



Published in final edited form as:

Clin Gastroenterol Hepatol. 2009 November ; 7(11): 1147–1150. doi:10.1016/j.cgh.2009.06.006.

Endoscopic ulcers are neither meaningful nor validated as a surrogate for clinically significant upper gastrointestinal harm

David Y. Graham, M.D.¹

¹ Department of Medicine, Michael E. DeBakey Veterans Affairs Medical Center and Baylor College of Medicine, Houston, Texas, USA

Keywords

Nonsteroidal anti-inflammatory drugs; ulcers; bleeding; complications; prevention; peptic ulcer; ulcer complications; proton pump inhibitor; celecoxib; surrogate

“A correlate does not a surrogate make”¹

Gastrointestinal complications are common among chronic NSAID/aspirin users. Despite the importance of this problem, randomized controlled trials with ulcer complications as the primary endpoint are rare. Theoretically, effective prevention strategies could be developed rapidly if a surrogate endpoint was available that reliably correlated with the true clinical outcome. In this issue of CGH, a group of opinion leaders make the argument that endoscopic ulcers might serve as a surrogate endpoint for clinically significant ulcer complications (abbreviated as “UGI harm”)². Their manuscript was based on a “white paper” intended for submission to the US Food and Drug Administration and initially produced by a commercial vendor under a contract from Pfizer followed by input from the authors. The authors note that because evidence examining a direct relationship between endoscopic ulcers and UGI harm was unavailable, they sought “indirect evidence from a range of interventions and risk factors relating to the direction and magnitude on endoscopic ulcers and clinical outcomes”. Among the studies they examined, they found consistent and reproducible effects on UGI harm that paralleled endoscopic ulcers and recommended endoscopic ulcers as a possibly surrogate endpoint for UGI harm.

Surrogate endpoints

Because expensive large randomized clinical trials are generally needed to identify differences in infrequent events, there has long been an interest in identifying alternative outcomes, or surrogate endpoints that would allow smaller and more efficient studies. There are many examples of successful surrogate markers. For example, blood pressure is an important risk factor for cardiovascular-related mortality (ie, among hypertensive patients a 5% reduction in cardiovascular-related mortality and a 10% reduction in stroke is obtained for every 1 mm Hg

Corresponding author: David Y. Graham, M.D., Michael E. DeBakey Veterans Affairs Medical Center, RM 3A-320 (111D), 2002 Holcombe Boulevard, Houston, TX 77030, Phone: 713-795-0232, FAX: 713-790-1040, dgraham@bcm.tmc.edu.

Potential conflicts of interest

In the last three years Dr. Graham has received small amounts of grant support and/or free drugs or urea breath tests from Meretek, and BioHit for investigator initiated and completely investigator controlled research. Dr. Graham is a consultant for Novartis in relation to vaccine development for treatment or prevention of *H. pylori* infection and a paid consultant for Otsuka Pharmaceuticals, manufacturer of the ¹³C-urea breath test. Dr. Graham will receive royalties on materials related to the US version of the ¹³C-urea breath test until October 2009. Dr. Graham is also an unpaid member of the executive committee of the PRECISION trial which is supported by Pfizer and is designed to compare the cardiovascular safety of celecoxib, naproxen, and ibuprofen in higher cardiovascular risk patients.

reduction in blood pressure)^{3, 4}. Almost 20 years ago criteria to validate surrogate endpoints in phase 3 trials were proposed⁵. Essentially these criteria required that the surrogate must be a correlate of the true clinical outcome and fully capture the net effect of treatment on the clinical outcome. Many proposed surrogate markers have failed to meet these criteria^{1, 6, 7}. Fleming identified four reasons for failure including 1) The surrogate is not in the causal pathway of the disease process, 2) Of several causal pathways of disease, the intervention affects only the pathway mediated through the surrogate, 3) The surrogate is not in the pathway of the intervention's effect or is insensitive to its effect, and 4) The intervention has mechanisms of action independent of the disease process⁶. Twaddell added an additional explanation: "even if an intervention has an effect on a surrogate marker and that marker is clearly in the causal pathway of the clinical end point, the effect may not persist long enough for the drug to alter the long-term clinical outcome. The drug may seem to be efficacious because of its short-term effect on the surrogate marker, but have no effect on the clinical outcome"⁷. An example of failure of a generally reliable surrogate is the effect of calcium channel blockers in hypertensive patients. Reduction in blood pressure have proven to be a reliable surrogate endpoint for the evaluation of diuretics, however the favorable antihypertensive effects of calcium channel blockers were offset by other mechanisms making the surrogate fail⁶.

Candidate surrogate markers are typically identified in phase 2 trials designed to identify whether a new intervention is biologically active and for guiding decisions about whether the intervention is promising enough to justify large definitive trials with clinically meaningful outcomes. Candidate surrogate endpoints must be rigorously validated in phase 3 trials. In the absence of a very well validated surrogate endpoint, the primary endpoint should be the true clinical outcome⁶.

The authors proposing endoscopic ulcer be considered as a surrogate rely on correlations to support their hypothesis². Possibly because of space limitations, each item of evidence was covered very briefly with little or no discussion of the weaknesses or exceptions. Importantly, the potential surrogate marker itself commonly used (a gastric or duodenal lesion ≥ 3 mm with significant depth) is in actuality often difficult to separate from an erosion. "Significant depth" has never been defined^{8, 9}. One can ask, why 3 mm and not 5 mm, 10 mm, or some other number? How is the size to be measured (ie, in any one direction (ie, 3 mm \times 0.5 mm?), minimal for any direction, etc)? One person's erosion may be another's ulcer^{8, 9}. For example, a blinded analysis of a teaching tape used to train investigators for detection of endoscopic ulcers for a pharmaceutical company sponsored study was performed by a group of experienced NSAID researchers. The experienced endoscopists disagreed that many of the lesions described as endoscopic ulcers were in fact ulcers¹⁰.

Other studies suggest that the endoscopic determination of the presence or absence of an endoscopic ulcer, at least as done in large multicenter trials, has a definite and not insignificant false-positive rate. For example, although actual ulcers are extremely rare among *H. pylori* negative asymptomatic patients not taking any gastrotoxic drugs, 4.2% of volunteers who initially had negative endoscopies and were receiving placebo were found to have endoscopic ulcers in a 12 week trial (ie, an annual incidence of 16.8%)¹¹. It is indeed rare for large multicenter trials to "calibrate" the observers to ensure consistency in diagnosis. In addition, video or still photographic images are rarely taken and review to confirm the reported findings. The authors noted that increasing age was associated with an increased prevalence of endoscopic ulcers and ulcer complications. Most correlations are incidental and not causative. For example, it has long been known that UGI bleeding increases with age independently of whether the patient is receiving NSAIDs^{12, 13}. As noted above, the validity of a surrogate can only be confirmed by clinical trials with the endpoint being the true clinical outcome. The outcome of most interest is the presence of UGI harm defined as bleeding, perforation, or obstruction. The authors noted that one potential weakness of their analyses was that it remains

unclear whether or how often endoscopic erosion/ulcers themselves transform into chronic or clinical ulcers and if so, which features allow one to predict such an occurrence². Although this outcome is different from UGI harm as defined above, there are data that endoscopic ulcers and chronic ulcers are not directly related. For example, one of the first animal experiments producing ulcers with drugs was the production of gastric ulcer following the chronic administration of cinchophen. In that model the course of gastric injury was first a typical acute severe mucosal damage developed, followed by recovery of the mucosa despite continuing the drug. Then, a small fraction of dogs then went on to develop chronic gastric ulcer {Bollman, 1938 20560/id}. It has long been known that frequency of gastric mucosal damage seen in acute studies in man is much greater than noted in endoscopic surveys of chronic aspirin users. In acute studies, the damage is almost universally present and is usually diffuse and moderately severe. In contrast, in chronic studies the damage is usually mild except for a small percent with true ulcers and is reported in only 20–50% of subjects)^{15, 16}.

Ulcers associated with UGI harm tend to be chronic and to recur in the same location suggesting that the process is different from that which produces endoscopic ulcers, especially when using locally gastrotoxic drugs¹⁷. If UGI harm is typically the result of transformation of a particular lesion, how can that lesion be identified and separated from the often diffuse injury seen in a typical study using an NSAID with both topical and systemic actions?

If one accepted the surrogate maker, how would one use it?

The authors suggest that there is a quantitative aspect of endoscopic ulcers (ie, COX-2 inhibitors cause fewer than traditional NSAIDs and appear safer, antisecretory drug therapy reduces bad outcomes and endoscopic ulcers)². First, is the hypothesis restricted to the few NSAIDs examined or universal? There are many papers ranking NSAIDs in terms of risk for UGI complications and the risk has been found to vary greatly^{18–21}. Prior to the introduction of COX-2 inhibitors, the prodrug sulindac was noted to rarely cause endoscopic ulcers. According to the endoscopic ulcer surrogate hypothesis, it should be a safe NSAID and early studies were very encouraging that this would be the case. For example, short term blood loss studies using chromium-tagged red blood cells showed that sulindac was no more damaging than placebo and contrasted markedly to results with other available NSAIDs²². Endoscopic studies using normal volunteers also showed that sulindac rarely caused erosions or significant gastric mucosal injury^{23, 24}. Subsequent studies provided a biologic basis for these observations; as a prodrug and the active form was not present in the stomach. In addition, both products, sulindac sulfoxide and sulindac sulfide, were almost insoluble in acidic gastric contents thus preventing topical injury²⁴. Unfortunately, the apparent lack of gastrotoxicity in short term studies did not reliably predict that sulindac was a particularly safe NSAID for when UGI bleeding within 30 days of exposure to one of seven NSAIDs was examined in 88,044 patients, sulindac users had the highest rate of UGI tract bleeding²⁵. Another NSAID thought to have a low propensity to cause endoscopic ulcers, nabumetone,²⁶ was given to high risk patients who had previously suffered NSAID-associated UGI bleeding. No evidence of protection was seen and bleeding was common²⁷. If the endoscopic ulcer hypothesis does not hold for all orally administered NSAIDs, one would like to know whether it would be *meaningful* for alternate routes of drug administration such as rectal, parenteral, or enteric coated or for drugs other than NSAIDs (ie, iron, pivampicillin, microencapsulated or wax matrix potassium) which are also associated with endoscopic ulcers^{28?}

Can the effect on endoscopic ulcer prevention be used to identify the best therapy?

In the original omeprazole vs. low dose misoprostol and the omeprazole vs. low dose ranitidine studies, low dose misoprostol was more effective for ulcer prevention among those without

H. pylori infection than was omeprazole; omeprazole and ranitidine were not statistically different²⁹. Low dose misoprostol was also superior to omeprazole for healing endoscopic gastric ulcers among *H. pylori* negative NSAID users²⁹. Standard dose misoprostol proved superior to lansoprazole in preventing recurrence of endoscopic ulcers in a head to head comparison³⁰. Using the proposed surrogate endpoint one might conclude that low dose misoprostol would be the preferred gastroprotective agent. In epidemiology studies, misoprostol as the combination drug Arthrotec® has consistently been associated with reduced events and deaths³¹ and is the only drug shown in large randomized clinical trials to reduce UGI harm among chronic NSAID users³². Epidemiologic studies suggest that PPIs are somewhat effective although the percent reduction remains unclear; there is no evidence suggesting a dose response effect³¹. H₂-receptor antagonists appear to be marginally useful³¹. Thus, despite the large number of studies using endoscopic ulcers as an endpoint, the lack of true outcome data prevents clinicians from making a reliable judgment regarding which gastroprotective strategy is superior or to assess their relative effectiveness.

The major gastroprotective strategies have been evaluated in single center studies using the clinical endpoint UGI harm (ie, bleeding). These studies have been done in high risk patients defined as those who had bled from NSAID ulcers. In that population, misoprostol showed essentially no benefit²⁷ and omeprazole proved to be clinically ineffective both for preventing recurrent ulcers³³ or UGI bleeding^{27, 34}. Celecoxib alone was also not clinically effective^{33, 34}. Clearly, for high risk patients, the proposed surrogate marker failed to correlate to actual outcome confirming again the admonition against using surrogates that have not been rigorously validated.

Where do we go from here?

A useful surrogate endpoint must be able to be reliably identified, and highly and consistently reproducible, with minimal interobserver variability. Endoscopic ulcers as currently used meet none of these prerequisites. If standardization were to be accomplished, then rigorous validation would still be needed to confirm that its use was applicable to different interventions. Such validation will require actual clinical trials and the results may not be applicable for drugs with different mechanisms of action (eg, NSAIDs with systemic and topical activity vs. systemic only). The authors point out it may be difficult or impossible to currently do a placebo controlled trial because of ethical considerations². Although, it may or may not be unethical not to use placebo, it may not be necessary in that misoprostol has been shown to reduce UGI harm making non-inferiority studies a viable option. Such trials are unlikely to be funded by PPI manufacturers which, as we noted previously, are particularly vulnerable to class effects, such that any new indications would usually be generalizable to all PPIs³¹. The Obama administration has instituted a government program to compare different therapies and thus may provide a mechanism for studies to compare preventive strategies and possibly identify valid surrogate markers. Considering all the problems and uncertainties currently surrounding endoscopic ulcers, we believe it is unlikely they will ever become more than a marketing tool of uncertain relevance. We doubt endoscopic ulcers will disappear as an endpoint for as Haynes and Haynes recently commented “Unfortunately, when medical hypotheses are disproved, they act more like zombies than corpses, revived by the sorcerers of Mammon, aided and abetted by the inertia of medical practice”³⁵.

Acknowledgments

Support: This material is based upon work supported in part by the Office of Research and Development Medical Research Service Department of Veterans Affairs. Dr. Graham is supported in part by Public Health Service grant DK56338 which funds the Texas Medical Center Digestive Diseases Center and R01 CA116845. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the VA or NIH.

References

1. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. *Health Aff (Millwood)* 2005;24:67–78. [PubMed: 15647217]
2. Moore A, Bjarnason I, Cryer B, Garcia-Rodriguez L, Goldkind L, Lanas A, Simon L. Evidence for endoscopic ulcers as meaningful surrogate endpoint for clinically significant upper gastrointestinal harm. *Clin Gastroenterol Hepatol.* 2009
3. Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, Godwin J, Qizilbash N, Taylor JO, Hennekens CH. Blood pressure stroke and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 1990;335:827–838. [PubMed: 1969567]
4. Collins R, Peto R, Godwin J, MacMahon S. Blood pressure and coronary heart disease. *Lancet* 1990;336:370–371. [PubMed: 1975346]
5. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431–440. [PubMed: 2727467]
6. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996;125:605–613. [PubMed: 8815760]
7. Twaddell S. Surrogate outcome markers in research and clinical practice. *Aust Prescr* 2009;32:47–50.
8. Graham, DY.; Genta, RM. Gastritis and gastropathy. In: Yamada, T.; Alpers, DH.; Kalloo, AN.; Kaplowitz, N.; Owyang, C.; Powell, DW., editors. *Atlas of Gastroenterology*. 4. Hoboken: Wiley-Blackwell; 2008. p. 251-260.
9. Graham, DY.; Horiuchi, A.; Kato, M. Peptic ulcer disease. In: Yamada, T.; Alpers, DH.; Kalloo, AN.; Kaplowitz, N.; Owyang, C.; Powell, DW., editors. *Atlas of Gastroenterology*. 4. Hoboken: Wiley-Blackwell; 2008. p. 237-250.
10. Sung JY, Lau JY, Chan FK, Graham DY. How often are endoscopic ulcers in NSAID trials diagnosed as actual ulcers by experienced endoscopists 2001;120(Suppl 1):A597.
11. Graham DY, Chan FK. Endoscopic ulcers with low-dose aspirin and reality testing. *Gastroenterology* 2005;128:807–808. [PubMed: 15765433]
12. Hernandez-Diaz S, Garcia Rodriguez LA. Cardioprotective aspirin users and their excess risk of upper gastrointestinal complications. *BMC Med* 2006;4:22. [PubMed: 16987411]
13. Hernandez-Diaz S, Rodriguez LA. Incidence of serious upper gastrointestinal bleeding/perforation in the general population: review of epidemiologic studies. *J Clin Epidemiol* 2002;55:157–163. [PubMed: 11809354]
14. Bollman JC, Stalker KL, Mann PC. Experimental peptic ulcer produced by cinchophen. *Arch Intern Med* 1938;61:119–128.
15. Graham DY, Smith JL. Aspirin and the stomach. *Ann Intern Med* 1986;104:390–398. [PubMed: 3511824]
16. Graham DY, Smith JL, Dobbs SM. Gastric adaptation occurs with aspirin administration in man. *Dig Dis Sci* 1983;28:1–6. [PubMed: 6600426]
17. 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. [PubMed: 12501222]
18. Garcia Rodriguez LA, Jick H. Risk of upper gastrointestinal bleeding and perforation associated with individual non-steroidal anti-inflammatory drugs [published erratum appears. *Lancet* 1994 Apr 23;343(8904):1048. [PubMed: 7909084] *Lancet* 1994;343:769–772. [PubMed: 7907735]
19. Henry D, Lim LL, Garcia Rodriguez LA, Perez Gutthann S, Carson JL, Griffin M, Savage R, Logan R, Moride Y, Hawkey C, Hill S, Fries JT. Variability in risk of gastrointestinal complications with individual non-steroidal anti-inflammatory drugs: results of a collaborative meta- analysis. *BMJ* 1996;312:1563–1566. [PubMed: 8664664]
20. Henry D, Dobson A, Turner C. Variability in the risk of major gastrointestinal complications from nonaspirin nonsteroidal anti-inflammatory drugs. *Gastroenterology* 1993;105:1078–1088. [PubMed: 8405852]

21. Langman MJ, Weil J, Wainwright P, Lawson DH, Rawlins MD, Logan RF, Murphy M, Vessey MP, Colin-Jones DG. Risks of bleeding peptic ulcer associated with individual non-steroidal anti-inflammatory drugs. *Lancet* 1994;343:1075–1078. [PubMed: 7909103]
22. Pemberton RE, Strand LJ. A review of upper-gastrointestinal effects of the newer nonsteroidal antiinflammatory agents. *Dig Dis Sci* 1979;24:53–64. [PubMed: 371938]
23. 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;77:19–24. [PubMed: 6465160]
24. 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. [PubMed: 3874038]
25. Carson JL, Strom BL, Morse ML, West SL, Soper KA, Stolley PD, Jones JK. The relative gastrointestinal toxicity of the nonsteroidal anti-inflammatory drugs. *Arch Intern Med* 1987;147:1054–1059. [PubMed: 3496062]
26. Roth SH, Tindall EA, Jain AK, McMahon FG, April PA, Bockow BI, Cohen SB, Fleischmann RM. A controlled study comparing the effects of nabumetone, ibuprofen, and ibuprofen plus misoprostol on the upper gastrointestinal tract mucosa. *Arch Intern Med* 1993;153:2565–2571. [PubMed: 8239849]
27. Chan FK, Sung JJ, Ching JY, Wu JC, Lee YT, Leung WK, Hui Y, Chan LY, Lai AC, Chung SC. Randomized trial of low-dose misoprostol and naproxen vs. nabumetone to prevent recurrent upper gastrointestinal haemorrhage in users of non-steroidal anti-inflammatory drugs. *Aliment Pharmacol Ther* 2001;15:19–24. [PubMed: 11136274]
28. Graham, DY. Effectiveness and tolerance of “solid” vs “liquid” potassium replacement therapy. In: Welton, PK.; Welton, A.; Walker, WG., editors. *Potassium in Cardiovascular and Renal Medicine*. New York: Marcel Dekker, Inc; 1986. p. 435-450.
29. Graham DY. Critical effect of *Helicobacter pylori* infection on the effectiveness of omeprazole for prevention of gastric or duodenal ulcers among chronic NSAID users. *Helicobacter* 2002;7:1–8. [PubMed: 11886469]
30. Graham DY, Agrawal NM, Campbell DR, Haber MM, Collis C, Lukasik NL, Huang B. Ulcer prevention in long-term users of nonsteroidal anti-inflammatory drugs: results of a double-blind, randomized, multicenter, active- and placebo-controlled study of misoprostol vs lansoprazole. *Arch Intern Med* 2002;162:169–175. [PubMed: 11802750]
31. Graham DY, Chan FK. NSAIDs, risks, and gastroprotective strategies: current status and future. *Gastroenterology* 2008;134:1240–1246. [PubMed: 18395101]
32. Silverstein FE, Graham DY, Senior JR, Davies HW, Struthers BJ, Bittman RM, Geis GS. Misoprostol reduces serious gastrointestinal complications in patients with rheumatoid arthritis receiving nonsteroidal anti-inflammatory drugs. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med* 1995;123:241–249. [PubMed: 7611589]
33. 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. [PubMed: 15480981]
34. Lai KC, Chu KM, Hui WM, Wong BC, Hu WH, Wong WM, Chan AO, Wong J, Lam SK. Celecoxib compared with lansoprazole and naproxen to prevent gastrointestinal ulcer complications. *Am J Med* 2005;118:1271–1278. [PubMed: 16271912]
35. Haynes B, Haynes GA. ACP Journal Club. What does it take to put an ugly fact through the heart of a beautiful hypothesis? *Ann Intern Med* 2009;150:JC3. [PubMed: 19306483]