

ASTRAZENECA v ALTANA PHARMA

Promotion of Protium

AstraZeneca complained about the promotion of Protium (pantoprazole) by Altana Pharma. The items at issue were two mailings and a clinical paper summary which compared Protium with AstraZeneca's product Nexium (esomeprazole).

AstraZeneca noted that the claims 'Endoscopic healing rates equivalent to esomeprazole 40mg', 'Endoscopic healing rates comparable to esomeprazole 40mg' and '40 mg pantoprazole and 40mg esomeprazole are equivalent in the healing of esophageal lesions' were referenced to Gillessen *et al* (2004), which was a non-inferiority study, comparing the endoscopic healing rates of pantoprazole 40mg (n=113) and esomeprazole 40mg (n=114) in oesophagitis. The study utilised a hierarchical test procedure assessing a difference initially of 15% down to 5% between the two arms. The results contained no power calculations or 95% confidence intervals. Therefore this study could not prove its primary end point in order to substantiate these claims. Statistical equivalence could not be inferred from this type of study.

Conversely the more recent EXPO study had shown that esomeprazole 40mg was superior to pantoprazole 40mg in terms of healing rates in oesophagitis (Labenz *et al* 2005). This was a much larger (n=3151), well-powered study than Gillessen *et al*. Labenz *et al* showed esomeprazole had statistically superior healing rates in oesophagitis at four and eight weeks compared with pantoprazole. In addition two systematic reviews had shown that esomeprazole had superior healing rates compared with other proton pump inhibitors (including pantoprazole) (Edwards *et al* 2006, Isakov and Morozov 2006). The EXPO study and the systematic reviews supported the overall balance of evidence that esomeprazole had superior healing rates compared with pantoprazole. The Code, required promotion to be based on an up-to-date evaluation of all the available evidence; it must not mislead or make exaggerated claims.

AstraZeneca alleged that the claims were incorrect, misleading and incapable of substantiation.

The Panel noted that three head-to-head studies of pantoprazole vs esomeprazole had been submitted (Gillessen *et al*, Labenz *et al* and Bardhan *et al*). The claims at issue had been referenced to Gillessen *et al* which was a study set up to determine whether two treatments were equivalent. The overall endoscopically proven healing rates for both treatment groups were 88% in the intention to treat population. The corresponding values for the per protocol population were 95% (pantoprazole) and 90% (esomeprazole). The authors stated that these figures demonstrated that there existed 'at least equivalence' of pantoprazole and esomeprazole therapy. At ten

weeks the healing rates were 91% in the pantoprazole group and 97% in the esomeprazole group. No significant differences between the pantoprazole and esomeprazole groups could be shown. The Panel did not accept that an inability to show a statistical difference between the groups proved that the two treatments were equivalent. Gillessen *et al* noted that prior to their study there existed no comparable clinical material that directly compared pantoprazole and esomeprazole.

The results of the EXPO study were published the year after Gillessen *et al*. This was a much larger study designed to compare esomeprazole 40mg (n=1562) with pantoprazole 40mg (n=1589) for healing in patients with erosive oesophagitis. After up to eight weeks significantly more esomeprazole-treated patients were healed (95.5%) compared with pantoprazole-treated patients (92%) (p<0.001).

The Panel noted the table of results from Bardhan *et al* given by Altana was stated to show the percentage of healing rates but the figures quoted were in fact the cumulative rates of complete remission as reported by the authors. (Complete remission was defined as both endoscopically confirmed healing and symptom relief as assessed by questionnaire.) Altana had shown for the last of these results (12 weeks) that Protium was statistically superior to Nexium; this was not so. At 12 weeks the authors had reported that pantoprazole was not inferior to esomeprazole. With regard to the healing of oesophageal lesions at 12 weeks, pantoprazole showed superior results compared with esomeprazole (98% v 94.4%) although the statistical significance of this result was not stated.

The Panel noted the sizes of the three studies cited and considered that the balance of evidence lay with the EXPO study ie that although in absolute terms the healing rates of both pantoprazole and esomeprazole were very similar there was a statistically significant difference in favour of esomeprazole.

The Panel thus considered that the claims that Protium 40mg was equivalent or comparable to esomeprazole in terms of healing were incorrect, misleading and not capable of substantiation as alleged. Breaches of the Code were ruled.

Upon appeal by Altana in relation to the claim 'Endoscopic healing rates comparable to esomeprazole 40mg', the Appeal Board considered that, in common parlance, if two medicines were described as comparable then prescribers and patients would generally not mind which one was used. The Code required material including comparisons to have a statistical foundation. Clinical relevance was an important consideration.

The Appeal Board noted how the parameters of Gilleson *et al* had changed as the study progressed and in that regard it considered that the results were not as robust as those from the EXPO study. The Appeal Board further noted that unlike the EXPO study, Gilleson *et al* had not included patients with Los Angeles grade D (ie more severe) oesophagitis. The EXPO study had shown that for both esomeprazole and pantoprazole there was a decline in healing rates with increasing baseline severity of oesophagitis. After 8 weeks of therapy the healing rates for esomeprazole 40mg were statistically superior to pantoprazole 40mg with LA grades B, C and D at baseline.

The Appeal Board considered that the claim 'Endoscopic healing rates comparable to esomeprazole 40mg' was too broad such that it was ambiguous. It implied that in patients with any grade of gastroesophageal reflux disease (GERD), healing rates observed with Protium had been shown to be statistically similar to those observed with Nexium which was not so. The claim was misleading in that regard. The Appeal Board upheld the Panel's ruling of a breach of the Code.

The Appeal Board noted that the EXPO study had shown that, overall, healing rates with Protium and Nexium were very similar in absolute terms. In that regard the Appeal Board thus considered that there was no breach.

AstraZeneca noted that the claim 'Once daily pantoprazole 40mg and esomeprazole 40mg have equivalent overall efficacy in relieving GERD-related symptoms' was referenced to Scholten *et al* (2003), a superiority study comparing the area under the curves (AUCs) for the symptom scores. There was no statistical difference ($p > 0.05$) between the two treatment groups. From this non-significant value it was concluded that pantoprazole and esomeprazole were equivalent with respect to symptoms. This was an incorrect conclusion; a non-significant p value for superiority did not imply equivalence. In order to show equivalence, a pre-specified equivalence margin had to be stipulated with construction of confidence intervals for the treatment difference. Equivalence was inferred if the confidence intervals fell entirely within the equivalence margins.

AstraZeneca submitted that differences that did not reach statistical significance must not be presented in such a way as to mislead. Thus this claim was misleading and incapable of substantiation.

The Panel noted that Scholten *et al* was designed to compare the efficacy of pantoprazole (40mg) (n=112) and esomeprazole (40mg) (n=105) in the treatment of GERD-related symptoms. The primary criterion of the study was to evaluate symptom load of GERD-related symptoms, defined as AUC for the symptom score. Over the 28 day treatment period the AUCs for the six typical GERD-related symptoms (heartburn, acid regurgitation, gastric complaints, pressure in the epigastrium, feeling of satiety and flatulence) were similar and comparable in the two treatment groups ($p > 0.05$). Thus the study was unable to show a

statistically significant difference between the two medicines. The results did not mean that the study had proven the two were equivalent. The Panel thus considered that the claim 'Once-daily pantoprazole 40mg and esomeprazole 40mg have equivalent overall efficiency in relieving GERD-related symptoms' was misleading and could not be substantiated as alleged. Breaches of the Code were ruled.

AstraZeneca noted that the claims 'Fast symptom control- 2 days faster than esomeprazole 40mg', 'daytime symptom relief - 2 days faster' and '2 days faster than esomeprazole 40mg' were referenced to the secondary end points of Scholten *et al*. As stated above, this study did not reach statistical significance in terms of the primary outcome (AUC of the GERD symptoms scores between esomeprazole 40mg and pantoprazole 40mg).

AstraZeneca believed that if there was an inconsistency in terms of the interpretation of the study from a secondary endpoint alone, the primary endpoint should be given sufficient clarity, such that the claim could be immediately seen in the context of the primary endpoint. AstraZeneca considered that it was misleading to use a secondary endpoint alone if it would lead the reader to draw a different conclusion to that of the primary end point.

AstraZeneca submitted that in this case, the secondary endpoint claims did not inform the reader of the primary outcome of the study (AUC of symptoms scores between esomeprazole 40mg and pantoprazole 40mg) and were not consistent with the result of the primary end point. In addition, as a secondary endpoint, the study would not have been appropriately powered to examine this measure, and was therefore at risk from statistical error.

In addition, the EXPO study showed that esomeprazole 40mg provided faster and more effective resolution of heartburn than pantoprazole 40mg. This was based on the time to sustained resolution of symptoms (defined as a period of seven consecutive days without heartburn). This was in contrast to the assessment of symptoms in Scholten *et al* that assessed time to adequate relief. In Scholten *et al* patients did not have to reach complete resolution of symptoms. Time to sustain a resolution of symptoms as shown by esomeprazole 40mg was much more clinically relevant as it was a period of prolonged improvement in contrast to achieving a period of partial symptomatic relief. Thus, the claims were misleading and did not reflect the available evidence.

The Panel noted that in Scholten *et al* patients recorded the perceived intensity of GERD-related symptoms (heartburn, acid regurgitation, gastric complaints, pressure in the epigastrium, feeling of satiety and flatulence). A five-point Likert scale was used to assess the intensity of each symptom: none (0), mild (1), moderate (2), severe (3) and very severe (4). Each symptom was assessed and scored and if the sum score fell below 5 for the first time, the patient was characterized as having reached adequate relief from GERD-related symptoms. The patients did not have to

reach complete symptom relief. The results of the study showed that for daytime, the first time to reach adequate relief of GERD-related symptoms in the pantoprazole group was 3.73 days and 5.88 days for the esomeprazole group ($p=0.034$). This was the result upon which the claims in question were based. The Panel noted, however, that the claims only referred to 'symptom relief' or 'symptom control', not 'adequate symptom relief control'. In the Panel's view the claims implied total symptom relief/control which was not so. The Panel further noted that the claims did not refer to 'first time' relief and in that regard there was an implication that sustained relief of symptoms was achieved with pantoprazole after 3.7 days. There was no data to show this.

The Panel thus considered that the claims at issue were misleading and did not reflect the available evidence as alleged. Breaches of the Code were ruled.

Upon appeal, the Appeal Board considered that it was unacceptable to use secondary data to claim an advantage for Protium over Nexium when the primary endpoint had been unable to show such a difference. The Appeal Board considered that the claims were misleading in this regard and did not reflect the available evidence as alleged. The Appeal Board upheld the Panel's rulings of breaches of the Code.

AstraZeneca UK Limited complained about the promotion of Protium (pantoprazole) by Altana Pharma Limited. The items at issue were two mailings (ref PAN208/071205/P and PAN291/020806/P) and a clinical paper summary (PAN202/291105/P) which compared Protium with AstraZeneca's product Nexium (esomeprazole).

- 1 Claims 'Endoscopic healing rates equivalent to esomeprazole 40mg' (PAN208/071205/P), 'Endoscopic healing rates comparable to esomeprazole 40mg' (PAN291/020806/P) and '40 mg pantoprazole and 40mg esomeprazole are equivalent in the healing of esophageal lesions' (PAN202/291105/P)

COMPLAINT

AstraZeneca noted that all of these claims were referenced to Gillessen *et al* (2004), which was a non-inferiority study, comparing the endoscopic healing rates of pantoprazole 40mg ($n=113$) and esomeprazole 40mg ($n=114$) in oesophagitis. The study utilised a hierarchical test procedure assessing a difference initially of 15% down to 5% between the two arms of the study. The results in this study contained no power calculations or 95% confidence intervals, which were the accepted methods to assess statistical relevance of the findings. Therefore this study could not prove its primary end point in order to substantiate these claims. This was further supported by a published letter to the editor of the journal which re-iterated that the study had insufficient power and sample size to reach a conclusion (Madisch *et al* 2005). Furthermore, statistical equivalence could not be inferred from this type of study.

AstraZeneca noted that in contrast the more recent EXPO study had shown that esomeprazole 40mg was superior to pantoprazole 40mg in terms of healing rates in oesophagitis (Labenz *et al* 2005). This was a much larger ($n=3151$), well-powered study than Gillessen *et al*. Labenz *et al* showed esomeprazole had statistically superior healing rates in oesophagitis at four and eight weeks compared with pantoprazole. In addition two systematic reviews had shown that esomeprazole had superior healing rates compared with other proton pump inhibitors (including pantoprazole) (Edwards *et al* 2006, Isakov and Morozov 2006). The EXPO study and the systematic reviews supported the overall balance of evidence that esomeprazole had superior healing rates compared with pantoprazole. AstraZeneca noted that the Code required promotion to be based on an up-to-date evaluation of all the available evidence; it must not mislead or make exaggerated claims.

AstraZeneca stated that there should be a sound statistical basis for all statistical claims and comparisons in promotional material, and that care should be taken to ensure that the information was not presented in such a way as to mislead. Thus, AstraZeneca alleged that the claims at issue were incorrect, misleading and incapable of substantiation in breach of Clauses 7.2, 7.3 and 7.4 of the Code.

RESPONSE

Altana submitted that Gillessen *et al* was a peer-reviewed article published in the Journal of Gastroenterology and as such both the study methodology and the clinical paper had been independently peer reviewed before publication. Furthermore the study design and statistical methods were approved by ten independent local ethics committees before the study started. This clearly demonstrated that the study design was robust and that the results achieved were both meaningful and clinically relevant. The study was designed to show non-inferiority using a hierarchical test procedure, testing the non-inferiority margin initially at 15%, then at 10% and finally at 5%. Therefore a lower 95% confidence interval of less than 5% would indicate non-inferiority. Whilst it was regrettable that this lower 95% confidence interval was not included in the original publication, the clinical research department at Altana AG (study sponsors) had confirmed that this figure was 4.88%, thus confirming the authors' conclusion that '40mg pantoprazole (Protium) daily and 40mg daily esomeprazole (Nexium) were equally effective for the healing of esophageal lesions'.

Altana submitted that the power calculations were not relevant to the outcome of the study. The letter from Madisch *et al* to the editor of the journal suggesting that the trial was underpowered and lacking in sample size was adequately refuted (Gillessen 2005a).

Altana noted that AstraZeneca had stated that the EXPO study and two review papers supported its position that Nexium was superior to Protium in terms of healing rates in erosive oesophagitis. Altana noted, however, that Edwards *et al* compared Nexium to 'other proton pump

inhibitors' (PPIs) which included omeprazole, lansoprazole and Protium. Therefore the Nexium versus 'combined PPI' summary findings had no relevance to this complaint when the data required was head-to-head comparisons of Nexium and Protium in the healing of erosive oesophagitis. Further, Edwards *et al* only included one Nexium versus Protium study (the EXPO study) in the set of six studies that were included in the final analysis. Thus in citing Edwards *et al* AstraZeneca had offered no further support to its position as it was, in effect, a repeat citing of the EXPO study.

Altana submitted that the Isakov and Morozov meta-analysis was also a combined analysis in which Nexium was compared to omeprazole, lansoprazole and Protium. This meta-analysis considered eight clinical papers, only three of which were trials of Nexium versus Protium. As stated earlier, this type of combined endpoint was not relevant to this complaint when the data required was head-to-head comparisons of Nexium and Protium in the healing of erosive oesophagitis.

Altana submitted the EXPO study was the only study cited by AstraZeneca to support a claim that Nexium had statistically superior healing rates in oesophagitis at four and eight weeks. However the absolute difference between the two treatments was very small, 3.5%, and both showed healing rates greater than 90%. Disparities in the distribution of less severe patients between the trial groups, which might have materially affected this very small absolute difference in favour of Nexium had been raised (Gillessen 2005b).

Equally the relevance of the absolute difference, 3.5%, observed in healing rates was of little clinical significance when both products had a success rate of over 90%.

Altana submitted that the claims in question were fully supported by a full review of the available evidence looking at healing rates in erosive oesophagitis in clinical trials of 40mg Protium versus 40mg Nexium.

Altana submitted a table that summarised the clinical trial results from three studies considering this matter (Gillessen *et al*, Labenz *et al* and Bardhan *et al* 2005). Whilst it would always be the case that individual studies would have a unique design the three listed all looked at endoscopically proven healing of erosive oesophagitis over time.

Altana submitted that the table supported its position that, upon an up-to-date analysis of all the available evidence, there was minimal difference between the two products in clinical terms for oesophageal healing rates. In different studies both Protium and Nexium had been shown to be statistically superior at different time points. However this was of no clinical relevance when the entire data set was reviewed and it was recognised that despite small inter-study variation the healing rates in every study were very closely similar.

Altana submitted that claims made in promotional material must not mislead and should reflect both the statistical and clinical relevance. Therefore this table of data strongly supported the terms 'equivalent' and

'comparable' as used in the claims at issue.

The term 'equivalent' was taken directly from the title of Gillessen *et al* and Scholten *et al* (2003) also used the term 'equivalent' in its title. These publications were in peer-reviewed journals and reflected the average physician's interpretation of the term 'equivalent' through its common or everyday meaning. In this clinical context 'equivalent' was understood to mean 'as effective as', and was not interpreted in a pure statistical manner.

Altana submitted the term 'comparable' was entirely appropriate and fully substantiated given the minimal absolute difference between the products in oesophageal healing rates in every study.

Altana denied breaches of Clauses 7.2, 7.3 and 7.4.

PANEL RULING

The Panel noted that three head-to-head studies of pantoprazole versus esomeprazole had been submitted (Gillessen *et al*, Labenz *et al* and Bardhan *et al*). The claims at issue had been referenced to Gillessen *et al* which was a study set up to determine whether the two treatments were equivalent. The overall endoscopically proven healing rates for both treatment groups were 88% in the intention to treat population. The corresponding values for the per protocol population were 95% (pantoprazole) and 90% (esomeprazole). The authors stated that these figures demonstrated that there existed 'at least equivalence' of pantoprazole and esomeprazole therapy. At ten weeks the healing rates were 91% in the pantoprazole group and 97% in the esomeprazole group. No significant differences between the pantoprazole and esomeprazole groups could be shown. The Panel did not accept that an inability to show a statistical difference between the groups proved that the two treatments were equivalent. Gillessen *et al* noted that prior to their study there existed no comparable clinical material that directly compared pantoprazole and esomeprazole.

The results of the EXPO study were published the year after Gillessen *et al*. This was a much larger study designed to compare esomeprazole 40mg (n=1562) with pantoprazole 40mg (n=1589) for healing in patients with erosive oesophagitis. After up to eight weeks significantly more esomeprazole-treated patients were healed (95.5%) compared with pantoprazole-treated patients (92%) (p<0.001).

The Panel noted that Altana had cited Bardhan *et al*. The table of results given by Altana was stated to show the percentage of healing rates but the figures quoted for Bardhan *et al* were in fact the cumulative rates of complete remission as reported by the authors. (Complete remission was defined as both endoscopically confirmed healing and symptom relief as assessed by questionnaire.) Altana had shown for the last of these results (12 weeks) that Protium was statistically superior to Nexium; this was not so. At 12 weeks the authors had reported that pantoprazole was not inferior to esomeprazole. With regard to the healing of oesophageal lesions at 12 weeks, pantoprazole showed superior

results compared with esomeprazole (98% v 94.4%) although the statistical significance of this result was not stated.

The Panel noted the sizes of the three studies cited and considered that the balance of evidence lay with the EXPO study ie that although in absolute terms the healing rates of both pantoprazole and esomeprazole were very similar there was a statistically significant difference in favour of esomeprazole.

The Panel thus considered that the claims that Protium 40mg was equivalent or comparable to esomeprazole in terms of healing were incorrect, misleading and not capable of substantiation as alleged. Breaches of Clauses 7.2, 7.3 and 7.4 were ruled.

APPEAL BY ALTANA

Altana appealed the Panel's rulings of breaches of Clauses 7.2, 7.3 and 7.4 of the Code with regard to the claim 'Endoscopic healing rates comparable to esomeprazole 40mg'.

Altana considered that the Panel's ruling appeared to be entirely inconsistent with the wording used within the text of the ruling. Altana submitted that the word 'comparable' was not a defined term with respect to statistics or medicine. Therefore the accepted use of this word in English should be used in this case, this being 'similar in size, amount or quality to something else'.

The ruling stated that 'The Panel noted the sizes of the three studies cited and considered that the balance of evidence lay with the EXPO study ie that although in absolute terms the healing rates of both pantoprazole and esomeprazole were very similar there was a statistically significant difference in favour of esomeprazole' (emphasis added by Altana).

Altana submitted that in view of the meaning of 'comparable', deeming that the word was 'incorrect, misleading and not capable of substantiation' in this instance appeared to be an illogical conclusion given that the Panel had agreed that there was almost no difference in absolute healing rates between the two products. This closely similar absolute healing rate represented the success rate that any physician might expect to achieve when using either product.

Altana submitted that by the Panel's own words it was clear that this statement was not misleading to the intended audience of health professionals. The healing rates of the two products were, without doubt, comparable when all the studies in the pool of evidence were considered.

Altana submitted that the balance of evidence showed that there was no difference between the two products in absolute healing rates, their effect was very similar and therefore use of the term comparable was appropriate and correct.

Altana submitted that it was improper, and in itself misleading, for the Panel to determine that the minimal

absolute difference in the EXPO study should be seen as a statistically superior advantage for Nexium given that two other well-powered studies showed contrary results. The balance of evidence strongly supported essential similarity between the products and justified use of the term 'comparable' in this context.

Altana submitted that large studies, such as the EXPO study might give rise to statistically significant results for clinically meaningless absolute differences. It was wrong to claim that the size of the study had any bearing on the balance of evidence. Studies were powered according to the study type (non-inferiority, superiority) and according to the magnitude of the difference between the treatments that was predicted to exist. Ethics committee review ensured patient enrolment into clinical studies was sufficient to demonstrate a real difference if the difference really existed. If the clinical difference between the products was predicted to be small many patients might be required as in the EXPO study.

Altana submitted that it was a flawed argument to suggest that the EXPO study should be given more credibility and weighting in the pool of available data than Gillessen *et al*, Achim *et al*, and Bardhan *et al* for the reasons given. A statistician would confirm that the size of a study did not relate to the relative merits of its outcome.

Altana submitted that there must be clinical relevance in the delivery of promotional claims or they were themselves misleading to the intended audience. For the Panel to express the opinion that the EXPO study carried more weight in the available evidence when Achim *et al* and Gillessen *et al* demonstrated non-inferiority and superiority for Protium over Nexium was not representative of the balance of evidence available.

Altana submitted that it had not claimed Protium superiority over Nexium because this would have misrepresented the entire data set and be misleading to health professionals. Equally the reverse was true. It could not be deemed by the Panel 'that although in absolute terms the healing rates of both pantoprazole and esomeprazole were very similar there was a statistically significant difference in favour of esomeprazole'. This was a misrepresentation of the entire data set available.

Altana submitted that the only possible outcome upon consideration of the whole data set, that would not mislead customers, was that Protium and Nexium had very similar or comparable healing rates. These considerations previously raised by Altana had not been adequately discussed in the Panel ruling to illustrate its reasoning and create a transparent response.

COMMENTS FROM ASTRAZENECA

AstraZeneca noted Gillessen *et al* used a hierarchical test procedure assessing a difference initially of 15% down to 5% between the two treatment arms. The study had several serious limitations due to poor statistical analysis

and inappropriate sample size in order to draw any meaningful conclusions.

- It did not follow the guidelines of the European Medicines Evaluation Agency in utilizing a pre-specified non-inferiority margin instead of shifting margins. Changing the non-inferiority margins would require a different sample size in order to prove the study hypothesis. The choice of the margin was critical in calculating the sample size and in the interpretation of the data.
- The authors did not describe any sample size and power calculations or 95% confidence intervals which was highly important for any non-inferiority study.
- If the study had planned a non-inferiority margin of 5% then more than 1000 patients would be required to test for non-inferiority at this level.
- Using a non-inferiority margin of up to -15%, was a difference too large to conclude that treatments were comparable in healing oesophagitis.
- Using the data presented, the 95% confidence interval (CI) for the intention to treat (ITT) difference might be calculated to -9 to +9%, clearly not significant at the non-inferiority limit of 5%. For the per protocol (PP) analysis the estimated difference was 4.4% and the 95% two-sided CI was -3 to +12%. Testing the PP treatment difference with Fisher's exact test gave $p=0.29$, which was clearly not statistically significant.
- The study was limited to patients with Los Angeles grade B and C oesophagitis and with treatment groups split into three strata, resulting in fewer than 40 patients per stratum. No results of this stratification were presented.

AstraZeneca alleged that Gillessen *et al* was unable to prove the primary endpoint of non-inferiority of pantoprazole 40mg to esomeprazole 40mg and thus the claim for comparable healing rates to esomeprazole 40mg could not be justified.

Statistical information should not be presented in a way to mislead the reader.

AstraZeneca alleged it had conclusively shown in a much larger ($n=3151$), well-designed study (EXPO) that was performed after Gillessen *et al*, that esomeprazole 40mg was indeed superior to pantoprazole 40mg for healing oesophagitis (Labenz *et al*).

AstraZeneca noted that Altana had claimed that the EXPO findings were not clinically important.

- Given the number of patients who were treated with PPIs, the statistically significant 3.5% improvement in healing rates with esomeprazole relative to pantoprazole was clinically important and represented a clear improvement over pantoprazole for patients with erosive oesophagitis.
- Moreover, the difference was substantially greater after 4 weeks of treatment and with increasing

severity of oesophagitis respectively.

- In addition, logistic regression analysis of EXPO clearly identified choice of PPI (esomeprazole vs pantoprazole - odds ratio 1.3) as an independent predictor of success in healing (Labenz *et al* 2006) and heartburn resolution (Labenz *et al* 2005).
- Furthermore, the EXPO study also provided greater therapeutic relevance because it assessed not only the acute treatment of oesophagitis, but also, in the same patient population, maintenance therapy with esomeprazole 20mg or pantoprazole 20mg (Labenz *et al* 2005).

AstraZeneca noted that Altana had referred to a study that was not used to support this claim in its promotional material. The abstract on healing, Bardhan *et al* and the combined analysis, Achim *et al* had not been published in a peer reviewed journal in order to assess their validity in determining sample size and statistical analyses. The authors used an integrated approach combining both endoscopic healing and symptom status. As this methodology combined two variables it could not be used to support the claim of 'comparable healing'.

AstraZeneca noted that in Achim *et al* the non-inferiority margin had been set at -15%; pending statistical validity, again such a large treatment difference could not justify the term 'comparable healing'.

AstraZeneca alleged that the claim 'comparable healing rates to esomeprazole 40 mg' could not be substantiated when it had been shown that esomeprazole was superior to pantoprazole in the healing of oesophagitis. Such a claim did not represent the balance of evidence.

APPEAL BOARD RULING

The Appeal Board considered that, in common parlance, if two medicines were described as comparable then prescribers and patients would generally not mind which one was used. The Code required material including comparisons to have a statistical foundation. Clinical relevance was an important consideration.

The Appeal Board noted how the parameters of Gillessen *et al* had changed as the study progressed and in that regard it considered that the results were not as robust as those from the EXPO study. The Appeal Board further noted that unlike the EXPO study, Gillessen *et al* had not included patients with Los Angeles grade D (ie more severe) oesophagitis. The EXPO study had shown that for both esomeprazole and pantoprazole there was a decline in healing rates with increasing baseline severity of oesophagitis. After 8 weeks of therapy the healing rates for esomeprazole 40mg were statistically superior to pantoprazole 40mg with LA grades B, C and D at baseline.

The Appeal Board considered that the claim 'Endoscopic healing rates comparable to esomeprazole 40mg' was too broad such that it was ambiguous. It implied that in patients with any grade of gastroesophageal reflux

disease (GERD), healing rates observed with Protium had been shown to be statistically similar to those observed with Nexium which was not so. The claim was misleading in that regard. The Appeal Board upheld the Panel's ruling of a breach of Clause 7.2. The appeal on this point was unsuccessful.

The Appeal Board noted that the EXPO study had shown that, overall, healing rates with Protium and Nexium were very similar in absolute terms. In that regard the Appeal Board thus considered that there was no breach of either Clause 7.3 or 7.4 and ruled accordingly. The appeal on these points was successful.

2 Claim 'Once daily pantoprazole 40mg and esomeprazole 40mg have equivalent overall efficacy in relieving GERD-related symptoms' (PAN202/291105/P)

COMPLAINT

AstraZeneca noted that the claim was referenced to Scholten *et al* (2003), which was designed as a superiority study comparing the area under the curves (AUCs) for the symptom scores of pantoprazole and esomeprazole. There was no statistical difference ($p>0.05$) between the two treatment groups. It was incorrect to conclude from this non-significant value that pantoprazole and esomeprazole were equivalent with respect to symptoms; a non-significant p value for superiority did not imply equivalence. In order to show equivalence, a pre-specified equivalence margin had to be stipulated with construction of confidence intervals for the treatment difference. Equivalence was inferred if the confidence intervals fell entirely within the equivalence margins.

AstraZeneca submitted that differences that did not reach statistical significance must not be presented in such a way as to mislead. Thus this claim was misleading, incapable of substantiation in breach of Clauses 7.2, 7.3 and 7.4

RESPONSE

Altana submitted that Scholten *et al* was designed as a non-inferiority study and not as a superiority study as stated by AstraZeneca. The study received prior independent ethics committee approval and was subsequently published in a peer-reviewed journal. The primary criterion of Scholten *et al* was to evaluate Protium and Nexium in terms of symptom load of GERD-related symptoms, defined AUC for the symptom score. The between group comparisons for the AUC was done by Wilcoxon rank-sum test (5% level, two-sided). The AUCs for the GERD-related symptoms were similar and comparable between the two treatment groups ($p>0.05$). This claim did not misrepresent the statistical outcome from this study.

Altana submitted that as in point 1 above, 'equivalent' was taken directly from the title of Scholten *et al*. Publication was in a peer-reviewed journal and reflected the average physician's interpretation of the term

'equivalent' through its common or everyday meaning. In this clinical context 'equivalent' was understood to mean 'as effective as', and was not interpreted in a pure statistical manner. This claim was not in breach of Clauses 7.2, 7.3 and 7.4.

PANEL RULING

The Panel noted that Scholten *et al* compared the efficacy of pantoprazole (40mg) ($n=112$) and esomeprazole (40mg) ($n=105$) in the treatment of GERD-related symptoms. The primary criterion of the study was to evaluate symptom load of GERD-related symptoms, defined as AUC for the symptom score. Over the 28 day treatment period the AUCs for the six typical GERD-related symptoms (heartburn, acid regurgitation, gastric complaints, pressure in the epigastrium, feeling of satiety and flatulence) were similar and comparable in the two treatment groups ($p>0.05$). Thus the study was unable to show a statistically significant difference between the two medicines. The results did not mean that the study had proven the two were equivalent. The Panel thus considered that the claim 'Once-daily pantoprazole 40mg and esomeprazole 40mg have equivalent overall efficiency in relieving GERD-related symptoms' was misleading and could not be substantiated as alleged. Breaches of Clauses 7.2, 7.3 and 7.4 were ruled.

3 Claims 'Fast symptom control - 2 days faster than esomeprazole 40mg' (PAN208/071205/P), 'daytime symptom relief - 2 days faster' (PAN202/291105/P) and '2 days faster than esomeprazole 40mg' (PAN291/020806/P)

COMPLAINT

AstraZeneca noted that the claims were referenced to the secondary end points of Scholten *et al* (time to adequate relief of GERD-related symptoms). As stated at point 2 above, this study did not reach statistical significance in terms of the primary outcome (AUC of the GERD symptoms scores between esomeprazole 40mg and pantoprazole 40mg).

AstraZeneca believed that it was appropriate to use secondary endpoints without the primary end point when the analysis of the secondary end point was consistent with the primary endpoint of the study. If there was an inconsistency in terms of the interpretation of the study from a secondary endpoint alone, the primary endpoint should be given sufficient clarity, such that the claim could be immediately seen in the context of the primary endpoint. AstraZeneca considered that it was misleading to use a secondary endpoint alone if it would lead the reader to draw a different conclusion to that of the primary end point.

AstraZeneca submitted that in this case, the secondary endpoint claims did not inform the reader of the primary outcome of the study (AUC of symptoms scores between esomeprazole 40mg and pantoprazole 40mg) and were not consistent with the result of the primary end point. In addition, as a secondary endpoint, the study would not have been appropriately powered to examine this

measure, and was therefore at risk from statistical error.

AstraZeneca considered that the Panel's ruling on a similar case, Case AUTH/1579/4/04, was relevant.

AstraZeneca stated that in addition, the EXPO study showed that esomeprazole 40mg provided faster and more effective resolution of heartburn than pantoprazole 40mg. This was based on the time to sustained resolution of symptoms (defined as a period of seven consecutive days without heartburn). This was in contrast to the assessment of symptoms in Scholten *et al* that assessed time to adequate relief. In Scholten *et al* patients did not have to reach complete resolution of symptoms. Time to sustain a resolution of symptoms as shown by esomeprazole 40mg was much more clinically relevant as it was a period of prolonged improvement in contrast to achieving a period of partial symptomatic relief. Thus, the claims were misleading, did not reflect the available evidence and were in breach of Clauses 7.2, 7.3 and 7.4.

RESPONSE

Altana submitted these claims were derived from a secondary endpoint stated in Scholten *et al*. With demonstration of the primary endpoint (as detailed in point 2 above), secondary endpoints that illustrated a meaningful clinical benefit to patients might be used without misleading the reader. Here a statistically superior and clinically relevant reduction in the time required to achieve pre-defined symptom relief was seen between the products, with Protium being superior to Nexium. No claims of superiority with regards to the primary endpoint had been made.

Altana stated that AstraZeneca's submission that 'as a secondary endpoint, the study would not have been powered appropriately to examine this measure, and was therefore at risk from statistical error' was incorrect. Power was defined as the probability to reject the null hypothesis in the case that a real difference existed. Therefore a statistically significant test result was not influenced by this parameter. In short, the power of Scholten *et al* had no influence on the conclusions drawn from the statistically significant difference seen in this secondary objective.

Altana noted that furthermore AstraZeneca alleged that as the EXPO study showed that esomeprazole 40mg provided faster and more effective resolution of heartburn than pantoprazole 40mg the claims were misleading and did not reflect the available evidence.

Altana submitted that Scholten *et al* focused on the treatment of GERD. Multiple definitions of GERD from wide-ranging parties existed (Vakil *et al* 2006, AstraZeneca website, NICE website). Although the precise definitions varied there was a common consensus that GERD was caused by the reflux of acidic contents from the stomach into the oesophagus leading to a variety of symptoms. Although heartburn was one of the most common symptoms there was growing evidence and consensus that many patients presented with a wide variety of GERD-related symptoms (regurgitation of gastric contents, chest pain, difficulty in

swallowing, wheezing, hoarseness etc) that were clinically significant and meaningful. This was also reflected in a very recent consensus publication, done by some of the leading experts in the field (Vakil *et al*). The approach taken by Scholten *et al* was in line with this and therefore reflected clinical reality. It attempted to gain a wide-ranging measure of GERD symptom relief on PPI therapy. This study looked at adequate symptom relief but did not require complete symptom resolution, reflecting that many patients might have mild intermittent symptoms during therapy but could be dramatically improved from their original symptoms. This was further supported by recent studies in individuals without GERD where it could be shown that they might also experience some mild symptoms that were commonly ascribed to GERD. This led to the introduction of a symptom threshold in contrast to a 'complete' symptom relief concept (Stanghellini *et al* 2005 and Stanghellini *et al* 2006).

Altana submitted that the EXPO study focused on heartburn only in terms of complete symptom control. Heartburn, although a symptom of GERD, did not represent the spectrum of symptoms associated with this disease. The EXPO study was based upon the time to sustained complete resolution of heartburn over a period of seven consecutive days.

Altana submitted that in summary;

- the EXPO study looked at oesophageal erosion healing rates and the absolute resolution of heartburn over time.
- Scholten *et al* studied the reduction in GERD symptom load over time (six different symptoms).

Altana submitted that these studies had thus considered different parameters measured by different methodologies. They could not be considered as similar and could not be compared. The concept as purported by AstraZeneca that the EXPO study might in some way negate or counter the claims made on the findings of Scholten *et al* was illogical on this basis. Altana denied that the claims were in breach of Clauses 7.2, 7.3 and 7.4.

PANEL RULING

The Panel noted that in Scholten *et al* patients recorded the perceived intensity of GERD-related symptoms (heartburn, acid regurgitation, gastric complaints, pressure in the epigastrium, feeling of satiety and flatulence). A five-point Likert scale was used to assess the intensity of each symptom: none (0), mild (1), moderate (2), severe (3) and very severe (4). Each symptom was assessed and scored and if the sum score fell below 5 for the first time, the patient was characterized as having reached adequate relief from GERD-related symptoms. The patients did not have to reach complete symptom relief. The results of the study showed that for daytime, the first time to reach adequate relief of GERD-related symptoms in the pantoprazole group was 3.73 days and 5.88 days for the esomeprazole group ($p=0.034$). This was the result upon which the claims in question were based. The

Panel noted, however, that the claims only referred to 'symptom relief' or 'symptom control', not 'adequate symptom relief control'. In the Panel's view the claims implied total symptom relief/control which was not so. The Panel further noted that the claims did not refer to 'first time' relief and in that regard there was an implication that sustained relief of symptoms was achieved with pantoprazole after 3.7 days. There was no data to show this. In that regard the Panel noted the results of the EXPO study which had shown that time to sustained resolution of heartburn, the most common GERD-related symptom, (defined as a period of seven consecutive days without heartburn) was statistically significantly shorter for patients treated with esomeprazole than for those receiving pantoprazole (6 days versus 8 days; $p < 0.001$).

The Panel thus considered that the claims at issue were misleading and did not reflect the available evidence as alleged. Breaches of Clauses 7.2, 7.3 and 7.4 were ruled.

APPEAL BY ALTANA

Altana appealed the ruling that the claims 'Fast symptom control - 2 days faster than esomeprazole 40mg', 'daytime symptom relief - 2 days faster' and '2 days faster than esomeprazole' were in breach of Clauses 7.2, 7.3 and 7.4.

Altana rejected the Panel's decision that Scholten *et al* and the EXPO study were suitable for direct comparison as they were based upon entirely different study designs, in different populations and with entirely different endpoints.

Altana submitted that as previously stated, the EXPO study looked at oesophageal erosion healing rates and the absolute resolution of heartburn over time. Scholten *et al* looked at the reduction in GERD symptom load over time - six different symptoms typical of GERD including acid regurgitation, gastric complaints, pressure in the epigastrium, feeling of satiety, flatulence and heartburn. Altana submitted the following as further supporting material reflecting the latest thinking in GERD, which made a comparison of these studies misleading in the extreme.

Altana submitted that an understanding of current medical thinking on GERD was vital in considering why the two studies were radically different in design and therefore could not be compared.

These studies considered different medical conditions and used different methodologies. They could not be considered as studying the same endpoint and thus could not be directly compared. Indeed the area under the curve (AUC) symptom load table (Scholten *et al*) illustrated that in endoscopically proven GERD, heartburn contributed less than 25% of the symptom load during the study.

Amongst others the Montreal Definition and Classification of Gastroesophageal Reflux Disease published in 2006 (supported by AstraZeneca) confirmed that GERD was considered to be a disease with a wide

range of both oesophageal and extra-oesophageal symptoms not just a disease of heartburn. Modlin *et al* (2007) (in press) reiterated the movement away from studying heartburn as a single symptom of GERD and the importance of considering the broad range of oesophageal and extra-oesophageal symptoms that patients experienced.

Altana submitted that the design of Scholten *et al* reflected this modern clinical interpretation of GERD. It looked for improvement in a range of six GERD related symptoms and did not focus entirely on heartburn. It defined a successful clinical outcome as a reduction in total symptom score to below a pre-defined level. This did not require complete symptom resolution.

Altana submitted that Stanghellini *et al* (2005 and 2006) discussed this concept of GERD symptom reduction to a lower threshold but not to zero. Individuals without evidence of GERD experienced low levels of symptoms commonly ascribed to GERD. The background incidence of GERD-type symptoms in a healthy population was not zero although a few individuals within the broader population might experience zero symptoms. This had been confirmed by two clinical studies with more than 1500 healthy volunteers. Stanghellini *et al* (2005) (national German study) eligible for analysis, $n=385$ and Stanghellini *et al* (2006) (international study) eligible for analysis, $n=1,167$.

Altana submitted that therefore, it followed that a study designed to illustrate complete symptom resolution (zero symptoms) in GERD would expect to fail. Thus at best one might hope to reduce the symptoms of GERD within a study population to reach the expected background incidence. However a pre-determined clinically meaningful benefit might be defined. This benefit would reduce the burden of symptoms to a clinically relevant threshold above the background level. This was what Scholten *et al* achieved.

Altana submitted that however, it was possible to achieve complete resolution of heartburn, as illustrated by the EXPO study, if only heartburn was considered.

Altana submitted that thus what was claimed to be 'complete symptom resolution' (zero heartburn) seen with the EXPO study could not be logically compared with the symptom load reduction seen in Scholten *et al*, which because of the applied threshold concept could never achieve complete symptom resolution. The study designs logically did not allow for comparison. Indeed the claim of complete symptom resolution made for the EXPO study was in itself misleading.

Altana thus disagreed with the Panel's ruling that the terms 'symptom control' and 'symptom relief' were misleading. For studies looking at symptom load reduction in GERD these phrases were entirely appropriate - symptom control/relief could not reach zero for the reasons stated above.

Furthermore Altana contested the Panel's assertion that 'there was an implication that sustained relief of symptoms was achieved with pantoprazole after 3.7 days'.

Altana submitted that an understanding of modern GERD clinical study design should have invalidated AstraZeneca's claim in its complaint that 'Time to sustain a resolution of symptoms as shown by esomeprazole 40mg was much more clinically relevant as it was a period of prolonged improvement in contrast to a achieving a period of partial symptom control'. AstraZeneca was factually incorrect as the EXPO study measured treatment of heartburn not resolution of symptoms as previously shown.

Altana concluded that Scholten *et al* represented the more modern methodology and more clinically relevant interpretation of GERD, assessing the broad spectrum of GERD symptoms. It could not be compared with older methodologies, such as the EXPO study measuring heartburn only. To this end the assertions in the complaint should carry no weight with the Panel nor influence the interpretation of Altana's claims, which should be viewed in isolation from any argument derived from the non-comparable EXPO study.

Altana submitted that its claims only referred to the time of onset of symptom relief in the Scholten *et al* head-to-head comparator study measuring GERD symptom load. A statistically significant difference between the two products was seen for this parameter in favour of Protium. This was stated. There was no claim of prolonged relief. The claims were entirely in line with the time to event analysis used to determine this outcome and suitably referenced.

COMMENTS FROM ASTRAZENECA

AstraZeneca noted that Scholten *et al*, a direct comparison study, evaluated the primary outcomes (AUCs for GERD symptom scores) between esomeprazole 40mg and pantoprazole 40mg. As stated in the results section there was no statistical difference ($p>0.05$) between the two treatment groups, ie the study did not meet its primary endpoint and was thus inconclusive.

The claims at issue 'Faster symptom control - 2 days faster than esomeprazole 40mg', 'daytime symptom relief - 2 days faster' and '2 days faster than esomeprazole' related to the secondary end points of Scholten *et al*. AstraZeneca alleged that as this study did not meet its primary endpoint it was not appropriate to use secondary endpoints that were inconsistent with the primary outcome of the study. This point was addressed in the European Medicines Evaluation Agency guidance.

AstraZeneca alleged that differences that did not reach statistical significance must not be presented in such a way as to mislead. Non-significant p values across the primary parameters equated with the negative results in the study irrespective of the results from secondary parameters. Secondary endpoints could not be used to 'salvage' an otherwise non-supported study. Results from secondary parameters might suggest new parameters that need to be explored as primary outcomes in a trial.

AstraZeneca therefore alleged these claims to be misleading, as the use of the secondary endpoints alone would lead the reader to draw a different conclusion if they were unaware of the primary outcome of the study. In addition there was no indication what type of symptoms were controlled/ improved and that partial symptom resolution was needed to be achieved in the study. These matters were addressed in the Panel's rulings.

AstraZeneca alleged furthermore, that the '2 day difference' was based on calculating the mean, which was a biased estimate for Kaplan-Meier analysis due to censored observations. The standard summary statistic should be the median, which was two days for both treatment groups.

In addressing the issue raised by Altana relating to a broader definition of GERD-related symptoms' which also included gastric complaints, feeling of satiety, and flatulence, AstraZeneca was concerned that these were not generally accepted as specifically related to GERD. The most important and predominant symptoms were heartburn and acid regurgitation as discussed in the Montreal definition (Vakil *et al*). In Scholten *et al* these symptoms were experienced by 77% of the patients.

AstraZeneca alleged that utilizing a much broader spectrum of GERD symptoms, that included elements of irritable bowel syndrome, raised uncertainty as an improvement in a patient's overall symptom score (eg driven by improvements in symptoms such as flatulence) could mask deterioration in a more troublesome symptom such as heartburn. The EXPO study showed that esomeprazole 40mg provided faster resolution of heartburn than pantoprazole 40mg. This was based on the time to sustained resolution of heartburn (defined as a period of seven consecutive days without heartburn). This was also addressed in the Panel's rulings.

APPEAL BOARD RULING

The Appeal Board noted the claims at issue relied upon secondary end point data from Scholten *et al*, a study which had failed to show a statistically significant difference between Protium and Nexium with regard to the primary endpoint. The failure to satisfy the primary end point was not made clear in the material. In such circumstances the Appeal Board considered that it was unacceptable to use secondary data to claim an advantage for Protium over Nexium when the primary endpoint had been unable to show such a difference. The Appeal Board considered that the claims were misleading in this regard and did not reflect the available evidence as alleged. The Appeal Board upheld the Panel's rulings of breaches of Clauses 7.2, 7.3 and 7.4. The appeal on this point was unsuccessful.

Complaint received	4 January 2007
Case completed	8 June 2007