The Risks and Benefits of Searching for Incidental Findings in MRI Research Scans

Jason M. Royal and Bradley S. Peterson

The question of how to handle incidental findings (IFs) has sparked a heated debate among neuroimaging researchers and medical ethicists, a debate whose urgency stems largely from the recent explosion in the number of imaging studies being conducted and in the sheer volume of scans being acquired. Perhaps the point of greatest controversy within this debate is whether the magnetic resonance imaging (MRI) scans of all research participants should be reviewed in an active search for pathology and, moreover, whether this search should be performed by a radiologist. Resistance to routine readings performed by radiologists, as opposed to selective review of those scans on which investigators have spotted a possible IF, has been fueled in part by the obvious and enormous cost — financial and logistical — of engaging radiologists to read massive numbers of scans. This cost would be especially burdensome, even prohibitive, to investigators who are not affiliated with a medical center, because of their limited access to radiologists and other medical expertise.

Nevertheless, others have argued that having radiologists routinely read all research scans provides the surest means of protecting participants from unidentified but potentially life-threatening conditions that may appear as IFs. Such readings therefore would fulfill the ethical obligation of researchers to maximize benefits and minimize risks to participants. Many — and perhaps the majority — of neuroimagers accept to some degree the validity of this ethical argument, even though virtually all acknowledge the practical difficulties of instituting routine readings by radiologists, as well as the probability that the costs of such a policy might determine which investigators can conduct research in the future, and in which settings. The net effect of this controversy thus has been some movement toward routine radiologist readings as an implicit policy across the field, but with considerable "heel dragging" — nearing the point of impasse — over the feasibility of its implementation.

Indeed, the general acceptance of the ethical justification for routine readings places the "burden of proof" upon the critics of this policy, who must show that the practical hurdles to routine readings are so extensive and intractable that they effectively outweigh what

Jason M. Royal *is currently a doctoral student in clinical psychology at the City University of New York. He served until recently as Scientific Editor in the Pediatric Brain Imaging Lab at the New York State Psychiatric Institute.* **Bradley S. Peterson, M.D.**, *is the Suzanne Crosby Murphy Professor in Pediatric Neuropsychiatry and Director of MRI Research in the Department of Psychiatry at Columbia College of Physicians & Surgeons and the New York State Psychiatric Institute.*

would seem to be an ethical obligation. The argument for routine readings as an ethical imperative, however, rests on a number of presumptions that have remained largely unquestioned. These presumptions magnify the potential benefits to participants of routine radiological readings while downplaying both the limitations in what research-quality scans can reveal and the difficulty of distinguishing reliably and consistently among pathology, benign anomalies, and visual artifacts in the scans themselves. Accepting these presumptions at face value thus exposes participants to a number of risks related to the limitations of research-quality scans, including the risks of false-positive and false-negative readings, as well as the risks of treating IFs of unknown clinical significance. These risks, moreover, are incurred without any certainty of yielding substantial benefit to the participants. Presumptions about the benefits and risks of having radiolo-

images, in contrast, are much more useful clinically than are T1-weighted images, because they can show the presence of tissue edema and necrosis, processes more directly related to tumors themselves than is the distortion of surrounding tissue detected by T1-weighted scans. T2-weighted images, however, are the least commonly used type of image in neuroimaging research and are therefore infrequently available for reading by a radiologist. Echoplanar images, another type of research-quality scan used for functional MRI (fMRI), offer extremely poor resolution and are highly prone to artifacts, making them virtually useless for clinical evaluation. Thus, when radiologists read research-quality scans as part of a study protocol, they most often read scans of limited or no clinical value.

Indeed, a neurological examination or a thorough review of possible symptoms caused by medical illnesses is likely to be more sensitive and specific than is

> Perhaps the point of greatest controversy within this debate is whether the magnetic resonance imaging (MRI) scans of all research participants should be reviewed in an active search for pathology and, moreover, whether this search should be performed by a radiologist. Resistance to routine readings performed by radiologists, as opposed to selective review of those scans on which investigators have spotted a possible IF, has been fueled in part by the obvious and enormous cost — financial and logistical — of engaging radiologists to read massive numbers of scans.

gists routinely search for pathology in research-quality scans therefore must be considered carefully before instituting any policy regarding the reading of participant scans.

One such presumption is that having a radiologist actively search for abnormalities of concern in research-quality scans will maximize the benefit to participants by maximizing the chance of detecting lesions, both symptomatic and asymptomatic, that can then be treated. Another presumption is that the chance of missing serious abnormalities will be minimized as well. The known limitations of research-quality scans, however, reduce their utility as a tool for detecting significant lesions (Figure 1). T1-weighted images, the type of scan most commonly used for research purposes, are extremely limited in their clinical utility in that they can detect little more than anatomical distortions associated with large, space-occupying lesions or hydrocephalus. Imaging protocols that use only T1-weighted images are thus unlikely to detect lesions that are clinically silent. T2-weighted

the reading of research-quality scans for the detection of clinically significant, symptomatic lesions. A review of possible symptoms therefore reasonably could be included in the screening of potential research participants during recruitment. Individuals who report significant symptoms could then be sent immediately to a physician for follow-up and, if appropriate, be excluded from the study. A symptoms review thus benefits symptomatic participants by reducing any delay to their seeking treatment — a delay that may well occur if detection is contingent on scheduling a scan and awaiting a radiological reading, if detection occurs at all. Screening out these individuals also will reduce the chance that clinically significant IFs will be identified within the sample, as well as the risk that a clinically significant lesion will be missed.

Although clinically manifest lesions may be identified more effectively through a review of symptoms, one might presume that having a radiologist read all scans would increase the likelihood that *asymptomatic* lesions would be identified and then treated

once detected. The benefit of actively treating clinically silent lesions, however, is far from certain. Estimated rates of the prevalence of clinically significant abnormalities within the general population, such as tumors[1] and aneurysms,[2] likely represent only lower bounds for the true prevalence of these abnormalities. Their true prevalence rates, although impossible

The treatment for many such abnormalities, however, can involve substantial risk. For example, a recent aggregate analysis of treatments for unruptured aneurysms found that endovascular coiling and surgical clipping — two standard treatments for unruptured aneurysms — had rates for adverse outcomes (including death) of 8.8 percent and 17.8 percent, respectively.[3] Treatment in some cases may therefore pose a greater risk to the health and psychological well-being of the participant than does the abnormality itself.

Even when the choice is made not to treat immediately such clinically silent lesions, a program of periodic monitoring may be recommended, which in itself may be stressful for the participant, who essentially will be "waiting for the other shoe to drop." This monitoring may even include substantial changes in lifestyle that impair the participant's quality of life. For instance, clinicians are likely to recommend periodic monitoring of numerous IFs whose clinical significance is difficult

Three Types of MRI Scans Commonly Used in Neuroimaging Research

(a) T1-weighted   (b) T2-weighted   (c) Echoplanar

(a) T1-weighted images, the most common in neuroimaging research, are used to investigate the anatomical structure of various regions of the brain. T1-weighted images can reveal only comparatively large, space-occupying tumors and hydrocephalus. The presence of a tumor appears in a T1-weighted image as a distortion of the anatomical structures that surround the space occupied by the tumor. Tumors themselves, however, cannot be seen. (b) T2-weighted images can reveal visually processes that are directly related to tumors, such as tissue edema and necrosis. However, this type of image is used far less frequently in neuroimaging research that is the T1-weighted image. (c) Echoplanar images, which are used in functional MRI, are of virtually no clinical value because of their extremely low resolution and poor signal-to-noise characteristics.

to determine, are probably higher, with a considerable number of abnormalities remaining silent for the rest of the individual's life. Because these abnormalities may remain undetected, we know very little about their natural history, and we therefore have no way of predicting their outcome upon detection as an IF. Many of these lesions, if left alone, may never require treatment. Thus, no clear justification exists for treating clinically silent abnormalities as a matter of course once they have been detected as IFs.

Detection of such a lesion nevertheless may compel the individual to seek treatment. Indeed, treatment may be recommended strongly by health care providers whom the individual consults as a safeguard against a health crisis in the future, despite the fact that the likelihood of such a crisis actually occurring is unknown.

to determine initially, but that often ultimately prove benign. IFs of uncertain clinical significance are common and include "unidentified bright objects" (UBOs) (Figure 2a), low-level to moderate cortical atrophy, moderate enlargement of the ventricles (Figure 2b), cysts of various kinds, borderline Chiari I malformations (Figure 2c), and small, isolated demyelinating plaques. We detected a substantial number of these types of IFs in brain images acquired from 641 participants over the course of some 15 years of research (Tables 1 and 2, Figure 2).

Finding Chiari I malformations (Figure 2c), for instance, may lead to recommended prohibitions on certain types of physical activities, such as contact sports, the loss of which can be especially disruptive to the lives of physically active young people. These

individuals may also be warned of their increased susceptibility to the dangers of neck trauma, as can occur in a motor vehicle accident, and then they will have to cope with worrying about accidents that they cannot avoid. The benefit of such warnings and prohibitions to the participant may be questionable, in that little is known of the relationship between the degree of tonsillar ectopia (the lowering of the tonsils of the cerebellum below the protective encasement of the skull, the defining characteristic of Chiari I malformation) and the appearance of symptoms.[4] Thus, detecting and either treating or monitoring clinically silent IFs — abnormalities that otherwise may have remained asymptomatic forever — can have decidedly adverse effects on the lives of participants without the certainty of providing them with a substantial benefit. Indeed, even when a clinically silent abnormality does indicate the presence of a life-threatening condition, we cannot assume that detection and treatment will necessarily benefit the participant by altering the natural course of that condition. Detecting a malignant tumor that has metastasized from another part of the body, for example, may not benefit the health of the participant because treatment may have little or no impact on an advanced cancer.

For those clinically silent lesions for which detection and treatment could help to alter the course of an illness, the presumption is that having a radiologist review all scans will maximize the possibility of identification and subsequent treatment. However, again, limitations in the quality of research scans make the detection of such abnormalities far from reliable. Indeed, imaging protocols that use T1-weighted images only — the majority of protocols used in neuroimaging research — may well fail to detect lesions that are clinically silent. The rate of false-negative readings thus is likely to be high, particularly in protocols that use only T1-weighted images. False-negative

Table 1

Incidental Findings (IFs) in a Sample of 641 Research Participants

		Patient Participants[1]	Normal Controls	Total
Sample Size		397	244	641
Age Groups (in years)	0-17	253	119	372
	18-50	123	106	229
	51+	21	19	40
Gender	Male	237	110	347
	Female	159	132	291
Scan Type	T1	94	90	184
	T1 & T2	303	154	457
Subjects with an IF		138/397 (34%)	66/244 (27%)	204/641 (32%)
Referals[2]		11/388 (2.8%)	12/231 (5.1%)	23/619 (3.7%)
	Routine[3]	8/388 (2.0%)	11/231 (4.7%)	19/619 (3.0%)
	Urgent[4]	3/388 (0.7%)	1/231 (0.4%)	4/619 (0.6%)

Selected Analyses: Patient participants were significantly more likely to have an IF detected than were healthy controls ($F_{1,638}$=4.2, p=0.04). However, the rate of IFs across the 6 age groups differed significantly ($F_{5,635}$=3.9, p-value=0.002), with control participants 51 years old and older being most likely to have an IF detected. In addition, IFs within the brain were significantly more likely to be detected when T2-weighted images were read than when T1-weighted images alone were read ($F_{1,139}$=16.3, p-value=0.0001).

[1]*Diagnosed with a neuropsychiatric disorder, including Tourette Syndrome, Obsessive-Compulsive Disorder, Attention Deficit-Hyperactivity Disorder, Anxiety Disorder, and Major Depressive Disorder.*
[2]*We had referral information for 619 out of 641 total participants.*
[3]*A classification of "routine" was given for conditions that were unlikely to be clinically serious but that warranted follow-up, such as Chiari I malformations, mastoid disease, evidence of past trauma, and various nonspecific anomalies of indeterminate etiology.*
[4]*We included in the "urgent" category IFs of unknown etiology or significance whose characteristics could suggest the presence of worrisome lesions (e.g., tumors or vascular malformations). Three of these "urgent" referrals were for patients—2 children and 1 adult under age 50. The fourth "urgent" referral was for a control adult under age 50. Follow-up information for one of these participants indicated that no worrisome lesion was present (see Figure 3). Readings of the scans from this sample of 641 research participants found no obvious and definitive evidence of tumors, aneurysms, or arteriovascular malformations.*

readings can lead, in turn, to an erroneous assumption by participants of a "clean bill of health." For those participants who would already be reluctant to seek treatment, this mistaken assumption may delay treatment further.

Even greater is the substantial risk of false-positive findings. The images used for research often contain artifacts, visual imperfections in the scan itself that do not represent anything physiological in the participant and that are caused most commonly by participant movement during the imaging procedure. Such artifacts can be difficult to distinguish from lesions, and this uncertainty, potentially combined with concerns about medicolegal liability from failing to identify a serious IF, often predisposes radiologists to err on the "safe side" by recommending further consultation. The uncertainty engendered by such recommendations, however, can cause considerable emotional distress for participants, not least because many individuals tend to assume the worst — that the IF in question represents a clinically serious abnormality — despite assurances by either researchers or radiologists that such an anomaly may well be benign or even nothing at all. This emotional distress often is particularly intense in the parents of child participants to whom suspicious IFs have been disclosed. Added to the possibility of substantial emotional distress is the potential for imposing on the participant considerable cost — in terms of both money and time — from follow-up consultation, especially because the follow-up typically will include not only an examination by a physi-

cian but also the expense of an additional MRI scan that is calibrated to detect clinically significant abnormalities, unlike the scans used for research. This cost is most worrisome for disadvantaged populations who participate in research, along with others who may not have adequate access to health insurance and other

> False-negative readings can lead, in turn, to an erroneous assumption by participants of a "clean bill of health." For those participants who would already be reluctant to seek treatment, this mistaken assumption may delay treatment further.

support. Participants thus may be exposed to the risk of anxiety and expense of various kinds, only to find that nothing is wrong with them. In addition, false positives may increase the rate of participant withdrawal from research, given that participants in whom an IF is identified arguably may be more likely to drop out of a study in order to seek clinical evaluation for the IF. False-positive findings — a substantial possibility because of the frequent occurrence of imaging artifacts in research-quality scans — thus may have a decidedly troubling impact on participants, as well as on the research itself.

The risk of false-positive readings, along with all of the other risks that we have discussed, apply not only to the routine reading of research-quality scans by radiologists, but also to the IFs identified by researchers themselves during the normal course of their work with images. The risks are magnified, however, when

Table 2

Detection Rates of Incidental Findings (IFs) with Indeterminate Clinical Significance

Incidental Findings*	Patient Participants	Normal Controls	Total
Unidentified Bright Object (UBO)	40/397 (10.0%)	20/244 (8.1%)	60/641 (9.3%)
Chiari I/Ectopia/Herniation	12/397 (3.0%)	6/244 (2.4%)	18/641 (2.8%)
Atrophy/Volume Loss	8/397 (2.0%)	7/244 (2.8%)	15/641 (2.3%)
Cyst	9/397 (2.2%)	6/244 (2.4%)	15/641 (2.3%)
Ventriculomegaly	5/397 (1.2%)	0/244 (0%)	5/641 (0.7%)

Selected Analyses: The detection rates of indeterminate IFs did not differ significantly between patients and controls ($F_{1,639}$=0.83, p-value=0.36). However, detection rates of indeterminate IFs across six age groups (patients and controls, each divided into three age groups: 0-17, 18-50, and 51+) did differ to a degree that approached statistical significance ($F_{5,635}$=2.2, p-value=0.054), with detection of indeterminate IFs most likely in controls aged 51 years or older. Controls aged 18-50 were the least likely to have one of these IFs detected. In addition, clinically indeterminate IFs were significantly more likely to be detected when T2-weighted images were read than when T1-weighted images alone were read ($F_{1,639}$=19.7, p-value=0.0001).
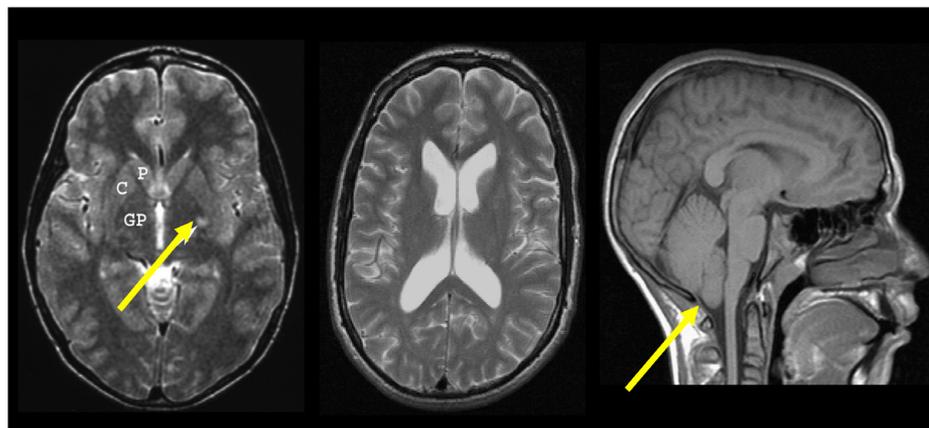
The detection rate for each IF includes occurrences in both participants who received a referral and those who did not. More than one finding could occur in a single individual.

radiologists or even researchers actively hunt for pathology in research-quality scans, because they are more likely to identify as IFs anomalies that are small and isolated, or artifacts that have a much lower probability of being clinically relevant than do anomalies that are larger and that appear obvious to researchers who are not hunting for IFs. Despite the increased

We assert that most, if not all, abnormalities of clinical significance that occur in research-quality scans (as opposed to scans that are optimized for clinical diagnosis) are likely to be obvious to investigators who have substantial experience reviewing MR images, without those investigators routinely hunting for IFs. Such abnormalities will be evident particularly to those investigators who have considerable experience working with anatomical images, given that the only IFs that currently can be identified are structural lesions of brain tissue rather than any relatively more subtle disturbances of brain function. Thus, searching for pathology in research-quality scans actually may increase the risk to participants, without affording them greater protection.

Incurring the increased risks that are involved in routine readings by a radiologist may be appropriate in specific populations that ultimately are found to be at an *a priori* greater risk for clinically silent but serious lesions. The decision whether to institute routine radiologist readings, however, must be founded on a careful cost-benefit analysis of the risks associated with actively

Figure 2

Incidental Findings (IFs) with Indeterminate Clinical Significance

(a) Unidentified Bright Object   (b) Ventriculomegaly (Enlarged Ventricles)   (c) Borderline Chiari I Malformation

(a) Unidentified bright objects (UBOs), which appear as bright spots on T2-weighted images (see yellow arrow in panel (a)), rarely can indicate the presence of vascular malformations, demyelinating plaques of multiple sclerosis, infarctions, or other entities. They also can appear as nonspecific, clinically indeterminate IFs, however, primarily in older populations, as well as in younger populations with various neurological or psychiatric illnesses. (b) Ventriculomegaly, or enlargement of the ventricles, is a nonspecific finding that has numerous possible causes, including destruction of tissue around the ventricles or obstruction of the flow of cerebrospinal fluid. (c) Chiari I malformation is a condition in which a portion of the cerebellum (the cerebellar tonsils in particular) has dropped below the protective encasing of the skull (through the foramen magnum), a process known as "ectopia" (see yellow arrow in panel (c)). Symptoms usually appear in midlife, but Chiari I malformations and ectopia often are identified as IFs in adolescent populations in neuroimaging studies. Severe ectopia can put an individual at risk for potentially life-threatening injury to the cervical spinal cord, which regulates a number of vital bodily functions. The clinical significance of mild or borderline ectopia, however, is unclear, and individuals with mild ectopia usually remain entirely asymptomatic. The identification of such IFs with indeterminate clinical significance can expose participants to the burden and risks of treatment or ongoing clinical monitoring, which may be unnecessary and of little clinical benefit.

risk of false-positive findings, some feel that an active search for IFs by a radiologist continues to be justified because radiologists alone are qualified to identify abnormalities on research-quality scans that warrant further clinical attention. Others argue that, barring radiologist involvement, researchers themselves should actively and systematically search for pathology in the scans of all research participants in order to maximally protect the health interests of participants.

searching for pathology in research-quality scans compared to data on the *a priori* risk of detecting serious IFs in a given population. That *a priori* risk could be defined, for example, by clinical diagnosis, age, or both. Such cost-benefit analyses also should take into account the use of specific types of scans (e.g., T1-weighted vs. T2-weighted), given the differences across scan type in what anomalies can and cannot be detected.

Few data have been published, however, that would allow for comparative analysis of the rates of serious IFs across diagnoses, age, and scan type. Recent studies of IFs in neuroimaging research have focused on healthy children and adults,[5] with few exceptions.[6] Thus, data on individual populations with specific diagnoses are limited. No studies to our knowledge have analyzed rates of IFs across scan type. One study examined the occurrence of IFs across age groups within a sample of 151 healthy adults.[7] This study found that adults aged 60 years and over were more likely than adults aged 18 to 59 years to have an IF detected, a finding that is somewhat supported by earlier studies.[8] The three "urgent" IFs that were identified, however, occurred in the younger group, perhaps broadly suggesting that IFs may tend to be more serious in this group when they occur at all, or perhaps that the radiological readings that generated the IFs tended to be more conservative when the images were from younger persons. Comparison of the rates of serious IFs detected in existing studies of healthy children and adults ideally could help to distinguish further between levels of *a priori* risk across age within healthy individuals. Comparing results across these studies, however, is complicated by differences among studies in the criteria used for categorizing IFs as "clinically significant" or "urgent."

Ranges for the detection rates in the general population of clinically significant and urgent IFs — estimated, respectively, as 2-8 percent[9] and 0.5-2 percent[10] — are based on combined data from healthy children and adults and therefore, by their nature, cannot help us to distinguish differences in the *a priori* risk of silent but serious IFs across either specific diagnoses or across age. Moreover, a lack of clinical follow-up in a number of studies, both to confirm the presence of IFs and to determine the clinical outcome of abnormalities, makes impossible any distinction between false- and true-positive IFs, or even between IFs of confirmed urgency and those that prove to be harmless over time. These statistical ranges thus represent the rates of *detection* of IFs that are prospectively thought to be significant or urgent, but not rates of *confirmed* occurrence. Indeed, a recent study of IFs in a sample of healthy children[11] shows clearly how clinical follow-up can reduce a high rate of perceived serious or "urgent" IFs within a sample to zero. In this case, all IFs in question turned out to be either a nonspecific anomaly that was likely benign or an artifact in the scan.

Some investigators have compared the rates of serious abnormalities (e.g., tumors and aneurysms) that have been detected as IFs within a given sample with the estimated prevalence rates for those abnormalities within the general population.[12] By definition, however, estimated prevalence rates of specific abnormalities within the general population exclude abnormalities that remain silent for the life of the individual — precisely the kind of abnormality that arguably is detected by radiological readings. Estimated prevalence rates for the general population therefore cannot serve as a meaningful standard against which we can interpret rates of serious IFs in samples of either patients or healthy individuals, unless they are regarded as providing an estimate only of the lower bound of the true population rates of IFs.

We analyzed our data from a sample of 641 research participants (Table 1) to determine whether the detection rate for IFs differed across diagnostic status, age groups, or the types of scan acquired. Our sample includes individuals who have been recruited into our research because they have been diagnosed with a neuropsychiatric disorder ("patients") and individuals who have been recruited because they have been deemed healthy based on the results of a recruitment screening process (normal "controls"). We found that the patient participants in our sample were significantly more likely to have an IF detected than were healthy controls ($F_{1,638}=4.2$, p=0.04). However, when we categorized both patient participants and control participants by age into one of three groups — 0-17 years, 18-50 years, and 51 years and older — we found that the rate of IFs across the six groups differed significantly ($F_{5,635}=3.9$, p-value=0.002), with control participants 51 years old and older being most likely to have an IF detected. Controls 18-50 years were the least likely to have an IF detected. Patients in the three age groups did not differ significantly from each other in the rates of IFs detected ($F_{2,394}=1.8$, p-value=0.16). We also compared the number of participants who had an IF detected when only T1-weighted images were read with the number of those with an IF for whom T2-weighted images were read. Only abnormal findings within the brain were included as IFs in this analysis. We excluded anatomical variants deemed normal and abnormalities located outside of the brain itself, such as sinus disease, prominent adenoids, and mastoiditis. We found that IFs within the brain were significantly more likely to be detected when T2-weighted images were read than when T1-weighted images alone were read ($F_{1,139}=16.3$, p-value=0.0001).

We also performed analyses to determine the detection rates for IFs of indeterminate clinical significance. This category comprises IFs for which the natural history and the clinical significance are unknown, and the identification of which may increase participants'

risk of unnecessary treatment and monitoring. In our sample, these IFs included UBOs, Chiari I malformations, atrophy or volume loss, cysts, and enlarged ventricles (Table 2). We found that the rates of indeterminate IFs across the six age groups (three age groups of patient participants and three age groups of normal controls) differed to a degree that approached sta-

cal significance, with T2-weighted images being much more likely to yield an IF than T1-weighted images alone. This dramatic difference in the *a priori* rates of IFs by scan type suggests that the value of a clinical reading is substantially lower when only T1-weighted images are acquired, as is most commonly the case in neuroimaging studies. The increased clinical value of reading T2-weighted images, however, may be counterbalanced by the fact that reading T2-weighted images increases the risk to participants of detecting IFs of indeterminate significance, the treatment or monitoring of which may be unnecessary or even harmful.

In addition, a broad interpretation of our results supports the conclusions of previous studies regarding differences in the rates of IFs across ages. Our findings suggest that the detection of IFs of any kind may be most likely in healthy participants who are over 50 years of age. All "urgent" IFs, however, were detected in children and in adults under age 50, which could suggest that IFs are more likely to be deemed of concern

Figure 3

Limitations of Research-Quality Scans Can Increase Risk and Burden to Participants

A radiologist identified in the T2-weighted brain scan of this 49-year-old woman an abnormality that appears as a bright spot (a "hyperintensity") located inferior to the sphenoid sinus and superior to the nasopharynx (indicated by an arrow). The abnormality initially was thought most likely to be a benign and relatively common type of cyst (termed a "Thornwaldt" cyst). This diagnosis could not be confirmed, however, based on a reading of our research-quality scans. Because of the outside possibility that the anomaly could have been a tumor or some other worrisome lesion, the participant was referred for subsequent clinical examination and MRI. Further consultation revealed no evidence of a dangerous tumor or lesion. The participant nevertheless was advised to undergo a course of periodic examination and additional MRIs — a "wait and watch" approach. In this instance, the limited information in our research-quality scan, compared with the more conclusive information that a detailed and definitive clinical scan provided, drew the participant into an open-ended program of clinical monitoring that may be neither necessary nor beneficial.

tistical significance ($F_{5,635}$=2.2, p-value=0.054), with detection of IFs in this category most likely in controls aged 51 years or older. Controls aged 18-50 were the least likely to have one of these IFs detected. In addition, we found that clinically indeterminate IFs were significantly more likely to be detected when T2-weighted images were read than when T1-weighted images alone were read ($F_{1,639}$=19.7, p-value=0.0001).

Our findings, along with previous findings, suggest that the rates — and therefore the *a priori* risks — of IFs of varying degrees of clinical significance do indeed differ across the domains of age, scan type, and clinical status. Scan type has a considerable impact on the rate of detecting IFs, including IFs of uncertain clini-

when they occur in these younger age groups. Any interpretation of these data is highly provisional, however, because we lack information on clinical follow-up for three of the four participants who received an "urgent" referral. Indeed, follow-up on the fourth "urgent" referral showed that the anomaly in question was not, in fact, likely to be a dangerous lesion (Figure 3). This case illustrates both the limited information afforded by research-quality scans, and the crucial role that clinical follow-up must play in evaluating the ultimate significance of any IF.

Further investigation clearly is needed to determine whether to institute routine radiological readings or any other active search for pathology in specific popu-

lations of participants in neuroimaging studies. Future investigations of the IF problem should distinguish between clinically silent and clinically manifest IFs. They also ideally should include multiple age groups and diagnoses of differing illnesses, as well as differing scan types, in order to analyze the rates of urgent,

<div style="color:red">

The question, it turns out, may not be whether financial and practical obstacles outweigh an ethical imperative to search for pathology in the MRI scans of participants. Rather, we must ask under what conditions the search itself is justified, ethically, given the substantial risks to participants that this practice likely incurs.

</div>

clinically significant, and clinically ambiguous IFs across these domains. Although difficult to achieve, criteria for categorizing IFs as "clinically significant," "routine," and "urgent" should be standardized to a greater degree, thereby facilitating a more meaningful comparison of findings across studies. In addition, inclusion of the clinical outcomes of follow-up evaluations for identified IFs will be essential to determining rates of false-positive findings, as well as the rates of true-positive findings that remain clinically silent. Finally, research must be conducted on the emotional and financial ramifications of false-positive findings for participants, as well as on the effects that notifying participants of the likelihood of detecting IFs has on research recruitment, given the high likelihood of false positives when reading research-quality scans.

The adverse effects of these risks on participants and on study recruitment can be contained, to some degree, through appropriate strategies for disclosing to potential participants the risks related to identifying IFs that are either true- or false-positives. Participants in neuroimaging research should be informed of all risks, explicitly and in clear terms, both in consent forms and verbally during the consent process. In addition, given the limitations in the quality of research scans and the clinically relevant information they contain, researchers should consider the option of disclosing to participants only IFs that are obviously life-threatening, whether such lesions are detected through a radiologist's reading or an investigator's reading during the normal course of research. If this option for managing IFs is selected, then participants must be made aware of it in the consent process. Alternatively,

participants could be allowed to opt out of having disclosed to them any IFs except those that are obviously life-threatening.

We cannot assume that the review of participant scans by a radiologist — or any active search for pathology in research-quality scans — inevitably will be beneficial to participants. Reviewing scans that are calibrated for research is, in all probability, less effective in detecting clinically manifest lesions than is a neurological exam or a thorough review of symptoms. In addition, the limitations of research-quality scans make the detection of clinically *silent* lesions far from certain, and treating the clinically silent abnormalities that ultimately are detected on research-quality scans may expose participants to risks that outweigh those posed by the abnormality itself. In addition, an active search for pathology in participant scans likely increases the risks to participants that are inherent in any attempt to identify IFs. These risks include the possibility of false-positive and false-negative findings, as well as the risk of drawing participants into a stressful, potentially costly course of clinical monitoring for anomalies that may remain silent and benign indefinitely. We argue that the presence of these risks and limitations drastically undermines the position that actively searching for pathology in research-quality scans, as a matter of course, maximizes benefit and minimizes risks to participants. Indeed, the opposite may be true.

These risks and limitations, although acknowledged previously, have remained relatively unexamined. The full extent of their import consequently has been overlooked in the ongoing debate on how to handle IFs in neuroimaging research. Our aim has been both to enumerate these risks and limitations and to outline logically the potentially considerable consequences that they may have for participants. The question, it turns out, may not be whether financial and practical obstacles outweigh an ethical imperative to search for pathology in the MRI scans of participants. Rather, we must ask under what conditions the search itself is justified, ethically, given the substantial risks to participants that this practice likely incurs.

References

1. F. G. Davis et al., "Prevalence Estimates for Primary Brain Tumors in the United States by Behavior and Major Histology Groups," *Neuro-Oncology* 3, no. 3 (2001): 152-158.
2. G. J. Rinkel et al., "Prevalence and Risk of Rupture of Intracranial Aneurysms: A Systematic Review," *Stroke* 29, no. 1 (1998): 251-256; J. M. Wardlaw and P. M. White, "The Detection and Management of Unruptured Intracranial Aneurysms," *Brain* 123, Pt. 2 (2000): 205-221.
3. T. Lee et al., "Aggregate Analysis of the Literature for Unruptured Intracranial Aneurysm Treatment," *American Journal of Neuroradiology*, 26, no. 8 (2005): 1902-1908.
4. J. Meadows et al., "Asymptomatic Chiari Type I Malformations Identified on Magnetic Resonance Imaging," *Journal of Neurosurgery* 92, no. 6 (2000): 920-926.
5. J. Illes et al., "Ethical Consideration of Incidental Findings on Adult Brain MRI in Research," *Neurology* 62, no. 6 (2004): 888-890; G. L. Katzman et al., "Incidental Findings on Brain Magnetic Resonance Imaging from 1000 Asymptomatic Volunteers," *JAMA* 282, no. 1 (1999): 36-39; B. S. Kim et al., "Incidental Findings on Pediatric MR Images of the Brain," *American Journal of Neuroradiology* 23, no. 10 (2002): 1674-1677; S. Kumra et al., "Ethical and Practical Considerations in the Management of Incidental Findings in Pediatric MRI Studies," *Journal of the American Academy of Child and Adolescent Psychiatry* 45, no. 8 (2006): 1000-1006; F. Weber and H. Knopf, "Incidental Findings in Magnetic Resonance Imaging of the Brains of Healthy Young Men," *Journal of the Neurological Sciences* 240, nos. 1-2 (2006): 81-84.
6. D. I. Lubman et al., "Incidental Radiological Findings on Brain Magnetic Resonance Imaging in First-Episode Psychosis and Chronic Schizophrenia," *Acta Psychiatrica Scandinavica* 106, no. 5 (2002): 331-336.
7. See Illes et al., *supra* note 5.
8. F. E. de Leeuw et al., "Prevalence of Cerebral White Matter Lesions in Elderly People: A Population Based Magnetic Resonance Imaging Study: The Rotterdam Scan Study," *Journal of Neurology, Neurosurgery & Psychiatry* 70, no. 1 (2001): 9-14; O. Salonen et al., "MRI of the Brain in Neurologically Healthy Middle-Aged and Elderly Individuals," *Neuroradiology* 39, no. 8 (1997): 537-545.
9. J. Illes et al., "Discovery and Disclosure of Incidental Findings in Neuroimaging Research," *Journal of Magnetic Resonance Imaging* 20, no. 5 (2004): 743-747.
10. J. Illes et al., "Incidental Findings in Brain Imaging Research," *Science* 311, no. 5762 (2006): 783 784.
11. See Kumra et al., *supra* note 5.
12. See Weber and Knopf, *supra* note 5.