CURRENT ONCOLOGY
LETTER TO THE EDITOR

Mammography, Martin Yaffe, and me: response and appreciation

The Editor
Current Oncology
7 August 2015

I thank Dr. Martin Yaffe for his many constructive comments in his thoughtful review[1] of my previous invited editorial[2], providing valuable insights into the complex and controversial issues of the current mammography debate. Although I take up here several important issues that he raises, the serious reader is strongly advised to give his response the closest of readings to uncover the many other perceptive observations he brings to bear on the debate beyond the scope of this modest response and reflection.

Whom Can You Trust?—Informed Decisions and Tainted Sources

In one of his many penetrating observations, Yaffe suggests with some irony the potential hidden upside of what I documented to be the inattention of the ultimate stakeholder, screening-eligible women, to the ongoing mammography debate: namely, that they will fortunately miss much of the "one-sided viewpoint on the issue of screening: against"—prominently instantiated, as he observes, by media reporters such as Health and Science reporter Ms. Gina Kolata of *The New York Times*.

Yaffe is on to something very important here—namely, the disruptive role of naïve and often downright ignorant media coverage (although, interestingly I should note, some screening critics see the opposite—a pro-screening bias in the popular media—as witness the posture of Dr. Donald Berry[3]). But the point remains that much of popular media coverage of the mammography debate often borders on the appalling while also being steeped in unresolved potential biases. Yaffe's highlighting of Kolata is on target: Kolata is herself a breast cancer survivor whose cancer went undetected by mammographic screening, which could compromise full objectivity, something *The New York Times* Health editors have been less than scrupulous about (with many required corrections[4]). Kolata recently extolled the virtues of the 25-year follow-up to the Canadian National Breast Screening Study (CNBSS) as "one of the largest and most meticulous studies of mammography ever done," which, as personal opinion, is her right, but neglects to mention the dozens of intensive critiques that were already available at the time of writing, which is not her right in any professionally responsible, balanced, and objective journalism. Hence, Yaffe correctly concludes that missing such egregious misinformation could in fact be cause to break out the champagne.

However, this leaves screening-eligible women with simply no reliable and, at the same time, *readable* sources to critically assist them in making screening decisions, and as I have already argued[5], medical professionals themselves might in fact offer little accurate clarification and assistance. Let us remember the lesson that Dr. David Eddy[6] taught us more than 30 years ago: Confronted with the goal of determining the probability that a positive mammographic screening finding in a symptom-free 52-year-old woman is really breast cancer—which is known as the positive predictive value—from known prevalence (1%), sensitivity (80%), and specificity (10%) data for this group of symptom-free women, 95% of all physicians surveyed answered that the positive predictive value (that is, the probability of a true positive) would be approximately 75%, which is off by 1000% (true positive predictive value being 7.5%), a result also holding true for medical students and lay audiences[7,8]. Thus, professionals and lay public both betray well-attested large degrees of innumeracy.

But exceptions to benighted media coverage on the mammography debate do occasionally occur. Appearing in a bit more learned source than newspapers and other popular media, Dr. Paul Taylor's masterful account of the core critical issues of the debate in the *London Review of Books*[9] is one such exception, but still admittedly a read not for the faint-hearted. Should Dr. Mary Costanza, another contributor to the debate in this issue of the journal, consider adopting her piece for popular media, I am absolutely confident that she would succeed impressively at both reliability and readability of coverage, as she has in fact done here. But there are, lamentably, only a handful of Paul Taylors and Mary Costanzas.

Overdiagnosis: Definitions, Disagreements, and Real-World Mammography

Elsewhere in his response, Yaffe properly underlines that all estimations of overdetection are challenging, in part because of the lack of a common scale for estimations of the benefits on the one hand, and for harms and limitations on the other—reminding us shrewdly that, in any weighing of the potential harms of overdetection and associated recall, the other comparator is always a life saved, an excellent corrective perspective for much of the overvaluation of the high price paid for overdetection. I would argue that here too, as noted in my previous discussions of normalization[2,5], we can to some extent normalize overdetection. In almost all the studies that were included in the EUROSCREEN working group review, there was no normalizing adjustment (to the risk of

DOI: http://dx.doi.org/10.3747/co.22.2883

cancer and advancement at diagnosis), leading to wide variation in estimates up to 41% (invasive only) and 57% (invasive plus *in situ*)[10]; however, after normalization adjustment, the rate drops to 10%[11]. In addition, some part of the disagreement and confusion over overdiagnosis is definitional, with no real consensus across evaluators and studies. And so we have distinctions across overdiagnosis, overdetection, and false positives compared with misdiagnosis, recalls, and overtreatment—these terms meaning different things to different investigators[12]. But in the final analysis, as I've noted, women appear to operate under a principle of "regret minimization," which entails that they will consistently and overwhelmingly assign higher value (in a game-theoretic sense) to acceptance of overtreatment than to undertreatment.

Yaffe also notes, and I agree[5], a limitation of much data on overdiagnosis in the failure to discriminate invited versus attendant participants, true of both much of randomized controlled trial (RCT) data and many observational studies. At least some studies have deployed individual-level data capturing actual and not just presumed (invited-only) screening exposure[13–17], all determining overdiagnosis to be in the range of 2%–17% for all disease, and significantly lower for invasive-only disease, with a South Australian study[18,19] finding overdiagnosis to be 8% for invasive breast cancers. More recently, Dr. Stephan Feig concluded that the rate of overdiagnosis is clinically negligible[20] (no more than 0%–5%), and that, despite claims that mammography captures significant proportions of non-evolutive cancers, most screening-detected cases of ductal carcinoma *in situ* (DCIS) were in fact of medium and high grade, and hence had substantial invasivity potential (a point anticipated earlier by Dr. Steven Narod).

Dueling Protocols: RCTs and Observational Data

Yaffe also reflects on the evidence provided by RCT compared with observational data. Here, however, we must be fair in acknowledging that each modality has its own set of strengths and limitations. Randomized controlled trials typically underestimate the benefit to women actually attendant at screening: Some women in the control arm actually receive screening, and RCTs notoriously have difficulty in monitoring the effects of screening over time. In contrast, observational or service screening trials, although they frequently observe significantly higher benefit rates in the screened arm, often stumble in controlling for lead-time and length biases, although it can be agreed that they remain more appropriate for monitoring and comparing the effects of various screening programs. No unilateral argument can be made for the superiority of either the RCT or the observational protocol design over the other; they have substantively different roles to play[21,22] and different capacities for handling overdiagnosis.

A recent important systematic review of studies on screening overdiagnosis[12], although concluding, after an evaluation of risk of bias and strength of evidence, that ecological and cohort studies—when conducted well (which, it noted, was uncommon)—stand as the most appropriate for monitoring and quantifying overdiagnosis. Nonetheless, of eighteen such breast cancer studies reviewed, only five ecological studies met minimal criteria of non-high-risk of bias, unbiased analysis, and fair-to-good time frames, while still being rated poor in consistency, with a not highly confidence-inspiring moderate risk of bias, leaving thirteen studies rated as failing even those minimal requirements. Under strict assessment as to the adequacy, integrity, and consistency of deployed methodology, there are no saints among RCTs, modelling studies, and observational or ecological or cohort studies; only the weight of all the aggregated relevant data, regardless of source and type, when systematically reviewed and critically appraised and methodology-score-assessed, can determine what the balance of the evidence actually determines.

Yaffe further expresses his quite reasonable reservations about the feasibility of another mammography screening RCT emerging to weigh in more decisively on residual controversies (reservations I shared by the qualification "if still feasible in this age" in my discussion). Still, at least some such trials appear to be in various degrees of planning, especially in the Middle East, and it remains to be seen whether sufficient recruitment will be an absolute barrier, especially in low and middle-income countries and amid populations with little enthusiasm and rather deep skepticism for screening (such as Iran, among others).

Randomization Integrity and the Issue of Advanced Cancers

Yaffe reintroduces the issue, first argued in his exchange with Narod, of the effect of exclusion of prevalence-screen cancer deaths on the derived hazard ratio in CNBSS, an issue whose substance I did not in fact address, but only its exemplification of two arguably unresolvable competitive narratives. Here, I add some further perspectives.

First, the differential effects and assumptions for prevalence (first) compared with incidence (subsequent) screening rounds are complex and not typically well addressed or controlled. For example, screening length bias itself is variable and depends on detection occurring at prevalence or at incidence screening; and for biennial screening, the prevalence round detection of breast cancer can be more than twice that of subsequent incidence rounds[23]. In addition, despite the need for reliable assessment of the distribution of lead times in those cancers that would have become symptomatic, separate lead time assumptions for prevalence compared with incidence screening rounds are rarely seen in studies[24].

Second, and more to my point, is that the Narod–Yaffe exchange finds a parallel in concerns about the number of advanced (node-positive) cancers in the mammography arm compared with the usual care arm of the CNBSS trials. In the exchange between Dr. Anthony Miller and Dr. László Tabár, the latest response by Tabár[25] notes that there remains an excess of advanced breast cancers in the mammography arm compared with the control arm, and if the numbers derived by Miller's team are used, they unsurprisingly entail no mortality-reductive benefit. In contrast, with the Tabár numbers (derived from CNBSS data, with plausible extrapolation) and upon exclusion of the prevalence screen tumours, invitation to screening is associated with a 10% reduction in breast cancer mortality. But because the CNBSS trials are known to have been underpowered to detect any mortality reduction under 40%, the debate over

the duelling sets of advanced cancer numbers is unlikely to bear any clinically relevant fruit, however ingenuous the contending arguments. With the CNBSS trials powered at a 40% detection threshold, the proper conclusion is *not* that *no* significant mortality-reductive benefit is derivable from mammographic screening, only that none is discoverable in a study with a 40% relative risk reduction floor (a rather high floor). But, on the other side of the debate, it must also be acknowledged that, given those boundaries of study power, teasing out a 10% reduction from a study floored at 40%—and similarly noting a shift of 19% in the hazard ratio when prevalent-screen cancer deaths are excluded (as Yaffe notes), yet still failing statistical significance (acknowledged)—might perhaps not much advance the conversation, hanging too much weight on such a slender thread. Admittedly, the decisive settlement of these "number" disputes would appear to be of less consequence to Yaffe and Tabár than is the implicit casting of further suspicion on the integrity of the randomization protocol used—not a trivial goal, but one that would be more convincing if chance were not lurking in the background as a viable alternative explanation.

However, the discussion serves indirectly to bring out an important observation: that besides its traditionally accepted benefits in detection of earlier-stage disease, mammography also contributes to substantial reductions in the incidence of advanced or metastatic breast cancer[26–34] as was noted in a review of mammography screening RCTs[35]. As Tabár's team demonstrated, trials reducing such advanced stage disease by at least 20% confer a 28% mortality reduction in screening-invited women, translating to an approximately 40% mortality reduction in screening-attendant women, compared with trials effecting a less than 10% reduction in advanced cancer, which were associated with no significant breast cancer mortality reduction[27,36]. On this view, screening appears to be significantly mortality-reductive only if a substantial reduction in incidence of advanced cancers is secured. Because the CNBSS trials manifestly failed to achieve those thresholds, regardless of which set of numbers is used, it is unsurprising that no significant mortality-reductive effect was detected, an instance of trial power limits (but the CNBSS trialists did not claim greater power).

The State of the Mammography Debate, and Looking Forward

In some highly perceptive opening words, Yaffe reflects on the current state of the mammography debate and also expresses some skepticism about what he considers the potentially "murkier and more complex" issues about biomarkers, a greater focus on which I argued for in my invited editorial[2]. But I would hold that any divergences of opinion on biomarkers for non-evolutive cancers—and biomarkers in general of prognostic and predictive consequence—are different in both kind and degree from those afflicting the mammography debate. Such markers are independently desirable and have never generated much in the way of controversy upon proper validation through guideline authorities such as the American Society of Clinical Oncology (ASCO) and others, and the goal of achieving such markers is a recommendation of numerous leading evidenced-based guideline authorities. Think of *KRAS, EGFR* gene mutation (but not expression), *EML4–ALK,* the Oncotype DX genomic assay, and endocrine receptor status and levels, all approved across the board by the leading relevant guideline panels, including the American Society of Clinical Oncology (ASCO), the U.S. National Comprehensive Cancer Network (NCCN), the European Society for Medical Oncology (ESMO), and the European Group on Tumor Markers (EGTM). Although some opinion differences have been aired, they have not been of the quantitative order of the hundreds and thousands of studies on the mammography screening controversy, nor of the qualitative order of the often vociferous and strident exchanges in the mammography debate.

And here, the goal is minimization of harm from overdiagnosis and overtreatment through superior differentiation between malignancy-progressive subtypes of ductal carcinoma *in situ* (and atypias) and lesions lacking that potential invasivity[37,38]. A biomarker—whether one of pathology, tumour biology, molecular imaging, or molecular signature—that is reliably predictive of what I call tumour militancy, that can differentiate treatment-mandatory evolutive cancers (including DCIS) from non-evolutive ones, and that would be fundamentally reductive of overdiagnosis would carry a high value and hence would positively rebalance the benefit–harm ratio in favour of mammographic screening.

As to Yaffe's perspective on the current state of the mammography debate, it suggests that, at the level of properly rigorous scientific discourse, the debate should be over, and further, that few anti-screening partisans still doubting the mortality-reductive benefits of mammographic screening remain. "To God's ear," I would say. Many intelligent and dedicated professionals still see this otherwise, although in one sense, Yaffe and I are in agreement—namely, as to what the *weight* of the evidence determines once properly normalized (as detailed in my review[5]), systematically reviewed and critically appraised, and assessed as to quality of methodology (trial consistency and integrity, and metrics of screening persistence and compliance, among the many criteria discussed).

As to the list of the principal anti-screening holdouts, it is perhaps not quite as short (Dr. Peter Gøtzsche, Dr. Anthony Miller) as suggested, and includes Drs. Karsten Juhl Jørgensen, Per-Henrik Zahl, Archie Bleyer, Judith Walsh, H. Gilbert Welch, Mette Kalager, Melania Maria Ramos Amorim, Susan Bewley, Cornelia Barnes, Philippe Autier, and dozens of others (just of those active in the recent past). They, in turn, are in large agreement with the guideline authorities of Australia, Denmark, Finland, France, Ireland, Israel, Italy, Luxemburg, Netherlands, Norway, Portugal, and Switzerland (biennial screening of women 50± years of age); the United Kingdom (triennially for those 50–70 years of age, and under phased extension, 47–73 years of age); and Japan, Korea, New Zealand, Spain, and Taiwan (biennially and inclusive of at least part of the 40s age group)[39,40]; and further extended by guideline authorities such as the U.S. Preventive Services Task Force and the World Health Organization holding against annual 40± screening. For every seemingly decisive finding (Yaffe cites the Independent U.K. Panel on Breast Cancer Screening), there are countervailing critics, some

remarkably ingenuous[16]. My own sense is of no appreciable diminution in the "volume" and often the ingenuity (but not necessarily correctness) of the anti-screening advocates, who, like the pro-screening advocates, believe they are advocating for the greater good of screening-eligible women. So, on either side, there is no shortage of sincerity, or even ferocity.

Except for small conversions (Gøtzsche, who initially found no reliable evidence supporting breast screening[41], but later[42,43] concluded for a 15% mortality reduction), few opponents are convinced out of their fold. In this debate, there are armies of the faithful, and only a disappointing scattering of moderators and peacemakers. Despite, as I have argued, far more convergence than is otherwise apparent (as when using normalization strategies), few "see the light." It could be, as I originally expressed more cynically than is my typical temperament, that the mammography debate makes cynics of us all, if we are really paying attention, and full *resolution* of the debate might only come, as I first discussed[5], with its *extinction* through use of the new screening modalities of abbreviated breast magnetic resonance imaging (ABMRI), enriched by the ultraFAST and TWIST protocols, and of digital breast tomosynthesis (DBT). Here, promise is large and becoming measurable. For example, tomosynthesis is advancing strongly and steadily toward being a primary and not just an adjunct (to conventional mammography) screening modality. At some centres, such as the Lynn Sage Comprehensive Breast Center (Chicago, IL), 90% of patients are already being imaged with tomosynthesis, obsoleting much of what we now see as core issues of the mammography debate, realizing hopefully sooner rather than later what I optimistically posited in the title of my Perspectives paper: Beyond the Mammography Debate[5].

Constantine Kaniklidis
Research Director
No Surrender Breast Cancer Foundation[a]
PO Box 84, Locust Valley
New York 11560 U.S.A.
edge@evidencewatch.com

CONFLICT OF INTEREST DISCLOSURES

I have read and understood *Current Oncology*'s policy on disclosing conflicts of interest, and I declare that I have none.

REFERENCES

1. Yaffe MJ. Response to: "Beyond the mammography debate: a moderate perspective." *Curr Oncol* 2015;22:e401-3.
2. Kaniklidis C on behalf of the No Surrender Breast Cancer Foundation. Through a glass darkly: the mammography debate. *Curr Oncol* 2015;22:171–3.
3. Berry DA. Failure of researchers, reviewers, editors, and the media to understand flaws in cancer screening studies: application to an article in *Cancer*. *Cancer* 2014;120:2784–91.
4. Kolata G. In Spinal Test, Early Warning On Alzheimer's. *The New York Times* [New York edition] 2010; 9 August: A1. [Correction published 16 September 2010; available online with revised headline at: http://www.nytimes.com/2010/08/10/health/research/10spinal.html; cited 7 August 2015]
5. Kaniklidis C on behalf of the No Surrender Breast Cancer Foundation. Beyond the mammography debate: a moderate perspective. *Curr Oncol* 2015;22:220–9.
6. Eddy DM. Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, England: Cambridge University Press; 1982: 249–67.
7. Ghosh AK, Ghosh K, Erwin PJ. Do medical students and physicians understand probability? *QJM* 2004;97:53–5.
8. Hoffrage U, Koller M. Chances and risks in medical risk communication. *Ger Med Sci* 2015;13:Doc07.
9. Taylor P. Breast cancer screening. *London Review of Books* 2014;36:30–32. [Available online at: http://www.lrb.co.uk/v36/n11/paul-taylor/breast-cancer-screening; cited 30 June 2015]
10. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;339:b2587.
11. Paci E on behalf of the EUROSCREEN Working Group. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen* 2012;19(suppl 1):5–13.
12. Carter SM, Rogers W, Heath I, Degeling C, Doust J, Barratt A. The challenge of overdiagnosis begins with its definition. *BMJ* 2015;350:h869.
13. Puliti D, Zappa M, Miccinesi G, Falini P, Crocetti E, Paci E. An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. *Eur J Cancer* 2009;45:3166–71.
14. Puliti D, Miccinesi G, Zappa M, Manneschi G, Crocetti E, Paci E. Balancing harms and benefits of service mammography screening programs: a cohort study. *Breast Cancer Res* 2012;14:R9.
15. Falk RS, Hofvind S, Skaane P, Haldorsen T. Overdiagnosis among women attending a population-based mammography screening program. *Int J Cancer* 2013;133:705–12.
16. Njor SH, Olsen AH, Blichert-Toft M, Schwartz W, Vejborg I, Lynge E. Overdiagnosis in screening mammography in Denmark: population based cohort study. *BMJ* 2013;346:f1064.
17. Coldman A, Phillips N. Incidence of breast cancer and estimates of overdiagnosis after the initiation of a population-based mammography screening program. *CMAJ* 2013;185:E492–8.
18. Beckmann KR, Lynch JW, Hiller JE, et al. A novel case–control design to estimate the extent of over-diagnosis of breast cancer due to organised population-based mammography screening. *Int J Cancer* 2015;136:1411–21.
19. Beckmann K, Duffy SW, Lynch J, Hiller J, Farshid G, Roder D. Estimates of over-diagnosis of breast cancer due to population-based mammography screening in South Australia after adjustment for lead time effects. *J Med Screen* 2015;:[Epub ahead of print].
20. Feig SA. Overdiagnosis of breast cancer at screening is clinically insignificant. *Acad Radiol* 2015;22:961–6.
21. Harris RP. How best to determine the mortality benefit from screening mammography: dueling results and methodologies from Canada. *J Natl Cancer Inst* 2014;106:pii:dju317.
22. Irvin VL, Kaplan RM. Screening mammography & breast cancer mortality: meta-analysis of quasi-experimental studies. *PLoS One* 2014;9:e98105.
23. Chamberlain J, Clifford RE, Nathan BE, Price JL, Burn I. Repeated screening for breast cancer. *J Epidemiol Community Health* 1984;38:54–7.
24. Cox B, Sneyd MJ. Bias in breast cancer research in the screening era. *Breast* 2013;22:1041–5.

[a] The No Surrender Breast Cancer Foundation is a U.S.-based 501(c)3 not-for-profit organization providing high-quality critically reviewed and appraised information and guidance to the breast cancer community.

25. Tabár L, Chen HH, Yen AM, *et al.* Response to Miller *et al. Breast J* 2015;21:459–61.
26. Gangnon RE, Sprague BL, Stout NK, *et al.* The contribution of mammography screening to breast cancer incidence trends in the United States: an updated age-period-cohort model. *Cancer Epidemiol Biomarkers Prev* 2015;24:905–12.
27. Tabár L, Yen AM, Wu WY, *et al.* Insights from the breast cancer screening trials: how screening affects the natural history of breast cancer and implications for evaluating service screening programs. *Breast J* 2015;21:13–20.
28. Foca F, Mancini S, Bucchi L, *et al.* on behalf of the IMPACT Working Group. Decreasing incidence of late-stage breast cancer after the introduction of organized mammography screening in Italy. *Cancer* 2013;119:2022–8.
29. Barchielli A, Paci E. Trends in breast cancer mortality, incidence, and survival, and mammographic screening in Tuscany, Italy. *Cancer Causes Control* 2001;12:249–55.
30. Hofvind S, Sørum R, Thoresen S. Incidence and tumor characteristics of breast cancer diagnosed before and after implementation of a population-based screening-program. *Acta Oncol* 2008;47:225–31.
31. Anttila A, Sarkeala T, Hakulinen T, Heinävaara S. Impacts of the Finnish service screening programme on breast cancer rates. *BMC Public Health* 2008;8:38.
32. Swedish Organised Service Screening Evaluation Group. Effect of mammographic service screening on stage at presentation of breast cancers in Sweden. *Cancer* 2007;109:2205–12.
33. Schouten LJ, de Rijke JM, Huveneers JA, Verbeek AL. Rising incidence of breast cancer after completion of the first prevalent round of the breast cancer screening programme. *J Med Screen* 2002;9:120–4.
34. Fracheboud J, Otto SJ, van Dijck JA, Broeders MJ, Verbeek AL, de Koning HJ on behalf of the National Evaluation Team for Breast Cancer Screening. Decreased rates of advanced breast cancer due to mammography screening in the Netherlands. *Br J Cancer* 2004;91:861–7.
35. Smith RA, Duffy SW, Gabe R, Tabár L, Yen AM, Chen TH. The randomized trials of breast cancer screening: what have we learned? *Radiol Clin North Am* 2004;42:793–806.
36. Smith RA, Manassaram-Baptiste D, Brooks D, *et al.* Cancer screening in the United States, 2015: a review of current American cancer society guidelines and current issues in cancer screening. *CA Cancer J Clin* 2015;65:30–54.
37. Partridge AH, Elmore JG, Saslow D, McCaskill-Stevens W, Schnitt SJ. Challenges in ductal carcinoma *in situ* risk communication and decision-making: report from an American Cancer Society and National Cancer Institute workshop. *CA Cancer J Clin* 2012;62:203–10.
38. van Ravesteyn NT, Stout NK, Schechter CB, *et al.* Benefits and harms of mammography screening after age 74 years: model estimates of overdiagnosis. *J Natl Cancer Inst* 2015;107:pii:djv103.
39. Smith RA. International programs for the detection of breast cancer. *Salud Publica Mex* 2011;53:394–404.
40. Ponhold L, Bickel H, Pinker K, Helbich TH. Mammography screening and follow-up of breast cancer. *Hamdan Med J* 2012;5:5–17.
41. Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000;355:129–34.
42. Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2011;(1):CD001877.
43. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2013;6:CD001877.