

REVIEW

Endoscopic ulcers as a surrogate marker of NSAID-induced mucosal damage

R Andrew Moore*

Abstract

The characteristic of a biomarker that makes it a useful surrogate is the ability to identify a high risk of clinically important benefits or harms occurring in the future. A number of definitions or descriptions of surrogate definition have been put forward. Most recently the Institute of Medicine of the National Academy of Sciences in the USA has put forward an evaluation scheme for biomarkers, looking at validation (assay performance), qualification (assessment of evidence), and utilisation (the context in which the surrogate is to be used). This paper examines the example of endoscopy as a surrogate marker of NSAID-induced mucosal damage using the Institute of Medicine criteria. The article finds extensive evidence that the detection of endoscopic ulcers is a valid marker. The process of qualification documents abundant evidence showing that endoscopic ulcers and serious upper gastrointestinal damage are influenced in the same direction and much the same magnitude by a variety of risk factors and interventions. Criticisms of validation and qualification for endoscopic ulcers have been examined, and dismissed. Context is the key, and in the context of serious NSAID-induced upper gastrointestinal harm, endoscopic ulcers represent a useful surrogate. Generalisability beyond this context is not considered.

Introduction

Establishing a biomarker as an effective surrogate, something measurable now but indicative of some later important clinical event, is both important and difficult. The often-quoted ideal of a surrogate marker is the blood cholesterol level. We know that if this level is elevated,

there is an increased risk of future serious cardiovascular harm, including death, and that reducing cholesterol levels reduces that risk of serious harm. For example, the Scandinavian Simvastatin Survival Study randomised patients with clinically established coronary heart disease to 5 years of simvastatin or placebo [1]. The statin produced significant and large reductions in blood lipids, as well as a 24% reduction in coronary mortality over 10 years, but with a number-needed-to-treat of about 50, meaning that there were 2% fewer coronary deaths when a statin was used [2].

In 2001 a National Institutes of Health working group defined surrogacy based on prediction of a more serious outcome from epidemiologic, therapeutic, pathophysiologic or other scientific evidence [3]:

- A biological marker (biomarker) is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention.
- A clinical endpoint is a characteristic or variable that reflects how a patient feels, functions or survives.
- A surrogate endpoint is a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm, or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic or other scientific evidence.

A more recent publication by the Institute of Medicine of the National Academy of Sciences has examined how biomarkers may be evaluated for effective use in chronic disease [4]. This document argues that the biomarker evaluation process should consist of three steps:

1. Analytical validation: analyses of available evidence on the analytical performance of an assay.
2. Qualification: assessment of available evidence on associations between the biomarker and disease states, including data showing effects of interventions on both the biomarker and clinical outcomes.
3. Utilisation: contextual analysis based on the specific use proposed and the applicability of available evidence to this use. This analysis includes a determination of whether the validation and qualification conducted provide sufficient support for the use proposed.

*Correspondence: andrew.moore@ndcn.ox.ac.uk
Pain Research, Nuffield Division of Anaesthetics, Nuffield Department of Clinical Neuroscience, University of Oxford, The Churchill, Oxford OX3 7LE, UK

The document is mainly aimed at determining how the US Food and Drug Administration should consider evidence around the use of biomarkers or surrogates. The publication has tested the methodology on several possible surrogates: tumour size for cancer, C-reactive protein, troponin, β -carotene, and low-density and high-density lipoprotein for cardiovascular risk. The message from each case study is that context is the key; utility for one purpose may well not mean utility for all.

For low-density lipoprotein cholesterol, 'the high probability that lowering LDL [low-density lipoprotein] for several interventions decreases risk of cardiovascular disease, and LDL, although not perfect, is one of the best biomarkers for cardiovascular disease' [4]. What is important here is that the panel considered low-density lipoprotein cholesterol a useful surrogate despite the fact that beneficial changes may occur for the surrogate in most patients, but benefits in terms of clinical events occur only for a few.

Determining the true value of a surrogate is not easy. Some approaches have taken a distinctly statistical approach [5,6]. Other studies are more philosophical [7-9]. The message that context is the key can be read into all of the various approaches to defining what is a surrogate, and how to evaluate whether a putative surrogate really is one.

The fact that evaluation is needed can be seen from a short search of the literature demonstrating the extent of interest in surrogate endpoints. Of the (almost) 3,600 papers with surrogate in the title found using PubMed, fewer than 100 also mention validation or validity in the title or abstract. Those papers that did examine the validity of potential surrogate markers or markers actually used as surrogates often found the evidence lacking. For example, surrogate endpoints used in liver surgery trials were generally not validated [10], and a simple walking test used in hypertension trials did not explain the treatment effect [11].

Endoscopy for nonsteroidal anti-inflammatory drug-induced mucosal damage

A wide-ranging systematic review evaluated the evidence that endoscopic ulcer may be a useful surrogate for more serious clinical harm from NSAIDs, and concluded that it was a strong candidate [12]. Other researchers disagreed [13]. This paper revisits the evidence in light of the Institute of Medicine report [4].

It is widely believed that there is a biological progression from lesser to more severe gastrointestinal damage with NSAIDs: from dyspepsia and other gastrointestinal symptoms, through endoscopic erosion and asymptomatic ulcers detected endoscopically, to ulcer complications (bleeding and perforation), and even to death [14,15]. Asymptomatic ulcers can also bleed.

Endoscopic ulcers may be an early step in a biological progression from mucosal injury to symptomatic ulcer and ulcer complication. These complications include the following:

- Obstruction complicating peptic ulcers: this is a function of the ulcer's anatomical location in which lesions involving the pylorus are more likely to present with obstruction than those in the gastric corpus.
- Perforation complicating peptic ulcers: like obstruction, this also depends on anatomical location; most perforating ulcers occur in the duodenum.
- Bleeding: this is less predictable, but is increasingly seen in association with anti-thrombotic agents.
- Death: with improved resuscitation and endoscopic therapy techniques, this complication largely depends on comorbidity. But while mortality from upper gastrointestinal bleeding has fallen substantially over recent years, bleeding associated with NSAIDs retains a higher mortality, above 10% [16].

The Institute of Medicine process for valuating a surrogate marker

The three parts of the process involve validation, qualification, and utilisation. There are broad general requirements in each of these three sections; different candidate markers will have different characteristics, but the aims can briefly be stated as follows:

Analytical validation

Analytical validation is as an assessment of assays for the biomarker and their measurement performance characteristics, determining the range of conditions under which the assays will give reproducible and accurate data. Put simply for our purposes, is upper gastrointestinal endoscopy an accurate and reliable test for the development of serious upper gastrointestinal events?

Qualification

Qualification is a factual description of the levels and types of available evidence. The aim is an objective analysis based on a reproducible, systematic assembly, and review of the evidence, which should include evidence of whether the biomarker is on a causal pathway in the disease pathogenesis and that interventions targeting the biomarker in question impact the clinical endpoints of interest. If the biomarker-clinical endpoint relationship persists over multiple interventions, it is thought to be more generalisable. The analysis should include addressing some or most of the elements of criteria for causation outlined by Hill [17]:

1. Strength of association: large relative risk, or odds ratio.
2. Consistency: relationship seen in different populations or circumstances.

3. Specificity: exposure causes only specific effect.
4. Temporal relationship: exposure precedes the event.
5. Biological gradient: a dose–response relationship.
6. Plausibility: biological plausibility.
7. Coherence: the cause-and-effect interpretation of data should not seriously conflict with the generally known facts of the natural history and biology of the disease.
8. Experiment: does removing the exposure lessen the effect?
9. Analogy: comparison between weaker and stronger evidence, or strong evidence of causality between another exposure and similar effect.

Utilisation

Utilisation is a contextual analysis of the available evidence about a biomarker with regard to the proposed use of the biomarker. This part considers how the surrogate marker will be used in very specific circumstances. If the circumstances change, so might the evaluation of the biomarker. In other words, a useful surrogate marker in one circumstance may not be a useful surrogate in another. Generalisation has always to be justified.

Evaluating endoscopic ulcers as a surrogate marker of nonsteroidal anti-inflammatory drug-induced mucosal damage

Analytical validation

Perhaps because endoscopy is a simple test, with the operator seeing a lesion, these methods have not been subjected to the same intense scrutiny that might have accompanied a new blood test, for instance. The common definition used has been that an endoscopic ulcer has to be a gastric or duodenal lesion ≥ 3 mm (sometimes ≥ 5 mm) with significant depth, although the depth is not defined.

Commentators have cast doubt on the reliability of methods for detection of endoscopic ulcers [13,18], in part because of the paucity of data demonstrating inter-observer accuracy and precision. Studies on gastroduodenal ulcer scars examined endoscopically and reporting differences between operators might be seen as supporting interobserver disagreement as a problem [19]. There are two main lines of criticism – that experienced endoscopists disagree about whether endoscopic ulcers are real ulcers, and that there has been a shift in prevalence of endoscopic ulcers over time because of a lack of training and consistency between endoscopists.

The first criticism [13] comes from a short abstract describing three experienced endoscopists viewing a training tape in a blinded fashion [20]. There was a 100% agreement with obvious ulcers and trivial lesions. The endoscopists considered that only one-third of endoscopic ulcers (≥ 3 mm in greatest dimension, with depth) were actual ulcers. That is a fair criticism at one level, but

in a sense it misses the point – the definition of an endoscopic ulcer for the purposes of acting as a surrogate need not be the same as that of a real ulcer.

The second criticism was that, in a systematic review of endoscopic ulcers in placebo arms of NSAID trials [21], early studies had no endoscopic ulcers while later studies had a significant rate of endoscopic ulcers, pointing to a failure of training [18]. The authors of the review have made their own eloquent defence [22], but there are other powerful arguments against the criticism.

The main defence lies in the difference of two populations given placebo. One group consisted of healthy subjects, mostly in small, older short studies of 1 or 2 weeks' duration (all ≤ 4 weeks), and mostly, although not exclusively, using Lanza scoring of various sorts. The other group consisted of patients with osteoarthritis or rheumatoid arthritis, in much larger studies lasting 6 or 12 weeks, and using a variety of scoring systems to identify lesions ≥ 3 mm with unequivocal depth. The results can be seen in Figure 1, which redraws data from the systematic review and adds data from two more recent studies [23,24]. For healthy subjects, most studies (including most of those performed since 2000) record endoscopic ulcer prevalence rates $\leq 2\%$. For patients with arthritis, most studies (all since 1998) report prevalence rates $\geq 2\%$.

The picture is one of a low prevalence of 1.1% in 559 healthy subjects given placebo in short-duration studies, and a 3.8% prevalence in 2,368 patients with osteoarthritis or rheumatoid arthritis given placebo in long-duration studies – typically six times longer than those for healthy subjects. The variability in Figure 1 is a reflection of chance effects in small populations [25], and given the small size of many of these patient groups the picture is one of consistency. The result for healthy subjects is similar to the 1% recorded in 619 healthy controls in northern Norway more than 20 years ago [26].

Consistently higher incidence rates of endoscopic ulcers are seen with both coxibs (mean rate 5.1% in 4,691 patients) and a variety of different NSAIDs (mean rate 17% in 3,915 patients) in individual treatment arms from clinical trials of celecoxib [27], valdecoxib [28], rofecoxib [29-31], and etoricoxib [32]. Figure 2 shows the individual studies, using different scales to Figure 1, and Figure 3 shows a pooled analysis of event rates with NSAID, coxib, and placebo both with and without the presence of low-dose aspirin.

Another general criticism of endoscopic ulcer measurement relates to the usual choice of definition: an ulcer ≥ 3 mm with unequivocal depth. The argument is that a 3 mm ulcer is trivial, and depth may be an uncertain quantity for such a small ulcer. Fortunately at least three studies of endoscopic ulcers have reported results using both the ≥ 3 mm and the ≥ 5 mm definitions [29,31,32].

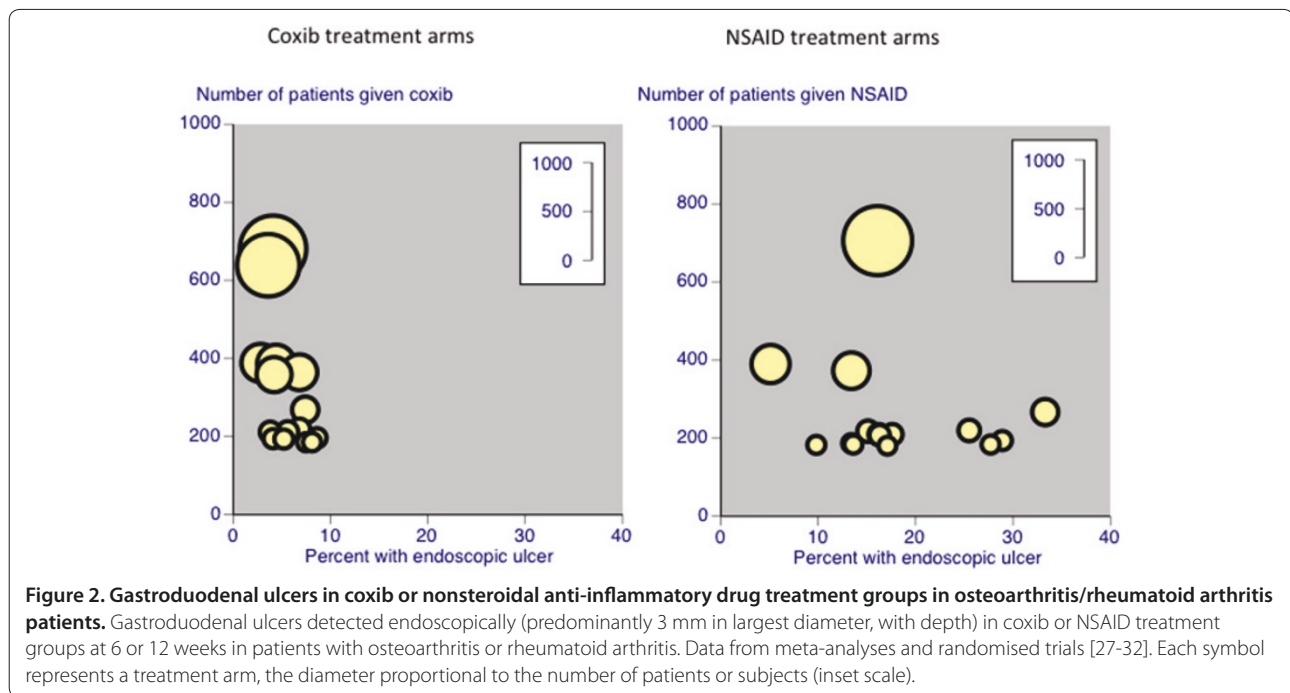
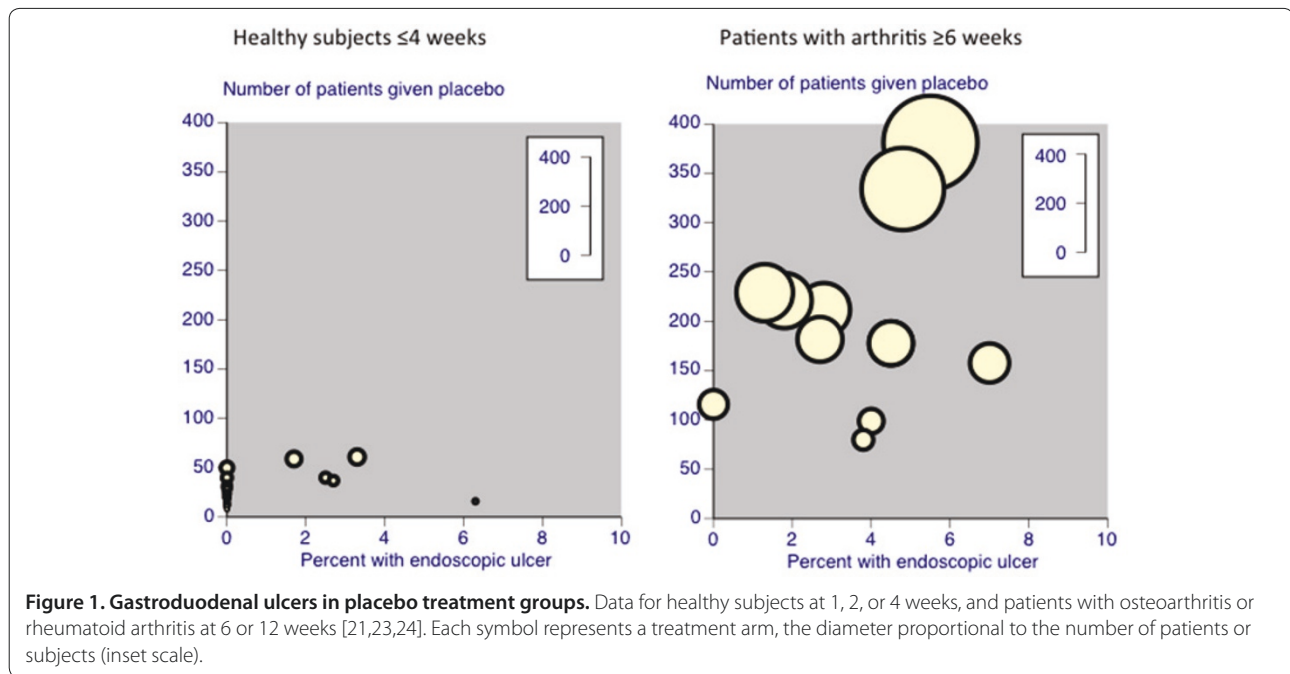


Figure 4 shows that the results for ≥ 5 mm were very close to those for ≥ 3 mm, and on average about 70% of endoscopic ulcers were ≥ 5 mm. The larger size gives the ulcers much greater relevance.

Despite a lack of formal tests for accuracy and precision for the endoscopic detection of ulcers, the weight of evidence from large amounts of data is that there is no cause for concern about the test.

Qualification

A review examining endoscopic ulcers as a surrogate endpoint for bleeding ulcers has collated the evidence for qualification, concluding that those factors having an effect on serious gastrointestinal harm (mainly bleeding ulcers) can affect endoscopic ulcers in the same direction and to much the same extent [12]. Table 1 summarises the consistent effects of various risk factors (age, previous

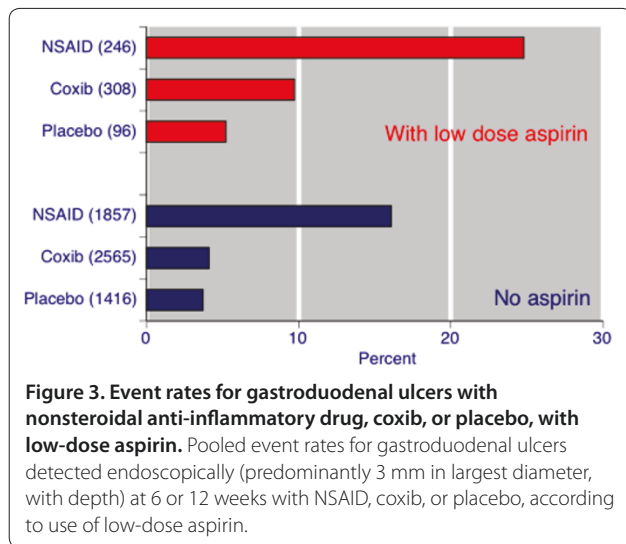


Figure 3. Event rates for gastroduodenal ulcers with nonsteroidal anti-inflammatory drug, coxib, or placebo, with low-dose aspirin. Pooled event rates for gastroduodenal ulcers detected endoscopically (predominantly 3 mm in largest diameter, with depth) at 6 or 12 weeks with NSAID, coxib, or placebo, according to use of low-dose aspirin.

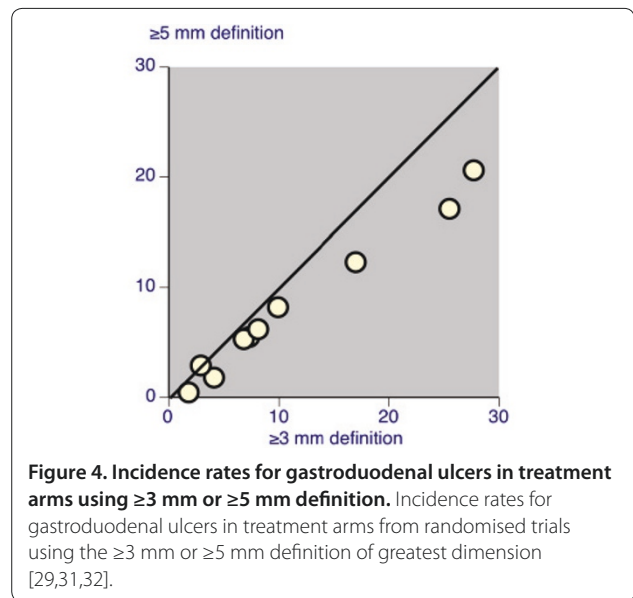


Figure 4. Incidence rates for gastroduodenal ulcers in treatment arms using ≥ 3 mm or ≥ 5 mm definition. Incidence rates for gastroduodenal ulcers in treatment arms from randomised trials using the ≥ 3 mm or ≥ 5 mm definition of greatest dimension [29,31,32].

ulcer or bleeding ulcer history, *Helicobacter pylori* infection, and the effects of aspirin alone or with NSAIDs) and ulcer prevention strategies with NSAIDs (misoprostol, histamine-2 receptor antagonists, proton pump inhibitors, and coxib substitution).

For each of these factors, the evidence from randomised controlled trials, from meta-analyses of randomised controlled trials, and from observational studies demonstrates effects in the same direction and of similar magnitude; details are not provided here but are available from the original review [12]. Often there are several studies that provide information; and for the larger, better, studies, consistency of findings is an important feature. An example is the similar findings in two meta-analyses of epidemiological studies of the association between NSAIDs and upper gastrointestinal bleeding, one being conducted using studies in the 1990s [33] and the other using studies between 2000 and 2008 [34]. Table 2 shows the large magnitude of the increased risk (about fourfold overall) and similarity in the results.

Endoscopic ulcers and actual bleeding events can occasionally be measured together, as in one of a series of randomised trials from Hong Kong [35-38]. These trials were performed in similar groups of older patients (mean ages 64 to 70 years) with an endoscopically proven healed upper gastrointestinal bleeding event but who needed to continue taking NSAIDs for pain relief. Participants were randomised into different treatment groups that included *H. pylori* eradication, NSAID plus proton pump inhibitor, celecoxib, or celecoxib plus proton pump inhibitor (Figure 5). The primary outcome was usually recurrent gastrointestinal bleeding over 6 or 12 months.

Figure 5 shows consistency in 6-month incidence rates in this population of patients in different studies; three studies had consistent gastrointestinal bleeding rates of 4

to 5% for celecoxib. Recurrent bleeding rates for other therapies varied from 19% for naproxen in the absence of any effective gastroprotective strategy to about 6% for diclofenac plus omeprazole, and 0% for celecoxib plus omeprazole. This observation is interesting, of course, because it shows how different strategies influence potentially serious harm.

This observation is also interesting because one of these trials measured both bleeding events and endoscopic ulcers in the same study [37]. The definition of bleeding was prespecified. The endoscopic evaluations were carefully done: a single operator performed all endoscopic examinations in a treatment-blinded fashion to avoid between-observer variation. An ulcer was defined as a circumscribed mucosal break ≥ 5 mm in diameter with a perceptible depth. With diclofenac plus omeprazole, endoscopic ulcer incidence was 1.4 times higher than for celecoxib; the incidence of recurrent bleeds was 1.3 times higher. This again argues for consistent effects of different therapies on both endoscopic ulcers and gastrointestinal bleeding events. This study is important for three key reasons:

- The study is the only one in which both the clinical outcome and surrogate marker were measured together.
- Upper gastrointestinal bleeding was determined against prespecified criteria.
- A single operator determined the presence of gastroduodenal ulcers using a larger size than is the norm, meaning that these endoscopic ulcers were not trivial.

In populations where risk of bleeding is lower, the number of bleeding events is so small that it is impractical to measure them in the same study. High-risk populations

Table 1. Summary of the consistent effects of various risk factors and ulcer prevention strategies with NSAIDs

Influence on outcomes	Effect on endoscopic ulcers	Effect on upper gastrointestinal bleeding
Risk factors		
Age	RCTs show incidence of ulcers to increase with age, more than threefold over five decades	Patients over 75 years old, 2.5-fold increased risk of upper gastrointestinal complications
History of previous ulcer or bleeding	Previous history increases risk of ulcer fourfold in RCTs	Previous history increases risk of bleeding fivefold in RCTs; observational studies support this
<i>Helicobacter pylori</i>	Meta-analysis of RCTs shows 60% decrease in ulcers	Meta-analysis of RCTs shows 80% decrease in ulcers
Aspirin alone	Dose-related increase in ulcers over range of 81 to 325 mg daily	Low-dose aspirin is associated with increased risk of bleeding events
Aspirin plus coxibs or NSAIDs	Low-dose aspirin increases ulcer rates with placebo, coxib, and NSAID in RCTs	Low-dose aspirin increases bleeding rates when added to coxibs or NSAIDs
Ulcer prevention strategies used with NSAIDs		
Misoprostol	Misoprostol reduced ulcers by 70% in a meta-analysis of RCTs	Misoprostol reduced bleeds by 40% in a meta-analysis of RCTs
Histamine-2 receptor antagonists	Histamine-2 receptor antagonist therapy reduced ulcers by 60% in a meta-analysis of RCTs	Histamine-2 receptor antagonist therapy reduced bleeds by 30 to 40% in two observational studies
Proton pump inhibitors	Proton pump inhibitor therapy reduced ulcers by 60% in a meta-analysis of RCTs	Observational studies support reduced risk of upper gastrointestinal complications and bleeding with proton pump inhibitor
Coxib use	Pooled analysis across RCTs indicates that coxibs reduce ulcers by about 70% compared with NSAIDs	Pooled analysis across RCTs indicates about a 40 to 50% reduction in ulcer complications with coxibs; observational studies show a consistent 50% reduction or more in upper gastrointestinal bleeding events

NSAID, nonsteroidal anti-inflammatory drug; RCT, randomised controlled trial.

Table 2. Increased risk of upper gastrointestinal bleeding with NSAIDs in two meta-analyses of observational studies

Drug	Relative risk or odds ratio	
	Hernández-Díaz and Rodríguez [33] (1990s, ≥80,000 patients)	Massó González and colleagues [34] (2000 to 2008, ≥40,000 patients)
Ibuprofen ≤2400 mg	2.1 (1.6 to 2.7)	2.7 (2.4 to 3.0)
Diclofenac ≤100 mg	3.1 (2.0 to 4.7)	4.0 (3.5 to 4.4)
Naproxen ≤1,000 mg	3.5 (2.8 to 4.3)	5.2 (4.3 to 6.2)
Piroxicam ≤20 mg	5.6 (4.7 to 6.7)	9.3 (7.5 to 11)
Current NSAID use	4.2 (3.9 to 4.6)	4.6 (4.3 to 4.9)

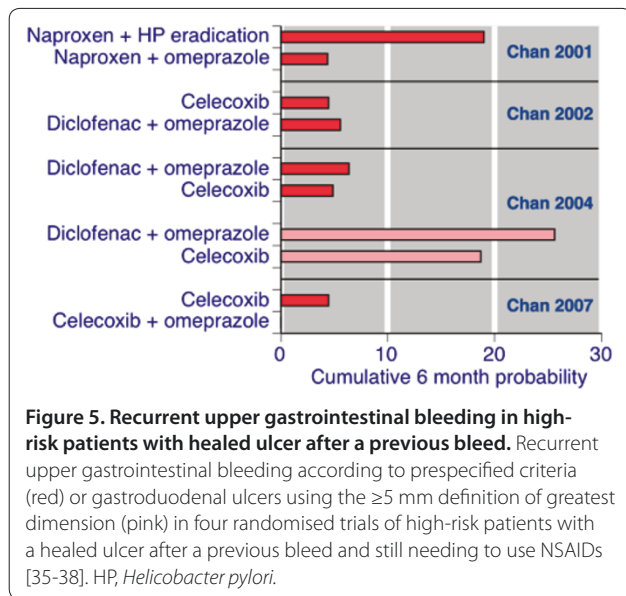
Consistently large magnitude of the increased risk of upper gastrointestinal bleeding with nonsteroidal anti-inflammatory drugs (NSAIDs) in two meta-analyses of observational studies [33,34]. Results show the risk of use of NSAID compared with nonuse, adjusted for confounders.

such as this offer an ethical approach to confirming links between a putative surrogate and a clinical endpoint.

One final piece of evidence linking endoscopic ulcers and bleeding events comes from an analysis of NSAID-induced harm [14]. This analysis examined a number of NSAID-related outcomes, from endoscopic ulcers to clinically diagnosed ulcers, to bleeding events and to death, and demonstrated a consistent effect with NSAIDs for all of these outcomes over a very wide range of event rates (Figure 6). The links at each stage of the process from symptoms such as dyspepsia, to endoscopic ulcers, to serious bleeding events, and even to death, are clear.

For example, patients taking low-dose aspirin or clopidogrel suffering dyspepsia have a spectrum of findings on endoscopic evaluation, including ulcer, erosions, and haemorrhagic spots (Figure 7) [39].

There do not appear to be any black swans – evidence contradicting the general finding that influences on bleeding events and endoscopic ulcers are coherent in direction and magnitude. A suggestion that sulindac, a nonsteroidal anti-inflammatory prodrug, elevates bleeding events but not endoscopic ulcers is only weakly supported. There is good evidence of sulindac elevating bleeding events [33]; the evidence for a lack of effect on



endoscopic ulcers derives from one study lacking data [40] and from another on 15 healthy subjects given sulindac for 7 days [41].

Finally, the association between upper gastrointestinal endoscopic ulcers and serious clinical events is strong. A positive view on each of the nine Hill criteria can be supported by the evidence we have – and although not all is covered in this article, Table 3 summarises what we know.

Utilisation

All of the evidence put forward to justify the surrogate nature of gastroduodenal endoscopic ulcers is in the context of harm from the use of aspirin or NSAIDs. The use of endoscopic ulcers as a surrogate would be justifiable in the context, for example, of new preventative measures being used with established NSAIDs, especially those known already to be associated with either serious upper gastrointestinal clinical events, or endoscopic ulcer, or both. Examples would be any of the new combination products of traditional NSAID plus proton pump inhibitor, as in naproxen plus esomeprazole [42], or NSAID plus histamine-2 receptor antagonist, as in ibuprofen plus famotidine [43].

With increasing use of gastroprotection in the community, and guidance that gastroprotection with proton pump inhibitors should be used even with coxibs [44], the incidence of bleeding events may fall to the point where clinical trials without gastroprotection become unethical. In Japan, a large increase in the use of proton pump inhibitors has resulted in a precipitous fall in bleeding rates, and particularly deaths from a bleeding event [45].

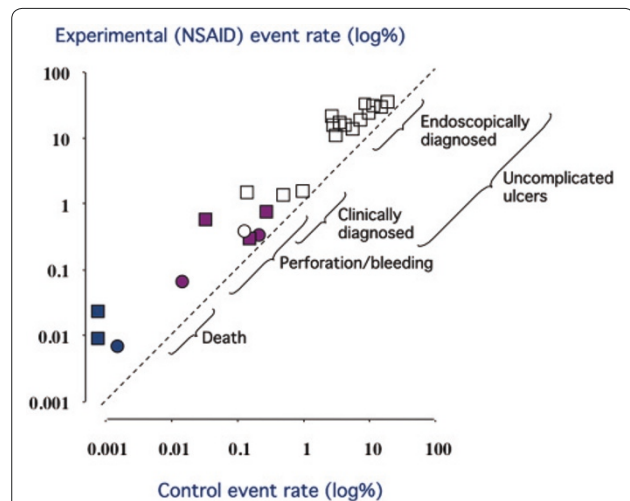
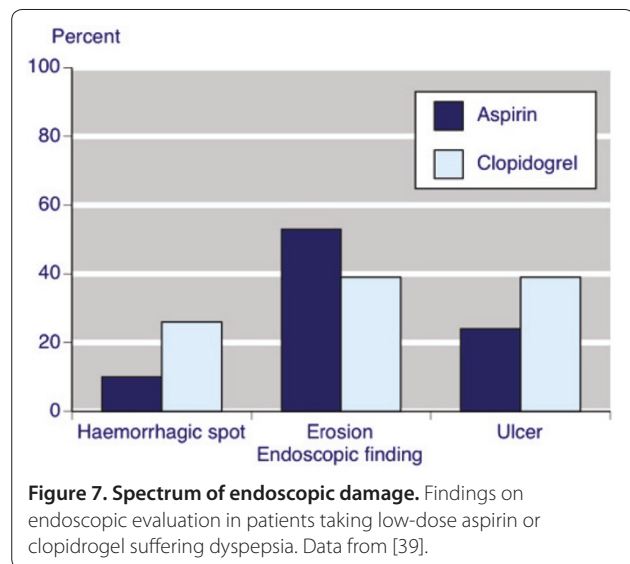


Figure 6. Rate of gastroduodenal complications with nonsteroidal anti-inflammatory drugs compared with control. Rate of gastroduodenal complications (events) with NSAIDs (includes aspirin) compared with control (no NSAID, or placebo, or NSAID + mucosal protection therapy) in 15 randomised controlled trials (RCTs; squares) and three cohort studies (circles). White symbols, uncomplicated peptic, gastric or duodenal ulcer; grey symbols, ulcer bleed or perforation; black symbols, death attributable to a bleeding or perforated ulcer. Several trials reported several levels of harm; that is, several events. Death outcomes in two RCTs had a control event rate of 0%; these were set a control event rate of 0.001% for graphical purposes. Dotted line, the line of equality. From [14] with permission.



Whether gastroduodenal endoscopic ulcers could justify a surrogate status in any other context could only be considered on a case-by-case basis. Each of the various stages of validation, qualification, and utilisation would need to be revisited for that specific context, and the Hill criteria also revisited.

Table 3. Overview of the evidence for each of the Hill criteria

Hill criterion	Supportive evidence
Strength of association	There is a strong association between use of aspirin or NSAIDs and the development of both endoscopic ulcers and clinical bleeding events. Protective strategies have a large effect in preventing both events
Consistency	There is a consistent effect across a range of different risk factors and interventions
Specificity	Exposure to aspirin or NSAIDs causes a spectrum of gastrointestinal harm, but these are found without exposure. The link between aspirin and NSAIDs is specific only because of the elevation of the incidence rates
Temporal relationship	Exposure to aspirin or NSAIDs precedes harm
Biological gradient	There is a consistent dose response with aspirin and NSAIDs, with higher doses and longer use increasing the incidence rates of the harms
Plausibility	There is a biological underpinning for upper gastrointestinal harm with aspirin and NSAIDs
Coherence	The consistent effect of aspirin or NSAIDs on a broad spectrum of upper gastrointestinal harms, from symptoms, to endoscopic findings, to serious bleeding events, is evidence of coherence
Experiment	A broad range of preventative therapies (misoprostol, histamine antagonists, proton pump inhibitors, coxibs) with different mechanisms of action all demonstrate significant reduction of harm with aspirin or NSAIDs
Analogy	Detection of ulcers endoscopically in circumstances where aspirin or NSAIDs are not causative (for example where there may be infection with <i>Helicobacter pylori</i>) would be regarded as a marker of high risk for developing more serious ulcer disease with bleeding

NSAID, nonsteroidal anti-inflammatory drug.

Discussion

This paper has sought to re-examine whether the evidence we have justifies using gastroduodenal endoscopic ulcers as a surrogate for serious upper gastrointestinal bleeding events within the context of the use of aspirin or NSAIDs. The article has followed a pathway for the evaluation of biomarkers and surrogate endpoints in chronic disease, building on a previous review that predated this pathway. Two important conclusions stand out.

There is a strong case for considering endoscopic ulcers as a surrogate for NSAID-induced gastrointestinal harm. These are valid measurements, supported by a wealth of evidence linking endoscopic with more serious upper gastrointestinal harm, within the context of the use of aspirin or NSAIDs. Criticisms of the original findings have been considered, and rejected. The weakness originally identified – the absence of an observation of the direct progression from endoscopic ulcers to ulcer complications – remains.

The structure suggested in the Institute of Medicine report has provided a constructive and focused way of examining this particular example of putative surrogacy. Separating the validity of the measurement, the qualification of the evidence of association, and the context or contexts in which the assumption of surrogacy is valid represents an important methodological statement that has worked well in this case, as it did in case studies in the report.

The Institute of Medicine evaluation process considered that if the biomarker–clinical endpoint relationship persisted over multiple interventions, it may be

thought to be more generalisable. The evidence of links with age, previous ulcer history, and *H. pylori* infection do not involve NSAIDs, and offer the prospect of generalisability to other contexts. This evaluation examined links with a number of risk factors and interventions, but within the context of NSAID use. Context is the key, and those other contexts require their own separate evaluations.

In the end, and despite efforts to the contrary, decisions on whether any marker is a useful or justifiable surrogate retain an element of subjectivity or even bias. Even Austin Bradford Hill, in his influential 1965 address to the Royal Society of Medicine that examined differences between association and causation, admitted that: ‘In asking for very strong evidence I would, however, repeat emphatically that this does not imply crossing every “t”, and swords with every critic, before we act’ [17]. The Institute of Medicine advice offers a mechanism to be systematic in assessing the strength and nature of evidence before the swords come out.

Conclusion

In the context of NSAID-induced upper gastrointestinal harm, endoscopic ulcers appear to be a valid surrogate marker. The Institute of Medicine evaluation schema has been proven to give direction and focus to evaluating a surrogate marker.

Key messages

- Determining whether a biomarker is a useful, acceptable, or valid surrogate for a future beneficial or

harmful event is complex, has been the subject of a number of approaches, and retains a degree of subjectivity.

- Surrogates are useful when they are relatively common or early in a biological pathway, but the clinical event is rare and/or late.
- The Institute of Medicine of the National Academy of Sciences in the USA has put forward an evaluation scheme for biomarkers, looking at validation (assay performance), qualification (assessment of evidence of association), and utilisation (a description of the context in which the surrogate has utility).
- This evaluation scheme has been applied to the example of endoscopy (particularly gastroduodenal ulcers) in the context of NSAID-induced gastrointestinal harm.
- Considerable evidence indicated that endoscopic ulcers were a valid measure.
- Considerable evidence indicated that there was a strong association between endoscopic ulcers and serious gastrointestinal harm, with a variety of risk factors and interventions influencing them in the same direction and with a similar magnitude.
- In the context of NSAID-induced upper gastrointestinal harm, endoscopic ulcers appear to be a valid surrogate marker.
- The Institute of Medicine evaluation schema proved to give direction and focus to evaluating a surrogate marker.

Abbreviations

NSAID, nonsteroidal anti-inflammatory drug.

Competing interests

RAM has received research grants and consulting or lecture fees from pharmaceutical companies, including AstraZeneca, Eli Lilly, Flynn Pharma, GlaxoSmithKline, Grünenthal, Horizon, Menarini, MSD, Pfizer, and Reckitt Benkiser, for work involving analgesic drugs.

Acknowledgements

This review has drawn heavily on previous work with other authors [15], and the author recognises their contribution.

Declaration

This article has been published as part of *Arthritis Research & Therapy* Volume 15 Suppl 3, 2013: 'Gastroprotective NSAIDs'. The full contents of the supplement are available online at <http://arthritis-research.com/supplements/15/S3>. The supplement was proposed by the journal and developed by the journal in collaboration with the Guest Editor. The Guest Editor assisted the journal in preparing the outline of the project but did not have oversight of the peer review process. The Guest Editor serves as a clinical and regulatory consultant in drug development and has served as such consultant for companies which manufacture and market NSAIDs including Pfizer, Pozen, Horizon Pharma, Logical Therapeutics, Nuvo Research, Iroko, Imprimis, JRX Pharma, Nuvon, Medarx, Asahi. The articles have been through the journal's standard peer review process. Publication of this supplement has been supported by Horizon Pharma Inc. Duexis (ibuprofen and famotidine) is a product marketed by the sponsor.

Published: 24 July 2013

References

1. Scandinavian Simvastatin Survival Study Group: Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet* 1994, **344**:1383-1389.
2. Strandberg TE, Pyörälä K, Cook TJ, Wilhelmsen L, Faergeman O, Thorgeirsson G, Pedersen TR, Kjekshus J; 4S Group: Mortality and incidence of cancer during 10-year follow-up of the Scandinavian Simvastatin Survival Study (4S). *Lancet* 2004, **364**:771-777.
3. Biomarkers Definitions Working Group: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharm Ther* 2001, **69**:89-95.
4. Institute of Medicine: *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*. Washington, DC: The National Academies Press; 2010.
5. Pryseley A, Tilahun A, Alonso A, Molenberghs G: Information-theory based surrogate marker evaluation from several randomized clinical trials with continuous true and binary surrogate endpoints. *Clin Trials* 2007, **4**:587-597.
6. Pryseley A, Tilahun A, Alonso A, Molenberghs G: An information-theoretic approach to surrogate-marker evaluation with failure time endpoints. *Lifetime Data Anal* 2011, **17**:195-214.
7. Aronson JK: Research priorities in biomarkers and surrogate end-points. *Br J Clin Pharmacol* 2012, **73**:900-907.
8. Qu Y: Evaluation of a surrogate marker: validity and efficiency. *Stat Med* 2012. doi: 10.1002/sim.5672.
9. Fleming TR, Powers JH: Biomarkers and surrogate endpoints in clinical trials. *Stat Med* 2012, **31**:2973-2984.
10. Mpabanzi L, van Mierlo KM, Malagó M, Dejong CH, Lytras D, Olde Damink SW: Surrogate endpoints in liver surgery related trials: a systematic review of the literature. *HPB (Oxford)* 2012, **15**:327-336.
11. Gabler NB, French B, Strom BL, Palevsky HI, Taichman DB, Kawut SM, Halpern SD: Validation of 6-minute walk distance as a surrogate end point in pulmonary arterial hypertension trials. *Circulation* 2012, **126**:349-356.
12. Moore A, Bjarnason I, Cryer B, Garcia-Rodriguez L, Goldkind L, Lanás A, Simon L: Evidence for endoscopic ulcers as meaningful surrogate endpoint for clinically significant upper gastrointestinal harm. *Clin Gastroenterol Hepatol* 2009, **7**:1156-1163.
13. Graham DY: Endoscopic ulcers are neither meaningful nor validated as a surrogate for clinically significant upper gastrointestinal harm. *Clin Gastroenterol Hepatol* 2009, **7**:1147-1150.
14. Tramer MR, Moore RA, Reynolds DJ, McQuay HJ: Quantitative estimation of rare adverse events which follow a biological progression: a new model applied to chronic NSAID use. *Pain* 2000, **85**:169-182.
15. Hawkey CJ: Non-steroidal anti-inflammatory drugs and peptic ulcers. *BMJ* 1990, **300**:278-284.
16. Straube S, Tramèr MR, Moore RA, Derry S, McQuay HJ: Mortality with upper gastrointestinal bleeding and perforation: effects of time and NSAID use. *BMC Gastroenterol* 2009, **9**:41.
17. Hill AB: The environment and disease: association or causation? *Proc R Soc Med* 1965, **58**:295-300.
18. Graham DY, Chan FK: Inaccurate endoscopy: a better explanation for placebo-associated endoscopic ulcers. *Aliment Pharmacol Ther* 2009, **30**:955-957.
19. Amano Y, Uno G, Yuki T, Okada M, Tada Y, Fukuba N, Ishimura N, Ishihara S, Kinoshita Y: Interobserver variation in the endoscopic diagnosis of gastroduodenal ulcer scars: implications for clinical management of NSAIDs users. *BMC Res Notes* 2011, **4**:409.
20. Sung JY, Lau JY, Chan FK, Graham DY: How often are endoscopic ulcers in NSAID trials diagnosed as actual ulcers by experienced endoscopists. *Gastroenterology* 2001 **120**(Suppl 1):A597.
21. Yuan YH, Wang C, Yuan Y, Hunt RH: Meta-analysis: incidence of endoscopic gastric and duodenal ulcers in placebo arms of randomized placebo-controlled NSAID trials. *Aliment Pharmacol Ther* 2009, **30**:197-209.
22. Yuan YH, Wang C, Yuan Y, Hunt RH: Inaccurate endoscopy: a better explanation for placebo-associated endoscopic ulcers: authors' reply. *Aliment Pharmacol Ther* 2009, **30**:955-963.
23. Sakamoto C, Kawai T, Nakamura S, Sugioka T, Tabira J: Comparison of gastroduodenal ulcer incidence in healthy Japanese subjects taking celecoxib or loxoprofen evaluated by endoscopy: a placebo-controlled, double-blind 2-week study. *Aliment Pharmacol Ther* 2013, **37**:346-354.
24. Moberly JB, Harris SI, Riff DS, Dale JC, Brees T, McLaughlin P, Lawson J, Wan Y, Xu J, Truitt KE: A randomized, double-blind, one-week study comparing effects of a novel COX-2 inhibitor and naproxen on the gastric mucosa.

- Dig Dis Sci* 2007, **52**:442-450.
25. Moore RA, Gavaghan D, Tramèr MR, Collins SL, McQuay HJ: **Size is everything – large amounts of information are needed to overcome random effects in estimating direction and magnitude of treatment effects.** *Pain* 1998, **78**:209-216.
 26. Bernersen B, Johnsen R, Straume B, Burhol PG, Jenssen TG, Stakkevold PA: **Towards a true prevalence of peptic ulcer: the Sørreisa gastrointestinal disorder study.** *Gut* 1990, **31**:989-992.
 27. Moore RA, Derry S, Makinson GT, McQuay HJ: **Tolerability and adverse events in clinical trials of celecoxib in osteoarthritis and rheumatoid arthritis: systematic review and meta-analysis of information from company clinical trial reports.** *Arthritis Res Ther* 2005, **7**:R644-R665
 28. Edwards JE, McQuay HJ, Moore RA: **Efficacy and safety of valdecoxib for treatment of osteoarthritis and rheumatoid arthritis: systematic review of randomised controlled trials.** *Pain* 2004, **111**:286-296.
 29. Laine L, Harper S, Simon T, Bath R, Johanson J, Schwartz H, Stern S, Quan H, Bolognese J: **A randomized trial comparing the effect of rofecoxib, a cyclooxygenase 2-specific inhibitor, with that of ibuprofen on the gastroduodenal mucosa of patients with osteoarthritis. Rofecoxib Osteoarthritis Endoscopy Study Group.** *Gastroenterology* 1999, **117**:776-783.
 30. Hawkey C, Laine L, Simon T, Beaulieu A, Maldonado-Cocco J, Acevedo E, Shahane A, Quan H, Bolognese J, Mortensen E: **Comparison of the effect of rofecoxib (a cyclooxygenase 2 inhibitor), ibuprofen, and placebo on the gastroduodenal mucosa of patients with osteoarthritis: a randomized, double-blind, placebo-controlled trial. The Rofecoxib Osteoarthritis Endoscopy Multinational Study Group.** *Arthritis Rheum* 2000, **43**:370-377.
 31. Hawkey CJ, Laine L, Simon T, Quan H, Shingo S, Evans J; Rofecoxib Rheumatoid Arthritis Endoscopy Study Group: **Incidence of gastroduodenal ulcers in patients with rheumatoid arthritis after 12 weeks of rofecoxib, naproxen, or placebo: a multicentre, randomised, double blind study.** *Gut* 2003, **52**:820-826.
 32. Hunt RH, Harper S, Watson DJ, Yu C, Quan H, Lee M, Evans JK, Oxenius B: **The gastrointestinal safety of the COX-2 selective inhibitor etoricoxib assessed by both endoscopy and analysis of upper gastrointestinal events.** *Am J Gastroenterol* 2003, **98**:1725-1733.
 33. Hernández-Díaz S, Rodríguez LA: **Association between nonsteroidal anti-inflammatory drugs and upper gastrointestinal tract bleeding/perforation: an overview of epidemiologic studies published in the 1990s.** *Arch Intern Med* 2000, **160**:2093-2099.
 34. Massó González EL, Patrignani P, Tacconelli S, García Rodríguez LA: **Variability among nonsteroidal antiinflammatory drugs in risk of upper gastrointestinal bleeding.** *Arthritis Rheum* 2010, **62**:1592-1601.
 35. Chan FK, Chung SC, Suen BY, Lee YT, Leung WK, Leung VK, Wu JC, Lau JY, Hui Y, Lai MS, Chan HL, Sung JJ: **Preventing recurrent upper gastrointestinal bleeding in patients with Helicobacter pylori infection who are taking low-dose aspirin or naproxen.** *N Engl J Med* 2001, **344**:967-973.
 36. Chan FK, Hung LC, Suen BY, Wu JC, Lee KC, Leung VK, Hui AJ, To KF, Leung WK, Wong VW, Chung SC, Sung JJ: **Celecoxib versus diclofenac and omeprazole in reducing the risk of recurrent ulcer bleeding in patients with arthritis.** *N Engl J Med* 2002, **347**:2104-2110.
 37. Chan FK, Hung LC, Suen BY, Wong VW, Hui AJ, Wu JC, Leung WK, Lee YT, To KF, Chung SC, Sung JJ: **Celecoxib versus diclofenac plus omeprazole in high-risk arthritis patients: results of a randomized double-blind trial.** *Gastroenterology* 2004, **127**:1038-1043.
 38. Chan FK, Wong VW, Suen BY, Wu JC, Ching JY, Hung LC, Hui AJ, Leung WK, Lee VW, Lai LH, Wong GL, Chow DK, To KF, Leung WK, Chiu PW, Lee YT, Lau JY, Chan HL, Ng EK, Sung JJ: **Combination of a cyclo-oxygenase-2 inhibitor and a proton-pump inhibitor for prevention of recurrent ulcer bleeding in patients at very high risk: a double-blind, randomised trial.** *Lancet* 2007, **369**:1621-1626.
 39. Tsai TJ, Lai KH, Hsu PI, Lin CK, Chan HH, Yu HC, Wang HM, Lin KH, Wang KM, Chang SN, Liu CP, Hsiao SH, Huang HR, Lin CH, Tsay FW: **Upper gastrointestinal lesions in patients receiving clopidogrel anti-platelet therapy.** *J Formos Med Assoc* 2012, **111**:705-710.
 40. Lanza FL: **Endoscopic studies of gastric and duodenal injury after the use of ibuprofen, aspirin, and other nonsteroidal anti-inflammatory agents.** *Am J Med* 1984, **77**(1A):19-24.
 41. Graham DY, Smith JL, Holmes GI, Davies RO: **Nonsteroidal anti-inflammatory effect of sulindac sulfoxide and sulfide on gastric mucosa.** *Clin Pharmacol Ther* 1985, **38**:65-70.
 42. Goldstein JL, Hochberg MC, Fort JG, Zhang Y, Hwang C, Sostek M: **Clinical trial: the incidence of NSAID-associated endoscopic gastric ulcers in patients treated with PN 400 (naproxen plus esomeprazole magnesium) vs. enteric-coated naproxen alone.** *Aliment Pharmacol Ther* 2010, **32**:401-413.
 43. Laine L, Kivitz AJ, Bello AE, Grahn AY, Schiff MH, Taha AS: **Double-blind randomized trials of single-tablet ibuprofen/high-dose famotidine vs. ibuprofen alone for reduction of gastric and duodenal ulcers.** *Am J Gastroenterol* 2012, **107**:379-386.
 44. National Collaborating Centre for Chronic Conditions: *Osteoarthritis: National Clinical Guideline for Care and Management in Adults.* London: Royal College of Physicians; 2008.
 45. Miyamoto M, Haruma K, Okamoto T, Higashi Y, Hidaka T, Manabe N: **Continuous proton pump inhibitor treatment decreases upper gastrointestinal bleeding and related death in rural area in Japan.** *J Gastroenterol Hepatol* 2012, **27**:372-377.

doi:10.1186/ar4176

Cite this article as: Moore RA: Endoscopic ulcers as a surrogate marker of NSAID-induced mucosal damage. *Arthritis Research & Therapy* 2013, **15**(Suppl 3):S4.