Evidence Base for Quality Control Activities in Cardiovascular Imaging

Mehdi Eskandari, MD,a,b Christopher M. Kramer, MD,c Harvey S. Hecht, MD,d Wael A. Jaber, MD,e Thomas H. Marwick, MBBS, PhD, MPH

ABSTRACT

Quality control is pervasive in most modern business, but, surprisingly, is in its infancy in medicine in general—and cardiovascular imaging in particular. The increasing awareness of the cost of cardiovascular imaging, matched by a desire to show benefits from imaging to patient outcome, suggests that this deficiency should be reassessed. Demonstration of improved quality has been proposed to require a focus on several domains: laboratory organization, patient selection, image acquisition, image interpretation, and results communication. Improvement in these steps will require adoption of a variety of interventions, including laboratory accreditation, appropriate use criteria, and continuous quality control and enhancements in reporting, but the evidence base for the benefit of interventions on these steps has been sparse. The purpose of this review is to evaluate the current status and future goals of developing the evidence base for these processes in cardiovascular imaging. (J Am Coll Cardiol Img 2016;9:294–305) © 2016 by the American College of Cardiology Foundation.

The initial adoption of scientific methods of quality control (QC) from industry to medicine started >50 years ago (1). Despite sporadic interest in QC, several markers point toward ongoing limitations of health care QC, including inappropriate care (2), disagreements among experts (3), geographic and provider variations in practice and care (4), and medical injuries to patients (5). Fortunately, the possibility of harm is limited in imaging (although there are potential risks from stress testing, contrast agents, radiation exposure, or misinterpretation of tests), but the other markers are prevalent in imaging practice.

A series of influential frameworks have sought to address these concerns and to encourage evidence-based medicine (6). Outside of the assessment of process measures, the efficacy of current strategies to improve care remains a subject of ongoing research. The field poses a number of challenges, not the least of which is that the role of the randomized controlled trial—the conventional approach to studying causal relationships and incremental benefit/harm—has limitations in the evaluation of complex social and interpersonal systems that characterize the interaction of imaging services with clinical practice.

The growth of cardiovascular imaging has had a sizable economic impact, but the contribution of imaging to changes in disease outcomes is unclear. Defining the contribution of existing and new tests to patient outcome and building an effective cardiovascular imaging QC process is an important goal (7). This paper reviews the components of imaging QC (including laboratory organization, patient selection, image acquisition, image interpretation, and results communication), the reported experience with QC in the imaging laboratory (including the assessment of ventricular function and valvular disease), and considerations about safety. The purpose

From the aMenzies Institute for Medical Research, University of Tasmania, Hobart, Australia; bKing’s College Hospital, London, United Kingdom; cUniversity of Virginia Health System, Charlottesville, Virginia; dMount Sinai School of Medicine, New York, New York; and the eCleveland Clinic, Cleveland, Ohio. The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Manuscript received August 3, 2015; revised manuscript received November 6, 2015, accepted November 11, 2015.
of this review is to facilitate the wider adoption of the QC process.

QUALITY CONTROL

The ultimate goal of imaging is to provide a single, appropriate test at the right time and to the right patient that is performed, interpreted, and integrated correctly into patient management (Central Illustration) (7). The following sections seek to define the evidence base for the 4 defined domains that affect patient outcomes (7) as well as the often-neglected but critical link of appropriate decision-making with outcome.

LABORATORY ORGANIZATION. Setting up the right processes is perceived as having a pivotal role in offering high-quality studies. An accreditation program can ensure that cardiovascular imaging laboratories identify and address potential problems on a regular basis. In the United States, the Intersocietal Accreditation Commission (IAC) provides such a program. Although the process is voluntary, it is recommended by professional bodies (e.g., American Society of Echocardiography [ASE], American Society of Nuclear Cardiology, Society for Cardiovascular Magnetic Resonance [SCMR], and other imaging societies) and linked to the reimbursement by a number of payers, including Medicare. The process of accreditation oversees the physical environment, facility and equipment, technical and medical staff, examination, and procedures; assuring that the laboratory meets minimum requirements and has a QC model in place. IAC stipulates that laboratories should have medical directors preferably with level 3 training (or equivalent), technical directors and technical staff with appropriate credentials, and interpreters at level 2 training or higher (8–11). However, the variation in stipulated training levels between jurisdictions (Table 1) (12–15) is a reflection of their limited or absent evidence base. The accreditation process is also variable, being voluntary and only provided for echo in Europe (15), whereas Australia lacks a formal assessment for laboratory accreditation. An optimal model in QC in a large laboratory would include the presence of a specific position to facilitate regular assessment of QC measures, organize regular QC meetings, and assure recording and appropriate follow-up of the findings. Optimally, this QC leader would be highly trained and experienced, but most importantly would be knowledgeable about the principles of QC. In most instances, this person would be the technical or medical director.

A second component of the laboratory environment is infrastructure. Funding arrangements in Australia involve differential reimbursement of current and older equipment. With the incorporation of 3-dimensional (3D) echocardiography and myocardial strain in guidelines (16,17), an echocardiography laboratory lacking this equipment or expertise may not be considered “state of the art.” Likewise, because image quality is suboptimal in 10% to 15% of echo-dendrograms and as many as 30% of critically ill patients (18), failure to use ultrasound contrast agents is a marker of suboptimal examinations and the proportion of studies involving contrast is a potential marker of quality. Each laboratory should have a list of indications for contrast, including poor endocardial delineation, suspected left ventricular (LV) thrombus, apical hypertrophic cardiomyopathy, LV noncompaction, and enhancement of suboptimal spectral Doppler signals (19). Similarly, the provision of appropriate equipment for dose minimization for cardiac computed tomography (CT) is likewise an essential marker of quality infrastructure (10).

An accreditation process also assures that academic laboratories involved in training programs have the expertise to offer quality training. The current training task force report mandates that an echocardiography laboratory in which training of cardiology fellows is undertaken should be supervised by a physician with level 3 training (13). For cardiac magnetic resonance (CMR), trainers should be at level 2 or 3 (the latter preferred) (20). The European Association of Cardiovascular Imaging (EACVI) recommends that echo laboratories involved in research and training should be at the “advanced standards” level (15).

Although many of these suggestions are logical, this process would be strengthened if evidence could be gathered to support the impact of these laboratory measures on patient outcome. This is particularly the case in relation to the application (and more importantly mandating) of this process in smaller laboratories and cardiology practices.

PATIENT SELECTION. The initial step to improve patient selection has been the development of appropriate use criteria (AUC). The growth of cardiovascular imaging has been an important catalyst to the development of these guidelines, and although their uptake has been slow outside of North America, this problem is not limited to just that jurisdiction. Thus, although the presence of different workflows may inhibit the implementation of exactly the same model, it seems likely that similar guidance will be needed in other regions of the world.
Most reports indicate >80% appropriateness for transthoracic echocardiography (TTE) and transesophageal echocardiography (TEE) (21). Although there has been a temporal improvement in appropriate use for TTE and TEE (21), it is unclear to what degree this change has contributed to reduction in cardiac imaging. The appropriate use of stress echocardiography is much less, and has also not improved (21).

Although a recent report suggested that 92% of CMR referrals at a single center were appropriate (22), another single-center study of 300 stress perfusion CMR studies showed 50% were deemed appropriate, 37% maybe appropriate, and 13% rarely appropriate (23). Ischemia was demonstrated 3 to 4 times more commonly in the appropriate or maybe appropriate studies than in the rarely appropriate ones.

An excellent example of what can be achieved through continuous quality improvement initiatives on appropriateness of imaging is provided by the Advanced Cardiovascular Imaging Consortium (24). This statewide initiative of 47 centers in Michigan sponsored by Blue Cross Blue Shield/Blue Care Network provided a prospective, observational study of a quality improvement program in 25,387 patients undergoing coronary computed tomography angiography (CTA). The program included continuing medical education, Clinical Champions presenting AUC at grand rounds and letters to participating physicians regarding imaging overuse, and was conducted 1 year pre-intervention during 2 years of intervention and 6 months post-intervention. In contrast to other studies, part of this strategy involved the threat of losing reimbursement in the absence of a definite, measurable change in appropriate use. Compared with the pre-intervention period, there was a 23% increase in appropriate (61% to 80%, p < 0.0001), 60% decrease in inappropriate (15% to 6%, p < 0.0001), 41% decrease in uncertain, and 42% decrease in unclassifiable scans during follow-up. The same Consortium performed a prospective, controlled, nonrandomized study aimed at radiation dose reduction in 4,862 patients undergoing CT with 2-month control, 8-month intervention, and 2-month follow-up periods (25). The interventions included lectures on dose reduction techniques, protocol customization, scanner manufacturer delineation of scanner-specific issues, and a physician and technologist for each site to implement the best practice program. The estimated mean radiation and effective radiation doses were reduced by 53% compared with the control period without loss of image quality or decrease in the number of interpretable studies.

On the other hand, although studies on impact of AUC on single photon emission computed tomography (SPECT) have shown improvement in prognostic value of the test in appropriately selected patients (26–29), it appears that educational intervention on AUC alone is unlikely to reduce inappropriate rate (21). There are also limited data to support the effect of AUC-based educational interventions on the echocardiography-ordering behavior of physicians. To our knowledge, only 1 paper has shown the number of ordered inpatient TTE and inappropriate TTE per day in an academic hospital were significantly reduced during an educational intervention to interns in training (30), and in this, the effect of intervention was not sustained (31). The evidence that AUC is able to reduce numbers of requested TTE is limited (32). Unfortunately, the coding of AUC has limited reproducibility and this subjectivity makes it susceptible to bias (21). A partial solution to this at a local level might be for the QC leader to code the
AUC on the basis of medical notes rather than the request form. An alternative and perhaps more feasible approach to retrospective AUC audit is to seek the common inappropriate requests at the point of service in the echocardiography laboratory. A checklist can be used to identify such at-risk requests with a view to further discuss the appropriateness with the ordering provider (33). Further studies are required to assess if this type of approach would have a more meaningful impact on reduction of unnecessary imaging.

Review of AUC as part of a laboratory QC program has been recommended; for example, IAC Echo accreditation recommends the review of a minimum of 30 consecutive studies for each modality (trans-thoracic, transesophageal, and stress echocardiography) each year (8). Although the process may facilitate an understanding and measurement of referral pathways with a view to informing educational interventions with ordering providers to reduce rarely inappropriate referrals, it is unlikely that such numbers would be informative. In most laboratories, the appropriate use rate is >80%, so in such a process, only 6 such rarely inappropriate referrals would be identified. Interestingly, only 65% of interns are aware of existence of AUC and only 45% consider the cost implications of tests they order routinely (20).

**IMAGE ACQUISITION.** Technological knowledge of how to improve image quality is the first step in image acquisition. In nuclear cardiology, IAC has well-defined protocols for cameras and image acquisition and tracer dosimetry and QC (generators, cyclotron, unit dose) (9). The latter aspects overlap the Nuclear Regulatory Commission and state and city regulations.

In echocardiography, the use of a study template of examination elements outlined by major society guidelines is an important step (19,34), but has two shortcomings. First, with different pathologies, a variety of additional images may be needed, so the template should be considered as minimum criteria that are inadequate without supplementation. Second, such a template would be strengthened by evidence supporting a clear policy of what to measure in each study and how to do it. This is essential because the assessment of many pathologies is multi-parametric and requires a structured approach at the time of image acquisition. The SCMR has established CMR imaging protocols for the major diagnoses for which CMR is ordered (20), and many of the scanner manufacturers have implemented the SCMR protocols as standard on their magnets.

The least problematic allocation of time for performance of each study is with stereotyped acquisitions including SPECT, positron emission tomography (PET), and cardiac CT. Echocardiography is at the opposite end of the spectrum of variation, with a wide range in examination duration from a simple peri-cardial effusion assessment in a young person to a complex multivalvular evaluation. For echocardiography, the European guidelines recommend a duration of at least 30 to 40 min, and the American guidelines recommend a duration of 55 to 60 min (15,19,34). In general, we lack evidence in support of this guidance. Similarly, some echocardiographic laboratories propose the use of a senior interpreter to review all the acquired images before the patient leaves the laboratory. This is logical in terms of selection of contrast or more advanced parameters, but the evidence to support the cost of this investment is lacking.

The number of imaging tests performed on an annual basis to maintain competency is also a potential QC tool, but this lacks a suitable evidence base because of variable requirements in different countries (8-11,14,15).
The translation of these principles can be ensured by involvement of the QC leader in regular audits of the technical quality of the studies and completion of the imaging protocol. For echocardiography, most of the indices in Table 2 can be tracked at intervals and assessed separately for each sonographer on an annual basis.

**IMAGE INTERPRETATION.** Adequate interpretation of imaging tests takes time, and this is often a challenge for the interpretation of echocardiograms because these are done in large volumes. Providing guidance about interpretation time is difficult; clearly, this varies with reader expertise and study complexity. We are unaware of evidence linking interpreting volume and accuracy, although inconsistencies in reporting have been associated with the rate at which studies are finalized (39). For CMR, image analysis guidelines have been published for all of the major imaging pulse sequences including function, perfusion, flow, and late enhancement (36). However, the majority of evidence regarding interpretation QC has been gathered with echocardiography.

The involvement of physicians in test interpretation implies that a physician QC leader (who may or may not be the laboratory medical director) is necessary, in addition to a technical QC leader. Relevant interpretive quality parameters include continuing medical education requirements and interpreting volume. Minimum requirements set by IAC and other relevant bodies are variable (8-11,12,15) and lack supportive evidence.

These challenges in controlling quality from an administrative standpoint may be overcome by consideration of accuracy (comparison with a reference standard) and precision (defined by reproducibility). The assessment of accuracy in patients studied with multiple modalities (e.g., CMR for LV function or fractional flow reserve for stress imaging tests) is often proposed as an effective empirical

### Table 2

**Markers of QC for Image Acquisition, as Exemplified by Echocardiography**

<table>
<thead>
<tr>
<th>Imaging Step</th>
<th>Quality Marker</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laboratory organization</td>
<td>Physical environment, facility, and equipment; technical and medical staff, examination and procedures</td>
<td>Recommendations by IAC Echo and EACVI</td>
</tr>
<tr>
<td></td>
<td>CME credits for sonographers and physicians</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percentage of echo studies performed/ reported by a certified sonographer/clinician as a QC marker for an echo laboratory (EACVI)</td>
<td></td>
</tr>
<tr>
<td>Patient selection</td>
<td>AUC</td>
<td>Recommendations by IAC Echo, ASE, and EACVI</td>
</tr>
<tr>
<td></td>
<td>Application of AUC to a minimum of 30 consecutive studies per modality/year (IAC Echo)</td>
<td></td>
</tr>
<tr>
<td>Patient preparation</td>
<td>Record of blood pressure, height, and weight</td>
<td>Recommendations by IAC Echo, ASE, EACVI</td>
</tr>
<tr>
<td></td>
<td>Recommendations by ASE, IAC Echo, EACVI</td>
<td></td>
</tr>
<tr>
<td>Image acquisition</td>
<td>Percentage of completed studies (adherence to template), documented per laboratory and individual sonographer quarterly (ASE)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Annual review of 5 to 10 studies per sonographer (goal of at least 90% of images) for adherence to template (ASE)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uninterpretable studies (when contrast is used appropriately, less than 5% and 10% of studies should be labeled nondiagnostic for LV function and RWMA, respectively) (ASE)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use of ultrasound contrast agent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percentage of successful intubation for TEE (EACVI)</td>
<td></td>
</tr>
<tr>
<td>Image interpretation</td>
<td>Concordance exercises (e.g., joint reading sessions, comparison with reference)</td>
<td>Recommendations by IAC Echo, ASE, and EACVI</td>
</tr>
<tr>
<td></td>
<td>Assessment of EF/RWMA, stenotic, and regurgitant lesions in 2 random cases per modality should be reviewed in laboratory meetings on a quarterly basis (IAC Echo)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quarterly selection of at least 2 echo for each reader/modality (TEE, TEE, stress echo) for blind interpretation by another echocardiographer (ASE)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Goals of reducing variability could be 10% for EF and a difference of 1 grade or less for valve regurgitation (ASE)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temporal variability (e.g., serial studies)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study review process (%) to 2% of studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation with other modalities (accuracy) (e.g., EF with cardiac MRI, stress echo with coronary angiography, TEE with other modalities or operative findings)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quantitative result of 4 TTE per interpreter per year should be compared with another test (ASE)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Annual meetings to discuss studies with significant variances in the presence of all members of echo laboratory (ASE)</td>
<td></td>
</tr>
<tr>
<td>Reporting</td>
<td>Presence of key report data elements</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percentage with critical parameters reported (e.g., LVEF should be reported in at least 90% of studies [ASE])</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Review of 10 random cases per year (IAC Echo)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum measurements</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Timeliness targets</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Critical value responses; documentation of interpreter to requesting physician communication (ASE)</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- ASE = American Society of Echocardiography; AUC = appropriate use criteria; CME = continuing medical education; EF = ejection fraction; IAC = Intersocietal Accreditation Commission; LV = left ventricular; MRI = magnetic resonance imaging; QC = quality control; RWMA = regional wall motion abnormality; other abbreviations as in Table 1.
approach for the purposes of QC. Unfortunately, accuracy is not an ideal quality marker because of the lack of a true comparator for the most relevant parameters (e.g., ejection fraction [EF], fractional flow reserve [FFR], measurements of valve regurgitation). The use of a chosen modality as a reference standard for EF cannot account for intermodality differences in normal ranges (37), although ranking of EF remains valid despite these differences. For this reason, comparison of LVEF with corresponding CMR images can be used to improve the visual estimation of EF (38), justifying an iterative process for calibrating echocardiographic LVEF estimates to LVEF by CMR. However, comparisons with FFR are more problematic because discordance between perfusion imaging and FFR can be expected on a pathophysiologic basis in up to 40% of the cases (39).

Precision is an important QC parameter that is often assessed as the repeatability of measurements on a single image, for the purposes of image interpretation. However, in a combined acquisition/interpretation QC process that is relevant to sequential follow-up, it should also be considered as the test/retest variation of measurements on multiple examinations in the patient in stable condition. In some situations, in which a good reference standard is lacking for definition of accuracy, the measurement of precision is especially valuable. Useful tools for this purpose include assessment and feedback regarding the evaluation of reference cases (40), or an over-reading process. The required number of cases is undefined: some laboratories undertake this in 1% to 2% of studies. Discordance should lead to self-directed learning, for which internet-based education and teaching, including digital libraries such as WikiEcho from EACVI, can be used to reduce variability in image interpretation. Individualized retraining programs have been shown to reduce variability and improved reproducibility with long-term sustainability (41).

The use of quantitation is an important means of reducing variability, and its use varies from uniform application with CMR to inconsistent use in SPECT and echocardiography. In the Advance MPI (Study of Regadenoson Versus Adenoscan in Patients Undergoing Myocardial Perfusion Imaging [MPI]) trial, variability of serial studies with visual assessment exceeded quantitative analysis (42). Likewise, 3D echocardiography provides less variable measurements than 2-dimensional (2D) echocardiography (43), and global longitudinal strain may be of particular value in detection of subclinical disease.

RESULTS COMMUNICATION. The accuracy, comprehensibility, and timeliness of results communication have a direct impact on the referring physician’s decision-making and are therefore perhaps the most critical step of imaging process on patient care (40). Reports in imaging laboratories should be uniform and include key elements, with a common terminology used across all modalities and comparison with previous studies (19,44). One aspect that has received scant attention is the avoidance of over-reporting of extraneous material (e.g., minutiae of diastolic function assessment), which may be confusing to the nonspecialist.

Structured reporting, using a menu of drop-down statements to generate a report improves the consistency of interpreter comments and digitally stored data (45). Although this type of facilitated reporting is thought to avoid discrepancies in report transcription, a single-center study showed contradictory statements to be present in the final TTE report in 4.0% of TTE, 3.6% of TEE, and 7.1% of stress echocardiograms during 11 years (46). Use of the same tool to check completed reports in real time showed 83% of reports to have some error or inconsistency, prompting either mandatory or suggested amendment of report. This was related to the number of reports per hour, rather than reader experience (35). It is unclear as to whether such a process can lead to an improvement in report quality over time.

The American Society of Nuclear Cardiology has specific guidelines for reporting nuclear cardiology studies (47). Unfortunately, these have not been widely followed, and American Society of Nuclear Cardiology launched a voluntary registry for all nuclear tests performed in the United States; participating laboratories have to agree to follow quality measures, including in relation to reporting, that will allow comparison with their peers. The Image Guide data form also includes information related to appropriate use.

Guidelines on the proper reporting of individual imaging techniques have also been reported by SCMR (48), ASE, and EACVI (19,34). In general, imaging studies should have an immediate report that is available by the end of the same day, with finalization within 48 h. Every laboratory should have a list of urgent diagnoses (e.g., tamponade, mechanical complications of myocardial infarction, aortic dissection) that trigger expedited review, and direct communication of critical findings should be made to the referring physician. Interpreters should take the opportunity to use the final report as a communication tool to teach referrers as to what degree each imaging
modality would be able to answer their clinical questions.

The QC targets for results communication would be a random audit of reports for completeness and timeliness (19). Potential future developments include the use of machine learning processes to facilitate automated interpretation.

**APPROPRIATE MANAGEMENT RESPONSES.** Imaging research has generally focused on the investigation of tests, rather than diagnostic strategies. Thus, although the diagnostic and prognostic implications of individual tests are well characterized, the results of their integration into clinical management and patient outcome are often considered self-evident (7). Research in this area has mainly been limited to cardiac imaging in coronary artery disease. In a multicenter prospective registry of 1,073 patients with intermediate or high likelihood of coronary artery disease; referred for SPECT, PET, or coronary CTA; and followed for 90 days, test results had only a modest impact on referral for coronary angiography or change in medical treatment (49). Indeed, of 8% who had moderate to severe abnormality, <50% of the SPECT or PET and 62% of the coronary CTA groups were referred for coronary angiography. A change in medication occurred in one-half of these patients, whereas only 25% were referred for catheterization and had a change in their medication (Figure 1). The results of this study suggest that evaluation of physicians’ behavior and patient management in response to the cardiac imaging tests should be a component of QC in cardiovascular imaging (49).

**REPORTED EXPERIENCE WITH QC**

As discussed previously, the lack of evidence supporting quality assessment from administrative data has left empirical evaluation of accuracy and reproducibility at a laboratory level as the main tools for QC. The existing published data—mainly in echocardiography—suggests that both variability and accuracy can be improved by formative review process.

**EF.** The assessment of global LV function is among the most common reasons for the performance of echocardiography. However, with test-retest variability as high as 14% (50), it is apparent that inconsistent measurements could alter the decisions for device therapy, or medical therapy for heart failure. Accordingly, EF has been an important target for QC in echocardiography. The available studies emphasize the importance of a case-based process for reducing interobserver variability in qualitative and quantitative assessment of EF, even in experienced readers. Johri et al. (50) showed that a teaching intervention on visual estimation of EF significantly improved interobserver variability (Figure 2). Fourteen 2D echocardiograms representing a spectrum of LVEF range and image quality were shown to 25 readers for visual estimation of EF. Subsequently, new reference and baseline cases were discussed and compared with EF derived from the modified Simpson biplane measurements for each case by 2 senior readers in 3 case-based teaching sessions over a period of 3 months. Three months later, 14 new cases were shown to participants. Post-intervention results showed a 40% reduction in interobserver variability. Importantly, the improvement also noted at the mid-range of EF, which had the highest misclassification than the overall group. The results were sustained for more than a year when a subgroup of readers participated in a follow-up session. Notably, this study included a wide range of sonographers and physicians of different level of experience and improvement with teaching intervention was observed across all readers (50).

A more quantitative approach assessed the impact of self-directed teaching on reduction of interobserver variability in qualitative and quantitative assessment of EF, even in experienced readers. Johri et al. (50) showed that a teaching intervention on visual estimation of EF significantly improved interobserver variability (Figure 2). Fourteen 2D echocardiograms representing a spectrum of LVEF range and image quality were shown to 25 readers for visual estimation of EF. Subsequently, new reference and baseline cases were discussed and compared with EF derived from the modified Simpson biplane measurements for each case by 2 senior readers in 3 case-based teaching sessions over a period of 3 months. Three months later, 14 new cases were shown to participants. Post-intervention results showed a 40% reduction in interobserver variability. Importantly, the improvement also noted at the mid-range of EF, which had the highest misclassification than the overall group. The results were sustained for more than a year when a subgroup of readers participated in a follow-up session. Notably, this study included a wide range of sonographers and physicians of different level of experience and improvement with teaching intervention was observed across all readers (50).
undergoing CMR within 48 h (38). Participants received their own case-by-case variance from EF measured by CMR, and the 10 cases with the largest reader variability were discussed along with corresponding CMR images. Self-directed learning was undertaken by side-by-side review of echo and CMR images. Two months later, 20 new cases were shown to the same 31 readers, using the same methodology. There was a significant improvement in interobserver variability and decrease in EF misclassification and absolute difference between echocardiography and CMR calculation. A combined physician-sonographer EF estimate improved the precision of EF determination by 25% compared with individual readers.

A recent study used a different methodology to assess the impact of teaching intervention on improvement of reproducibility. Five readers of different levels of experience were asked to interpret 10 TTEs for a number of parameters including LV end-diastolic volume and EF (Table 3) (41). On the basis of published data review and previous studies, a range of acceptable difference for LV end-diastolic volume and EF was defined as 30 ml and 10%, respectively. Reproducibility was evaluated by pairwise comparison of the interpretative variability among readers for each parameter. All readers then underwent retraining involving group discussions, case illustrations, and open forum questions as well as individualized one-to-one sessions for those in whom the difference in measurements exceeded the prespecified acceptable value. Upon completion of the retraining program, readers with unacceptable reproducibility were given 10 TTE for re-evaluation. The process of training and retesting was continued if acceptable reproducibility was not achieved. To assess the durability of training program, the process was repeated with the same readers 1 year later. After the retraining sessions, readers demonstrated improved reproducibility, which was maintained on subsequent testing 1 year later (41).

**Assessment of Right Ventricular Function.** Accurate echocardiographic assessment of right ventricular (RV) size and systolic function is challenging, and qualitative measurements compromises both accuracy and reproducibility. The quantification of chamber function is of particular value in RV assessment by echocardiography. In a study of 15 readers evaluating the RV function of 12 patients, Ling et al. (51) documented the inaccuracy of visual estimation of RV function compared with CMR. The use of quantitative measurements increased accuracy and inter-reader agreement compared with qualitative assessment alone, especially in the distinction of normal and abnormal.

---

**FIGURE 2 Misclassification Rate of Visually Estimated Ejection Fraction**

Misclassification rate of visually estimated EF compared with expert-derived quantified EF before (green bars) and after intervention (pink bars) in 3 groups of all cases, mid-range EF (30% to 55%), and outside mid-range EF (<30% and >55%). Misclassification for all cases was reduced significantly after educational intervention. EF = ejection fraction.

**Table 3 Acceptable Difference for Echo Parameters Suggested as Quality Marker**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Acceptable Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVEDV, ml</td>
<td>30</td>
</tr>
<tr>
<td>Biplane EF, %</td>
<td>10</td>
</tr>
<tr>
<td>Mitral regurgitation</td>
<td>&lt;1 sequential grade</td>
</tr>
<tr>
<td>AR</td>
<td>&lt;1 sequential grade</td>
</tr>
<tr>
<td>LVOT diameter, cm</td>
<td>0.2</td>
</tr>
<tr>
<td>Aortic valve area, cm²</td>
<td>0.2</td>
</tr>
<tr>
<td>Aortic valve peak gradient, mm Hg</td>
<td>20</td>
</tr>
<tr>
<td>Aortic valve mean gradient, mm Hg</td>
<td>10</td>
</tr>
</tbody>
</table>

AR = aortic regurgitation; LVEDV = left ventricular end-diastolic volume; LVOT = left ventricular outflow tract; other abbreviations as in Table 1.
limitations may compromise the accuracy of measurements. Indeed, in regurgitant valve lesions, it is difficult for a single parameter to make a categorization of severity, and guidelines emphasize the importance of a multiparametric approach (54). Unfortunately, although this caters for variation in image quality, the lack of a hierarchical approach increases discordance between observers when results are discordant. Moreover, the recommended 2D measurements for assessing mitral regurgitation—vena contracta and proximal isovelocity surface area—have significant interobserver variability, which might result in misclassification of patients (55). The use of 3D echocardiography to measure proximal isovelocity surface area can reduce variability, but at the cost of additional processing time, and has not been adopted in most clinical laboratories.

In patients with aortic regurgitation (AR), a study of 20 randomly selected patients, graded by 17 level 3 readers, showed a high interobserver variability resulting from the lack of a uniform approach to combine the parameters. Readers were recalibrated after the formulation of a consensus strategy, validated against CMR in a separate group of 80 patients. This consensus strategy to categorize AR severity (Figure 3) was based on the combination of LV volume with AR-specific parameters (vena contracta width, jet height, and holo-diastolic flow reversal). The development of this consensus strategy improved concordance and also improved the accuracy of the test relative to CMR in a separate validation group of 80 patients (56). Similarly, a structured algorithm to differentiate between severe and nonsevere tricuspid regurgitation improved both inter-reader agreement and accuracy (57). This process of identifying the causes of discordance and seeking consensus about a hierarchy needs to be considered for many echocardiographic diagnoses that require multiparametric assessment.

**CMR.** The QC published data for other modalities is less well developed than for echocardiography. The use of systematic reading criteria for interpretation of adenosine perfusion CMR has been shown to reduce interobserver variability. In this work, 106 studies (46 positive and 60 normal, interpreted by an experienced radiologist) underwent visual assessment by a technician and 2 residents with different levels of experience (2 months and 2 years). One month later, the systematic use of reading criteria was applied to interpretation of the same studies in a different order. Overall kappa improved from 0.59 to 0.71, which was mainly resulted from improvement in least experienced reader (58).

Teletraining for QC has been studied in the CMR environment (59). A German network, which increased from 5 to 14 centers between 2009 and 2014, showed that network training reduced offsite training for new sites to only 5 weeks. Real-time remote supervision and scan control opportunities were used to increase the number of smaller and remote sites that were able to offer high-quality CMR.

**SAFETY**

The downside of echocardiography is more in the realm of diagnostic limitation/misinterpretation than medical harm. Poor image quality can be significantly improved with the use of ultrasound contrast agents. Anaphylactic shock has been a rarely reported severe
adverse reaction to ultrasound contrast agents (60). In October 2007, the U.S. Food and Drug Administration mandated labeling changes for ultrasound contrast agents following reports of severe adverse effects and death during or shortly after their administration (61). This resulted in contraindications for use of ultrasound contrast agents in patients in unstable condition including respiratory failure, worsening heart failure, unstable angina, and acute myocardial infarction as well as severe pulmonary hypertension (61). No clear causal relationship was defined between contrast agent and mortality, and subsequent studies showed no increase in acute mortality in critically ill patients undergoing contrast echocardiography, nor was there any increase in pulmonary artery pressures (62,63). In a large observational registry, Main et al. (63) showed 28% lower mortality at 48 h in critically ill patients who underwent contrast echocardiography, and a subsequent study of pulmonary artery pressure with ultrasound contrast agents revealed no change in baseline normal or elevated pulmonary artery pressures (62). The risk of ultrasound contrast seems small, considerably less than the risk of inadequate cardiac function assessment from failure to use ultrasound contrast agents in critically ill patients (18).

Nuclear-based cardiac imaging modalities, although safe in the short term, are potentially associated with potential long term hazards from exposure to ionizing radiation (64). In recognition of these risks, there have been recommendations advocating lower exposure to ionizing radiation in nuclear imaging (65,66). To reinforce patient-centric imaging, these statements recommended applying the AUCs, considering alternative tests without radiation when feasible, using SPECT and PET mostly in intermediate-risk patients, and avoidance of layered or serial testing as the best tools to limit radiation exposure and enhance patient safety. A proposal to set limits of reducing radiation exposure/test to <9 mSv in 50% of patients by 2014 remains to be independently verified. Similarly, concerns about hazards from CT radiation exposure have led to questions about net benefit (67,68). The cumulative dose of radiation from cardiac imaging modalities has certainly increased (67), and the potential risk of radiation exposure is often underestimated (69).

Risk of injury may arise from CMR if there is failure of systems to avoid exposure of ferrous materials to the scanner. Nephrogenic systemic fibrosis is a catastrophic but rare side effect associated with administration of gadolinium to patients with stage 4 or 5 chronic kidney disease (70).

The potential indirect harms of cardiac imaging include evaluation failure and misleading (false-negative and false-positive) test results, wasted time (especially prolonging length of stay), and cost inefficiency. The rapid growth in cardiac imaging technology offers an increasing range of diagnostic tests, but the potential financial consequences arising from overutilization and inappropriate testing has led to restriction on use of cardiac imaging modalities by payers (71). There are hopes that the implementation of guidelines and appropriate use criteria may minimize harm from this process. Additional outcomes research will be needed to better understand the balance between the harm, cost, and benefit of investigations.

Imaging research has generally focused on the investigation of tests, rather than diagnostic strategies. For example, the diagnostic and prognostic implications of individual tests are well characterized, but the results of their integration into clinical management and patient outcome, although considered self-evident, are less clear. A patient-oriented approach to imaging outcomes will provide this information. Future research on the impact of imaging on patient outcomes is crucial (7).

### CONCLUSIONS

There is a historical “lack of awareness” about the merits of QC in cardiac imaging. A systematic approach with defined domains and actions to confront the problem has been proposed by relevant organizations, but the uptake of the QC process has been patchy. Quite simple steps can be of value (Table 4), and ongoing efforts are needed to improve QC programs and develop an optimal model for widespread implementation. Such a process will be an important step in linking imaging quality to patient outcomes.

**TABLE 4 Strategies to Improve Imaging Quality**

<table>
<thead>
<tr>
<th>Imaging Step</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition</td>
<td>Quantitative measurements in protocols</td>
</tr>
<tr>
<td>Interpretation</td>
<td>Uniform hierarchy algorithms</td>
</tr>
<tr>
<td></td>
<td>Group teaching activities, concordance with other modalities</td>
</tr>
<tr>
<td></td>
<td>Internet based image references</td>
</tr>
<tr>
<td>Results verification</td>
<td>Cross-modality reporting data standards</td>
</tr>
</tbody>
</table>

**REPRINT REQUESTS AND CORRESPONDENCE:** Dr. Thomas H. Marwick, Menzies Institute for Medical Research, 17 Liverpool Street, Hobart, Tas 7000, Australia. E-mail: tom.marwick@utas.edu.au.
REFERENCES

1. Lembcke PA. Medical auditing by scientific methods, illustrated by major female pelvic surgery. JAMA 1956;162:646-55.


KEY WORDS pulmonary perfusion and coronary anatomy imaging roles in coronary artery disease