Botulinum toxin type A for the prevention of headaches in adults with chronic migraine

ERG Critique of Allergan Additional Analyses

March 16 2012

ERG comments on Allergan document of 24 Feb 2012: Botox for migraine

This document presents the ERG responses to the Allergan analyses submitted to NICE in the light of the ACD. The analyses of the Allergan submission are referred to as the original analyses, while those in response to the ACD are referred to as the revised analyses.

The additional Allergan analyses revise the model inputs for two main negative stopping rules:

- A negative stopping rule of 2 health states in 2 cycles.
- A negative stopping rule of a minimum 30% improvement in HDPM in 2 cycles

Given the other changes made to the base case the cost effectiveness can be presented for the various treatments of utilities.

1. Those for the 3+ prior medication patient group from the preferred MSQ mapping
2. As for (1) but averaging across the lowest two health states to avoid non-monotonicity
3. As for (1) but averaging between the arm for a given health state
4. All three of the above taken together

Point (1) corresponds to the analysis requested under point 1.3 of the ACD. Points (2), (3) and (4) correspond with the scenario analyses requested under point 1.4 of the ACD.

Revised cost effectiveness estimates

-ve stopping rule	2 health states in 2 cycles	30% improvement in HDPM
1. 3+ patient group	£16,507	£16,214
2. (1) + Lowest 2 HS pooled	£15,270	£14,999
3. (1) +Equalised between the arms	£23,952	£24,412
4. (1), (2) and (3)	£24,540	£24,939

Allergan has chosen to concentrate upon scenario analysis (2) that pools the quality of life values for the lowest two health states.

Base case modification

The revised base case is in line with that requested by the ACD with the following exceptions.

Administration costs

For the revised base case Allergan has applied the £140 neurologist OP cost for the cost per botox administration as requested in the ACD. But it has also applied £140 within the placebo arm which was not requested in the ACD.

The base case modelling previously applied half the administration costs within the placebo arm compared to that within the botox arm. This seemed justified on the basis of there being less frequent follow up, though formally within the modelling this was treated as £36.50 for 15 minutes follow up with a consultant compared to the £73.00 for 30 minutes for a botox appointment.

Incidentally, we note in one of the other responses to the ACD, from BASH, that “all patients require a consultation for at least 30 minutes” with Botox administration requiring an additional 10-15 minutes (BASH section b, first paragraph).

For the 2 health states in 2 cycles negative stopping rule, assuming placebo routine care costs are half that of botox at £70 per 12 weeks appears to worsen the ICER from £15,270 per QALY to £19,583 per QALY. Retaining the previous value of £36.50, which would also be similar to a 3 monthly GP visit, results in an ICER of £21,646 per QALY[1].

For the minimum 30% improvement in HDPM negative stopping rule, assuming placebo routine care costs are half that of botox at £70 per 12 weeks appears to worsen the ICER from £14,999 per QALY to £19,244 per QALY. Retaining the previous value of £36.50 results in an ICER of £21,275 per QALY.

The requirement for only a 30% reduction in headache days is based on expert opinion, (page 7, last para). In CM, this may be reasonable. In the 24+ band of headache days per month, a 30% reduction ranges from 7 to over 8 days, and this equates to a shift in two headache states. This does not apply to those who had 22-23 HDs at baseline, because they only have a one band shift.
It does apply again to those with baseline days 20 and 21.
So change from a stopping rule based on a 2-band shift, to one based on a 30% reduction in HDs, makes little difference to those who start in the worst two states. Though it should be noted that with a 30% reduction, none of those with baseline HDs of 22 or above will transfer from CM to episodic migraine.

Pooling of bottom 2 health state utilities to avoid non-monotonicity
Allergan has chosen to implement this for the revised base case, rather than as additional scenario analyses.

Routine care costs applying to those discontinuing botox
This appears to have been implemented, though as with the originally submitted model it remains difficult to read across between the botox cohort flow worksheets. But the ERG cross check model rebuild suggests these have been implemented as requested.

ERG cross check
There is good correspondence between the Allergan revised cost effectiveness estimate of £15,270 per QALY and the ERG cross check model rebuild of £15,166 per QALY when applying the same assumptions. The revised TPMs for the 30% negative stopping rule have not been implemented within the ERG cross check model rebuild.

Negative stopping rule

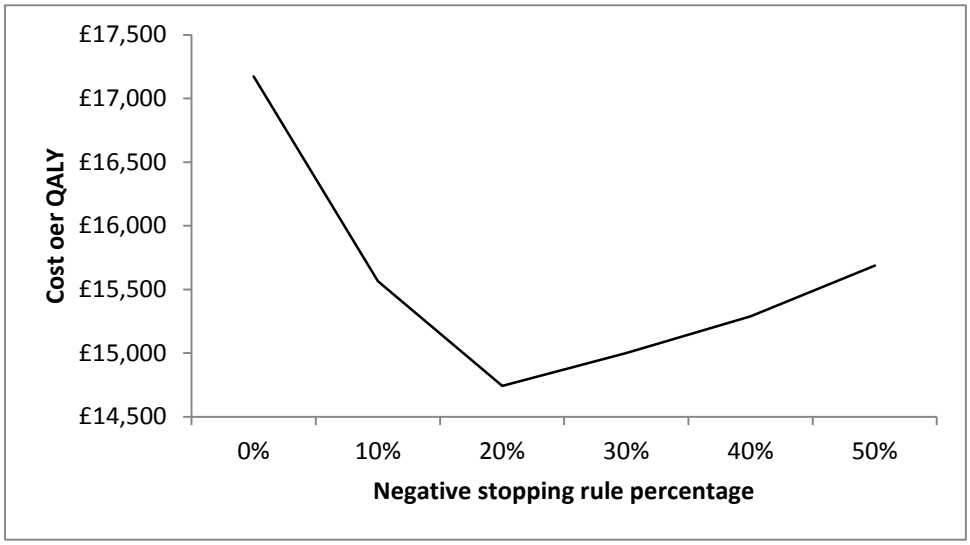
A number of additional TPMs have been implemented within the model corresponding to a range of revised negative stopping rules. These TPMs are summaries of patient level data and cannot be checked by the ERG.

Table 2 of the Allergan document suggests that with no negative stopping rule the cost effectiveness is £19,508 per QALY.
It should be noted that based on the trial data, 10.9% of patients discontinue Botox treatment prior to week 24 for a number of reasons, including adverse effects (2.8%), personal reasons (2.8%) loss to follow-up (1.9%), lack of efficacy (0.7%) and pregnancy (0.4%). (These figures are for the Botox arm. MS table5.6.)

If a stopping rule is applied to patients who worsen between baseline and week 24, and they discontinue botox, another 10.7% of patients discontinue and the cost effectiveness improves to £17,174 per QALY: an improvement of £2,334 per QALY (12%) compared to no

[1] Implemented within the *BOTOXcost* worksheet by changing the value within cell D12

stopping rule. This may be the most reasonable baseline from which to assess stopping rules.



Applying stricter negative stopping rules initially improves the cost effectiveness of botox, due to the patients who gain relatively little from botox having the additional costs of botox stopped from week 24 onwards.

But requiring an improvement of much more than 20% to 30% at week 24 results in the cost effectiveness estimate worsening. Even though the fewer patients continuing with botox are those benefitting most from it, and continue to experience further benefits between week 24 and week 60, those modelled as stopping botox at week 24 become an increasingly large proportion. Their benefits are less, and the fixed costs of treating them during the first two cycles may be weighing more heavily upon the cost per QALY calculation[2].

Probabilistic sensitivity analysis

See review of appendix 2 below.

The main revision stated by Allergan is the application of a nested beta for TPM sampling rather than the previous more bread and butter dirichlet implementation.

In the current implementation it is not possible to check the results when no priors are applied. There may be a concern around the priors permitting patients to cycle out of having discontinued, with this element being new to the structure of the electronic model. But this may be a misunderstanding on the part of the ERG economic reviewer.

It is currently unclear whether the move to the nested beta is required, and whether it is the source of the reduction in the non-linearity of the model.

Predictive accuracy of the model

This seems reasonable.

[2] In extremis a negative stopping rule could be applied to all but the best responding patient. While the other patients may have received some benefit by week 24, it is only for the one patient modelled as remaining on treatment there is a continued improvement modelled. But in cost effectiveness terms this one patient would have to justify a larger proportion of the fixed cost of all the other patients receiving botox to week 24 and the overall cost effectiveness of botox would be poor.

Time horizon

Given the available data the ERG viewed the 2 year time horizon of the model as reasonable. It was also noted that if the 2 year model was seen as one of a series of two year models it was possible that the balance of patients re-presenting for botox treatment in subsequent models might be tilted more towards those who responded to it previously. This could improve the subsequent cost effectiveness of botox.

Page 13 has modelling to 20 years, but this must be regarded as highly speculative since we don't have data beyond the PREEMPT trial timelines of 12 months. There is no evidence for "lingering benefit".

Effectiveness of placebo as treatment

Page 15, on placebo treatment, presents a sensitivity analysis wherein it is assumed that the placebo group experience no placebo effect. The reason for doing this is to deal with the situation whereby in routine care, no placebo would be given. The resulting ICER is therefore for Botox versus nothing. There is a problem with that approach because the benefits of Botox that drive the ICER consists of the placebo effect plus any pharmacological effect of Botox. The placebo effect will be responsible for most of the QALY gain.

UK IBMS resource use data

See review of appendix 3 below.

Update on the Rothrock data
The Allergan additional analyses document mentions, correctly, that the Rothrock abstract included only 100 patients. However it is one of the few pieces of evidence on long-term use. It has been updated by another abstract from the same group, with first author Hanlon, from the September 2011 meeting of the American Neurological Association. 123 consecutive patients have now been followed up for at least 2 years (range 24 to 61 months; mean 2.8 years). The patients in this study had all had a 50% or greater reduction in headache days. 10 of the 123 returned to CM and stopped Botox. 67% continued to respond but required continuing Botox injections at mean interval 3.4 months (range 3 to 6 months). Only 25% were able to stop Botox, after a mean of 4.8 sets of injections (range 2-8).

Thus suggests that the overall disposition of patients might be as follows;

- 51% do not respond as per the PREEMPT studies (table 5.32 of Allergan submission)
- Of those who respond 8% relapse (4% of the original cohort)
- Of those who respond, 25% stop Botox (12.5% of the original cohort)
- Implying that 33-34% continue on Botox long-term

However because of the small size of the Hanlon cohort, there would be wide confidence intervals around some of these figues. That around the 25% would be approximately 17 to 33%.

We note from the small Australian study, discussed below, that some patients remain on Botox for 5 years.

Appendix 1: Utilities

The ERG accepts that there are patient gains over and above the number of HDPM that arise from botox over placebo. This will give rise to additional patient gains and the possible differentiation of HRQoL between the arms for a given health state defined in terms of HDPM.

A concern remains about the extent of the differences in HRQoL between the arms for a given health state defined in terms of HDPM. Allergan has been asked for further and more explicit clarification around the calculation of the utilities underlying the analysis.
Table 15 of appendix 1 appears to confirm that the difference in utility values applied within the modelling for a given health state between the arms was derived from week 24 data. There may be some concerns given the short duration of the time horizon of the model about applying these differentials from baseline.

Table 15 also outlines some lack of monotonicity within the various measures as HDPM are increased:

- HIT-6 placebo arm HS 15-19 HDPM
- MSQ-RF restrictive botox arm HS 24+ HDPM
- MSQ-RF restrictive placebo arm HS 15-19 HDPM
- MSQ-RF preventative botox arm HS 24+ HDPM
- MSQ-RF preventative placebo arm HS 15-19 HDPM
- MSQ-RF preventative placebo arm HS 20-23 HDPM
- MSQ-RF preventative botox arm HS 24+ HDPM
- MSQ-EF placebo arm HS24+

The ERG been unable to apply the utility functions to the mean MSQ and HIT-6 data supplied within appendix 1 to arrive at any sensible results. It is unclear how the values have been calculated, and also how botox has been treated in relation to the headache medication coefficient of the MSQ utility equation.

Appendix 2: Revised PSA

Unfortunately, due to time constraints the ERG has not yet been able to fully assess the implementation of the probabilistic modelling. The Allergan method has changed from applying very peculiar priors in the context dirichlet sampling to more reasonable priors in the context of nested beta distribution sampling. ERG statistical input has confirmed that the nested beta is equivalent to a nested dirichlet, but is distinct from the more usually implemented dirichlet.

An additional continuity assumption that only applies the prior to TPM rows if any individual cell patient count within that row is less than 5 has also been applied. The reason for this continuity assumption is unclear and appears arbitrary. ERG work varying this value from 1 through to 9 suggests that it does not particularly affect results.

But within the current model implementation it has not yet been possible to set this to zero, and in effect not apply the priors. This may matter since the priors are also applied within the discontinuation row of the TPMs. This might suggest that within the probabilistic modelling patients can cycle out of having discontinued and re-appear elsewhere within the botox arm. But this may be an error of interpretation by the ERG economic reviewer.

There is also a TPM for those who have discontinued which appears to be new to the analysis and may suggest some further structural changes within the model other than just the move to the nested beta.

Note that the revised probabilistic modelling retains the oddity of summing the TPMs for cycles 3, 4 and 5 of the modelling in order to arrive at a single TPM for these periods and for extrapolation. This may have improved the deterministic ICER by around 5% to 10%. But it also serves to triple the number of patients within this TPM in the botox arm. This is likely to have somewhat reduced the uncertainty around the associated transition probability estimates applied from week 24 within the botox arm. As a consequence, it may also have reduced the degree of non-linearity between the deterministic and the probabilistic estimates.

Appendix 3: Other resource use data

Table 21 of appendix 3 of the Allergan additional analyses does not specify the time period that the resource use estimates relate to, but the ERG has assumed it to be 12 weeks. The text of Blumenfeld 2010 suggests that the IBMS respondents were asked about resource use during the previous 3 months.

It is not clear from Blumenfeld whether the IBMS respondents were asked to report all health care resource use data, or only resource use data related to headache and migraine but the companion paper by Stokes reports that they were asked about headache related resources use.
: *"Participants were asked about the frequency of resources used over the previous three months for health care professional visits, use of emergency department or urgent care clinic, overnight hospital stay, diagnostic tests, and any headache-specific treatments used including pharmacologic and non-pharmacologic interventions."*

If the resource use data collected during the IBMS was not specific to migraine and headaches, whether comorbidities differing between CM and EM patients could give rise to an additional concern around the resource use data is unknown.

Hospitalisation rates and lengths of stay

Among the 57 UK CM respondents 5 reported being hospitalised, with a range of length of stays of 0 to 15 days and an average length of stay of 4.40 days. This contrasts with the 1013 UK EM respondents, of whom 15 reported hospitalisations, with a range of 0 to 8 days and an average length of stay of 2.13 days. This LoS data yields averages of 0.39 days across the 5 UK CM IBMS respondents and 0.03 days across the 1013 UK EM IBMS respondents, which correspond with the values given in table 20 for the Allergan additional analyses.

Among the 5 UK CM respondents the stated lower bound of 0 for LoS will relate to a day case. The upper bound suggests that this resource use data is dominated by one patient who reported a LoS of 15 days. An average of 4.40 days LoS among the 5 respondents could be arrived at with:

- 1 patient 0 day LoS
- 1 patient 1 day LoS
- 2 patients 3 day LoS
- 1 patient 15 day LoS

Excluding the 1 patient with 15 days as an inpatient would reduce the average LoS from 4.40 days to 1.75 days.

Within any analyses based upon LoS, for the UK CM subset it would also seem prudent to undertake a sensitivity analysis that ignores the 1 patient reporting a LoS of 15 days.

Emergency room visits
Among the 57 UK CM respondents, 7 reported emergency room visits, with a range of 1 to 15 visits and an average of 4.14 visits. This contrasts with the 1013 UK EM respondents, of whom 35 reported emergency room visits, with a range of 1 to 10 visits and an average of 1.86 visits.

Among the 7 UK CM respondents, the upper bound suggests that this resource use data is dominated by one patient who reported 15 emergency room visits. An average of 4.14 among the 7 respondents could be arrived at with:

- 1 patient with 1 visit
- 2 patients with 2 visits
- 3 patients with 3 visits
- 1 patient 15 visits

Excluding the 1 patient with 15 visits would cause the average number of visits to fall from 4.14 visits to 2.33 visits.

Whether the 1 respondent reporting 15 days average LoS within the hospitalisation data is the same respondent reporting 15 emergency room visits is unknown. If it is, it might also be the case that each emergency room visit led to an overnight admission.

Other costs
The ERG still has doubts about these costs. For example looking at the cost items in table 21;

- Physician assistants are not used in primary care in the UK
- "other specialist visits" are not explained
- The inclusion of diagnostic tests and blood tests is inappropriate in chronic migraine, especially with one patients reported to have had 12 blood tests
- The inclusion of unproven forms of treatment such as acupuncture and occipital nerve blocks is inappropriate. (Note that the original modelling included costs for GON, on the assumption that it would be used if Botox were not available. The ERG view is that GON is not yet proven to be cost-effective and so should not be used.)
- No costs are included for NHS Direct.

Conclusions
In the light of the above it is unclear quite how to interpret the resource use data from both Blumenfeld and Bloudek. Blumenfeld reports "hospital visits" in Table 6, but the text and other papers from IBMS suggest that these are actually admissions. No data on length of stay are provided, while Bloudek reports average LoS data in line with tables 20 and 21 of the Allergan additional analyses.

The Blumenfeld data reflect patient resource use in a set of very different health care systems in nine countries, but may be preferable for the analyses, because of the small respondent numbers reporting resource use within the UK CM IBMS subset.
Another paper from the IBMS (Stokes et al 2011) reports data for the USA and Canada, and reports that no patients with chronic migraine in the IBMS groups were hospitalised in those countries.

In the light of the small respondent numbers, any analysis based upon the UK CM IBMS subset should also explore the impact that excluding the outlying patient has upon results.

The Australian report by So What Research.
This report is of limited value, because the results presented are very likely to be biased;

- Only 10 patients were included
- They were selected because they had done well on Botox. One would expect that they were chosen because botox had been particularly successful.
- They were paid for taking part.
- There are oddities such as the use of pethidine for analgesia which seems not in keeping with UK practice

The report is anecdotal with few numbers. For example, there are comments such as "patients reported reduced ER admissions and hospitalisations" (pages 3, 27 and 29) but no numbers are given. Page 25 says "hospitalisation not uncommon" but no data are provided on how many patients were admitted or how often, except that one was admitted while pregnant because of vomiting.

There are some items of interest. The patients' accounts show how severely migraine can affect daily life.

Page 11, table 1, shows that patients had botox every 3 months, some for 5 years. So in this group of patients in whom botox appeared very successful ("appeared" because we don't know how much was placebo effect), no successful stopping rule was applied. One patient made a comment about being scared to stop Botox.

Page 26 – the dose used was 100 to 150 units. According to this study, that was highly effective, raising questions as to why a minimum of 155 units is being marketed in the UK.

March 17th.