10 Discussion

This study is the first of its type in Australia and no comparable research was identified in the international literature. It provides the first insight into the imaging orders placed by GPs, into the variance in ordering between GPs and into the relationship of tests ordered to the GP and patient characteristics and morbidity under management. These ordering patterns have been evaluated in light of the literature. The extent to which the current guidelines for imaging testing are followed by practising GPs and the extent to which they may be useful to them in deciding on the most appropriate tests have also been examined.

At least one imaging order was initiated by the GP at 6.7% of all encounters, an increase of 17.5% (from 5.7% of encounters) in 1990–91 (Bridges-Webb et al. 1992). The 17.5% increase over the last decade in the proportion of encounters generating at least one imaging order suggests a national increase of about 2.2 million encounters, from 4.7 million per year in 1990–91, to 6.9 million encounters in 1999–00.

GPs currently order, on average, seven imaging tests per 100 GP–patient encounters, with a standard deviation of 4.9 and the majority order between three and 11 tests per 100. However, the wide variance in ordering rates among the GP participants is notable, the range being zero to 62 tests per 100 encounters. The literature review provided a basis on which to assess the appropriateness of these ordering rates for the quality care of patients. However, as demonstrated in Chapter 4, several outliers with very high ordering rates should raise some question of appropriateness in these few cases. It is, however, possible that these outliers have specialised practices dealing with problems requiring radiological investigation.

For most problems managed by GPs, no imaging tests are ordered. For the 4.7% of problems for which they are ordered, there are 113 tests ordered per 100 tested problems (Section 4.1). This indicates that in by far the majority of cases only one imaging test was ordered per problem. In Section 4.1 it was demonstrated that for almost 90% of tested problems only one test was ordered but that the remaining 10% of tested problems accounted for almost 20% of total orders. One of the major exceptions to the one-test order pattern was in the management of breast lumps where both mammography and ultrasound were ordered for nearly 40% of tested cases.

The extrapolated estimate of eight million test orders from general practice across Australia includes some orders for mammography, many of which would not be counted in the MBS data because they would go through BreastScreen Australia. The data suggest that GPs are playing an active role in encouraging their patients to attend BreastScreen even though there is free, unreferred access to this service. GP provision of a referral to BreastScreen may have a positive impact on patient attendance.

The Health Insurance Commission data suggest there were 7.6 million radiology tests claimed through the MBS in approximately the same year. Considering there are about 103 million GP–patient encounters across the country per year, this ordering rate does not seem excessive. However, the data considered in this report suggest there are some areas in which increased use of a specific test might improve the quality of care, and other areas in which a decrease in testing would be unlikely to affect quality of care.

The extent to which the HIC data (drawn from the Medical Benefits record of payment to radiologists for tests instigated by GPs) reflect the pattern of tests actually ordered by the GPs was tested in this study at a very broad level. The comparison demonstrated very similar patterns in terms of the distribution of tests across major MBS groups, with slightly higher proportions of diagnostic radiology and computed tomography in the *BEACH* data and slightly lower proportion of ultrasounds. The extent to which these small differences reflect radiologists' decisions to conduct a test not ordered by the GP is not known.

However, in both datasets, diagnostic radiology accounted for by far the greatest proportion of all imaging tests (64.8% in *BEACH* and 62.4% in the MBS), followed by ultrasounds (26.1% and 28.4% respectively) and computed tomography (8.7% and 7.8% respectively). The very few MRIs and nuclear medicine tests recorded by, the GPs in *BEACH* may reflect those few cases in which the patient was covered by workers' compensation or some other insurance payment source, or where the patient chose to pay all costs for these tests, for they are not covered by the MBS when ordered by a GP. It is unlikely that these came from procedural GPs for these were spread across almost as many practitioners as there were tests ordered.

The relationship between ordering rates and GP characteristics was investigated at three levels. Initial analysis concentrated on mean ordering rates for specific GP groups. This suggested that female GPs order imaging tests at a significantly higher rate than their male counterparts and that ordering rates increase with practice size. The GPs were then divided according to their ordering rate into three groups—low, medium and high orderers. Female GPs and those practising in small rural and remote areas were over-represented in the high ordering groups while solo practitioners were less likely to be classed as high ordering GPs. The gender difference is supported by Rosen's findings that, in an outpatient setting female doctors were 40% more likely to order imaging than their male counterparts (Rosen et al. 1997).

In the analysis of variance, the factors that were significantly associated with ordering level were identified. These were GP sex, rurality and size of practice, sex of patients, rates of new patients and number of health care card holders seen, management rates of specific groups of morbidity, and the relative rate of management of problems described in symptomatic or ill-defined terms.

The subsequent multivariate analysis used general linear modelling to identify the best independent predictors of imaging order rates. The results indicated a significant relationship between high imaging order rates and practice size of 11–15 GPs and between high ordering and small rural or remote practice location (earlier noted in the descriptive analysis). Higher rates of new patient presentations and of the management of musculoskeletal, urinary and female genital problems, together with those described in symptomatic terms, were also significant predictors of high ordering rates in general practice.

The number of health care card holders seen by the GPs, the number of patients aged 25 to 44 years and the management rate of psychological, skin and general problems had an inverse relationship with imaging order rates. Note that GP gender earlier had disappeared from the model, having been overtaken by the better predictive value of the variables listed above. This suggests that while differences do exist in test ordering rates of male and female GPs these differences are better explained by other factors associated with the GP's practice that weave in with gender, rather than the gender factor itself.

The finding of a relationship between solo practice and low ordering rates is contrary to that of Njalsson et al. who found that ordering rates decreased with increase in practice size in Iceland (Njalsson et al. 1995). However, the Njalson study also revealed higher ordering rates in rural areas where clinicians often performed their own imaging within the practice. The higher ordering rates in rural areas of Australia demonstrated in the current study are likely to result from a combination of some rural GPs performing their own x-rays and GPs going through more detailed investigation prior to referral where access to specialists is poor.

These results suggest that the GPs most likely to be high orderers of imaging are those who practise in small rural and remote areas and those in large practices of 11–15 GPs, particularly those who see large numbers of new patients not holding a health care card and who manage large numbers of the problems of the type listed above. Together with Njalson's findings, these indicate that if corporatisation results in increased co-ownership of imaging facilities and general practices, higher imaging order rates could be expected in the future. In a situation of capped radiology rebates this may affect the distribution of available revenue among radiologists. Health planners need to keep this in mind in the development of future policy.

The finding of a relationship between the rate of problems that are described in symptomatic terms and the rate of imaging orders is supported by Klinkman, who found a higher level of general practice resource utilisation associated with diagnostic uncertainty or non-specific diagnosis in patients with abdominal pain in a United States family practice (Klinkman 1996, 75). This suggests that in many cases GPs rely on imaging together with other secondary services to assist them in reaching a diagnosis.

When imaging was ordered by the GP there were far less medications prescribed but significantly more therapeutic procedures undertaken, pathology tests ordered and referrals made to specialists. Secondary costs associated with imaging encounters would clearly be considerably higher than those of non-imaging encounters, above and beyond the cost of the imaging test itself. These results also align with Klinkman's findings (Klinkman 1996).

The extent to which GPs use imaging to assist in the diagnostic process was considered in respect of each specific morbidity group investigated in Chapter 8. The problem under management and its status to the patients were used to define whether the test was investigative in nature, used to assist management decisions or used to monitor the care of the patients. The differences in extent to which each test type reflected investigation rather than the other modes varied considerably. Whereas 60% of breast ultrasounds and 51% of non-obstetric pelvic imaging were investigative in nature as defined in this report, only 16% of ultrasounds of the shoulder were classed in this group.

The patients at encounters generating imaging orders differed from those attending non-imaging encounters only in their age distribution, there being a significantly higher proportion aged between 25 and 44 years and a significantly lower proportion of children. However, the multivariate investigation of the relationship between various GP characteristics, patient characteristics and morbidity managed to the number of imaging orders demonstrated an inverse relationship between ordering rates and the number of 25–44 year old patients seen during the recording period, after adjustment for the other factors in the model, including morbidity patterns. This suggests that when these younger adults do attend the GP they are more likely to be managed for musculoskeletal, psychological, skin, urinary and female genital problems but that when these problems are managed they are less likely to have imaging ordered than people in other age groups who have such problems.

While the proportion of health care card holding status of patients did not differ at imaging and non-imaging encounters, the multivariate analysis demonstrated an inverse relationship between the number of patients seen who held a health care card and the GP imaging order rate. That is, high imaging ordering rates were less likely the more health care cardholders the GP saw, probably due to the nature of the health problems of the population holding a health care card.

The literature review reported in this document demonstrated that, as with pathology, there is a great deficiency of high-quality research into the value of imaging, particularly in the general practice setting. Further, meta-analysis of the benefit of imaging tests is virtually non-existent. The currently available guidelines have been developed using a combination of the little available evidence and consensus of radiologists with an apparent lack of general practice input. There is little research into patient outcomes with or without the use of imaging. The little that is available is either based on the technical side of imaging or on patient selection. The latter work attempts to build predictive models of sensitivity and specificity of test results to give guidance on the selection of patients for whom the test should be undertaken. However, the majority of this research has been conducted either in hospitals or emergency settings with a different population base to that found in general practice. Revicki's suggestion that there is a need for focused assessment and guideline development on specific diseases or diagnostic problems rather than on the use of specific tests (Revicki et al. 1999) should be heeded.

While guidelines have successfully been developed in other areas, particularly in the area of therapeutic interventions, in light of the paucity of outcome-based research in the area of imaging tests, it is far harder to link the outcome of diagnostic investigation with patient outcome.

Two sets of guidelines have been used extensively in this report, those developed by the American College of Radiology (ACR) and the recently released revised edition of the RANZCR guidelines. Neither set of guidelines assist the clinician in deciding when not to use imaging. Rather they aim to assist the clinician in the selection of patients who should have imaging undertaken and/or guide in the selection of the appropriate imaging to be ordered in certain circumstances.

The ACR criteria are published with a well-summarised literature review that allows assessment of the evidence used in the preparation of the criteria. In common with other literature on diagnostic testing, little of the evidence has been gathered in general/family practice where the circumstance of low prevalence makes the predictive value of tests much lower. The extent to which such guidelines are 'portable' to general practice in Australia is therefore open to question. In the absence of better evidence, however, the ACR criteria probably represent the state-of-the-art advice on appropriateness of diagnostic imaging tests for the conditions that they cover.

The ACR guidelines are based on a mixture of consensus and scientific evidence. The philosophy of evidence-based practice is well supported by inclusion of a well-constructed literature review and references for each of the guidelines. They attempt to address the deficiencies in evidence by at least presenting the evidence that is available in a readable format. They then present a series of scenarios for each guideline and these are likely to be very effective as a learning tool. They also include flags, which aid the clinician in patient selection for a specific course of action. This is supported by a scoring system that allows the reader to weigh up alternative approaches in each individual circumstance.

The RANZC guidelines use an algorithm approach to illustrate the diagnostic choices in a wide range of circumstances. Some of the algorithms are supported with a brief list of references. Reference to the American College of Radiology Appropriateness Criteria and its

more comprehensive literature reviews is not uncommon. In an era where GPs are being encouraged to learn and apply critical review techniques it would seem more appropriate to include the evidence rather than references only. For example, in imaging of the shoulder, selection of the patient is just as important as selection of modality and in this area both the ACR and RANZCR guidelines are quite good. However, in other areas, such as x-ray of ankle injuries, the ACR guidelines rely heavily on the Ottowa rules to guide in the selection of patients for whom imaging is appropriate. In contrast, the RANZCR guidelines choose to ignore these rules.

Unlike the ACR criteria, no quantification of 'appropriateness' is provided to allow clinicians to consider the weighted alternatives. Instead, they provide simple branching algorithms to guide decision-making. Unfortunately, without the evidence and without the scoring system, the clinician is left with little to allow thoughtful discriminatory decision-making between alternative options for an individual patient. An example of this is in the investigation of back pain with radiculopathy where the ARC guidelines suggest that an MRI is the best choice. However, Medicare does not cover MRIs ordered directly by GPs. The ARC guidelines demonstrate that while the MRI is the preferred option (with a score of 8) the use of CTs is still acceptable with a score of 6. This information is extremely valuable. If the CT had received a score of zero for the investigation of this problem, the balance of choice for the GP would be completely different and the selection of the CT totally inappropriate.

This study suggests that the next revision of the RANZCR guidelines could benefit from the inclusion of evidence for each guideline. Consideration should also be given to the development of a scoring system similar to that used in the ACR Appropriateness Criteria. However, care must be taken to ensure that any advice to GPs is in line with the system limitations placed on GP ordering. Currently, the guidelines appear to have little input from general practitioners, yet they are the end users of the product. Inclusion of GPs on the editorial panel in future could be of benefit in the development of the most usable set of guidelines. Future revisions should also consider the need for more focused attention on specific diseases or diagnostic problems and guideline development rather than on the use of specific tests.

The evidence on which to judge the appropriateness of GP ordering behaviour and the impact of supplying imaging guidelines to GPs is scanty and equivocal. There was no published literature on the effect of the release of the earlier editions of the Australian guidelines on the ordering behaviours of GPs, but general conclusions regarding the effect of imaging guidelines can be drawn from research on those used in the United States, the United Kingdom and Canada. It would appear from some studies that the introduction of guidelines can effect a decrease in selected undesirable ordering behaviours. Other research suggests that guidelines alone are ineffective in changing behaviour and that guidelines need to be combined with feedback of individual behaviour in order to be effective. Other studies suggest that such interventions cannot overrule the important influence of patient expectations on GP behaviour.

If guidelines were to be combined with GP feedback about their ordering behaviour in an attempt to affect ordering patterns, the use of the HIC radiology data for feedback would be likely to have little impact. The current report and review of the guidelines demonstrate the importance of the relationship between morbidity and ordering.

Overall rates of ordering for all tests, or for a selected imaging test, fail to consider the age of patients being seen and their morbidity patterns—both factors which have been shown to have a strong predictive relationship to the level of testing undertaken by an individual GP.

More research needs to be undertaken into the impact of guidelines on performance, prior to any large investment in decision support systems in this area. There is a lack of evidence that their introduction will have any more impact on imaging ordering than currently available guidelines. Further, linking quality improvement in diagnostic imaging with more general quality improvement initiatives may be productive. The 'Building on Quality' project currently being undertaken by a consortium of six Divisions of General Practice with funding from the DHAC is an example of such initiatives.

In the majority of areas investigated, GPs appear to be ordering imaging in a manner that is consistent with the available guidelines. These include the ordering of chest x-rays, mammographies, breast ultrasounds and Dopplers, imaging of the kidney, imaging undertaken for abdominal pain, assessment of breast lumps, imaging of the shoulder, and management of head injuries. However, in some areas a lack of guidelines makes it difficult to assess the appropriateness of GP ordering behaviour. Consideration should be given to the development of guidelines in some problem groups where there are no guidelines currently available, such as osteoarthritis.

There are some areas investigated in this report in which there are systemic blocks to improvement in performance.

Echocardiography is the most sensitive, specific and accurate test for patients with cardio-respiratory symptoms, yet heart failure was ranked seventh in the morbidity associated with orders for plain chest x-ray. The literature clearly demonstrates that the rate of echocardiography in patients with congestive cardiac failure is well below optimum levels. There were 308,000 echocardiograms claimed against Medicare in 1999–00. This equates to a total of 16 tests per 1,000 population per year. A study in the United Kingdom suggests the appropriate level should be about 22 per 1,000. From the *BEACH* dataset in that year, the extrapolated estimate for the total echocardiograms ordered by GPs for congestive heart failure was 47,000. The remainder of the 308,000 tests would presumably have been ordered by cardiologists. Krum suggests that about 30% of patients with congestive cardiac failure managed in general practice in Australia have had an echocardiogram (Krum et al. 2001). Together, these data suggest that the quality of care provided to general practice patients could be improved with increased use of echocardiography.

GP access to echocardiography in Australia is usually through cardiologists' rooms. Krum suggests that a combination of access, the cost of the test to Medicare and a lack of knowledge about the test and its interpretation lead GPs to avoid echocardiography and rely on chest x-rays. These factors may limit application of the recently published National Heart Foundation – Cardiac Society guidelines (Krum 2001).

Mair demonstrated that the importance of early diagnoses and treatment of congestive heart failure is not fully appreciated by GPs (Mair et al. 1996). There may be insufficient education in selection of diagnostic interventions at both the undergraduate and postgraduate levels of medical training. Improved training of GPs in the early detection and treatment and the interpretation of echocardiographs may be of benefit. A greater understanding of the balance of costs between echocardiography and that of hospitalisation resulting from late diagnosis of the disease may also assist in improving the testing of congestive cardiac failure. Khunti et al. are currently conducting a randomised control trial of the cost-effectiveness of use of guidelines in improving the use of echocardiography in general practice (Khunti et al. 2000). Their results should be watched with interest.

The system limitations of the Australian health care system must also always be considered in the development of guidelines. For example, in several topics investigated in this report (e.g. suspected soft tissue injury of the shoulder and back problems) the guidelines and/or the literature suggested MRI was the appropriate modality. However, within the Australian health care system, the GP must refer to a specialist in order for MRI to be rebatable under the MBS. This blocking of direct access to MRIs means that the GP who does not feel a specialist referral to be appropriate at this stage is limited to investigations less appropriate than the MRI.

A further example is GP ordering of CT scans in the investigation of headache and head injury where a neurological cause is suspected. Both the ACR and RANZCR guidelines suggest that where vertebrobasilar problems are involved, MR angiography is the modality of choice. Since MRIs are barred to the GP, the use of CT scans is probably an appropriate substitution according to the guidelines, though only in situations where MR imaging is unavailable or inaccessible.

In the United Kingdom, studies of direct access by GPs to MRI studies have demonstrated generally appropriate use of this form of imaging (Apthorp et al. 1998; Chawda et al. 1997; Robling et al. 1998). We therefore conclude that removal of the system block for direct GP ordering of MRI, MR angiography and echocardiography should be considered for selected problems.

There are three areas in which the results of this study suggest that a decrease in imaging orders could be possible without having a negative impact on quality of care. The current level of 41.9 knee x-rays per 100 problem contacts for sprain/strain of the knee, while well below North American Emergency Department levels, could probably be significantly reduced. In the investigation of knee injury, plain x-rays have the lowest yield for diagnosing clinical significant fractures. The ACR Appropriateness Criteria for acute trauma to the knee synthesise the research into the use of decision rules for knee imaging (Stiell et al. 1995c). Introduction of these criteria in Australia has the potential to significantly reduce orders for knee radiology without losing sensitivity of fracture detection.

The evidence suggests that spinal x-ray is contraindicated within 14 days of onset of uncomplicated back pain with or without radiation. There is no scientific evidence to support the use of x-ray short of seven weeks from onset and even then the return was poor (Quebec Task Force on Spinal Disorders 1987). The guidelines all support the proposition that uncomplicated low back pain without 'red flags' will settle within 6–12 weeks in 90% of cases without investigation and with little more than supportive therapy (Agency for Health Care Policy and Research 1994; Anderson et al. 2000; RANZCR 2001). In this study, GPs ordered spinal imaging at the rate of 28.7 per 100 new back pain problems seen and 13.2 per 100 old back problems. While some of the patients with new back problems may have had symptoms for some time before presentation, these data suggest that an appropriate intervention could reduce the spine x-ray rate without decreasing the quality of patient care.

In this study GPs ordered imaging at a rate of 44.8 per 100 problems labelled as fractured ankle and at the rate of 37.7 per 100 problems labelled as ankle sprain/strain. In the case of ankle injury, the well-validated Ottawa decision rules for imaging adopted by the ACR as appropriate guidelines for the diagnosis of fracture. The RANZCR guidelines do not provide any guidance on the selection of patients for x-ray for suspected fracture of the ankle. Providing GPs in Australia with the Ottawa guidelines has the potential to decrease the x-ray level in this area.

This study has provided new insight into the relationship between GP characteristics and ordering behaviour, and between orders for specific imaging test types, the patients for whom the orders were placed and the problems under investigation. It has demonstrated

that in the majority of areas GP ordering behaviour follows the available guidelines. However, it has also highlighted some areas in which improvement would be desirable in both the guidelines and in GP test selection.

This report provides a baseline against which future practice can be compared so that changes over time in imaging order behaviour can be documented. It further provides a pre-measure of ordering patterns on which to test the impact of the new RANZCR guidelines on ordering behaviour of general practitioners in the future.