Intravenous proton pump inhibitors for stress ulcer prophylaxis in critically ill patients: determining statistical equivalence according to evidence-based methods

Andrea Messori¹, Valeria Fadda³, Dario Maratea³, Roberta Gatto¹, Sabrina Trippoli¹, Mauro De Rosa³, and Claudio Marinai²

¹HTA Unit, ²Department of Pharmaceutical Logistics, ESTAV Toscana Centro, Regional Health Service, Florence, and ³The BW Project, Sifact, Società Italiana di Farmacia Clinica e Terapia, Milan, Italy

Abstract. Background: Although intravenous proton pump inhibitors (PPIs) are considered at least as effective as H2-receptors antagonists for stress ulcer prophylaxis (SUP) in critically ill patients, there is no data on whether there is also the proof of no difference among these agents. Methods: The clinical material was the same as that reported in previous meta-analyses and included all trials comparing intravenous PPIs vs. H2-receptor antagonists for SUP in critically ill patients. Our methodology was a combination of meta-analysis and equivalence testing based on confidence intervals (CIs). The end-point was the rate of overt bleeding. All PPIs evaluated in the included trials were separately studied. The equivalence margins were derived from power calculation data of the original trials. Results: Our analysis involved 8 randomized trials for 851 patients. Two comparisons were made (pantoprazole vs H2-receptor antagonists and omeprazole vs H2-receptor antagonists). The following RDs were estimated: pantoprazole, RD = –1.2%, 95% CI: –3.5% to +1.2%; omeprazole, RD = –3.0%, 95% CI: –7.2% to +1.3%. The 95% CIs confidence intervals for RDs remained within the ± 6% margin. These results indicate that intravenous pantoprazole and intravenous omeprazole are equivalent. Conclusion: These two PPIs, when administered by intravenous route, are equivalent according to reasonable equivalence margins. This conclusion can be the basis to develop local acquisition tenders on these drugs. One advantage of this approach is that the feasibility of administrative decisions can directly be tested on clinical grounds and on the basis of standard evidence-based methods.

Introduction

In interpreting randomized trials, the differentiation between no proof of difference (an inconclusive result) and proof of no difference (demonstration of equivalence) is attracting an increasing interest in both theoretical [1, 2] and practical terms[3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. This differentiation can be made using trial-sequential analysis [3, 4, 5, 6, 7, 8, 9, 10, 11] or by application of statistical testings based on confidence intervals (CIs) [12, 13, 14, 15, 16]. Both of these approaches require that a margin (or delta) is pre-specified to define the minimal improvement in the end-point that is thought to be clinically relevant.

In the Italian national health system (NHS), local tenders have long represented the standard tool for medicines procurement. In December 2012, however, a national regulation (“Decreto Balduzzi” [17]) has been issued to accelerate, across different regions, the patients’ access to reimbursed medicines of our National Formulary. In practice, however, this regulation turned to be an obstacle for local tenderings run by the NHS; in this way, in fact, all tenderings involving “homogeneous” drug classes have been mandatorily subjected to a preventive technical authorization by our national Agency for Medicines. Hence, AIFA has been assigned the responsibility of declaring which drugs classes are “equivalent” and can therefore be managed through local tenderings involving different agents of the same pharmacological class.

This new regulation is a tough challenge in operational terms, but has also attracted the attention of evidence-based researchers because specific criteria are being sought for these demonstrations of equivalence. Initial
experiences have been published in which this demonstration of equivalence has been managed by a combined application of meta-analysis and equivalence testing [12, 13, 14, 15, 16, 17, 18]. In this framework, one critical point is the retrieval of the most appropriate margin for the specific analyses; this question has found its most obvious solution in the retrieval of Δ-values from the sample size calculations reported in the trials.

We present herein a further experience on the use of equivalence testing with reference to proton pump inhibitors (PPIs) administered by intravenous route.

Methods

The clinical material for our analysis was derived from published randomized trials comparing PPIs vs H2-receptor antagonists. Only intravenous agents were examined. A single clinical indication (stress ulcer prophylaxis in critically ill patients) was analyzed; bleeding during hospital stay was the clinical end-point. The indication of acute peptic ulcer treated with intravenous PPIs has already been evaluated in a separate study [18] and therefore was not considered in the present study.

The methods of our equivalence analysis were the same as previously described [12, 13, 14, 15, 16, 17, 18]. Briefly, the clinical material was obtained from the most recent and most comprehensive systematic reviews in this field. The literature search, based on PubMed, covered the last 5 years; only reviews, systematic reviews, and meta-analyses (according to PubMed definitions) were eligible for being the source of our clinical material. Studies employing an oral suspension administered by nasogastric tube were not eligible. A single search term (“stress ulcer”) was used. The “best” paper(s) for the purposes of our analysis were identified through the following procedure. Firstly, we identified all reviews/meta-analyses that were considered eligible for our purposes and we retrieved their full text. To identify the best article(s) from this subset of select publications, we adopted an evaluation form in which, for each article, four items were rated on a 0 to 5 scale: [a] number of clinical trials suitable for our analysis; [b] impact factor of the journal in which the article was published; [c] degree of literature update; [d] description of clinical trials not already reported in articles assigned a better rating for items [a] through [c] than the present one. Four authors (AM, VF, DM, ST) were involved in this process of literature selection and evaluation.

The meta-analytic values for the comparison of PPIs vs H2-receptor antagonists were directly obtained from the meta-analysis articles. However, since our analysis adopted risk difference (RD) as outcome measure, the outcomes expressed differently (e.g., as odd-ratio or risk ratio) were re-computed from the raw data and then expressed as RD. Meta-analytical values of RD were computed according to a standard random-effect model as implemented by the OMA Software (Open Meta-Analyst version 4.16.12, Tufts University, U.S., url http://tuftscaes.org/open_meta/).

As regards the statistical section of our study, the margins for equivalence testing were retrieved by examining all original trials selected as our source of clinical information. All CIs were estimated at the statistical level of 95%. Equivalence testing was based on these CIs. The α-level was at 2.5%.

Results

Our literature search (last query run on 1 February 2014) firstly identified a total of 89 citations. After reading the abstract of all of these articles and the full text of a subset of 39 studies among these 89 citations, we identified the two articles by Alhazzani et al. [18] and by Reveiz et al. [19], respectively, as the best source of clinical data for our analysis. Overall, this clinical material consisted of 8 randomized trials (3 for pantoprazole and 5 for omeprazole; 593 and 258 patients, respectively). Only 1 of these 8 studies involved pediatric patients.

Our supplementary material describes how we re-analysed the raw data from these randomized studies in order to generate our meta-analytic values of RD. These values of RD were: pantoprazole, −1.2% (95% CI: −3.5% to +1.2%); omeprazole, −3.0% (95% CI: −7.2% to +1.3%). The overall meta-ana-
The proof of no difference by equivalence testing, in most cases generates, the same conclusions as those reached in standard narrative reviews. Equivalence testing has however several advantages. Firstly, one point of strength of the quantitative approach lies in the standardized nature of the evidence-based methodology. Another advantage is that, in this way, the feasibility of administrative and/or regulatory decisions can directly be tested on clinical grounds. More importantly, a differentiation can be made between “no proof of difference” and “proof of no difference”. The main limitation is that the role of margins is still fraught with some uncertainties; for example, it is often postulated that equivalence and non-inferiority margins differ from superiority ones, but this assumption does not seem to have any specific basis [23]. In fact, margins are widely recognized to intrinsically have a certain degree of arbitrariness.

In the case of this specific analysis, one point of controversy might be the inclusion of one clinical trial carried out in pediatric patients; regardless, if this study is excluded the results remain essentially unchanged (data not shown).

Finally it should be noted that no evidence from randomized trials was retrieved for other intravenous PPIs (e.g., esomeprazole). No conclusion can therefore be made on whether esomeprazole is equivalent to pantoprazole or omeprazole.

Since our analysis will be submitted to AIFA with the perspective of running thereafter a regional tendering (at least in the Toscana region), the present experience will hopefully confirm the feasibility of managing a practical problem of medicine procurement by application of evidence-methods. Finally, the methodology presented herein can also be a model for the decisions on equivalence that our Agency will make in this class and in other pharmacological classes.

**Discussion**

![Equivalence testing based on meta-analytical values of RD in comparisons between PPIs and H2-receptor antagonists. The equivalence test is based on the area comprised between the two vertical dotted lines, that reflect the pre-determined equivalence margins (from -7.6% to +4.4%). Each horizontal bar indicates the two-sided 95% CI for the RDs (solid square) for the comparisons of pantoprazole vs H2-receptor antagonists (in green) and omeprazole vs H2-receptor antagonists (in blue). The margins were centered around the meta-analytical RD across all studies at -1.6% (vertical dashed line). The criterion for demonstrating equivalence is when both extremes of the 95% CI remain within the two vertical dotted lines. All values of RD are expressed as percentages. Negative values of RD favor the PPI and positive ones the H2-receptor antagonists. The complete bibliographies for the 8 randomized trials (3 for pantoprazole and 5 for omeprazole) can be found in references [20] and [21]. RD = risk difference.](image)

Our equivalence testing generated the results shown in Figure 1. As indicated by the CIs reported in the figure, our results demonstrated the equivalence between the two PPIs.

**Conflict of interests**

Although Andrea Messori is a member of a national Committee of AIFA (“Segretariato Ufficio Prezzi e Rimborsi”), the views pre-
sented in this paper do not necessarily represent those of the agency; Mauro de Rosa is the President of SIFACT (Milano, Italy); there are no conflict of interests for the other authors.

References


[23] Norman G, Monteiro S, Salama S. Sample size calculations: should the emperor’s clothes be off the peg or made to measure? BMJ. 2012; 345 (aug23 1): e5278.
Supplemental figure 1. Meta-analysis based on risk difference. Our meta-analysis based on risk difference (RD) re-examined the same information previously reported by Alhazzani et al. and by Reveiz et al. (references [18] and [19] of our published paper, respectively). The graph in Figure S1 shows the Forest plot with the values of RD for individual trials (■ with 95%CIs indicated by horizontal bars) and for the pooled analysis (diamond in blue and vertical dotted line in red). The pooled rates for omeprazole and pantoprazole are shown as yellow diamonds. All data are presented as absolute rates on a scale 0 to 1 (with 95%CI); I^2 is a measure of heterogeneity. Statistical calculations were performed by the OMA software (Open Meta-Analyzer version 4.16.12, Tufts University, U.S., url http://tuftscaes.org/open_meta/). Ev = number of events; Trt = number of patients receiving treatment. All trials were retrieved from the meta-analysis by Alhazzani et al. with the trial of Yildizdas et al. which was obtained from the meta-analysis by Reveiz et al.