

**COMMENTS AND RESPONSES**

**Effects of Three Therapies for Neck Pain**

**TO THE EDITOR:** In their study of manual therapy, physical therapy, and general practitioner care for patients with neck pain, Hoving and colleagues concluded, “In daily practice, manual therapy is a favorable treatment option for patients with neck pain compared with physical therapy or continued care by a general practitioner” (1). We consider this an overstatement of the significance of their results. We note that although the manual therapy outcomes showed “higher scores,” scores were later referred to as “not statistically different.” We also see terms like trend, a way of saying that the results did not impress the statistician but at least are pointing in the right direction. Strictly speaking, such trends could simply be the result of some unknown or random error. Hoving and colleagues stated, “The success rates for manual therapy were statistically significantly higher than those for physical therapy.” However, they then stated, “Manual therapy scored better than physical therapy on all outcome measures, although not all differences were significant.” It seems that to appreciate the “favorable” effect of manual therapy, one has to be very careful about which outcomes are selected, since some will not show differences but arbitrarily defined “success rates” will. Hoving and colleagues noted that patients receiving manual therapy had fewer work absences, but then they stated that this was not significant. Again, such an observation might be attributable to random or some unknown error.

While significant differences might increase a study’s chances of being published, statements like “Our results consistently favored manual therapy on almost all outcome measures” are never justified. Such statements no longer refer to statistical or clinical significance but rather address the more nebulous quality “consistently favored.” While heeding Bacchetti’s fair warning for reviewers (2), we believe Hoving and colleagues would have done well to mind their methods.

We wonder what would have happened to the results of this study if the patients were followed for 8 to 12 weeks. If the groups had equalized even more by then, and all of the “consistently favored” outcomes were lost, would the cost of manual therapy compared with general practitioner care have been worth it? Chronic pain costs society a great deal. Part of that cost is spent on therapies that remain unproven.

Oliver Kwan, PhD (Psy)
Jon Friel, PhD (Psy)
Edmonton, Alberta T5H 2L8, Canada

**References**


**IN RESPONSE:** The question at hand is the way in which one should interpret statistically nonsignificant results. Should one take the common point of view that without statistical significance there is no effect, or the more Bayesian view that combines new evidence of the magnitude and statistical significance of the effect with its prior probability (1)? In the latter case, the P value is less sacrosanct and the results of other outcome measures are taken into account. It is clear that we took the latter viewpoint in our paper. Although we agree with Drs. Kwan and Friel that our conclusion did not specifically refer to the size and potential uncertainty of the effect, we hold to our conclusion that manual therapy appears to be more effective than general practitioner care and probably than physiotherapy.

We can reassure Drs. Kwan and Friel that the effect of manual therapy persists after 6 weeks. Our recent cost-effectiveness analysis (2) shows that the total costs of manual therapy over 12 months were statistically significantly lower than the costs of physiotherapy and general practitioner care, confirming that manual therapy is a cost-effective intervention in the long term.

In our trial, we chose perceived recovery as a primary outcome for the very reason that it is a multifactorial scale, summing the effects of pain and function and other aspects that patients may consider important. Drs. Posner and Glew (3), in their accompanying editorial, equated perceived recovery to satisfaction and to nonspecific effects. We strongly disagree. Patients include in their perceived recovery score specific effects, such as pain reduction and functional improvement. Our trial was designed as a pragmatic trial, comparing three frequently applied interventions in primary care, because we were interested in the total effect (both specific and nonspecific) of these interventions. To study only specific effects, a placebo therapy is required. A double-blind trial of these interventions for neck pain would be extremely difficult to perform and has no relevance for primary care.

We have no idea why Drs. Posner and Glew consider us to be advocates of manual therapy. We are independent scientific researchers with epidemiologic backgrounds and a tradition of studying the effectiveness of frequently used primary care interventions for musculoskeletal disorders (4, 5).

Jan Luca Hoving, PT, PhD
Cabrini Medical Centre
3144 Victoria, Australia

Bart W. Koes, PhD
Erasmus Medical Center Rotterdam
3000 DR Rotterdam, the Netherlands

Henrica C.W. de Vet, PhD
Vrije Universiteit Medical Centre
1081 BT Amsterdam, the Netherlands

**References**


Subgroup Variation in Diagnostic Test Evaluation

TO THE EDITOR: The article by Mulherin and Miller on subgroup variation in diagnostic test evaluation is an important contribution (1). Although test performance may truly vary across subgroups, this variation might also be spurious—reference test bias masquerading as spectrum effect.

Whenever the reference test—the gold standard—is imperfect (a common occurrence), test characteristics (sensitivity, specificity, and likelihood ratios) will vary across subgroups that differ in prevalence of disease (2–4). Sensitivity will spuriously appear higher and specificity will spuriously appear lower in the patient subgroups with the higher prevalence, merely because the reference test is imperfect. This pattern assumes that the errors of the test and the reference standard are independent; if not, then the variation of sensitivity and specificity with prevalence can have a different pattern.

In their classic article, Lachs and colleagues (5) described spectrum bias in the dipstick test for urinary tract infections. They compared the test characteristics of the urine dipstick in two subgroups that differed greatly in disease prevalence (50% vs. 7%). In the high-prevalence subgroup, the sensitivity was markedly higher (92% vs. 56%) and the specificity was markedly lower (42% vs. 78%). Although the authors attributed this variation in test performance to spectrum effect, a plausible alternative explanation is that the variation was spurious, generated by an imperfect reference test (midstream urine culture; presence of “disease” defined as \( >10^5 \) colony-forming units/mL).

In their Table 1, Mulherin and Miller described a hypothetical example of spectrum effect that could not be attributed to reference test bias because the prevalence in the two subgroups was the same. An imperfect reference test cannot create spurious variation in test characteristics unless subgroups vary in prevalence of disease. In their clinical example of enzyme immunoassay for Chlamydia trachomatis, Mulherin and Miller described variation in test characteristics by age group but no variation by clinic type. The authors could have ruled out the possibility of reference test bias as an alternative explanation if the prevalence of disease was the same across age groups and different across clinics. But if prevalence of disease varied by age group, and if the reference standard (ligase chain-reaction assay) was imperfect (4), then reference test bias might have been masquerading as spectrum effect.

Mulherin and Miller found that enzyme immunoassay had a big difference in sensitivity but only a tiny difference in specificity across age groups. Such differences could occur if the reference test had imperfect specificity but perfect sensitivity and if the prevalence of disease was greater in the younger age group. Alternatively, the false-positive errors of the two tests (enzyme immunoassay and ligase chain-reaction assay) might be correlated.

It would be helpful if Mulherin and Miller could share information on the prevalence of disease among the various subgroups and information on any imperfections in their reference standard. More important, can they advise us on how we can avoid clinical error by distinguishing between true spectrum effect and spurious variation due to an imperfect reference standard?

Arthur T. Evans, MD, MPH
Cook County Hospital and Rush Medical College
Chicago, IL 60612

References

TO THE EDITOR: Mulherin and Miller (1) stated that “correct diagnosis is nearly as important as correct treatment.” The context for this statement was the clinical trial. I appreciate this context yet remain concerned about the ranking implied in Mulherin and Miller’s conclusion. Is correct treatment “more important” than correct diagnosis? Fundamentally, how does one correctly treat an incorrect diagnosis?

The Annals Information for Authors (2) describes Academia and Clinic papers as “descriptions and evaluations of innovations in concepts and procedures in medical education and training and in clinical practice.” A concern is that Mulherin and Miller’s nuanced statement suggesting a secondary role of diagnosis could be seen as an innovation in a concept or as medical education. I hope that neither implication was the authors’ intent. The fundamental principle is that correct treatment is predicated on correct diagnosis. The clinical trial setting, or any other setting, should not change that.

Ronald R. Scobbo, MD
West Virginia University
Morgantown, WV 26505-9190

References

IN RESPONSE: Dr. Evans raises an important point regarding the potential of reference test bias to masquerade as spectrum effect. Apparent spectrum effect may be observed under three conditions: 1) Test performance may actually differ across subgroups; 2) test performance may appear to differ across subgroups because of reference test bias and variation in prevalence; and 3) test performance may actually differ across subgroups but the observed effect may be exaggerated (or potentially reduced) by reference test bias and variation in prevalence. The magnitude of the bias in the last two circumstances depends on the prevalence in the subgroups, the performance of the new test and the reference test in the subgroups, and the presence of conditional dependence between the two tests. If the reference test is
imperfect, spectrum effects may be present for both the new test and the reference test.

In our clinical example, we did not attempt to account for reference test bias. Using ligase chain reaction assay as our reference standard for chlamydial infection, we found that the measured prevalences in the sexually transmitted disease and family planning clinic settings were 11.0% and 7.8%, respectively. By age group, the measured prevalence was 12.5% among women 24 years of age or younger and 3.3% among women older than 24 years of age. Ligase chain reaction assay is not a perfect reference test; its estimated sensitivity is approximately 90%, and its estimated specificity is approximately 99.0% (1).

For the clinician, one key to discriminating spectrum effect from reference test bias is the biological plausibility of the reported spectrum effect. For example, in many infectious diseases, asymptomatic infections may have lower organism burden and, consequently, tests may have lower sensitivity. For the clinical researcher, the potential magnitude of the reference test bias could be addressed in sensitivity analyses by using algebraic adjustment for the reference test bias under different conditions. Alternatively, methods such as latent class analysis that do not require a reference standard may be used to assess test performance in subgroups.

Dr. Scobbo’s point that correct diagnosis is an essential precursor to correct treatment is also well taken. The precise diagnosis of a disease should enhance treatment efficacy, although in some circumstances syndromic management is sufficient to treat the condition effectively. However, we concur with Dr. Scobbo’s conclusion that applying a rigid hierarchy that universally prioritizes either treatment or diagnosis would subvert both the art and the science of medicine.

Stephanie A. Mulherin, MSPH
William C. Miller, MD, PhD, MPH
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599

Reference

Postmenopausal Hormone Replacement Therapy

TO THE EDITOR: I was disappointed that in their review of postmenopausal hormone replacement therapy (HRT) and primary prevention of cardiovascular disease (CVD), Humphrey and colleagues (1) drew no distinction between treatment with conjugated equine estrogens, medroxyprogesterone acetate, and other semi-synthetic estrogen and progestin drugs, none of which are human hormones, and replacement of female hormones with molecule-for-molecule bioidentical copies of estriol, estradiol, and progesterone, a process commonly called natural hormone replacement. The article may have demonstrated that equine hormones and progesterone-like drugs can be harmful, but it said nothing about natural hormone replacement.

I feel it is a disservice to women who seek relief from menopausal symptoms and who seek the myriad potential health benefits of estrogen therapy or hormone replacement, including possible prevention of CVD, to state that “HRT use did not reduce CVD incidence and, in fact, suggests a small increase in risk.” By calling conjugated equine estrogens, medroxyprogesterone acetate, and such drugs “estrogen” and “progestrone,” as most authors seem to do so loosely, Humphrey and colleagues confused the issue and condemned natural hormone replacement by association. They did not acknowledge that bioidentical HRT with “real” estrogen and progesterone was not studied and may provide the benefits of HRT with fewer side effects and risks. Large studies of bioidentical hormone replacement are needed but are unlikely to be done because pharmaceutical companies, which fund the bulk of medical research, will find it difficult to market and profit from these human hormones. Perhaps the National Institutes of Health or another independent body will sponsor such research.

Steven J. Meyerson, MD
Miami, FL 33173

Reference

TO THE EDITOR: Humphrey and colleagues (1) somewhat blindly accept HRT and treatment with Premarin (Wyeth, Madison, New Jersey) as being interchangeable. They clearly are not, and the differences are sufficient to cause questioning of all of Humphrey and colleagues’ data. Premarin is conjugated equine estrogen. The estrogenic compounds are not native to humans. Some are more active in estrogenic assays, and some are less active. Actions on other pathways are not well known and may account for some of the events seen with Premarin. Drawing conclusions and basing therapeutic decisions on data from Premarin may have been acceptable when the native human estrogen, estradiol, was not available, but it is uncons- tructible in the 21st century. One would not consider basing treatment standards for thyroid replacement on data from trials using desiccated porcine thyroid anymore, although this was the standard for a long time. The same should be true with estrogen replacement. To draw valid conclusions, one should use data derived from true replacement.

Premarin is usually taken orally, while natural estrogen is released directly into the circulation. This has several ramifications. The most obvious is the first-pass effect in the liver. For example, compelling data indicate that C-reactive protein level is increased by oral Premarin and estradiol, but not by transdermal estradiol. All of the ramifications of the first-pass effect are unclear but may account for at least some of the adverse events seen with the two former agents. The possible confounding of data by nonphysiologic administration cannot be discounted.

More subtle is the time-constant effect. Oral preparations cause peaks and valleys of estrogen exposure, while transdermal administration results in an even, measurable blood level that allows in vivo dosages to be accurately assessed and compared. Clearly, time-constant assessment was not performed in any of the studies that Humphrey and colleagues reviewed.

There are many differences among Premarin and various other estrogen preparations. One must be very careful when drawing conclusions based on data that may be flawed, perhaps fatally. Merely because studies are large, have statistical significance, and used what was, rightly or wrongly, the standard at the time does not mean that the resulting data will always be acceptable.

For the record, I have no financial or professional interest in any

Downloaded From: http://annals.org/ on 10/27/2018


Arthur B. Chauser, MD, PhD
Dorn Veterans Affairs Medical Center
University of South Carolina
Columbia, SC 29209

Reference

Meir J. Stampfer, MD
Harvard School of Public Health
Boston, MA 02115

References

IN RESPONSE: We appreciate the opportunity to again discuss the limitations of the observational literature evaluating hormone therapy and CHD in women. Before the publication of the Heart and Estrogen/progestin Replacement Study in 1998, postmenopausal hormone therapy was widely recommended to women to prevent CHD, largely on the basis of data from observational studies. However, ascertainment of hormone therapy use in these studies was limited, as we discussed and described in Appendix Tables 1 and 2 of our paper. Very few studies evaluated the type of estrogen used, whether progestins were used, or the dose of either formulation. Despite limitations, these data were used to make widespread recommendations to women about the likelihood of CHD protection associated with hormone therapy. Our meta-analysis was an attempt to evaluate the observational studies judged most likely to provide valid findings and to determine whether, in the aggregate, the data were sufficient and of high enough quality to support this widespread recommendation. Our findings suggested that they were not. We agree with Drs. Meyerson and Chausmer that even among the best studies, very few data are available on the type of estrogen used and that generalizations about all estrogen or progestin formulations are inappropriate. However, we strongly believe that any recommendations of hormone therapy in women for the prevention of CHD should be made only on the basis of randomized, controlled trials. The concerns of Drs. Grodstein, Manson, and Stampfer are addressed in another publication (1). The important points of that response are as follows. First, the level of education sufficient to become a registered nurse can vary from nursing school diploma to bachelor’s degree to advanced degree, a range that is very likely to be associated with important variation in socioeconomic status. Thus, we did not include the Nurses’ Health Study in our original meta-analysis of the articles that accounted for socioeconomic status. We would have done so if we had assumed that including only registered nurses is equivalent to statistical adjustment for socioeconomic status. When we did include the Nurses’ Health Study, the summary figures changed only minimally, as described in more detail elsewhere (1).

Second, when we included studies that adjusted for either alcohol use or physical activity in our meta-analysis (including the Nurses’ Health Study), the pooled estimates did not suggest CHD.
benefit. The pooled estimates are reported elsewhere (1). Third, animal data and intermediate outcomes can suggest hypotheses or potential biological mechanisms of action, but only randomized, controlled trials of estrogens with or without progestins can prove that hormone therapy is cardioprotective.

Linda L. Humphrey, MD, MPH
Oregon Health & Science University
Portland Veterans Affairs Medical Center
Portland, OR 97201

Benjamin K.S. Chen, MS
Oregon Health & Science University
Portland, OR 97201

Harold C. Sox, MD
American College of Physicians
Philadelphia, PA 19106

Reference

Screening for Postmenopausal Osteoporosis

TO THE EDITOR: The U.S. Preventive Services Task Force’s unqualified recommendation of routine osteoporosis screening in women older than 80 years of age (1) does not seem justified by available evidence, including that presented by Nelson and colleagues (2). Although only one bisphosphonate, risedronate, has been evaluated in large numbers of women older than 80 years of age (3), no interventions have been shown to decrease the frequency of hip fractures in this group. Unfortunately, routine screening recommendations from a respected source are very likely to be interpreted as the standard of care by some physicians, institution administrators, practicing attorneys, and government agencies that monitor the quality of health care. This makes the related decisions of physicians interested in providing cost-effective medical care for their patients more difficult.

The fracture risk in these elderly women is obviously high, but bone mineral density measurements may be less likely to anticipate hip fractures among them than among younger women because of other factors contributing to fracture rate, such as muscle weakness, vertigo, poor balance, arrhythmias, postural hypotension, blindness, environmental hazards, and other age-related phenomena. In this setting, medications designed to decrease fracture rate by increasing bone mineral density might be expected to be relatively ineffective.

Women in their 80s who are physiologically younger than their peers clearly deserve both screening for osteoporosis and appropriate therapy when indicated. Routine screening of women in this age group, however, would assign a diagnosis of osteoporosis to a large percentage and would result in great societal pressure for treatment with medications not proven to be of benefit.

Harry W. Daniell, MD
University of California, Davis, Medical School
Redding, CA 96001

References

IN RESPONSE: There is indeed much less evidence on the effectiveness of bisphosphonates for women older than 80 years of age than for younger women. A trial of alendronate in elderly women in long-term care facilities (mean age, 78.9 years; range, 65 to 91 years) reported increased bone density at the spine and hip (1). Two analyses from the Fracture Intervention Trial (FIT) of alendronate indicated fracture benefit for older women. After an average of 2.9 years of follow-up, the relative risk for new vertebral fractures was 0.49 (95% CI, 0.35 to 0.68) for women younger than 75 years of age and 0.62 (CI, 0.41 to 0.94) for women 75 years of age and older. There was no apparent interaction between treatment and age (P > 0.2) (2). Risks for other clinical fractures were also reduced in both groups, although the CI crossed 1.0 for older women. Another analysis of FIT data after 4.3 years of follow-up indicated that the effect of alendronate on the incidence of multiple fractures was not affected by age (<75 years of age vs. =75 years of age) (3).

A large trial of risedronate indicated a significant reduction in hip fractures for women ages 70 to 79 years (relative risk, 0.6 [CI, 0.4 to 0.9]), but not for those 80 years of age and older (relative risk, 0.8 [CI, 0.6 to 1.2]) (4). However, the older women were selected by age and risk factor criteria, and bone density measurements were available for only 31%. Women 80 years of age and older in this study may not be comparable to women 70 to 79 years of age, who were selected by bone density criteria and had mean femoral neck T-scores of −3.7.

Although data are limited for women older than 80 years of age and interpretations may vary, the available data suggest possible benefit for older women with low bone density. Trials of other interventions, such as calcium and vitamin D supplementation and use of external hip protectors, indicate fracture benefit in appropriate candidates. However, trials of strength and balance training, improving vision problems, reducing fall hazards, and other similar interventions are lacking and were not reviewed in our evidence report on screening for postmenopausal osteoporosis. Factors besides low bone density, such as changes in bone connective tissue and functional frailty, become increasingly important for fractures as women age. A thorough clinician will consider multiple clinical factors when assessing an elderly woman for fracture risk and will design an individual management plan that may or may not include medical therapies.

Heidi D. Nelson, MD, MPH
Oregon Health & Science University
Portland, OR 97201

References
Breast Cancer Screening

TO THE EDITOR: Dr. Sox provided a concise summary of some of the issues surrounding the controversy about recommending mammographic screening for women 40 to 49 years of age (1). However, he forgot one of the most basic issues that continues to be ignored by those who oppose mammographic screening for women before the age of 50 years, namely, that this age has no biological or screening significance. There are, in fact, no data showing that anything biologically relevant occurs at age 50. None of the variables associated with screening (for example, recall rates, recommendations for biopsy, cancer detection rates, tissue density) change abruptly at that time (2). This age was initially chosen as a surrogate for menopause, but there are also no data that directly show any relationship between screening results and menopause.

The age of 50 years has been made to appear to have some significance because researchers have taken data that change gradually with increasing age and have analyzed them dichotomously around this particular time point (≤49 years of age vs. ≥50 years of age) (3). Changes that actually occur with aging are made to appear as if they occur suddenly at the age of 50 years (4), causing even “experts” to be misled (5). I suspect that some who discuss and write on the issue are aware that the age of 50 is arbitrary, but women and their physicians have been misled to believe that this age really has some meaning.

Dr. Sox’s admonition to keep women “informed as best we can” should be heeded. Women and their physicians should be informed that beginning mammographic screening at the age of 50 is purely arbitrary, with no scientific support, and that the screening data show a benefit from screening beginning at the age of 40.

Daniel B. Kopans, MD
Massachusetts General Hospital and Harvard Medical School
Boston, MA 02468

References

IN RESPONSE: Dr. Kopans raises a familiar question: Why do people refer to age 50 as a significant age for breast cancer screening when it has no biological significance? The answer is that this age is a proxy for the occurrence of natural menopause. Ideally, the clinical trials of breast cancer screening would have reported their results according to menopausal status at the time of screening but, to my knowledge, they did not. Dr. Kopans points out that there are no data that indicate a relationship between screening results and menopause. The reader should note that “no data” does not mean “no relationship.” Until someone tests the hypothesis that results differ by menopausal status, we will have to use age 50 as a proxy for menopause.

My editorial asserted that when the Canadian study (1) and the meta-analysis of the Swedish studies (2) counted breast cancer deaths in the same way, both showed a small effect of screening, with 95% CIs that included no effect. Dr. Kopans ends his letter by saying that we should inform women that screening is effective beginning at age 40, but he does not present evidence to contradict my conclusion and support his own.

Harold C. Sox, MD
Editor

References

Physicians and Patient Spirituality

TO THE EDITOR: I was astonished by Dr. Graner’s suggestion that it is “inappropriate” for religious icons to hang in individual physicians’ offices and “clearly inexcusable” to mention the word heaven to a sick child’s mother (1). Dr. Graner’s premise is that a physician’s office space is a site of medical care, representing “a segment of the secular, public space” (2), and reference to religion should be avoided because the office is not “uniquely personal” (3).

Ironically, the very same Annals issue in which Dr. Graner’s most recent letter appeared included an article titled “Do Ask, Do Tell,” which was written by a lesbian physician who openly displays her orientation by “having gay-friendly posters” in her office (4). The author stated, “I have gradually learned that it is advantageous to be open about my sexual orientation. Disclosure is empowering: It allows me to be myself, integrate my public and private lives, voice my opinions, celebrate all of my achievements, and work passionately to increase tolerance and acceptance.” On the other hand, being silent would be a denial of who she is, making “self-respect difficult to preserve” and limiting opportunities for collaboration “with sympathetic people from the mainstream who share [her] values and goals.”

Neither the lesbian physician nor the physician with religious icons in her office is necessarily ramming her beliefs down patients’ throats (although the former inches more than the latter in that direction). However, the implication is that it is somehow “correct” for a homosexual physician to nonverbally communicate who he or she is through office adornments but not “correct” for the Christian

References
physician or physicians of other faiths to do so by similar means. Is there a double standard here? Both Christianity and homosexuality are “lifestyles” that make up part of the fabric of the individual. A physician’s office space, although an extension into the secular world, remains an extension of the individuality of the physician. And if an office can be used to promote openness to the lesbian message, with how much more zeal should it be used to promote openness to the Christian message?

What Dr. Graner implies about religious icons goes far beyond the rights of physicians to adorn their offices. One of the hospitals in which I practice is Catholic and has religious icons almost everywhere. The nuns walk the halls and care for the sick while wearing habits. The operating room in which I do surgery has a crucifix in it. Shall the crucifixes be taken down and the nuns take off their habits because the hospital is dispensing secular, public care? Of course not! For the nuns and the other Catholic caretakers, the Catholic hospital is an extension of who they are, and the icons and symbols are an extension of the love they have for the sick and less fortunate. Few would take issue with the manner in which Mother Teresa conducted her public works of mercy in Calcutta. Taking the absurd further, should hospitals be forbidden from having Christmas trees in their very public and secular lobbies? Only a Scrooge or a Grinch could prohibit that! Just how cold and heartless do we intend to make the medicine of the future? Nothing has done more to disfigure the beautiful art of medicine than such nihilistic atheism.

Fritz Baumgartner, MD
Harbor/UCLA Medical Center
Torrance, CA 90509

References

IN RESPONSE: The majority of Dr. Baumgartner’s letter pertains to the office display of “gay-friendly posters,” which is a subject entirely separate from that addressed in my letters published over the years. Obviously, one’s sexual orientation is not the same as one’s religious orientation. If he somehow thinks he is referring to anything I have said, Dr. Baumgartner is sparring with an imaginary partner when he states, “The implication is that it is somehow ‘correct’ for a homosexual physician to nonverbally communicate who he or she is through office adornments.” I don’t agree with this practice any more than I do with the display of religious icons in the office. Why in the world should the physician use the office as a tool to “promote openness to the lesbian message,” the “Christian message,” or the elements of any other personal value system? Why not instead make the comfort of the average patient one’s major priority as a physician?

The issue of the display of religious icons in a hospital with an obvious religious affiliation is an interesting one, and I am surprised that it has not been raised previously in these discussions (I have to admit I was relieved that it wasn’t!). A “middle of the road” stance is that since this type of hospital always displays its religious heritage to the public in an obvious fashion, those who choose not to obtain care in such a setting receive “fair warning” and can go elsewhere. The issue becomes more problematic when an element of coercion (such as the offering of a service not easily obtained elsewhere) is involved, since no member of the public should be forced to receive care in a religious setting not of his or her choosing. The policy of one of our own religiously affiliated hospitals is to allow the patient to decide whether religious symbols will hang in her or his hospital room. The display of such icons in the corridors is to be expected and has never been an issue, nor has the religious dress of the workers, as long as the institution proclaims its religious affiliation in a publicly obvious manner. I realize this is not a perfectly satisfactory resolution of the issue, and that, I suppose, is why the debate continues.

John Graner, MD
Mayo Clinic
Rochester, MN 55905

RESEARCH LETTER

Hepatitis B Virus Reactivation after a Single Session of Transarterial Chemoembolization in Patients with Hepatocellular Carcinoma

TO THE EDITOR: Background: Hepatitis B virus (HBV) reactivation is a well-recognized complication in patients with cancer receiving systemic cytotoxic chemotherapy (1) and may cause varying degrees of liver damage. Immunosuppression due to cytotoxic therapy enhances viral replication and infection of the hepatocytes (2). Withdrawal of chemotherapy leads to restoration of immune function, with a consequent potential risk for rapid destruction of infected hepatocytes. Reactivation of HBV has been reported mostly in hepatitis B surface antigen (HBsAg) carriers with lymphoma after systemic chemotherapy or bone marrow transplantation (3). In some cases, HBV reactivation has been severe and has resulted in hepatic failure and death (4).

Objective: To report three cases of HBV reactivation in patients with liver cirrhosis and hepatocellular carcinoma treated with a single session of loco-regional (hepatic) transarterial chemoembolization.

Case Report: Patient 1 was a 63-year-old man with hepatocellular carcinoma (6-cm lesion of the right liver lobe), liver cirrhosis (Child–Pugh class B-7), and no ascites or encephalopathy. Test results were positive for HBsAg and negative for hepatitis B e antigen (HBeAg) and HBV DNA. The patient’s first session of transarterial chemoembolization consisted of epirubicin, 80 mg, followed by injection of lipiodol, 10 mL, in one branch of the right hepatic artery. On day 70 after chemoembolization, the patient developed malaise and dark urine; blood analysis showed an aspartate aminotransferase (AST) level of 746 U/L, an alanine aminotransferase (ALT) level of 1182 U/L, and a total bilirubin level of 48 μmol/L (2.8 mg/dL). Serum HBV DNA was also detected (5600 × 10^6 genome equivalents/mL). Despite lamivudine therapy (100 mg/d), progressive liver failure with ascites and encephalopathy developed. The patient progressed to renal failure and coma and died on day 82.

Patient 2 was a 40-year-old man with multifocal hepatocellular carcinoma and liver cirrhosis (Child–Pugh class A-6). Tests for HBsAg and anti--hepatitis D virus were positive, but HBeAg and HBV DNA were not detected. The patient’s first session of transarterial chemoembolization consisted of epirubicin, 90 mg, and gel-
foam embolization of a branch of the right hepatic artery feeding the tumor. His AST level started to increase on day 54, and jaundice was detected on day 63 (bilirubin level, 77 µmol/L [4.5 mg/dL]; AST level, 586 U/L; ALT level, 943 U/L). Hepatitis B virus DNA was detected in serum (7400 × 10^6 genome equivalents/mL). Lamivudine was started at a dosage of 100 mg/d, and liver function rapidly improved afterward. On day 80, the bilirubin level was 22 µmol/L (1.3 mg/dL), the ALT level was 72 U/L, and the AST level was 76 U/L. Test results for HBV DNA were negative on day 90 after chemoembolization. The patient was later able to receive subsequent sessions of transarterial embolization without concomitant use of chemotherapy.

Patient 3 was a 55-year-old man with hepatocellular carcinoma (two lesions, the largest 4 cm in diameter) and liver cirrhosis (Child–Pugh class A-5). Test results were positive for HBsAg and HBV DNA (800 × 10^6 genome equivalents/mL), but HBeAg was not detected. The patient’s first session of transarterial chemoembolization consisted of 90 mg of epirubicin followed by gel-foam embolization. On day 40 after chemotherapy, the patient presented with jaundice (total bilirubin level, 127 µmol/L [7.4 mg/dL]) and a peak ALT level of 1200 U/L. HBV DNA level also increased significantly (12 000 × 10^6 genome equivalents/mL). Therapy with lamivudine, 100 mg/d, was started immediately. Subsequently, jaundice disappeared (bilirubin level on day 60, 24 µmol/L [1.4 mg/dL]) and the ALT level improved progressively to a level of 78 U/L on day 70. Test results for HBV DNA were negative (undetectable HBV DNA levels by polymerase chain reaction) on day 82 after chemotherapy.

**Conclusion:** This is one of the first reports showing that transarterial chemoembolization represents a major risk factor for HBV reactivation in HBsAg-positive patients with cirrhosis and hepatocellular carcinoma. Death and morbidity associated with HBV reactivation are more common in patients with preexisting chronic liver disease than in patients without liver disease. The evidence of efficacy of preemptive lamivudine therapy in other populations, such as recipients of bone marrow transplants and patients with lymphoma treated with chemotherapy (5), justifies the use of this drug in HBsAg-positive patients with hepatocellular carcinoma who receive chemoembolization.

Giovan B. Vizzini, MD
Angelo Luca, MD
Ignazio R. Marino, MD
Istituto Mediterraneo per i Trapianti e Terapie ad Alta Specializzazione
90134 Palermo, Italy

University of Pittsburgh Medical Center
Pittsburgh, PA 15213

**References**

**CORRECTION**

**Correction: Reversal of Warfarin-Induced Excessive Anticoagulation**

An article on reversal of warfarin-induced excessive anticoagulation (1) contained an error. On page 887, in the first full sentence of the second column, the range of doses of recombinant factor VIIa that led to adequate hemostasis was 15 to 20 µg/kg, not 15 to 20 µg.

**Reference**
Hepatotoxicity Associated with a Dietary Supplement

TO THE EDITOR: We read with interest the article by Favreau and colleagues (1) describing a series of patients with severe hepatotoxicity possibly caused by the dietary supplement LipoKinetix (Syntrax, Cape Girardeau, Missouri). This series involved seven patients, including one who developed fulminant hepatic failure. We report a case of a 24-year-old woman who developed fulminant hepatic failure after using LipoKinetix and subsequently required orthotopic liver transplantation. This is one of the first reported cases of liver failure requiring transplantation after use of this over-the-counter supplement.

A previously healthy 24-year-old woman was admitted to the hospital with hepatic failure. She had a 6-week history of fatigue and malaise and a 2-week history of jaundice and nausea. Laboratory tests revealed an aspartate aminotransferase level of 449 U/L, an alanine aminotransferase level of 528 U/L, a total bilirubin level of 221 μmol/L (12.9 mg/dL), an albumin level of 26 g/L, and an international normalized ratio of 2.34. The patient had been taking LipoKinetix for 3 months before the onset of her symptoms. Results of work-up for viral, autoimmune, and metabolic diseases, as well as other toxic agents, were negative. Over the next 2 weeks, the patient’s coagulopathy worsened and she began to develop hepatic encephalopathy. She underwent successful orthotopic liver transplantation on day 15. The resected liver specimen showed extensive severe necrosis with shrunken hepatocytes containing abundant lipid material admixed with numerous lymphocytes and macrophages. The patient has since made a full recovery and is leading an active life.

Brennan A. Scott, MD
Christoph Troppmann, MD
Lorenzo Rossaro, MD
University of California, Davis, Medical Center
Sacramento, CA 95817

Reference