Cardiac Monitoring Guidelines in Clinical Trials and Post-Approval Surveillance for Patients Exposed to Anti-Cancer treatments: Do the Data Support the Recommendations?

Michael S. Ewer, MD, JD, PhD
Nicolas L. Palaskas, MD
Jay Herson, PhD

Running Title: Cardiac Monitoring Guidelines and Recommendations

From the Department of Cardiology, Internal Medicine Division, The University of Texas MD Anderson Cancer Center, Houston, TX USA (MSE and NLP)

and

The Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD USA (JH)

Corresponding author: Michael S. Ewer, Department of Cardiology, Internal Medicine Division, UT MD Anderson Cancer Center, Pickens Tower Ste 18.500, 1400 Pressler Street, Houston Texas 77030; Tel +1 713 792-6242; email: mewer@mdanderson.org

Number of pages: 25
Number of tables: 1
Number of references: 23
Text word count without references: 3864
Abstract word count: 249
Introduction word-count 556
Discussion word-count 691
Section assignment: Cardiovascular
ABSTRACT

Concerns regarding cardiac adverse events during and after cancer care include contractile dysfunction, dysrhythmia, and inflammation. Clinical trials and practice guidelines may require or recommend sequential ejection fraction determinations for early recognition of contractile dysfunction, bio-marker screening where inflammation or contractile dysfunction could be anticipated, and multiple electrocardiograms with timings of cardiac intervals. In some instances, surveillance schedules used in clinical trial protocols have been incorporated in recommendations without revision or critical scrutiny. When adverse events are rare and interpretative parameters imperfect, false positive results may lead to delay or interruption of vital cancer treatment, may suggest that further cardiac testing be undertaken, and may add to patient anxiety. The risks of excessive monitoring also include inconvenience, and increased cost. This paper looks at areas where surveillance recommendations may be problematic. Specifically, ejection fractions, cardiac biomarkers, and electrocardiographic monitoring. Changes reported following surveillance monitoring of cancer patients using these parameters may reflect true adverse events or clinically relevant future risk, but interpretative uncertainty or true physiologic change that is unrelated to the drug in question should be considered. Clinicians may not be sufficiently aware of the degree to which reported changes may reflect surveillance artifacts. A balance that incorporates both the likelihood of an event that could be prevented along with clinical implications is suggested. The authors recognize that differentiating among these variables is not always possible yet advocate
for modifying surveillance schedules to balance the frequency and severity of events that can be mitigated, based on reliable data.

Significance Statement:

The authors’ concerns regarding the predictive value of surveillance initiatives are explored. Confounding factors and false positive results may add to the expense of cancer care and/or compromise optimal therapeutic initiatives.
INTRODUCTION

The realization that anti-cancer treatments could affect the heart adversely came of age in the 1970s with the recognition that anthracyclines caused heart failure (Ewer et al., 2011). With the paucity of effective agents available at that time, strategies were introduced to define a therapeutic range where oncologic benefit would be balanced with acceptable cardiotoxicity. Additionally, attempts were made to identify those who could tolerate higher cumulative dosages, and surveillance was undertaken with multiple cardiac biopsies and serial cardiac ultrasound evaluations; the concept of identifiable risk factors emerged. While various cardiac risk factors were identified in numerous publications, it is recognized now that any individual who has a history of prior cardiac damage of any kind, or who have a diminished threshold for cardiac damage is at increased risk for cardiac events when exposed to toxic levels of some agents (Ewer 2019). With the introduction of monoclonal antibodies and tyrosine-kinase inhibitors, several factors emerged. The monoclonal antibody trastuzumab initially was thought to have cardiotoxicity similar to that seen with the anthracyclines. Concerns regarding the widespread use, especially in the adjuvant setting, were expressed, perhaps brought to attention by the early report of 27% toxicity of trastuzumab when included in a protocol with an anthracycline (Feldman AM et al., 2000). It was subsequently recognized that the dose-dependent myocyte destruction seen with the anthracyclines did not occur with trastuzumab, and that in many instances the observed changes appeared to be reversible. The classification of Type I (directly destructive to the myocyte that corresponds to the cumulative dose administered) and Type II (not
directly toxic to the myocyte but associated with myocyte dysfunction by other, usually secondary mechanisms) was introduced (Ewer and Lippman, 2005). The field of cardio-oncology expanded as a result of these observations, as well as the desire of the healthcare teams to provide optimal balance between cardiac risk and oncologic efficacy.

Notwithstanding these initiatives, several enigmas are now apparent: first, there is considerable diversity in the reported incidence of cardiotoxicity and the clinical implications for those who have been identified as having had a cardiac event. While it is well recognized that some cardiac imaging techniques, as usually undertaken in the setting of cancer care, are imperfect, the question of whether the weaknesses in the methodology have been fully appreciated and appropriately integrated into the algorithms of surveillance must be asked. Additionally, one must ask if those who market agents that have been considered cardiotoxic, or the regulatory agencies that approve their use, have considered updating their recommendations? In some instances, surveillance algorithms that were employed in pre-approval protocols have remained unchanged. False positive results may be a confounding factor for cardiotoxicity diagnosis (Lorenzini et al., 2017). A further potentially troubling phenomenon is our increasing ability to obtain large amounts of data that attempt to quantify the true extent of cardiotoxicity. While the use of such databases provides important information, the inability to access and adjudicate individual medical records, may result in the reporting of meta-analyses that do not reflect the true incidence of
clinically relevant cardiac events. Finally, different thresholds have been used to define toxicity and this may, at least in part, explain the diversity of reported results.

This paper attempts to bring some of these concerns to the attention of those treating cancer patients with agents that may have true or perceived cardiotoxicity. It addresses concerns related to ejection fraction determinations derived from cardiac ultrasound, electrocardiogram rhythm and interval changes, and cardiac biomarkers.

THE EJECTION FRACTION DETERMINED BY CARDIAC ULTRASOUND

The threshold for cardiotoxicity evolved by clinicians wanting to define both a meaningful drop in ejection as well as a value that suggested a decline into an abnormal range. In a number of trials, the criteria for a cardiac event were defined as a decline of greater than ten percentage points to a level < 50%. This somewhat arbitrary criteria were not based on mathematical attempts to differentiate those with a meaningful decline definition; they were not uniformly incorporated in many trials making comparisons of reported instances of cardiac adverse events problematic. As an example, the criteria used in evaluation the cardiac safety of lapatinib in 3689 patients enrolled in clinical trials used the criteria of a decrease of ≥ 20% relative to baseline and below the institution's lower limit of normal (Perez et al., 2008).

Considerable concern was raised regarding many of the newer agents, and these concerns resulted in recommendations for cardiac monitoring. Ejection fraction
determinations were recommended every three months during treatment, and at regular post-treatment intervals with the threshold of toxicity defined as a decline in ejection fraction of ≥10 percentage points to a value of less than 50%. The early trials included restrictive exclusion criteria resulting in patients with any prior cardiovascular disease not being included. In “real-world” clinical practice, patients with a myriad of cardiovascular comorbidities receive these potentially cardiotoxic agents and the cardiac monitoring proposed in the clinical trials is not strictly followed (Chavez-MacGregor et al., 2015).

Interestingly, and a fact that led to further analysis was the observation that cardiotoxicity of single agent and combinations that incorporated drugs considered to have potential cardiotoxicity resulted in the reporting of events that was remarkably consistent and was approximately 3-4 percent when monitoring was limited to a baseline determination and 4 subsequent determinations (Table 1). As is expected, in instances where there is a possibility of false positive adjudications, increased surveillance points with additional ejection fraction determinations resulted in higher incidences of reported events. A modeling initiative was undertaken to see if the reported incidence might be due, at least in part, to a surveillance artifact, a concept initially considered more than 20 years ago (Ewer et al., 1999). The modeling initiative made two assumptions: first, that the reading of cardiac ultrasound was imperfect, and based on reported variance, the deviation from a true ejection fraction would follow a normal distribution, with standard deviation of 2.5 percentage points; two standard deviations thereby falling within ± 5 percentage points of the true value. The second
assumption was that physiologic variations would also contribute to deviations from a patient’s baseline value. These deviations would be related to true changes in cardiac ejection caused by endocrine, metabolic, or neurologic variation that were not directly related to the drug under surveillance, and while less precisely predicted from the literature, were assumed to be of an extent like that of interpretative variation. Assuming an agent that was non-cardiotoxic was administered and monitoring undertaken according to the schedules undertaken in a number of clinical trials, the likelihood of false positive results fell remarkably close to what was observed in some clinical trials and was estimated to 3.6% (95% confidence interval of 2.66–4.54%) (Ewer and Herson, 2018). The confidence intervals for expected false positive results and reported toxicity overlapped in the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-31 trial, the Herceptin Adjuvant (HERA) trial, the trastuzumab + pertuzumab arm of the Cleopatra trial, and Berenice trial. The reasons for lack of overlap in Breast Cancer International Research Group 006 trial (BCIRG-006) may have been due to the lack of a threshold ejection fraction (EF) cut-off resulting in higher reported toxicity from the trial. For example, a decline in EF from 66% to 55% met criteria for toxicity despite still having an EF in the normal range. Similarly, in the more recent SUCCOUR trial the rate of toxicity observed was higher than expected from false positive determinations alone, but this may be due to the definition of cardiotoxicity that included a drop of only 5 percentage points in the left ventricular ejection fraction for those patients who are symptomatic to a cut-off level of 55% (Thavendiranathan et al., 2021). The time may be ripe to ask to what extent the reported cardiotoxicity was due to surveillance artifact, as a repeat determination following a reported false-positive estimation of ejection fraction,
would be unlikely. If reported cardiotoxicity was in part or largely due to surveillance artifact that was not invariably reproducible, recovery might also, at least in part, be an artifact (Ewer and Swain, 2022). Interestingly, some clinical trials now require that an event only be adjudicated as positive if it confirmed with a second determination. There are now anecdotal cases where trastuzumab has been given in the setting of metastatic disease for periods more than 10 years, further supporting the concept that dose-dependent direct toxicity, i.e., Type I toxicity, is not present for this agent. Yet some clinical trials still incorporate monitoring schedules established decades ago, prior to our recognition of Type II or secondary toxicity, and before the mechanisms of toxicity, the imperfect nature of the methodology and the role of confounding factors were sufficiently understood.

THE 12-LEAD ELECTROCARDIOGRAM

Changes in the electrocardiogram may be a harbinger of impending cardiac adverse events. Alteration of the electrocardiographic intervals has become a matter of increasing scrutiny, perhaps propelled by incidences of sudden death resulting from *torsades de pointe* following arsenic trioxide administration and the concern of atrial fibrillation associated with ibrutinib (Westervelt et al., 2001, Essa et al 2022, Ganatra et al., 2018). Electronic measurement of the corrected QT and the PR interval as well as intraventricular conduction delays has been hugely important in identifying these abnormalities, however concerns persist. In an attempt to identify errors in the electronic measurement of intervals as well as to identify incidental dysrhythmia, consecutive tracings are often undertaken as a series of three taken over a brief interval.
Electrocardiograms with documentation of the corrected QT interval have been incorporated into clinical trials. Some of these trials require triplicate electrocardiograms which entails performing three subsequent six second 12-lead electrocardiograms at an interval of one to five minutes between recordings. A number of trials require this as a baseline assessment in addition to the use of monitoring triplicate electrocardiograms during and after each drug administration, sometimes up to eight sets of triplicate electrocardiograms at 1-hour intervals. This results in the possibility that a patient has 24 electrocardiograms over an eight-hour time span; in some instances, they are all normal and unchanged. The intended goal of this monitoring is to mitigate variation and to detect abnormality and variation of measurable physiologic intervals. The recommendation for triplicate electrocardiographic monitoring was empiric, and presently there is no substantial evidence that this level of monitoring either provides a benefit to the individual patient who is the subject of this monitoring, or that it provides meaningful and useful data regarding adverse events or trends in randomized clinical trials.

**BIOMARKER MONITORING**

Monitoring of cardiac biomarkers has now been incorporated in clinical trials and surveillance of patients showing signs of myocyte apoptosis as a sign of Type I injury, as well as in patients treated with checkpoint inhibitors (Lakhani et al., 2021; Mahmood et al., 2018). The rationale for biomarker monitoring in patients who have received checkpoint inhibiting agents stems from the 25-50% mortality associated with
established myocarditis. Notwithstanding that the incidence of established myocarditis is low, and estimated to be approximately 1%, early recognition of cardiac inflammation sufficient to cause elevations in troponin allows earlier intervention. Increases in troponin levels, however, can come from many etiologies, and studies have shown that around 11% of patients on immune checkpoint inhibitors will have a rise in their troponin level, but only about 12% of the rises in troponin (approximately 1.3% of treated patients) have elevations that are due to myocarditis (Srinivas et al., 2021). Data-based algorithms have not yet been introduced, and it must be recognized that delays in cancer treatments and increased costly cardiac testing need to be balanced with improved outcome. Conclusive data as to the benefit of these determinations is eagerly awaited and should help define true toxicity as well as provide information of sub-clinical yet highly relevant cardiomyocyte injury or death (Januzzi et al., 2021).

**Analysis and Modeling of Left Ventricular Ejection Fraction Surveillance, Electrocardiographic Monitoring, and Biomarker Determinations**

As noted above, an estimate of false positive results for an agent without toxicity has been explored, but low levels of toxicity may exist for agents presently under investigation. The well-established Learned Hand inequality can help provide prospective (Grossman et al., 2006).
The Learned Hand inequality considers burdens on society and the individual. Total burdens, including indirect and financial burdens, associated with recognizing actual or impending toxicity (B) should be ≤ to an integration of (L) [likelihood of an event] x (S), the [severity of the event], or

\[ B \leq (L \times S) \].

Table 1 shows some variables than could be incorporated into this Learned Hand application. It must be recognized that in the case of very serious event, screening hundreds, or even thousands of patients to recognize and or prevent a rare but catastrophic adverse event is appropriate and necessary; severity of the event becomes a vital component.

Hand, in writing a legal opinion in the context of an accident, suggested, that when the product of the likelihood of an event and the severity of that event were greater than the investment required in precautions to prevent the event, there is a burden to act in a way to reduce the likelihood of the event and its consequences (United States v. Carroll Towing Co., Inc 2nd Cir. 1947). The concept is useful in helping us understand relationships between burdens and benefits of surveillance initiatives as we treat cancer patients.

**How the Hand equation relates to burdens and benefits in cancer patients**

It is understood that the incidence of cardiac dysfunction is elevated in the case of high-dose anthracycline exposure, and that the incidence and severity of cardiac events is
small or even very small in the case of some of the tyrosine-kinase inhibitors or monoclonal antibodies when administered in the absence of confounding factors. The severity of the consequence is also widely variable: anthracyclines cause myocyte destruction, and the end-result is actual or potential cardiac failure that, in extreme instances, may be fatal. In the case of more modest or even questionable toxicity the risk is small. While we cannot assign an absolute numeric value to the likelihood of risk or its severity, we can recognize that the product of two small numbers is qualitatively different from the case where the product of likelihood and severity is large. On the other side of the Land inequality, we have the investment required in our attempts to prevent an event: while cardiac ultrasound is not especially burdensome or costly, other factors must enter on that side of the equation and include the risk of premature interruption of important therapy that may occur in the face of false positive ultrasound reports, administration of cardiac drugs that may add to cost and have adverse events of their own, and are likely to add at least some degree of anxiety for our patients. And so, we may place the Learned Hand formula in the context of appropriate surveillance for cardiotoxicity. When the likelihood times severity side of the inequality meets or exceeds the burdens, surveillance is appropriate, necessary, and becomes an important part of clinical care for the cancer patient. When the burdens are clearly greater, i.e., when \( B > (L \times S) \) we must recognize that fact, and seek ways to reduce the burdens, either by less intense or less frequent testing; we must recognize when testing in a particular setting is disproportionate to the ultimate benefits when large cohorts of patients are considered. Some of the variables that contribute to burdens are depicted in table 1; guidance related to severity may derived from the Common Terminology
Criteria for Adverse Events v3.0 that allows for a perspective of how this relationship might be applied (Freites-Martinez et al., 2020).

It appears that regulatory bodies have not yet recognized the need to include such an analysis, and guidance related to cardiac monitoring sometimes has followed the schemata used in the original clinical trial where recognizing and quantifying events was a goal, rather than optimizing rational and appropriate monitoring in the after-market setting; revisiting these guidelines with the intent to revise recommendations to maximize patient benefit is both appropriate and timely.

In the case of cardiac ultrasound, we must look at how many cardiac ultrasound determinations must be undertaken to avert a single case of a meaningful cardiac event, and consider if reducing the present recommendations, taking into consideration the cost, the inconvenience, and the potential injury invoked by pre-mature interruption is reasonable. We ask if the evidence suggesting that patient safety will be varied or compromised by this change, and if a reduction in surveillance should be considered. It is time that regulatory bodies such as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) reevaluate the recommendations or requirement to provide confidence that amendments of recommendations have been adequately scrutinized and that modifications are not disproportionate with regard to risk.
Electrocardiography in clinical trials can detect hugely important changes. Ischemia, inflammation, and arrhythmia can all be detected using serial electrocardiograms, and abnormally prolonged corrected QT (QTc) is an important marker for ventricular ectopy that may be life-threatening. Beyond isolated or non-sustained rhythm disturbances that may only be seen in one of a series of three electrocardiograms taken over a five-minute period, the benefit of triplicate electrocardiography needs scrutiny and a better understanding of burdens and benefits. Triplicate electrocardiography done repeated per investigational protocols needs to be scrutinized to quantify benefit. While the risk and cost of triplicate electrocardiograms is comparatively small, the number of studies undertaken continues to be a potential burden for patients that can and should be balanced with what we as clinicians are able to mitigate in the form of adverse events and short- or long-term toxicity. We believe that evaluation of meaningful added benefit of triplicate electrocardiograms over single ones needs to be undertaken, and the practice rethought in the context of burdens and real event prevention.

Strategies for biomarker surveillance can be useful if rises in cardiac biomarkers allow for earlier diagnosis of specific cardiovascular conditions and allow for interventions that lead to improved outcomes. As mentioned previously, the burden of this surveillance can easily outweigh the benefit if the incidence of the cardiotoxicity is sufficiently low requiring a higher number of patients screened to detect cases of toxicity that have no proven short or long-term consequences. When utilizing biomarker surveillance and analyzing risks and benefits, clinicians should consider the poor specificity of cardiac biomarkers for specific conditions, such as myocarditis. In addition, mild asymptomatic
rises in biomarkers should not reflexively lead to diagnosis of cancer therapy related cardiac adverse events. Appropriate cardiac diagnostic workup and clinical judgement are paramount in this decision-making process. Cardiovascular disease has a high background prevalence and only properly designed studies accounting for potential confounders will allow clinicians to understand the incidence and severity of events associated with cancer therapeutics.

DISCUSSION

This paper looks at some of the problems encountered as increasing numbers of cancer patients undergo treatment with agents either thought or proven to cause cardiotoxicity. The extent of surveillance that presently is recommended, per guidelines or regulatory agencies releases may be problematic. Such guidance during and after treatment is often based on monitoring used in clinical trials or on inadequately supported consensus statements that many consider standards of care. Ideal monitoring should be data-driven and should balance our ability to prevent meaningful sequelae with direct and indirect burdens. The optimal balance may not have been defined with the present monitoring strategies. In considering how we might fine-tune our surveillance strategies several considerations come to mind: the extent of monitoring should be scrutinized and revised based on ongoing clinical data, and revisions, must be based on objective rather than anecdotal observations.
Widely divergent estimations of cardiac adverse events following cancer treatment that appear in the literature may well be the foreseeable result of variation in data thresholds, failure to identify confounding factors, and reporting of false-positive results, the latter being of special concern when the incidence of true positive results is small and the predictive value of the measuring tool sub-optimal. Varying reports based in unvalidated data may be a partial explanation as to why some clinicians taking care of cancer patients do not adhere to guidelines for cardiac surveillance (Chavez-MacGregor et al., 2015). Clinicians, at least in some instances, lack confidence in what has been proposed. They understand that reporting of events to databases without clinical review, especially when such databases have been derived from payment claims data, may be over-inclusive. Not recognizing the weakness of such reports may suggest the need for yet higher levels of surveillance. The higher costs as well as increased anxiety and inconvenience levels for patients may be factors that clinicians balance with the perceived benefit in the form of prevented events when they choose to follow or ignore recommendations. Clinicians prescribing cardiotoxic agents need and expect clear evidence that our ability to prevent meaningful late events through surveillance is proportionate to the burdens inflicted on our patients. Understanding and incorporating a balance suggested by the Hand formula offers an interesting opportunity to better understand these relationships, as well as the ambiguities associated with the present recommendations. Unfortunately, such data is not yet available.

Beyond cardiac ultrasound, extensive electrocardiography and the use of biomarkers raises similar concerns. Data demonstrating a meaningful benefit of triplicate
electrocardiograms and serial biomarkers should be sought, and criteria defining when such monitoring offers incremental benefit when taken together with the associate burdens. While sporadic dysrhythmia could be detected and documented by triplicate electrocardiograms, such monitoring is unlikely to demonstrate meaningful variation in physiologic intervals. A single, perhaps longer electrocardiographic recording might prove equally revealing and less burdensome. Interestingly, in a review of 8518 cardiograms undertaken in 525 patients in phase I trials not using triplicate electrocardiograms, no clinically significant information was provided from extensive monitoring (Naing et al., 2012).

Meaningful surveillance will vary according to individual patient risk, and as we move forward, we may expect huge benefits form new tools that will integrate individual characteristics to optimize the balance between preventive benefit and burdens for increasingly smaller cohorts of patients.

We have focused our concerns on the cardiac ultrasound, the standard 12-lead electrocardiogram, and biomarker blood levels, as these are the testing parameters that are most used for the routine surveillance of cancer patients being treated with agents deemed to be cardiotoxic. Imaging studies used in special situations or for the diagnosis of suspected abnormalities rather than screening, such as cardiac computerized tomography scans or magnetic resonance imaging have not been included in this discussion.
A second consideration in the burden/benefit balance is that on-going surveillance in the face of low event prevalence, when tested using techniques that are imperfect, may lead to much data being entered into databases. Much of this data cannot be adjudicated with a medical record review, and confounding data may be noted in such databases as events. Review of such data, and metanalysis thereof my suggest levels of toxicity sufficiently high to perpetuate the perceived need to continue or expand surveillance.

CONCLUSION

Present guidelines and recommendations for cardiac surveillance following anticancer treatment should undergo data-based revision and a coordinated undertaking to balance our goal of preventing cardiac sequelae with burdens associated with the surveillance. Event analysis should include both incidence and severity of events, while burdens should include direct risks, interruption of cancer therapeutics, and financial costs. These considerations apply to routine surveillance undertakings.
Table 1: Some of the components of burden that may be considered in applying the Learned Hand formula to balance cardiac surveillance during and after cancer treatment.

<table>
<thead>
<tr>
<th>Burdens associated with surveillance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs</td>
<td></td>
</tr>
<tr>
<td>Cost of the test</td>
<td></td>
</tr>
<tr>
<td>Transportation-related costs</td>
<td></td>
</tr>
<tr>
<td>Anxiety of testing and its implications</td>
<td></td>
</tr>
<tr>
<td>False positive results</td>
<td></td>
</tr>
<tr>
<td>Need for retesting</td>
<td></td>
</tr>
<tr>
<td>Possible alteration of treatment or treatment schedule or administered dose</td>
<td></td>
</tr>
<tr>
<td>False negative results</td>
<td></td>
</tr>
<tr>
<td>Continuing a toxic treatment</td>
<td></td>
</tr>
<tr>
<td>Reduced future surveillance</td>
<td></td>
</tr>
<tr>
<td>(delayed recognition)</td>
<td></td>
</tr>
<tr>
<td>Failure to initiate appropriate therapy</td>
<td></td>
</tr>
</tbody>
</table>
References


Authorship Contributions:

Wrote or contributed to the writing of the manuscript: Ewer, M.S., Palaskas, N.L., and Herson, J.

Financial Disclosures:

The authors received no funding related to this paper.

Conflicts of Interest Statement:

No author has an actual or perceived conflict of interest with the contents of this article.