, 2012.
[15] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc B Stat Methodol 2002;64:583–639.
[16] Chapple CR, Khullar V, Gabriel Z, et al. The effects of antimuscarinic treatments in overactive bladder: an update of a systematic review and meta-analysis. Eur Urol 2008;54:543–62.
[17] Novara G, Galfano A, Secco S, et al. A systematic review and meta-analysis of randomized controlled trials with antimuscarinic drugs for overactive bladder. Eur Urol 2008;54:740–64.
[18] Anger J, Weinberg A, Suttorp M, et al. Outcomes of intravesical botulinum toxin for idiopathic overactive bladder symptoms: a systematic review of the literature. J Urol 2010;183:2258–64.
[19] Welton NJ, Caldwell DM, Adamopoulos E, Vedhara K. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. Am J Epidemiol 2009;169:1158–65.
[20] Ades AE, Mavranezouli I, Dias S, et al. Network meta-analysis with competing risk outcomes. Value Health 2010;13:976–83.
[21] Spiegelhalter DJ, Thomas A, Best N, Lunn D. WinBUGS user manual version 1.4. January 2003. Upgraded to Version 1.4.3. Available from: http://www.mrc-bsu.cam.ac.uk/bugs. [Accessed January 20, 2013].
[22] Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 1998;7:434–55.
[23] Lambert PC, Sutton AJ, Burton PR, et al. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Stat Med 2005;24:2401–28.
[24] Ihaka R, Gentleman R. A language for data analysis and graphics. J Comput Graph Stat 1996;5:299–314.
[25] Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. J Stat Softw 2005;12:1–16.
[26] van Valkenhoef G, Lu G, de Brock B, et al. Automating network meta-analysis. Res Synth Methods 2012;3:285–99.
[27] Kanters S, Mills EJ, Thorlund K, et al. Antiretroviral therapy for initial human immunodeficiency virus/AIDS treatment: critical appraisal of the evidence from over 100 randomized trials and 400 systematic reviews and meta-analyses. Clin Microbiol Infect 2014;20:114–22.
[28] Hong H, Carlin BP, Shamliyan TA, et al. Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. Med Decis Making 2013;33:702–14.
[29] White I, Barrett J, Jackson D, Higgins J. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Res Synth Meth 2012;3:111–25.
[30] Nam I, Mengersen K, Garthwaite P. Multivariate meta-analysis. Stat Med 2003;22:2309–33.
[31] Achana FA, Cooper NJ, Dias S, et al. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. Stat Med 2013;32:752–71.