이학박사 학위논문

Diagnostic sensitivity and specificity of 2-mSv CT vs. conventional-dose CT in adolescents and young adults with suspected appendicitis: a *post hoc* subgroup analysis of the LOCAT

충수염 의증 청소년 및 젊은 성인에서 **2-mSv** **CT**와 기존 선량 **CT**의 민감도 및 특이도: **LOCAT**의 사후 하위그룹 분석

2019년 8월

서울대학교 대학원

융합과학부 방사선융합의생명전공

고 유 선

Ph.D. Dissertation in Science

Diagnostic sensitivity and specificity of 2-mSv CT vs. conventional-dose CT in adolescents and young adults with suspected appendicitis: a *post hoc* subgroup analysis of the LOCAT

충수염 의증 청소년 및 젊은 성인에서 2-mSv CT와 기존 선량 CT의 민감도 및 특이도: LOCAT의 사후 하위그룹 분석

August 2019

Graduate School of Convergence Science and Technology
Seoul National University
Biomedical Radiation Science Major

Yousun Ko

Diagnostic sensitivity and specificity of 2-mSv CT vs. conventional-dose CT in adolescents and young adults with suspected appendicitis: a *post hoc* subgroup analysis of the LOCAT

지도 교수 이 경 호

이 논문을 이학박사 학위논문으로 제출함
2019년 4월

서울대학교 대학원
융합과학부 방사선융합의생명전공
고 유 선

고유선의 이학박사 학위논문을 인준함
2019년 7월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

ABSTRACT

Diagnostic sensitivity and specificity of 2-mSv CT vs. conventional-dose CT in adolescents and young adults with suspected appendicitis: a *post hoc* subgroup analysis of the LOCAT

Yousun Ko

Program in Biomedical Radiation Science

Department of Transdisciplinary Studies

Graduate School of Convergence Science and Technology

Seoul National University

Introduction: To explore heterogeneity across patient or hospital characteristics in the diagnostic sensitivity and specificity of 2-mSv CT relative to conventional-dose CT (CDCT) in adolescents and young adults with suspected appendicitis.

Methods: We used the per-protocol analysis set of a large randomized controlled noninferiority trial conducted between Dec 2013, and Aug 2016, comparing 2-mSv

CT and CDCT (typically 7 mSv). The data included 2,773 patients (median age [interquartile range], 28 [21–35] years) and 160 radiologists from 20 hospitals. We tested for heterogeneity in sensitivity and specificity for the diagnosis of appendicitis across predefined subgroups by patient sex, body size, clinical risk scores for appendicitis, time of CT examination (i.e., working hours [typically 08:00–17:00 of working days] vs. after hours), CT machines, radiologists' experience, previous site experience in 2-mSv CT, and site practice volume. We drew forest plots and tested for additive or multiplicative treatment-by-subgroup interaction on sensitivity and specificity.

Results: The 95% CIs for the between-group differences, particularly for sensitivity, were wide due to small sizes (< 200) for the subgroups of extreme body sizes, high clinical risk score for appendicitis, newer CT machines, hospital with prior experience in 2-mSv CT, and hospitals with small appendectomy volume. Otherwise, the 95% CIs in most subgroups contained the previously reported overall between-group differences as well as null hypothesis value (i.e., 0). There was no significant additive or multiplicative interaction for either sensitivity or specificity.

Conclusions: We found no notable subgroup heterogeneity, which implies that 2-mSv CT can replace CDCT in diverse populations. Further studies are needed for the populations for which our subgroups were small.

Keywords: Diagnostic performance, Appendicitis, Computed tomography, Subgroup heterogeneity

Student Number: 2016-34869

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACR	American College of Radiology
AE	Adverse Event
ALARA	As Low As Reasonably Achievable
APR	Appendiceal Perforation Rate
AUC	Area Under the receiver-operating-characteristic Curve
BMI	Body Mass Index
CDCT	Conventional-Dose CT
CI	Confidence Interval
CRF	Case Report Form
CTCAE	Common Terminology Criteria for AEs
$CTDI_{vol}$	Volume CT Dose Index
DICOM	Digital Imaging and Communications in Medicine
DLP	Dose-Length Product
DRL	Diagnostic Reference Level
DSMB	Data and Safety Monitoring Board
GCP	Good Clinical Practice
ICH	International Conference on Harmonisation
IQR	Interquartile Range
IRB	Institutional review board
LDCT	Low-Dose CT
LOCAT	Low-dOse CT for Appendicitis Trial
NAR	Negative Appendectomy Rate
RERI	Relative Excess Risk due to Interaction

SAE	Serious Adverse Event

INTRODUCTION

Motivations of LOCAT

Acute appendicitis is one of the most common indications for emergency abdominal surgery (1). CT has assumed a paramount position in the disposition of adult patients with suspected appendicitis in the developed world, owing to its many advantages over other diagnostic tests, including ultrasonography (2, 3). Studies conducted in Korea (4) and the United States (5-7) have reported preoperative CT utilization rates ranging from 93% to 98% in patients undergoing appendectomy in 2007−2011. CT is highly accurate, readily available and rapid, easy to perform and to interpret, and rarely affected by the presence of bowel gas, severe abdominal pain, or extreme body habitus (8). Despite historical debate (9), several recent studies (6, 7, 10-13) have consistently shown that the increased use of CT coincides with a reduction in negative appendectomy rate (NAR) without an increase in appendiceal perforation rate (APR). NAR and APR are two important reciprocal measures of quality-of-care indicative of false-positive diagnosis and delayed diagnosis, respectively. The routine use of CT in patients suspected of having appendicitis has also been reported to be cost-effective through prevention of delayed or inaccurate diagnoses (14, 15).

There has been a surge in CT usage for diagnosing appendicitis during the last decade in the United States (6, 7, 9-13), indicating that the threshold for the decision to use CT may have declined. Over 300,000 appendectomies are performed in the United States each year (16), while approximately 90,000 were performed in Korea in 2017 (17). Most of these patients undergo CT examination preoperatively (6, 7, 10, 12, 13). Moreover, there are many additional patients with negative CT results,

who do not eventually undergo appendectomy. Factors contributing to these trends include improved CT technology, widespread availability, favorable reimbursement, and a general shift in the culture of medicine toward defensive medicine (18) and more dependence on imaging tests (19).

Many patients with suspected appendicitis are children or young adults (1), for whom CT radiation is of particular concern (20). Although debatable, there are increasing concerns that even a single typical abdomen CT examination may increase the risk of carcinogenesis (20-22). While such risk induced by an individual CT scan would be minute, multiplication by a large number of exposures may imply the real occurrence of cancer. With greater awareness of the carcinogenic risk (22, 23), it may no longer be certain if the benefits of CT in diagnosing appendicitis clearly outweigh the risk associated with the radiation doses traditionally used. It should be noted that the conventional radiation doses have historically been determined without robust scientific basis (24), with substantial variations in practice across hospitals (4, 25). Furthermore, while there is no rationale for using the same dose in young appendicitis patients and elderly patients with malignancies, attempts have rarely been made to properly differentiate the dose levels according to applications.

Results from several studies have suggested that reducing the radiation doses by 50-80% does not significantly impair the diagnosis of appendicitis (26-28), although the dose reduction causes a loss in the image quality. Recently, a single-center randomized controlled trial (29) demonstrated noninferiority of 2-mSv CT to 8-mSV CT with respect to NAR (3.5% vs. 3.2%; 95% confidence interval for the difference, -3.8 to 4.6 percentage points), an important measure of quality-of-care, in adolescents and young adults with suspected appendicitis. However, the study had

a potentially important limitation. While appendicitis is a very common disease encountered in emergency departments worldwide, it remains uncertain if the results of that particular study can be generalized to other institutions that are less experienced in using low-dose CT (LDCT). At the time of writing a doctoral dissertation, LDCT techniques have not been widely accepted as the standard-of-practice in many institutions. Our research group has therefore conducted a multi-center pragmatic trial, Low-dOse CT for Appendicitis Trial (LOCAT; clinicaltrials.gov number, NCT01925014; www.locat.org) (30), of similar study design to confirm the generalizability of the results of the previous single-center study. In addition to this primary research purpose, we aimed at the dissemination of the 2-mSv CT technique throughout the participating sites through the course of LOCAT.

Purposes of LOCAT

LOCAT (30) aimed for the effectiveness of 2-mSv CT as the first-line imaging test in regard to negative appendectomy rate in adolescents and young adults with suspected appendicitis were tested in hospitals with predominantly limited LDCT experience. The trial also compared the 2-mSv CT and conventional-dose CT (CDCT) groups in regard to appendiceal perforation rate, the proportion of the patients requiring additional imaging test(s) to diagnose or rule out appendicitis, delay in patient disposition, and diagnostic performance of the CT reports. LOCAT demonstrated that radiation dose of appendiceal CT for adolescents and young adults could be reduced to 2 mSv without impairing clinical and diagnostic outcomes in teaching hospitals.

Motivations of Dissertation Research

Evidence has been accumulated for the use of LDCT in diagnosing appendicitis. Lowering CT radiation dose is particularly important in adolescents and young adults who account for the majority of population suspected as having appendicitis. Prospective trials (29, 30) and retrospective studies (31-33) consistently reported that the radiation dose per CT examination could be reduced to 2 mSv without impairing diagnostic performance and clinical outcomes. However, the adoption of low-dose techniques in practice has been disappointingly slow, and the radiation dose of appendiceal CT often varies between hospitals by a factor of ten or greater (4, 34). A recent survey (34) attributed the underuse of LDCT to care providers' concern on potential misdiagnosis in some patient subgroups (e.g., extreme body sizes) despite the excellent overall diagnostic results in the previous studies.

If there are such doubts about the benefit of a new treatment in specific subgroups, subgroup analysis of previous study data is useful for better understanding the study data (35). Subgroup analysis is particularly valuable when the previous study was large enough to have sufficient data in each subgroup (36), and when the original study was in a noninferiority design. The absence of qualitative interaction (i.e., new treatment better than control treatment in one subgroup, but the reverse in the other subgroup) observed in subgroup analysis would mitigate the concern that such a qualitative interaction has led to the overall results that the two treatments have comparable effects (37). The data of LOCAT (30), owing to the large scale, provides an unprecedented opportunity for extensive subgroup analyses. LOCAT was a randomized controlled trial which proved the noninferiority of 2-mSv CT to institutional CDCT (typically 7 mSv) with respect to important clinical outcomes in patients aged 15–44 years. LOCAT involved 3,074 patients and 161

radiologists from 20 sites.

Large pragmatic trials such as LOCAT provide the most reliable data about treatment effects in the real world (35). However, deliberately broad and sometimes ill-defined eligibility criteria or study setting in pragmatic trials may lead to some heterogeneity of treatment effect across subgroups, and therefore, to difficulty in applying the overall study results to particular subgroups (38).

Prespecified subgroup analyses of the LOCAT data already showed consistent results across various subgroups regarding negative appendectomy and appendiceal perforation (30). The two co-primary endpoints were set (39) on the premise that they represent the clinical consequences of false-positive diagnosis and delayed (i.e., false-negative) diagnoses, respectively (40). However, researchers are now increasingly arguing that neither of the two outcome measures is an ideal indicator of the diagnostic quality, because of lack of consideration of spontaneously resolving appendicitis (41), loose coupling between appendiceal perforation and delayed diagnosis of appendicitis (42), and vague and inconsistent definition of appendiceal perforation (43).

Purposes of Dissertation Research

To this end, we performed *post hoc* subgroup analysis of the LOCAT data with respect to more direct measures of diagnostic performance (i.e., sensitivity and specificity as well as area under the receiver-operating-characteristic curve [AUC]) for appendicitis. The primary aim of our study was to explore any heterogeneity across patient or hospital characteristics in the diagnostic performance of 2-mSv CT relative to CDCT in adolescents and young adults with suspected appendicitis. The secondary aim was to identify specific subgroups of limited comparison because of

too small sample sizes. Our results would answer care providers in which patient subgroups 2-mSv CT can substitute CDCT.

BACKGROUND

Epidemiology of Appendicitis and CT utilization

Acute appendicitis is the most common cause of acute abdominal illness that requires surgical intervention. Across all age groups, the incidence rate is 8–19/10,000 persons-years in the United States (44), United Kingdom (45), and Korea (17). The lifetime cumulative incidence in California, the United States has been estimated as 9% (46). The highest incidence is observed in the age group of 10 to 19 years old with a slight male predominance (17, 44).

Imaging Utilization

Although there have been historical debates (9), ample evidence suggests that utilization of preoperative imaging, particularly CT, prevents negative appendectomy (i.e., unnecessary removal of uninflamed appendix) without increasing appendiceal perforation (8). The implication of negative appendectomy rate and appendiceal perforation rate as indices of diagnostic quality will be later discussed. Imaging tests, particularly CT, are also helpful in identifying other various abnormalities that can explain patients' abdominal symptoms (i.e., alternative diagnoses). The use of preoperative imaging tests in patients with suspected appendicitis is now accepted as the standard practice in many developed countries (15, 47). According to recent large multi-center studies, the utilization rate of preoperative imaging tests in patients undergoing appendectomy exceeds 90% in the United States (5) and Korea (4).

Interestingly, increased imaging utilization may have affected the pattern of the

disease presentation. According to epidemiologic studies, the incidence rate of nonperforated appendicitis, which had been decreasing until 2,000 (1, 42), showed reversely increasing tendency with the surge of preoperative imaging studies, particularly CT, in the United States while that of perforated appendicitis remained constant for the entire period (48). These trends provided circumstantial evidence supporting non-surgical treatment of uncomplicated appendicitis.

Popularity of CT

It is undebatable that ultrasonography is the most preferred imaging test for small children and pregnant women with suspected appendicitis (47). In non-pregnant late adolescents and young adults who account for the majority of the population requiring appendiceal imaging, geographic and institutional variations exist as to which, when, and how imaging tests should be used in adults with suspected appendicitis

In many countries, CT is the mainstay of imaging modality in adults with suspected appendicitis. Previous meta-analyses (2, 3, 49) drew a consistent conclusion that CT outperforms ultrasonography in the diagnosis of appendicitis. Due to its excellent diagnostic performance, CT is utilized 10–15 times more frequently than ultrasonography in the United States (12) and Korea (4, 50). In the United States, the use of preoperative CT is regarded to be cost-effective by preventing unnecessary appendectomies and hospital admissions (8, 14).

On the other hand, in some European regions (51-53), ultrasonography is widely used, and CT is reserved as the second-line diagnostic test used in case of inconclusive ultrasonography. Recently, magnetic resonance imaging has been introduced to the diagnosis of appendicitis, showing high reported diagnostic

sensitivities and specificities often exceeding 95% (54). However, it is yet to be determined if the promising results can be generalized to average hospitals and different healthcare systems.

CT Radiation

Although debatable, CT radiation has the risk of carcinogenesis in adolescents and young adults (20, 55-57). This raises a question as to whether the diagnostic benefits of CT genuinely outweigh the radiation-associated risk in patients with suspected appendicitis (29). The concern is based on the following epidemiologic knowledge. A vast number of patients undergo appendiceal CT throughout the world due to the high incidence of appendicitis and the popularity of CT. Importantly, among those who are exposed to the CT radiation, the number of patients who turn out to have normal appendix is greater than that of patients with appendicitis, particularly in the regions where appendiceal CT is popular. In a single-center observational study (58) in the United States that included 2,871 consecutive patients who underwent CT for suspected appendicitis, the prevalence of confirmed appendicitis was 23.5%. In a multi-center study from Korea (30) including 3,074 patients, the prevalence was 35.4%.

Even if the carcinogenic risk for an individual patient is assumed to be very small, the risk projected to such a large population may be significant for causing fatal cancers. Importantly, a substantial portion of the patients suspected as having appendicitis are adolescents and young adults who would otherwise have average life expectancies. These young patients are intrinsically more vulnerable to the carcinogenic risk compared to older patients.

Radiation Dose Level

In this article, “conventional dose” refers to a dose near the Diagnostic Reference Levels (DRLs) (the term DRL will be discussed below), while “low dose” refers to a dose considerably lower than the DRLs. Since the terms “low dose” and “conventional dose” are ambiguous (59), the radiation doses we refer will be specified whenever a relevant previous study can be cited. The unit of the radiation dose used in this dissertation paper is effective dose (in mSv). The effective dose is a general measure of detrimental effect from ionizing radiation, often used for comparing imaging studies or justifying the use of an imaging study (60). As our interesting range of effective dose lies in 1–10 mSv, minute difference of decimals will be regarded as unimportant, and decimals will thus be rounded to the nearest whole number when an mSv value is quoted. If utilized effective dose cannot be found or estimated from the cited article, we will instead quote the tube-current products (in mAs or effective mAs [defined as tube-current divided by a helical pitch]) and tube potentials.

Typical Radiation Dose for Multi-purpose Abdomen CT

CT radiation dose used for the evaluation of suspected appendicitis varies widely depending on region, hospitals, and CT machines. The term DRL has been used to refer to the representative dose of a given CT application in a population. DRL has been typically defined as the third quartile of the doses collected across CT machines and hospitals (61). We are not aware of any large-scale data on the DRL of CT examinations dedicated to the diagnosis of appendicitis. However, some data are available regarding DRLs used for multi-purpose abdomen CT examination in adults, which is presumably the same scanning protocol used for adults with

suspected appendicitis in many hospitals. The reported DRLs from various countries have ranged from 560 mGy·cm to 980 mGy·cm in the dose-length product (DLP), which corresponds to effective doses of 8–15 mSv with a conversion factor of 0.015 $mSv \cdot mGy^{-1} \cdot cm^{-1}$ (62). In general, CT radiation dose has been decreasing over the last decade due to advances in radiation-saving technology, greater awareness of radiation dose issues among radiologists, and auditing efforts (63, 64). Unfortunately, according to large-scale DRL data published in 2000–2018 (65-68), the dose-lowering trend below 10 mSv for abdomen CT has not been notable.

Even within a region or country, the radiation dose of abdomen CT varies widely across hospitals. In a study from the American College of Radiology (ACR) CT Accreditation Program (69), volume CT dose index for adult abdominal CT exceeded the ACR guideline (25 mGy, as of 2008) in 17.0% of the 600 CT machines investigated in 2002–2004 in the United States. Similar wide variations in dose exist in Europe (65) and Asia (70, 71).

Typical Radiation Dose for Appendiceal CT

There are only limited data available on the radiation dose range for appendiceal CT. In a survey involving 14 hospitals in 2004–2005 (25), a majority of the hospitals were using fixed tube-current time product ranging from 160 mAs to 380 mAs with peak tube potential of 120 kVp or 140 kVp, while fewer hospitals were using automatic exposure control techniques. A 2011 survey by Park and colleagues (4) involving 22 CT machines in 11 Korean hospitals found a surprisingly wide variation in the dose used across the hospitals, ranging from 2 mSv to over 20 mSv. The variation was partly attributable to the use of multiphase scanning in some hospitals.

Such a wide variation is associated with the fact that the utilized radiation doses

have not been determined on a robust scientific basis. With greater awareness of the potential carcinogenic risk associated with CT radiation, radiation doses should be adjusted according to its purpose and application. There is no rationale for using an identical dose for a young appendicitis patient and an elderly cancer patient.

Low Doses Explored in Research Settings

In 2004, Keyzer and colleagues (26) first reported the use of LDCT (around 2 mSv) for the diagnosis of appendicitis in adults. In subsequent studies comparing LDCT and CDCT (26-30, 72-74) (**Table 1**), the tested low dose ranged from 1 mSv to 4 mSv, while the tested conventional dose ranged from 5 mSv to 10 mSv, with 2–6 fold difference between the low and conventional doses within each of the studies. In recent comparative studies (29, 30, 74, 75), the tested low dose ranged 1–2 mSv, which was similar to the radiation level that Keyzer and colleagues (26) tested a decade ago. In non-comparative studies that explored the usefulness of LDCT in the diagnosis of appendicitis (76, 77), the tested low dose also ranged 1–2 mSv. Such dose, which is far below the DRLs for typical abdomen CT, is close to the worldwide average annual exposure to natural radiation sources (78) or the dose of three conventional abdominal radiographs (60). Park and colleagues (79) tested even lower dose (i.e., sub-mSv level) for appendiceal CT by using an iterative reconstruction technique.

Carcinogenic Risk Associated with CT Radiation

CT radiation is arguably carcinogenic.

Controversy

Extensive epidemiological studies (20, 55-57) have suggested that CT radiation is carcinogenic in children and adolescents. For example, a United Kingdom cohort study (20) reported that cumulative organ doses of 50–60 mGy triple the risk of leukemia and brain cancer. An Australian cohort study (55) showed that CT radiation exposure is associated with an increase of 20% or more in cancer risk. However, because these studies did not specify the indications for CT examinations, the reported carcinogenic risk may have been overestimated if the CT examinations were performed for suspected cancer (i.e., reverse causation) or other conditions related to higher cancer risk (i.e., confounding bias) (80). A French cohort study (81) claimed that further adjustment for confounders (i.e., cancer-predisposing factors) might reduce overestimation of carcinogenic risk associated with CT radiation. The upcoming European study EPI-CT (82) will offer an opportunity for better understanding of the potential risk of CT radiation in children and adolescents. In terms of adults, it is even more controversial whether CT radiation induces cancer (83).

A large epidemiologic study is underway on whether CT for the diagnosis of appendicitis increases cancer risk in Korean adolescents and young adults (unpublished data). It is difficult to conduct such a study since the carcinogenic risk, even if it truly exists, would be too small to be measured in a subpopulation of patients having a single disease (i.e., suspected appendicitis).

Alternatively, a risk projection model could be used for the estimation of the carcinogenic risk. In the risk projection, organ-specific radiation doses are calculated (84), and then sex- and organ-specific lifetime excess incidence of radiation-induced cancer is estimated using risk models such as the Biological Effects of Ionizing

Radiation VII model (85). Similar simulation method has been used in landmark studies that estimated the carcinogenic risk associated with CT radiation (22). However, it should be noted that the risk projection involves unverified assumptions and that the methods of estimating the carcinogenic risk are still evolving.

Using the risk projection model, Kim and colleagues (29) estimated that an exposure to 2-mSv appendiceal CT at the age of 30 years would result in a lifetime excess risk of 14 and 16 cancers per 100,000 male and female patients, respectively, while that to 8-mSv CT would result in 63 and 72 cancers, respectively. These estimates imply that using 2-mSv instead of 8-mSv in estimated 2,000 male or 1,800 female patients of 30 years of age would eventually prevent one case of cancer.

On the contrary, a decision analysis study (86) by Kiatpongsan and colleagues suggested that the choice of imaging modality between CT and others (combined ultrasonography and CT, or magnetic resonance imaging) for the diagnosis of appendicitis would affect life expectancy only minimally. The researchers estimated the life-expectancy loss attributable to surgical mortality, missed appendicitis, radiation-induced cancers, and competing for the age- and sex-based mortality risks. For example, a 20-year-old man would have a life-expectancy loss of 5.8, 6.8, and 8.2 days by choosing magnetic resonance imaging, combined ultrasonography and CT, and CT, respectively. The small differences in the life-expectancy loss among different imaging modalities would be attributable to the very low incidence of radiation-induced cancer. If the fatality is diluted in a large base of the population, the average individual loss in life expectancy would be small.

ALARA Principle

Setting aside the debate, there is no reasonable basis to insist on using radiation

dose of multi-purpose abdomen CT for the diagnosis of appendicitis, particularly in adolescents and young adults. Without definitive data to prove otherwise, it would be prudent to assume that the radiation-induced carcinogenic risk exists and to adhere to the "as low as reasonably achievable" (ALARA) principle. For appendiceal CT, as in all other CT applications, two essential components of the ALARA principles are: first, to ensure that the examination is performed only when clinically necessary; and second, to keep the radiation dose as low as possible in individual patients.

Efficacy and Effectiveness of LDCT Compared to CDCT

Herein we briefly review the results of the relevant original studies neutrally and critically. We limit our review to the studies that compared LDCT to CDCT in a head-to-head manner in adolescents and young adults (**Table 1**). Those results should be critically appraised with three important viewpoints that were not adequately addressed in the previous reviews (31-33).

First, most of the studies were prone to potential biases intrinsic to their retrospective nature. Therefore, the biases must have affected three meta-analyses regarding the diagnostic performance (31-33). Only two of the original studies were prospective randomized controlled trials (29, 30). The remaining studies mostly featured multi-reader multi-case design (i.e., a couple of participating radiologists retrospectively reviewed the images from all patients included in each study) for intra-patient pair-wise comparison of two serial CT scans (26, 28, 73, 74, 87, 88) or dose simulations (72). One study (27) with a before-and-after design retrospectively analyzed official CT reports.

Second, all but one study were single-center studies conducted by a small

number of expert radiologists motivated toward the use of LDCT, which raises concern for the generalizability of the study results. In a pragmatic clinical trial (30, 39), 20 hospitals with little prior experience in LDCT delivered 2-mSv CT practice successfully to over 1,500 adolescents and young adults. Even for the trial, the generalizability of the study results is still uncertain from a strict viewpoint: all the participating sites were teaching hospitals, only a third of the eligible patients were randomly assigned, and the catchment area was limited to Korea.

Third, clinical outcomes were assessed in only two randomized controlled trials (29, 30). Other studies were limited to an assessment of subjective image quality, diagnostic performance, or inter-observer agreement, which are all intermediate outcomes that are often decoupled from a more ultimate outcome (89). For instance, a study that compared filtered back-projection and iterative reconstruction for 2-mSv appendiceal CT (75) showed that the difference in subjective image quality is not linked to the difference in diagnostic performance. Undoubtedly, the measure of clinical outcomes would give a more definitive insight than the measure of subjective image quality or diagnostic performance; however, the former requires more research resources than the latter.

Table 1. Retrospective studies comparing LDCT and CDCT the first-line imaging test in adults with suspected appendicitis

Study	Year	Study design*	Comparison		Population/sample		Results			
			LDCT	CDCT	Age (years)	Sample size†	Outcome	LDCT	vs.	CDCT
Keyzer C *et al*. (26)	2004	Scan each patient twice MRMC (2 readers)‡	1–2 mSv	5–7 mSv	16–74	29/95§	AUC Sensitivity Specificity	0.92–0.93 97–100% 80–94%	vs. vs. vs.	0.91–0.93 97–100% 82–94%
Keyzer C *et al*. (72)	2009	Dose simulation MRMC (2 readers)‡	30 mAs^{eff} and 120 kVp (simulated‖)	100 mAs^{eff} and 120 kVp	18–87	33/131¶	With IV contrast Sensitivity Specificity Without IV contrast Sensitivity Specificity	 76%–88% 98%–99% 82%–91% 90%–95%	 vs. vs. vs. vs.	 91% 97%–99% 79%–82% 95%
Seo H *et al*. (28)	2009	Scan each patient twice MRMC (2 readers)	4 mSv	8 mSv	15–83	78/207§	AUC Sensitivity Specificity	0.98–0.99 98.7–100% 95.3–96.9%	vs. vs. vs.	0.97–0.98 100% 93–96.9%
Platon A *et al*. (73)	2009	Scan each patient twice MRMC (2 readers)‡	1–2 mSv	7–10 mSv	18–96	37/86§	Sensitivity Specificity	95% 96%	vs. vs.	100% 96%
Kim SY *et al*. (27)	2011	Prospective image interpretation Before-and-after design	2 mSv	8 mSv	15–40	95/257**	AUC Sensitivity Specificity	0.96 90% 92%	vs. vs. vs.	0.97 89% 94%

Table 1. Retrospective studies comparing LDCT and CDCT the first-line imaging test in adults with suspected appendicitis (continued)

Study	Year	Study design*	Comparison		Population/sample		Results			
			LDCT	CDCT	Age (years)	Sample size†	Outcome	LDCT	vs.	CDCT
Kim SH *et al*. (74)	2015	Scan each patient twice MRMC (2 readers)‡	2 mSv	4 mSv	15–82	58/102§	AUC	0.96–0.97	vs.	0.93–0.97
Yun SJ *et al*. (88)	2016	Scan each patient twice MRMC (2 readers)	2 mSv	7 mSv	≥ 18	141/270§	AUC	0.95	vs.	0.98–0.99
							Sensitivity	87.9–89.4%	vs.	97.9–99.3%
							Specificity	88.4–90.7%	vs.	95.4–98.6%
Chang CC *et al*. (87)	2016	Scan each patient twice MRMC (6 readers)	4 mSv	10 mSv	≥ 18	48/101§	AUC	0.82–0.97	vs.	0.88–0.97
							Sensitivity	56.3–93.8%	vs.	60.4–97.9%
							Specificity	81.1–96.2%	vs.	83.0–94.3%

Note.—AUC = area under the receiver-operating-characteristic curve, CDCT = conventional-dose CT, IV = intravenous, LDCT = low-dose CT, mAs^{eff} = effective mAs, MRMC = multireader multicase. *All studies were single-center studies. †Number of confirmed appendicitis/number of patients undergoing CT. ‡Patients inclusion was included prospectively, but images review was reviewed retrospectively. §Each patient underwent LDCT and then CDCT. ‖LDCT was simulated by adding noise to original images. ¶Patients were randomized into oral-contrast or no-oral-contrast group. Each patient underwent CDCT before and after IV contrast enhancement. **Each patient underwent either LDCT or CDCT.

Clinical Outcome

The primary clinical outcomes measured in the two randomized controlled trials (29, 30) were NAR (i.e., the percentage of uninflamed appendices out of all non-incidental appendectomies) and APR (i.e., the percentage of perforated appendicitis out of all cases of confirmed appendicitis). Negative appendectomy indicates the clinical consequence of false-positive diagnosis of appendicitis, whereas appendiceal perforation is associated with the delayed (or false-negative) diagnosis. The two reciprocal endpoints have been commonly used as the quality indices in the management of patients with suspected appendicitis (40), and the patient access to emergency medical care (90).

The two randomized controlled trials (29, 30) showed that 2-mSv CT is comparable to CDCT in the clinical outcomes. The single-center trial (29) compared 2-mSv CT and 8-mSv CT in 891 Korean adolescents and young adults, 358 of whom underwent appendectomy. The 2-mSv CT group was non-inferior to the 8-mSv group regarding negative appendectomy rate (3.5% vs. 3.2%; 95% confidence interval for the difference, -3.8 to 4.6 percentage points), which was the primary endpoint. The two groups did not differ significantly in the appendiceal perforation rate (26.5% vs. 23.3%) or in the proportion of patients who needed additional imaging tests (3.2% vs. 1.6%). To test the generalizability of the single-center trial results, the Korean researchers conducted another multi-center pragmatic trial (30) including 20 hospitals with little prior experience with LDCT. 3,074 adolescents and young adults were randomized to undergo 2-mSv CT or CDCT (≤ 8 mSv). Again, the two groups were comparable for negative appendectomy rate (3.9% vs. 2.7%; 95% confidence interval for the difference, -0.8 to 3.3 percentage points), appendiceal perforation rate (34.7% vs. 31.2%; -2.1 to 9.1 percentage points), the need of additional imaging

tests, or the delay in patient disposition.

However, researchers are now increasingly arguing that neither NAR nor APR is an ideal indicator of the diagnostic quality. First, NAR does not take into account the small fraction of patients with appendicitis that resolves without appendectomy (41). In the multi-center randomized trial (30), which is probably the largest prospective study on the diagnosis of appendicitis, more appendectomies tended to occur in the CDCT (≤ 8 mSv) group than in the 2-mSv CT group. The imbalance was attributed to a small number of cases of incipient appendicitis that were undetected with CT and then resolved without appendectomy. Such cases may have occurred more frequently with 2-mSv CT than with CDCT, because of the limited capability of 2-mSv CT in depicting subtle inflammation. Second, loose coupling exists between appendiceal perforation and delayed diagnosis of appendicitis, because perforated appendicitis and non-perforating appendicitis may be two discrete entities with differing pathophysiology (42). Third, the definition of appendiceal perforation has often been vague and inconsistent across studies. While some studies counted either of surgical (i.e., surgical records) or pathological (i.e., pathologic reports) documentation as appendiceal perforations, other studies regarded only the surgical documentation as clinically meaningful (43).

Diagnostic Performance

We were able to find ten previous studies that directly compared the diagnostic performance between LDCT and CDCT as the first line imaging test in adolescents and young adults with suspected appendicitis. The studies were conducted by six different researcher groups from Korea (27-30, 74, 88), Taiwan (87), and Europe (26, 72, 73). Importantly, the studies consistently reported that LDCT is comparable to

CDCT with respect to AUC, sensitivity, and specificity in the diagnosis of appendicitis.

The largest study was the Korean multi-center randomized controlled trial (30). Of the 161 radiologists (median seven from each of the 20 hospitals) involved in the trial, 71% had little prior experience with low-dose appendiceal CT. Residents, instead of attending radiologists, made the initial CT reports for 40% of the included patients. Even in this pragmatic setting, the area under the receiver-operating-characteristic curve, sensitivity, and specificity in the 2-mSv CT (n = 1,459) and CDCT (≤ 8 mSv) groups (n = 1,429) were 0.983 vs. 0.986, 97.1% vs. 98.0%, and 95.8% vs. 94.0%, respectively, showing very small differences. These results were consistent with those of the single-center trial that compared 2-mSv CT and 8-mSv CT (29). Smaller non-randomized studies that used different CT imaging protocols have also shown similar results (26-28, 72-74, 87, 88) (**Table 1**), as summarized in the meta-analyses (31-33).

However, the reported diagnostic performances may have been inflated due to verification biases, because in all the studies histopathologic confirmation of appendicitis was obtained selectively in patients with positive CT results (29, 30). Furthermore, the bias may have occurred differently in the LDCT and CDCT groups even in the randomized controlled trials, due to the imbalance mentioned above in the number of appendectomies (30).

Inter-observer Agreement

Retrospective studies (26, 28, 72, 73) of multi-reader multi-case designs reported excellent between-radiologist agreements regarding LDCT (1–4 mSv) interpretation for the diagnosis of appendicitis. The reported kappa values (0.93–

0.97) were comparable to those for CDCT (8–10 mSv) (28, 73).

Differentiation between Complicated vs. Uncomplicated Appendicitis

It is increasingly supported that uncomplicated appendicitis does not always progress to complicated (i.e., irreversible due to perforation or gangrene) appendicitis and that some of the uncomplicated appendicitis resolve without surgery (42, 48). Recent clinical trials (91) have shown the potential of non-surgical treatment (i.e., antibiotics or observation) in treating a considerable portion of patients with appendicitis. These ground-breaking attempts now require using CT not only to diagnose the presence of appendicitis but also to differentiate complicated from uncomplicated appendicitis (92). The motivation is to triage patients, thereby applying non-surgical treatment selectively to patients with uncomplicated appendicitis while performing immediate appendectomy in patients with complicated appendicitis. Published data are scarce regarding whether LDCT can be helpful in such patient triage, and further research is warranted in this area.

For successful non-surgical treatment, sensitivity is more important than specificity in diagnosing complicated appendicitis, because the false negative diagnosis may result in not only failure of non-surgical treatment but also serious complications such as abscess or peritonitis (93). A recent systematic review (92) summarized CT findings indicative of complicated appendicitis (e.g., appendiceal wall defect), and suggested that many of the helpful findings are highly specific but not sensitive. The investigators thus recommended averting non-surgical treatment if any of the specific CT findings are present. However, none of the 23 original studies included in the systematic review has addressed LDCT technique, and it was therefore uncertain whether the results of the systematic review can be applied to

LDCT images. A subsequent retrospective multi-reader study (94) showed that the sensitivity of the suggested diagnostic criteria (i.e., regarding the presence of any of specific CT findings as complicated appendicitis) was over 90%, which was higher than radiologists' gestalt assessment, at the cost of lower specificity.

No study has formally measured the diagnostic performance of LDCT in predicting complicated appendicitis. Even the most recent largest multi-center trial (30) has not addressed this question yet, although the prespecified secondary endpoints included the diagnostic performance for appendiceal perforation (not including appendiceal gangrene). The data analysis is now underway. A subgroup analysis in the multi-reader study (94) using the sensitive diagnostic criteria showed that 2-mSv CT is comparable to 8-mSv CT for the sensitivity and specificity in the diagnosis of complicated appendicitis. For predicting appendiceal perforation (not including appendiceal gangrene), three studies (27-29) from the same Korean researcher group reported the diagnostic performance of LDCT in comparison to CDCT. The single-center trial that compared 2-mSv CT and 8-mSv CT (29) showed the sensitivity of 36% vs. 55% and the specificity of 91% vs. 88%. The other two retrospective studies (27, 28) reported higher sensitivities for both LDCT (2–4 mSv) and CDCT (8 mSv). Although these studies found no significant difference between the LDCT and CDCT groups, none of these studies were designed to compare the radiation doses in the diagnosis complicated or perforated appendicitis. All of these studies had limited numbers of patients, particularly for the calculation of sensitivity (with denominators not greater than 44), which raises the need for a more extensive study.

Image Quality

The assessment of clinical outcomes, which are the most desirable endpoints for comparative trials, requires considerable resources. Image quality, on the other hand, is an alternative endpoint that can be assessed less expensively. In earlier studies that introduced new CT techniques for abdominal applications, researchers (95-97) measured objective image-quality indices such as contrast-to-noise ratio (98) or rated subjective image quality using predefined scoring criteria such as the European Guidelines for Quality Criteria for CT (99). However, the objective metrics are not the ideal index for medical image quality because they are too simplistic to convey the clinical context in the images. Subjective image quality rating is unavoidably subject to intra- and inter-observer variations, radiologists' adaptation to noisy images, and arbitrariness and ambiguity in the descriptors used for image quality.

These esthetic approaches have been rare in studies on the diagnosis of appendicitis. Since ample evidence regarding the clinical and diagnostic outcomes is already available, the lack of previous study regarding the image quality should not be a significant problem in establishing the clinical efficacy of LDCT in comparison to CDCT.

Visualization of the Appendix

Some investigators have used appendiceal visualization as an index of the image quality of LDCT. This analysis has been typically limited to patients confirmed as not having appendicitis since LDCT can visualize virtually all inflamed (and therefore enlarged) appendices. The reported rate of partial or complete visualization of normal (i.e., uninflamed) appendix with LDCT (1–4 mSv) has

ranged from 53% to 96% (27-29, 100, 101). The wide range is probably associated with subjectivity in rating the visualization, and inconsistent definition of the "visualization" across the studies. According to head-to-head comparison studies (27-29, 100), the visualization of normal appendix tended to be slightly compromised with LDCT (2–4 mSv) compared to that with CDCT (8–11 mSv) (53%–96% vs. 68%–98%). The difference was significant in a randomized controlled trial (78% vs. 87%) (29).

Alternative Diagnoses

Regarding the debate on the use of LDCT instead of CDCT in patients with suspected appendicitis, a common question raised by care providers is whether LDCT works well also for alternative diagnoses (i.e., conditions that clinically mimic appendicitis). Although there has been no study designed to answer this critical question directly, relevant data are available from studies that compared LDCT and CDCT. Unfortunately, all those data had inevitable limitations including a small number of patients in each disease category, the incompleteness of reference standards, and subjectivity in adjudicating final diagnoses. Specifically, final diagnosis often had to be adjudicated based on the CT results, as CT was practically the most accurate test for the diagnosis (e.g., ureteral stone). Due to the concern of these limitations, the two randomized controlled trials that compared LDCT and CDCT (29, 30) did not specify alternative diagnosis as an endpoint.

In the multi-center randomized controlled trial (30), final diagnoses other than appendicitis were adjudicated in 673 (43.8%) patients in the 2-mSv CT group and 687 (44.6%) patients in the CDCT (≤ 8 mSv) group. In both groups, the most common five alternative diagnoses were, in descending order, nonspecific

gastroenterocolitis, right colonic diverticulitis, pelvic inflammatory disease, ureter stone, and complicated adnexal cyst. The 2-mSv CT group tended to have more diagnoses of gastroenterocolitis but fewer diagnoses of gynecological or urinary diseases compared to the CDCT group, although the between-group differences were minute. Otherwise, the two groups had a similar distribution of alternative diagnoses, which was consistent with the single-center trial results (29). The disease distributions reported in these two Korean trials (29, 30) were also similar to that in a United States cohort (102), except for the fact that Korean studies showed a higher prevalence of right colonic diverticulitis. In the retrospective study of a before-and-after design (27), the diagnostic sensitivity for alternative diagnosis was similar between the 2-mSv CT and 8-mSv CT groups (80% vs. 81%). Other smaller retrospective studies that compared LDCT (1–4 mSv) and CDCT (7–10 mSv) showed similar results (28, 73, 74).

Importantly, the three largest (27, 29, 30) of the studies discussed above included only adolescents and young adults roughly ranging 15–44 years in age. In general, patients in this age group who present with suspected appendicitis rarely prove to have serious chronic or malignant disease (29, 30). Evidence has accumulated for LDCT as an adequate alternative to CDCT for diagnosing urinary stones (103), colonic diverticulitis (104), or Crohn's disease (105). Other common alternative diagnoses, such as gastroenterocolitis, pelvic inflammatory disease, complicated adnexal cyst, or urinary tract infection, should be made primarily based on clinical findings or diagnostic tests other than CT.

In summary, available data strongly suggest that LDCT is comparable to CDCT for alternative diagnoses in adolescents and young adults. However, this optimism may not apply to an older population that has a higher prevalence of serious chronic

or malignant diseases that can mimic appendicitis or cause secondary appendicitis (102).

Step-wise Multimodal Diagnostic Approach Incorporating LDCT

The promising study results on LDCT offer further opportunities for developing radiation-efficient algorithmic approaches in the diagnosis of appendicitis. In these approaches, clinical examination, laboratory tests, and imaging tests can be combined so that the use of radiation can be reserved for patient subgroups that can genuinely benefit from it, while still maintaining the excellent overall diagnostic performance. A variety of algorithms could be conceivable, depending on the regional or institutional practice pattern and available diagnostic resources. There could be largely two different approaches, which are not mutually exclusive, in incorporating LDCT into the step-wise diagnostic algorithm: selective utilization of LDCT following another first-line test(s), or first-line LDCT followed by an additional imaging test(s) in selected patients.

Patient Subgroups Less Benefited from LDCT

In developing a diagnostic algorithm, some knowledge is required as to which patient subgroups would benefit less from using LDCT instead of CDCT. Such knowledge can aid the selection of a patient subgroup for which CDCT should be used instead of LDCT. The radiation-saving advantage of using LDCT instead of CDCT is doubtful in patients with shorter life expectancies due to old age or serious comorbidity. CDCT may be more appropriate in such patients since they have a higher prevalence of critical intra-abdominal abnormalities as an alternative

diagnosis or incidental finding at CT (102). In our centers, we adhere to 2-mSv CT protocol in adolescents and young adults but are more flexible for older patients.

Owing to a large number of included patients (3,074 adolescents and young adults), the multi-center randomized controlled trial (30) could provide extensive prespecified subgroup analyses, which showed consistent results of important clinical outcomes across various subgroups. The subgroups were stratified by sex, body size, clinical risk scores for appendicitis (i.e., Alvarado score and Appendicitis Inflammatory Response score) (106, 107), time of CT examination (working hours vs. after hours), CT machine, radiologist's experience, site practice volume, and site experience with LDCT. These consistent results are probably due to small event rates of undesirable clinical outcomes.

We recognize the concern that the diagnostic accuracy of LDCT may be compromised in patients with obesity. However, the physical principle that image noise increases with increasing body size does not directly project to modern CT machines equipped with automatic exposure control that can keep consistent image quality across different body sizes. Patients with obesity tend to have more intra-abdominal fat, which in fact helps to visualize the appendix on CT images (108). We are unaware of any published data suggesting that larger body size limits the performance of LDCT in the diagnosis of appendicitis or alternative diagnoses. Unfortunately, none of the studies on LDCT included a sufficient number of obese patients to answer the question.

On the contrary, smaller body size can arguably limit the performance of LDCT. Two small studies (73, 76) have suggested that 1–2-mSv CT may have limited diagnostic sensitivity in patients with small body sizes or sparse pericecal fat. However, many other studies (26-30, 76) including the two randomized controlled

trials showed no notable effect of body mass index or pericecal fat (100) on the diagnosis of appendicitis at 1–4 mSv CT. Regardless of these data, ultrasonography instead of CT should be used primarily for such slender patients, as ultrasonography would be accurate and technically easy in these patients who generally have good sonic window.

Selective Utilization of LDCT

Many studies that did not specifically address the use of LDCT have instead proposed selective utilization of CT in patients with suspected appendicitis, thereby saving the total radiation dose used in the population. The motivation of the studies is to identify patient subsets who are very likely or very unlikely to have appendicitis, reserving CT to clinically or ultrasonographically equivocal cases.

In 2007, a single-center prospective randomized clinical trial (109) including 152 patients reported that NAR and APR were lower in the mandatory-CT group than in the selective-CT group. However, more recent non-randomized studies (110-112) have consistently reported promising results in identifying patient subgroups who would not require preoperative CT, by using a clinical scoring system such as Alvarado score (106) or Appendicitis Inflammatory Response score (107). Other non-randomized studies (51, 113, 114) have reported that step-wise diagnostic algorithm of using CT selectively in case of negative or inconclusive ultrasonography yielded high overall sensitivity (93%–100%) and specificity (86%–99%) for the diagnosis of appendicitis or other urgent abdominal conditions.

It is worthwhile to consider introducing these selective-CT approaches in regions where CT utilization is liberal (or excessive), such as the United States (5) and Korea (4), particularly by incorporating LDCT into the diagnostic algorithms.

In the multi-center randomized controlled trial (30), clinical outcomes were consistent across the subgroups stratified by the clinical risk scores for appendicitis, implying that 2-mSv CT is comparable to CDCT (≤ 8 mSv) in clinically equivocal cases as well. Poletti and colleagues (77) prospectively evaluated a step-wise algorithm using first-line ultrasonography, second-line 1-mSv CT, and then finally 7–10-mSv CT. Of the 183 patients who initially underwent ultrasonography, 99 required 1-mSv CT, and 18 eventually required 7–10-mSv CT. The step-wise algorithm showed excellent overall sensitivity (99%) and specificity (97%).

Additional Imaging Test(s) Following LDCT

Inconclusive CT results can occur slightly more often with LDCT than with CDCT (29, 30). In the multi-center randomized clinical trial (30) in which site radiologists interpreted CT images prospectively using a 5-grade Likert scale, grade 2–4 readings occurred in 20.7% and 16.7% of the 2-mSv CT and CDCT (≤ 8 mSv) group patients. In most such inconclusive cases, appropriate patient disposition could be made eventually through further in-hospital clinical observation or additional imaging tests (29, 30).

In the two randomized controlled trials (29, 30), the study protocols allowed the care providers to add an imaging test (i.e., ultrasonography or CDCT) if the diagnosis of appendicitis remained undetermined after the initial CT and clinical observation. Only a tiny fraction of patients in the 2-mSv CT groups required additional abdominal imaging test. In the multi-center trial (30), the 2-mSv CT and CDCT (≤ 8 mSv) groups were comparable for the proportion of patients who required an additional imaging test (2.5% vs. 2.7%). Since most of the required additional imaging tests were ultrasonography, few patients in the 2-mSv CT group underwent

additional CDCT. The single-center trial showed similar results (29).

Imaging Techniques for LDCT for Suspected Appendicitis

Review articles (115, 116) are available for CT radiation-dose reduction techniques and related physical principles. Here we limit our review to the technical aspects of low-dose appendiceal CT in adolescents and young adults. Combining two or more of the following dose-reducing techniques often show synergy. For example, a tube-current reduction can be most effective when used together with sliding-slab averaging technique (39, 117).

Intravenous Contrast Enhancement

From our experience, we believe that intravenous contrast enhancement is essential to compensate for the low image quality of LDCT. Although debatable, the ACR Appropriateness Criteria (47) now recommends using intravenous contrast enhancement. However, this guideline was based on earlier studies that used CDCT, and there have been little investigations on the need for intravenous contrast enhancement in LDCT. Several studies (26, 28, 72, 73) have reported that pre-contrast LDCT (1–4 mSv) was comparable to contrast-enhanced CDCT (5–10 mSv) in the diagnosis of appendicitis. However, these studies also found that unenhanced LDCT may be limited for alternative diagnoses (28, 72), diagnostic confidence for appendicitis (28), or visualization of the normal appendix (28).

Contrast-enhancement Phase

We recommend obtaining portal venous phase images only. We are not aware of any published evidence suggesting that any additional pre-contrast, arterial phase,

or delayed phase images are helpful for patients with suspected appendicitis. Guidelines including the ACR Appropriateness Criteria (47) have not specifically addressed this issue. The 2011 survey by Park and colleagues (4) surprisingly showed that 10 of the 11 sites were routinely acquiring either of pre-contrast or arterial phase images, or both in addition to portal venous phase images in appendiceal CT.

Radiologists and referring physicians or surgeons are often reluctant to abandon pre-contrast CT or multiphase scanning. The reluctance is often due to the concern of missed diagnosis of urinary stone, which is an important alternative diagnosis (102). A recent study (118) reported that three radiologists' retrospective reading of portal-venous-phase images showed sensitivities ranging 92%–96% in detecting urinary stones larger than 3 mm. Smaller stones were more prone to be missed, although they generally do not require any invasive treatment procedures (119).

Enteric Contrast

We recommend not using enteric contrast in LDCT. As the ACR Appropriateness Criteria (47) stated, the evidence is trending against the use of enteric contrast for intravenous contrast-enhanced CT. A large observational study (120) from the United States showed that the use of enteric contrast does not improve the diagnosis of appendicitis. These guideline and study results were based on data obtained using CDCT, and there have been little investigations on the need for enteric contrast in LDCT.

Anatomical Coverage

Several researchers (121-123) have proposed limiting scan coverage to the

pelvis in appendiceal CT. However, this "focused" CT was criticized by other researchers (124, 125) who showed that it could lead to some missed diagnosis of appendicitis or other critical abnormalities outside the pelvis. The ACR Appropriateness Criteria (47) recommends scanning the upper abdomen as well as the pelvis in patients with suspected appendicitis.

In practice, scanning beyond intended range (i.e., "image creep") often occurs in abdomen CT (126), which leads to higher doses. In most CT machines, the scan range cannot be set as a part of an automated scan program in CT machines but needs to be adjusted manually by technologists. Therefore, we recommend specifying anatomical landmarks for determining the scan coverage from scout images and using those landmarks consistently. For example, we set the scan range from 4 cm above the liver dome to 1 cm below the ischial tuberosity.

Tube Current

Reducing tube current has been the mainstay of dose-reducing techniques in previous LDCT studies. As stated above, those studies have proved excellent diagnostic and clinical outcomes with LDCT (1–4 mSv) using reduced tube currents. In our centers, we set the reference value for effective tube-current–time product as 45–110 mAs, aiming at effective radiation doses of 2 mSv. This wide range is primarily due to variation in the tube potentials used in individual CT machines and hospitals. We recommend activating all available automatic exposure control techniques in the automated scan program that is saved in each CT machine. In spite of proven advantages of automatic exposure control techniques (127), the 2011 survey by Park and colleagues (4) showed that one of the 11 participating hospitals inadvertently failed to use the automatic exposure control technique in some patients.

Tube Potential

The tube potential for standard abdomen CT for adults has been typically 120 kVp or 140 kVp. It is now widely accepted that lower tube potentials (80–100 kVp) can be used for smaller adults (128). While it is desirable to individualize tube potential by patient size automatically (95) or manually, the 2011 survey by Park and colleagues (4) has shown that all 11 participating sites were using fixed tube potentials: 120 kVp at ten sites and 100 kVp at one site.

Iterative Reconstruction

Many studies advocated that using an iterative reconstruction instead of a filtered back-projection can allow considerable dose reduction without significant sacrifice in image quality (129). However, few studies have investigated whether an iterative reconstruction is truly helpful in low-dose appendiceal CT. Park and colleagues (75) retrospectively compared a filtered back-projection and an iterative reconstruction in 107 patients who underwent 2-mSv CT for suspected appendicitis. Interestingly, the researchers did not find any notable advantage of the iterative reconstruction over the filtered back-projection in the diagnostic performance or diagnostic confidence, although radiologists assigned higher subjective image-quality scores for the iterative reconstruction than for the filtered back-projection. In a more recent prospective study (79), the same researcher group showed that the radiation dose of appendiceal CT could be potentially lowered to 0.5 mSv by using a new-generation iterative reconstruction technique.

Image Reconstruction Thickness

As the appendix is a small structure, it has been believed that thinner sections are advantageous in depicting the normal or inflamed appendix. A decade ago, Johnson and colleagues (130) reported that appendiceal visualization improved with decreased section thickness from 5 mm to 2 mm at CDCT (using effective tube-current–time product of 200 mAs and tube potential of 120 kVp). With LDCT using dose around 2 mSv, however, the conventional wisdom that thinner sections are advantageous may not be valid. As image noise is inversely correlated with the number of X-ray photons that contribute to the formation of that image (131), decreasing section thickness increases image noise further. Considering the trade-off between z-axis spatial resolution and image noise, we recommend 3–5 mm as the viewing thickness (i.e., section thickness) for LDCT with a dose around 2 mSv, based on our experience from the two large randomized controlled trials (29, 30). In our centers, we reconstruct two transverse image datasets from each helical scan: 4-mm thickness with 3-mm interval and 2-mm thickness with 1-mm interval. We primarily review the 4-mm-thick images and occasionally use the 2-mm-thick images for multiplanar sliding-slab averaging review that we will discuss later.

Coronal Reformation

In a retrospective study using CDCT (tube-current–time product of 350 mA and tube potential of 140 kVp), Paulson and colleagues (132) reported that coronal reformations used in addition to transverse images enhanced radiologists' diagnostic confidence but did not improve diagnostic performance for appendicitis significantly. We are not aware of any study that formally measured the advantage of additional coronal reformations in LDCT. The advantages of additional coronal reformations

may be theoretically more pronounced for LDCT, given that better appendiceal visualization by additional coronal reformations may compensate for the low image quality of LDCT.

Sliding-Slab Averaging Technique

The sliding-slab averaging technique is a real-time image rendering technique that is useful for rapidly reviewing large thin-section datasets. While the viewing slab slides through the volume along a viewing direction in a small increment, the overlapping slabs create an illusion of image-to-image continuity, thereby preserving high through-plane spatial resolution that is inherent to a thin-section dataset. With the flexibility that allows a reviewer to arbitrarily choose the slab thickness and viewing direction, the dynamic navigation technique is theoretically more advantageous than adding simple coronal reformations, particularly in tracing small tortuous tubular structures such as the appendix. Lee and colleagues (133) have introduced the sliding-slab averaging technique in appendiceal CT. In their retrospective study using CDCT (unspecified tube current and tube potential of 120 kVp), sliding-slab averaging review of 2-mm-thick sections outperformed regular stack review of 5-mm-thick transverse sections in radiologists' diagnostic confidence, although the difference in diagnostic performance did not reach a statistical significance.

Similar results were found with LDCT in a subsequent retrospective study (117) by the same researcher group. In theory, the sliding-slab averaging technique may be particularly helpful for LDCT because averaging voxels within the slab improves the quality of the rendered images by canceling out the image noise out of the thin-section source images. As mentioned above, we recommend keeping slab thickness

as 3–5 mm in reviewing LDCT, since a very thin slab would have too much image noise.

Image Interpretation and Reporting for LDCT

There have been few studies that revealed any difference in image interpretation or reporting for LDCT and CDCT.

Diagnostic Criteria for Appendicitis

The diagnostic criteria for appendicitis on CDCT have been mostly consistent across previous studies, showing only minor variations. The criteria used in our centers are as follows. The primary diagnostic criterion is appendiceal enlargement with mural thickening and periappendiceal fat stranding. An appendix larger than 6 mm in diameter is considered potentially abnormal. Secondary diagnostic criteria include abnormal mural enhancement, appendicolith, phlegmon, and abscess. While we adopted these criteria from review articles (134) that were published before LDCT era, the criteria were found to show excellent diagnostic performance also with LDCT in several studies including the two randomized controlled trials (29, 30) as we stated earlier. Minor revision of the criteria may be needed, although previous small studies (26) have not found a notable difference between LDCT and CDCT in the CT findings of appendicitis.

On CDCT, appendiceal non-visualization can be construed as a sign to rule out appendicitis when an experienced radiologist interprets CT images. In previous studies (135, 136) using CDCT (the used radiation doses were not clarified), a tiny fraction (around 2%) of patients with appendiceal non-visualization proved to have appendicitis. However, on LDCT, radiologists may be uncertain if appendiceal non-

visualization is attributable to small size of uninflamed appendix or to compromised image quality. As we stated earlier, the visualization of the normal appendix tended to be slightly compromised with LDCT compared with CDCT, showing a significant difference in a randomized controlled trial (29). Nevertheless, our data (unpublished data) show that, LDCT has a reasonably high diagnostic performance for appendicitis even in patients with limited (i.e., not identified, or unclearly or partially visualized) appendiceal visualization, and that those patients rarely prove to have appendicitis.

Structured Reporting

Referring physicians or surgeons are often highly dependent on CT results in determining the disposition of patients with suspected appendicitis. Accurate and efficient communication of CT results is essential, particularly when CT results are not conclusive, because miscommunication of the diagnostic certainty (or uncertainty) may lead to an inappropriate or delayed patient disposition. The diagnostic certainty needs to be explicitly delivered, ideally in a highly-standardized manner, particularly for LDCT for which inconclusive reports occur more often than with CDCT (29, 30). A structured reporting form based on a Likert scale would be an effective means of successful standardization (137, 138). Appendiceal CT in adolescents and young adults has unique features suitable for structured reporting. First, the diagnostic task is simple (i.e., the likelihood of appendicitis) and report conclusion can be efficiently summarized using a Likert scale. Second, unexpected complex cases are rare, since most alternative diagnoses are limited to a small number of non-serious diseases in the target patient population as we discussed previously. Third, standardized communication is essential because interdisciplinary

collaborative decisions need to be made urgently, often involving less-experienced care providers such as rotating trainee doctors who would use various descriptors for diagnostic certainty.

We have developed (101) such a structured reporting form for the likelihood of appendicitis using a 5-point Likert scale and deployed it across the 20 hospitals through the course of the randomized controlled trial (30). A survey (138) that was conducted at the final phase of that trial, including 594 care providers from the 20 hospitals, showed that the care providers preferred structured reporting to free-text reporting. However, we were unable to prove whether standardized communication using a structured reporting form can lead to better diagnostic and clinical outcomes. Such an investigation regarding ultimate outcomes would require an unrealistically large study sample.

Other Practical Issues in Implementing LDCT

As we mentioned earlier, a vast disparity exists between science and practice for CT radiation dose (103). Acknowledging this challenge, we designed the multi-center randomized clinical trial (30) as a pragmatic trial, with the intention that the participating sites would eventually embed 2-mSv CT into their usual care by implementing the trial protocol (39). First, the eligibility criteria (i.e., patients aged 15–44 years undergoing CT due to suspected appendicitis) were broad and largely depended on the judgment of individual care providers. Second, we minimized the requirements for the CT imaging and interpretation protocol. Third, all co-interventions (i.e., diagnostic and therapeutic procedures other than the initial appendiceal CT) followed the standard practice of each site without using extra resources.

Despite extensive efforts over the years of the trial design and conduct, follow-up results regarding LDCT adoption in the trial sites were not very satisfactory according to a survey (34) conducted during the final phase of the trial. The survey of 579 care providers from the 20 trial sites showed that 7.9% of the care providers were still unwilling to use 2-mSv CT, while the remaining care providers supported consistent (27.3%) or selective (e.g., during working hours) (64.8%) use of 2-mSv CT. The survey showed that many of the care providers were still concerned that the low image quality of LDCT might lead to incorrect diagnoses. It is disappointing that those care providers were still unaware of or disregarded previous study results showing that LDCT is comparable to CDCT regarding diagnostic performance (31-33) and clinical outcomes (29). A follow-up survey (34) conducted six months after the trial completion showed that six of the 20 participating sites were using the standard-of-care radiation doses of 4 mSv or higher, while the remaining 14 hospitals lowered the dose to 2 mSv. These survey results are partly disappointing given that all the 20 sites were highly-resourced teaching hospitals that voluntarily participated in the trial (29). Our experience shows difficulties in implementing LDCT practice in the reality. In addition to an understanding of the theories and imaging techniques of LDCT, real challenges lie in the practical issues that we discuss below. For successful implementation of LDCT practice, it is helpful to organize a team of a radiologist, referring physician or surgeons, and CT technologists, each of whom can champion the change toward LDCT and educate the colleagues in their field (116).

Dedicated Protocol for Appendiceal CT

For successful implementation of LDCT practice, we strongly recommend first setting up a dedicated appendiceal CT protocol in the hospital information system

and the corresponding automated scan program in each CT machine. This automation and standardization are particularly crucial in hospitals wherein the workflow does not allow radiologists to determine scanning protocol for each of individual patients, or in large hospitals, wherein not all care providers are enthusiastic about dose reduction. Setting up the dedicated CT protocol can be a starting point to identify which components should be reinforced or revised in the CT examination cycle, spanning from the order entry to the report of the results. For example, a simple query to hospital information system can identify care providers who are reluctant to the shift from the general-purpose CDCT to the dedicated appendiceal LDCT. Those reluctant care providers could be the primary target of further education and encouragement, as we discuss later.

The 2017 survey (34) conducted six months after the completion of the multi-center randomized controlled trial (30) showed that only four of the 20 trial sites were consistently using the dedicated appendiceal CT protocol for adolescents and young adults with suspected appendicitis. Six sites were selectively using the dedicated protocol, and ten abandoned the dedicated protocol from their usual practice. Although partly disappointing, these results were still a remarkable progression from the 2011 survey by Park and colleagues (4), which showed that only one of the 11 participating hospitals had dedicated appendiceal CT protocol.

Education for Referring Physicians and Surgeons

It is understandable that some referring physicians, surgeons, and even radiologists are not enthusiastic or even reluctant towards dose reduction. Care providers' actions are often unfortunately influenced by the concern of malpractice litigation. In the United States, appendicitis is one of the most common medical

conditions associated with litigation against emergency department physicians, with up to one-third of cases ending up with claims paid to patients (139, 140). The risk of an inaccurate diagnosis of appendicitis due to degraded image quality by using inadequately low radiation dose can immediately affect care providers as well as patients.

On the contrary, the potential risk of carcinogenesis due to excessive radiation is so small and unlikely to be immediate that the risk may rarely affect the care providers' choice of the CT examination. Therefore, it is essential to create higher-level evidence supporting the dose reduction and to educate colleague physicians, surgeons, and radiologists on such evidence. The education can occur through lectures, printed material, institutional and societal websites, individual consultations by radiologists or physicists to referring clinicians, and use of decision support in order entry (116).

Kim and colleagues recently conducted a survey (34) of 579 care providers from the 20 hospitals that participated in the multi-center randomized controlled trial (30) for their willingness to use 2-mSv CT. The results showed that the willingness was significantly associated with a care provider's belief that there is compelling evidence on the carcinogenic risk of CDCT radiation. As we discussed earlier, extensive epidemiological studies (20, 55-57) have suggested that such risk exists in children and adolescents, while the risk is more controversial for adults.

Education for Radiologists

With a higher image-noise level, LDCT images are often less straightforward to interpret than CDCT, especially for inexperienced radiologists. Two studies (27, 29) reported that radiologists' diagnostic confidence tended to be lower with LDCT

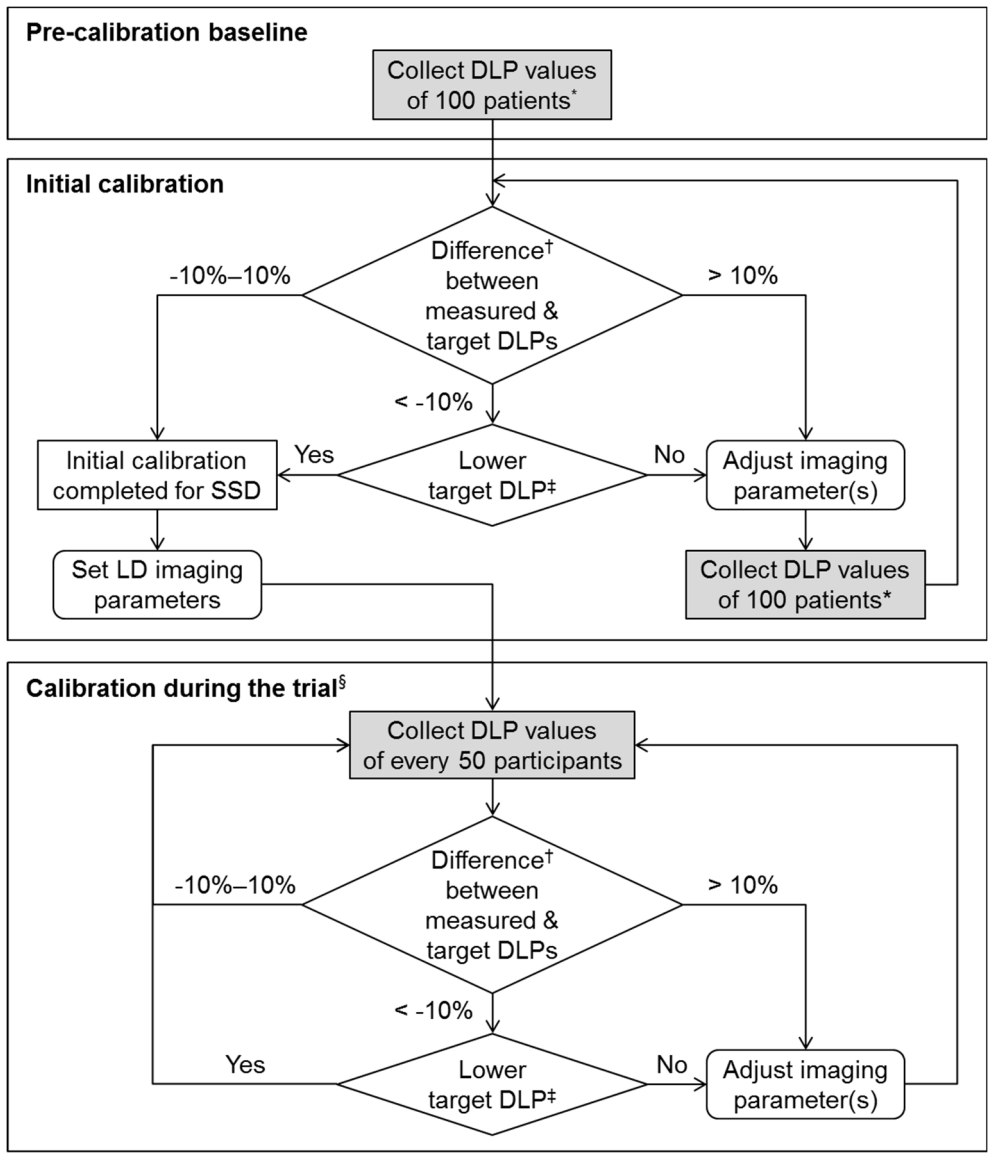
than with CDCT, although the observed difference did not reach statistical significance. In a prospective study by Yang and colleagues (141), 63 attending radiologists and 166 radiologist residents with little prior experience with LDCT from 22 hospitals completed an online training course of 30 case challenges of 2-mSv CT with direct feedback. Interestingly, these data did not show notable intra-reader learning curves over the 30 cases. Instead, the diagnostic performance was affected rather by readers' years of overall clinical experience and prior experience with appendiceal CT regardless of radiation dose. As the diagnostic performance for the 30 cases was reasonably high for the attending radiologists and senior residents (with pooled AUC of 0.92–0.94), the investigators suggested that the clinical implementation of 2-mSv CT would be feasible in many hospitals without further education, assuming that qualified site radiologists carefully supervise the practice.

Dose Calibration and Monitoring

While there has been some literature on the principle of low-dose scan techniques, they have rarely addressed the specific step-by-step procedures on how to adjust scanning parameters to reach and maintain the desired dose. Because different CT machines use different mechanisms of dose adjustment and automatic exposure control, there cannot be a single correct guideline. Here we introduce the dose calibration and monitoring procedures that we originally developed as a part of the protocol (39) of the multi-center randomized clinical trial (30) (**Fig. 1**).

Figure 1. Radiation dose calibration procedures for each CT machine.

DLP = dose-length product, LD = low dose, SSD = site standard dose. *In regular abdomen CT examinations for various purposes in patients not enrolled in the trial. †Difference (%) = (measured median DLP – target DLP) / target DLP × 100. ‡At the discretion of the lead radiologist. §For each of 2-mSv CT and CDCT groups.



Since we use automatic exposure control techniques, the actual radiation dose varies substantially with individual patient size. For each patient, the modulated radiation dose in terms of volume CT dose index ($CTDI_{vol}$, based on the use of 32-cm diameter reference phantom) and DLP is recorded as a text table in a Digital Imaging and Communications in Medicine image. If an additional scan is performed for any reason (e.g., rescan for a non-diagnostic initial scan, machine failure, or extravasation), then the DLP for each helical scan is recorded. For an average-size patient, we initially set the target DLP as 130 mGy·cm for each scan, which corresponds to an effective dose of 2 mSv with a conversion factor of 0.015 $mSv \cdot mGy^{-1} \cdot cm^{-1}$ (62). As we discussed earlier, we chose this initial dose level based on the previous studies that directly compared LDCT and CDCT. In each CT scanner, scan parameters such as reference tube-current–time products (or noise level) and tube potential are adjusted aiming at the target DLP value (142), and the parameter set is saved as an automated scan program. In general, DLP value is roughly proportional to reference tube-current–time products but is nonlinear to the change of tube potential.

The target dose level can gradually decrease to some extent over time with advances in CT technology and radiologists' adaptation to noisy images. Therefore, we have a unidirectional standpoint in resetting the target radiation dose: being flexible toward a lower dose while being strict against a higher dose (**Fig. 1**). For each CT machine, we draw a box-and-whisker plot of the DLP distribution in a sizable group (e.g., 50) of consecutive patients to ensure appropriate calibration. We calculate the median DLP while excluding outliers caused by inappropriate scan techniques or technical failures. If the median DLP value is less than 90% or greater than 110% of the predefined target DLP (i.e., out of the error range of ±10% from

the target DLP), we readjust the scan parameters (e.g., reference tube-current–time product or noise level) as appropriate. The calibration and monitoring processes are then iterated for every 50 patients for each CT machine.

MATERIALS AND METHODS

Study Overview

The institutional review boards of the 20 sites approved this study and waived patient informed consent. All data collection was planned (39) before any patient recruitment. All subgroup analyses of the present study were planned after the data collection. Group allocation for the trial was concealed from the patients and outcome assessors. Care providers could not be masked because of the obvious between-group difference in the CT image texture. No masking procedure was implemented for site pathologists and data collectors. It was expected that the 2-mSv CT and CDCT groups would receive virtually the same management by the same care providers including radiologists, except for the main intervention (i.e., radiation dose) which was also the index test.

We have detailed elsewhere (30, 39) the trial procedures regarding site recruitment, patient eligibility criteria, patient identification, randomization, index test (e.g., radiation dose calibration), co-interventions, reference standards, sample size estimation, and data monitoring. LOCAT was designed as a pragmatic trial (143) recruiting sites inexperienced in LDCT to overcome the limited generalizability of previous single-center studies (29, 31-33) conducted in hospitals with expertise in LDCT. Therefore, the trial procedures were standardized for the minimum protocol requirements regarding history taking, CT technique, CT report form, pathological criteria for appendicitis, and telephone follow-up. Otherwise, all study procedures, including the allocation of care providers to patients, followed the individual site's usual practice (4, 39). The practice pattern could vary across the sites or care providers. We wrote this report in line with reporting guidelines (35, 144-146).

Practice Setting

The principal investigator of the trial (Dr. Kyoung Ho Lee, a radiologist with 19 years of research experience) invited hospitals through the network of the Korean Society of Abdominal Radiology. All the 20 sites were teaching hospitals and used CT as the primary imaging modality for adults with suspected appendicitis (4). Only the lead site (Bundang Seoul National University Hospital), where the previous single-center trial (29) was conducted, had previous experience of the systematic implementation of 2-mSv CT for appendicitis.

Pre-registration Procedures

It was mandatory to develop and implement Patient Screening and Enrollment Standard Operating Procedure to standardize the screening procedures at each site. Site research coordinator or emergency department physicians on service, including attending physicians, fellows, and residents, were to identify eligible patients and asked the patients or legally acceptable representatives to participate in LOCAT. It was recommended, although not mandatory, that all eligible patients were consecutively screened 24 hours per day, 7 days per week so that the enrolled participants could be representative of the patient population undergoing CT because of suspected appendicitis at each site.

In addition to the standard procedures in patient history taking at each site, it was strongly recommended, although not mandatory, to ask each patient to complete Patient Questionnaire soliciting chief complaint and past medical history at the time of entrance to the emergency department. The use of the Patient Questionnaire standardized the history-taking procedures whether the patient was eventually

enrolled in LOCAT or not. The Patient Questionnaire served as a source document when patients successfully entered LOCAT.

Study Organization and Site Recruitment

The organization of LOCAT, including the office, committees, data center, participating sites and investigators, DSMB, and funding sources, is described in a separate document, LOCAT Organization.

Site recruitment procedures were as follows. The LOCAT office regularly sent out an invitation letter to all members of the Korean Society of Abdominal Radiology (a nationwide society of abdominal radiologists), which had 303 members from 119 hospitals at the time of writing the initial version of the LOCAT protocol. The initial requisites for study participation stated in the invitation are: (a) at least 50 nonincidental appendectomies were performed annually, and (b) multidisciplinary services were provided for patients with suspected appendicitis 24 hours per day, seven days per week. If any hospital expressed interest in the study participation, the LOCAT office regularly contacted the potential site lead investigator to inform the site lead investigator of additional requisites for study participation and to solicit information regarding site practice patterns. The LOCAT office and site lead investigator worked together to prepare study procedures in the site through gradually completing the appendix documents.

It should be noted that the site recruitment procedure primarily based on the voluntary participation of subspecialist gastrointestinal radiologists could limit study generalizability. The participating sites might tend to be teaching hospitals and to have higher study motivation and greater hospital resources than hospitals not participating.

Site Activation

To ensure effective participant recruitment and complete data collection, all site lead investigators, site lead radiologists, site research coordinators, site lead CT technologists, and data center research associates were required to review the LOCAT protocol and appendix documents thoroughly. They were required to complete a review course which consists of several hours of didactic lectures and group discussions. Before registering the first participant, each site lead investigator submitted Site Lead Investigator's Agreement and accompanying documents including Site Preparation Checklist to the LOCAT office.

Before the registration of the first participant, the site lead investigator completed a rehearsal of the study procedures in at least 20 patients, who were not included in the sample size or final analyses. The rehearsal patients underwent CDCT only regardless of the results of random assignment (sham-randomization). Otherwise, the rehearsal included all the study procedure in the LOCAT protocol, including submission of the study data to the LOCAT office. Through the rehearsal, site investigators could prove that they had been able to identify potential participants in sufficient numbers and focused their attention on how this could be achieved. The LOCAT office restricted participation to those sites achieving reasonably good performance in the rehearsal.

LOCAT procedures were piloted thoroughly at three sites (Seoul National University Bundang Hospital, Daejin Medical Center Bundang Jesaeng General Hospital, Soonchunhyang University Bucheon Hospital) for months before participant enrollment was rolled-out to other centers. Problems in the pilot sites were identified and solved so that other sites later joining could benefit from

procedures that had already been proven to work in practice. The rollout was implemented in a step-wise fashion, rather than starting all other sites simultaneously. This allowed LOCAT office to concentrate on one or two new sites at a time. LOCAT office hosted regular meetings for research staff members to share problems and solutions.

Patients

The eligibility criteria of the trial were patients aged 15–44 years who were referred for CT examination due to suspected appendicitis. Of 8,593 patients considered eligible, 3,074 were randomized to undergo either the 2-mSv CT or CDCT group from December 2013 through August 2016. The data for 28 participants were discarded or not collected due to inappropriate enrollment or withdrawal from the trial. We excluded 126 patients who did not adhere to the trial protocol (39) regarding eligibility or radiation dose (**Table 2**). In 147 participants, reference standards for the presence of appendicitis were incomplete, as defined in the trial protocol (39). Therefore, our analyses finally included the remaining 2,773 patients (median age [interquartile range], 28 [21–35] years) including 1,392 in the 2-mSv CT group and 1,381 in the CDCT group (**Fig. 2**). They were 1,516 females (27 [21–35] years) and 1,257 males (29 [22–36] years).

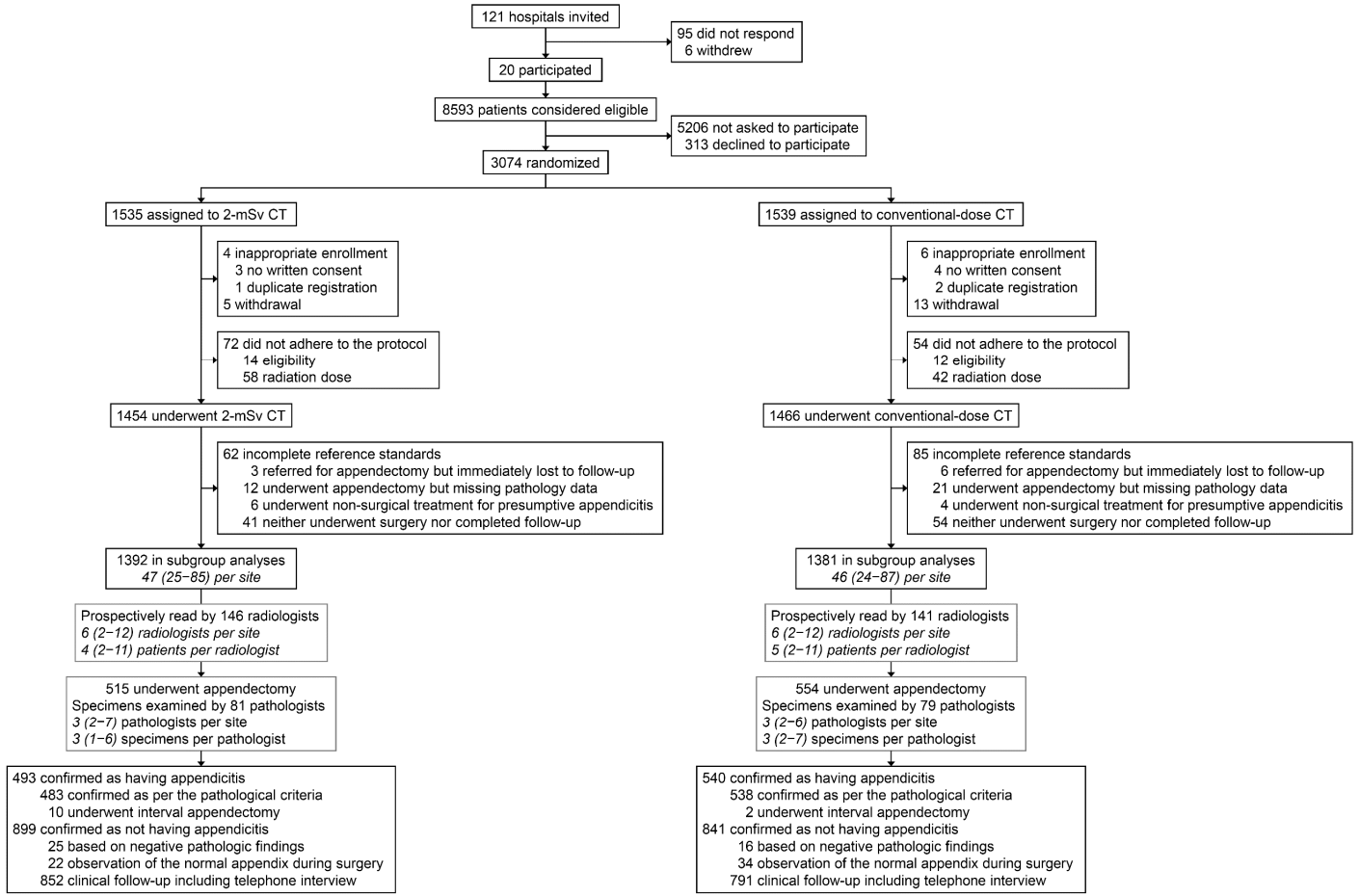
Table 2. Protocol non-adherences in LOCAT

Protocol non-adherences	Total (n = 163)	2-mSv CT group (n = 109)	CDCT group (n = 54)
Eligibility criteria	26	14	12
Age was < 15 years or > 44 years	22	11	11
Prior history of appendectomy	2	2	0
Prior CT to evaluate the presenting symptoms	2	1	1
Radiation dose	137	95*	42
Additional arterial phase scan	75	39*	36
Assigned to 2-mSv CT group but mistakenly underwent CDCT	48	48*	0
Predefined automated scanning program was not used	8	4	4
Repeat scan due to severe motion artifacts	6	4	2

Note.—Data are numbers of instances of non-adherence. CDCT = conventional-dose CT.

*Thirty-seven patients had both events due to wrong parameter setting in a CT machine.

Figure 2. Patient flow diagram.



Eligibility Criteria

Site research coordinators or emergency physicians identified eligible patients aged 15–44 years who were referred from the emergency departments for intravenous contrast-enhanced CT due to suspected appendicitis. The clinical suspicion of appendicitis and the need for CT examination were determined by individual care providers. The details criteria for eligible patients were as following:

- 15-44 years of age.
- Emergency department visit with suspected symptoms and signs of acute appendicitis.
- Intravenous contrast-enhanced computed tomography examination[*†] requested due to suspicion of appendicitis.
- Willing to provide telephone or cell phone numbers for follow-up.
- Signed informed consent provided before study entry.

[*]Patients were generally recommended to undergo ultrasonography instead of CT if they had slender body shape (body mass index less than 18.5 kg/m^2) (147), prior history of allergy to iodinated intravenous contrast materials, or prior history of renal insufficiency, although none of these was an absolute exclusion criterion.

[†]Negative pregnancy status was confirmed before CT examination in all female patients of child-bearing potential as required by the standard-of-care at each site. Patients deemed pregnant underwent ultrasonography and/or magnetic resonance imaging instead of CT.

Patients were not eligible if they underwent prior cross-sectional imaging test to evaluate the presenting symptoms and signs or had a prior history of appendectomy.

Clinical Suspicion for Appendicitis

While clinical suspicion for appendicitis was raised by known symptoms and signs including right lower quadrant pain, migration of pain, vomiting, tenderness and/or rigidity (148), we chose not to create any specific clinical criteria to define the "suspicion for appendicitis" in LOCAT. Instead, the clinical suspicion for appendicitis in LOCAT was left to the discretion of the emergency department physicians on service. In general, patients with appendicitis had diverse presentations (149-151) and the clinical assessment often unavoidably involved many physicians with different expertise, due to the high prevalence of the disease. Therefore, the adoption of any fixed clinical criteria could be compromised the generalizability of the study findings. Wagner *et al*. (148) reviewed 10 previous "high-quality" studies and found that all of the studies used inclusion criteria of "suspected appendicitis" or "abdominal pain" without further definition. This was understandable, as fixing clinical criteria to investigate diagnostic clinical features was prone to circular logic.

The Need for CT Examination

The need for CT examination was also determined at the discretion of the emergency department physicians on service, who could proceed with imaging tests unless symptoms subside or an alternative diagnosis was established during the observational period. While it may be debatable in Western countries (152) whether patients with typical presentations of appendicitis require a preoperative imaging test or not, such patients mostly underwent CT examination in Korea where right-sided colonic diverticulitis, which often clinically mimics typical appendicitis (153, 154), was a common alternative diagnosis (29).

While the need for CT examination or any other diagnostic testing should be individualized for each patient at the discretion of the physician, the decision threshold to utilize CT as the initial imaging modality may not have been uniform among sites or emergency department physicians (2, 152). Although the indication of performing CT examination in previous studies could be largely divided into 2 thresholds of atypical versus all suspected (including both typical and atypical) patients, a meta-analysis (152) had suggested that the dichotomy of atypical versus typical presentations was merely theoretical. With the broad eligibility criteria in LOCAT, patients with either typical or atypical presentations could be enrolled (3, 152).

Even though imaging tests such as CT were originally intended for the cases with atypical presentations historically, the use of CT has been expanded to include patients with more typical presentations, which showed better overall clinical outcomes such as lower NAR (155).

Generalizability

Again, the LOCAT protocol did not impose any fixed criteria for (a) the clinical suspicion for appendicitis or (b) the need for CT examination. The broad eligibility criteria were to reflect the normal practice pattern in the sites and presumably in many other institutions, which kept a reasonably sensitive standpoint in raising the clinical suspicion of appendicitis and then used imaging tests to confirm or rule out appendicitis. To determine the generalizability of the LOCAT results, the data of participant baseline characteristics were thoroughly collected through standardized procedures such as Patient Questionnaire. In the final report of the main LOCAT results, the participant baseline characteristics were described particularly using

Alvarado score (106) and appendicitis inflammatory response score (107). These were the two most popular scoring systems to rate the overall likelihood of appendicitis based on symptoms, signs, and laboratory findings. It has been remained to be debatable if the scoring systems could be used as a guide to determine which patients required further observation or imaging studies. None of the scoring systems has been adopted into widespread clinical practice (140, 156).

The eligibility criteria differed from those in the previous studies measuring the effect of preoperative CT on NAR (6, 7, 9-13) in that the LOCAT protocol limited the participants to adolescents and young adults, for whom the long-term risks of cumulative radiation were more relevant (20).

Representativeness of Study Sample

The followings were essential in estimating the representativeness of the study sample at each site.

- A lead radiologist in the site was privileged to create and control dedicated "appendix" CT order(s) (Study Description, DICOM [Digital Imaging and Communications in Medicine] tag number: 0008,1030) in the Hospital Information System. These orders were consistently used for all potentially eligible patients undergoing CT evaluation of the appendix, whether they were eventually enrolled in LOCAT or not. In this way, the total number of patients of 15–44 years in age undergoing CT examinations to diagnose or rule out appendicitis during the study period at the site could be counted at the end of the study in a retrospective manner.
- Likewise, the lead radiologist was privileged to create and control dedicated "appendix" ultrasonography order(s). These orders were consistently used for

all potentially eligible patients undergoing ultrasonographic evaluation of the appendix, whether they were eventually enrolled in LOCAT or not. In this way, the total number of patients of 15–44 years in age who underwent ultrasonography examinations to diagnose or rule out appendicitis during the study period in the site could be counted in a retrospective manner.

- The site lead investigator made all possible efforts to use such orders consistently for all potentially eligible patients at the site.
- The site lead investigator was able to provide the total number of nonincidental appendectomies in patients of 15–44 years in age performed at the site during the study period in a retrospective manner.

Withdrawal Criteria

All participants had the right to withdraw at any point during the study without risk of future prejudice. Any Investigators or medical staffs could discontinue the study procedure(s) in any participant at any time if medically necessary. When the study procedures were discontinued for any participant, the reason must have been recorded, and the LOCAT office must have been notified promptly. Participants had never been replaced in LOCAT.

Randomization

Participants who had given consent was assigned an Enrollment Number and then randomly assigned to undergo either 2-mSv CT or CDCT of the abdomen and pelvis at a 1:1 ratio. Details of stratification/blocks were confidential since it was theoretically possible to predict allocation for some participants if past allocation and block site had been known (157). Sequentially numbered (i.e., the with the

Enrollment Number printed), opaque, sealed envelopes containing computer-generated random assignments were prepared for each site by the data center and kept at an agreed-upon location at the CT unit of each site. The time point of the initiation of the study procedure was defined as the time point of opening the envelope. Randomization happened at the time of the CT examination. To enter a participant into LOCAT, the radiology technologist on duty opened the next consecutively numbered envelope. Before opening the envelope, the radiology technologist had to write the date and his/her signature on the front of the envelope. Randomization did not occur until after the arrival of the participant at the CT unit, but could occur before the availability of laboratory test results. After randomization, if a participant was found to have contraindications to intravenous contrast-enhanced CT, the participant was treated according to the standard-of-care at each site and remained in the study.

While the medical staff members were not able to be blinded to the allocation because of obvious differences in the image texture (dependent on the CT radiation dose), the participants and outcome assessors were kept blinded to the allocation.

Index Test

Intravenous contrast-enhanced abdomen and pelvis CT examinations were performed using 22 CT machines (**Table 3**). For the experimental group, target effective dose was 2 mSv, which was roughly the lowest dose explored in previous studies (31-33). For the control group, target effective dose was individualized for each CT machine following the institutional normal dose, ranging from 3 mSv to 8 mSv (median 7 [interquartile range, 6–7] mSv) across all CT machines. This flexibility was in line with the goal of the trial which was lowering the radiation dose

in usual practice through the course of the trial.

In comparison, the mean dose for general-purpose abdomen CT in national surveys from different countries has ranged from 7 mSv to 11 mSv (158). The median DLP were 131 (interquartile range, 117–147) and 481 (389–554) mGy·cm for the two groups, respectively. Image section thickness had to be 5 mm or thinner with 20% or greater overlap. The use of iterative reconstruction was recommended but not required. Otherwise, there was no restriction regarding the imaging technique. Other imaging techniques were identical between the two groups for each CT machine (**Table 4**).

Table 3. CT machines

Characteristics	CT machine* (n = 22)
Manufacturer	
Siemens Healthcare	12
Philips Healthcare	5
GE Healthcare	3
Toshiba Medical System	2
Number of channels	
16-channel	3
64-channel	7
128-channel	9
256-channel	2
640-channel	1
Target effective dose (2-mSv CT vs. CDCT)†	
2 mSv vs. 3 mSv	1
2 mSv vs. 4 mSv	0
2 mSv vs. 5 mSv	1
2 mSv vs. 6 mSv	6
2 mSv vs. 7 mSv	9
2 mSv vs. 8 mSv	5

Note.—Data are numbers of CT machines. CDCT = conventional-dose CT. *Two of the sites operated two CT machines per site. The remaining 18 sites operated a single CT machine per site. †While the effective dose for 2-mSv CT was aimed at 2 mSv for all CT machines, the target effective dose for CDCT was individualized for each CT machine following the institutional normal dose.

Table 4. CT imaging parameters

Imaging parameters	CT machine (n = 22)	
	2-mSv CT	CDCT
Tube potential (kVp)		
Fixed		
100	5	4
110	1	1
120	15	15
Automatic selection		
80–120	1	0
80–140	0	1
100–140	0	1
Automatic tube current modulation		
Reference tube current-time product (mAs)*	55 (50–70)	200 (180–220)
Noise level (HU)†	32 (16–33)	13 (12–15)
Iterative reconstruction		
Used	8	4
Available, but not used	4	8
Not available	10	10

Note.—Data are numbers of CT machines or median (IQR). CDCT = conventional-dose CT. *For machines of Siemens Healthcare or Philips Healthcare. †For machines of GE Healthcare or Toshiba Medical Systems. Noise level was defined as the standard deviation of CT numbers of a region-of-interest in a water phantom.

Site radiologists (n = 160) made CT reports prospectively as a part of daily practice. They could access medical records and generally confer with the referring physicians. They had to use a predefined structured report form (138) indicating the likelihood of appendicitis on a five-point Likert scale (**Table 5**). During the working hours (typically 08:00–17:00 of working days), 50 attending radiologists made 791 reports in the 2-mSv CT group, while 51 attending radiologists made 771 reports in the CDCT group. The remaining after-hour reports were made by 96 on-call radiologists in the 2-mSv CT group and 90 on-call radiologists in the CDCT group. All but three of the on-call radiologists were residents. Attending radiologists later

revised the after-hour reports. However, we did not include the addenda in our analysis. 71% (n = 114) of the 160 involved radiologists had little prior experience with low-dose appendiceal CT (34 radiologists with ten cases or less, and 80 with no experience) (**Table 6**). To ensure patient safety, we invited all potentially involved radiologists to an online training course including 30 cases of 2-mSv appendiceal CT (141) before they conducted any trial procedure. 89% (n = 143) of the involved radiologists completed this course.

Table 5. Minimum requirement for CT report and scoring criteria

Analyzed findings and scoring criteria
Visualization of the appendix[*]
Grade 0. Not identified.
Grade 1. Unsure or partly visualized.
Grade 2. Clearly and entirely visualized.
Likelihood of appendicitis[†]
Grade 1. Definitely absent. Clinical observation is recommended.
Grade 2. Probably absent. Clinical observation is recommended.
Grade 3. Indeterminate. Clinical observation or surgical exploration is recommended.
Grade 4. Probably present. Surgical exploration is recommended.
Grade 5. Definitely present. Surgical exploration is recommended.
Appendiceal perforation[‡]
Grade 1. Unlikely present
Grade 2. Equivocal
Grade 3. Likely present
Periappendiceal abscess that needs a drainage procedure
Absent or present
Alternative diagnosis

Note.—Other findings could be added to CT reports following site policy or radiologist preference. [*]In cases with phlegmon or abscess, Grade 2 was assigned if there was clear continuity between the lesion and the remaining appendiceal base, indicating that the lesion had originated from the appendix. [†]The primary diagnostic criteria were appendiceal enlargement with mural thickening and periappendiceal fat stranding. An appendix larger than 6 mm in diameter was considered potentially abnormal. Secondary diagnostic criteria included abnormal mural enhancement, appendicolith, phlegmon, and abscess (134, 159). The lack of appendiceal visualization in conventional-dose CT is a highly predictive sign in ruling out appendicitis (135, 136). In February 2016, the trial protocol was amended to adopt new study results (unpublished data) that appendiceal non-visualization is a highly predictive sign in ruling out appendicitis also in 2-mSv CT. [‡]Based on findings of extraluminal gas or appendicolith, periappendiceal fluid or phlegmon, severe periappendiceal fat stranding, and defect in the appendiceal wall (160, 161).

Table 6. Care providers

Experience	Either group	2-mSv CT group	CDCT group
Emergency physicians	…	…	…
Attending physicians	…	…	…
Clinical experience (years)	…	…	…
Trainees	…	…	…
Radiologists	160, *7 [3–13]*	146, *6 [2–12]*	141, *6 [2–12]*
Attending radiologists	53, *3 [1–3]*	50, *2 [1–3]*	51, *2 [1–3]*
Clinical experience (years)	10 [4–15]	11 [4–15]	10 [4–15]
On-call radiologists or trainees	107, *9 [2–10]*	96, *8 [4–9]*	90, *8 [3–9]*
Prior experience (number of cases)			
In appendiceal 2-mSv CT			
0	80, *3 [1–8]*	71, *2 [1–8]*	73, *3 [1–7]*
1–10	34, *2 [1–3]*	30, *2 [1–3]*	27, *1 [1–2]*
≥ 11	43, *2 [1–4]*	42, *2 [1–4]*	38, *2 [1–4]*
Missing data	3, *1 [1–1]*	3, *1 [1–1]*	3, *1 [1–1]*
In appendiceal CT regardless of radiation dose			
≤ 10	40, *3 [2–5]*	35, *2 [1–5]*	37, *2 [2–5]*
11–100	61, *4 [2–5]*	55, *3 [2–5]*	47, *4 [3–5]*
≥ 101	56, *3 [1–4]*	53, *2 [1–4]*	54, *2 [1–4]*
Missing data	3, *1 [1–1]*	3, *1 [1–1]*	3, *1 [1–1]*
Patients per radiologist	8 [3–18]	4 [2–11]	5 [2–11]
Attending surgeons*	146, *7 [6–10]*	118, *6 [5–9]*	121, *6 [3–9]*
Clinical experience (years)	7 [4–11]	7 [4–12]	7 [5–12]
Appendectomies per surgeon	3 [2–7]	2 [1–4]	2 [1–4]
Attending pathologists†	91, *4 [3–7]*	81, *3 [2–7]*	79, *3 [2–6]*
Clinical experience (years)	12 [4–21]	12 [4–22]	12 [4–21]
Appendectomy specimens per pathologist	5 [2–10]	3 [1–6]	3 [2–7]

Note.—Unless otherwise specified, data are medians [IQR] for the numbers of care providers or patients. Data in italic are medians [IQR] per sites. Data of experience are those at the time of first involvements of individual care providers. A small number of care providers were counted twice due to job position changes within or across sites during the study period. Ellipses indicate that data could not be obtained due to the complexity related to team-based approach and rotational shiftwork. CDCT = conventional-dose CT. *Performed or supervised appendectomy. †Verified pathologic report for appendectomy specimens.

CT Image Acquisition and Archiving

The LOCAT protocol in regard to CT techniques was implemented collaboratively by each site team and the CT technique coordinator (Bon Seung Gu, R.T., M.S.). While the number of the site team members could vary depending on practice size, the basic membership typically included the lead radiologist and a lead CT technologist.

To ensure the consistency in image acquisition and archiving, the lead radiologist was privileged to create and control dedicated CT imaging protocols for suspected appendicitis according to the LOCAT protocol. It was mandatory to set up dedicated automated scanning and reconstruction programs for both 2-mSv CT and CDCT protocols in each CT machine and that all participants were scanned with the dedicated programs, hence requiring minimal operation by CT technologists in the image acquisition (116). Typically, the lead CT technologist set up and appropriately updated the dedicated automated scanning and reconstruction programs, and also educated other site CT technologists.

Single breath-hold, intravenous contrast-enhanced, helical scans were obtained during the portal venous phase using 16- or higher detector-row CT machines. It was strongly recommended to reconstruct and archived from each helical scan two transverse image data sets with different section thicknesses: thick (3 to 5 mm) and thin (≤ 2 mm). The reconstruction of the thin-section images was essential, particularly in sites where a thin-client image distribution solution had been available. The technical advantages of this two-tier (thick and thin) image reconstruction have been previously described (162). In sites where the thin-client image distribution solution had not been available, the reconstruction of thin-section images was not

mandatory. Instead, those sites were strongly recommended to reconstruct and archive both axial and coronal images with a section thickness of 3−5 mm and a reconstruction interval of 2 mm or less, at least for 2-mSv CT. The reconstruction of the transverse thin-section images or the coronal images was essential to ensure the availability of the multiplanar imaging capability with a reasonably high through-plane resolution. As long as other imaging parameters conformed to the LOCAT protocol, a minor deviation from the suggested section thickness or reconstruction interval was not regarded as a protocol non-adherence, and the participants remained for the analysis in the group to which they had originally been assigned.

We strongly recommended the use of an up-to-date iterative reconstruction, whenever it was available, particularly for 2-mSv CT. The iterative reconstruction is an evolving technique potentially reducing the required radiation dose (96). According to a recent study (75), the use of an iterative reconstruction did not significantly improve radiologists' diagnostic performance and confidence when compared with a conventional filtered back projection in diagnosing appendicitis at 2-mSv CT, while the iterative reconstruction exhibited higher subjective image quality than the filtered back projection. As iterative reconstruction was not used as the current standard-of-care in many sites, it was left to the discretion of the lead radiologist whether or not to use iterative reconstruction, and whether the iterative reconstruction should be used as a replacement for or an adjunct to conventional filtered back projection. Whichever policy the lead radiologist had chosen, it was mandatory to archive all the images obtained by using the filtered back projection or iterative reconstruction.

Otherwise, there was no restriction regarding scanner type or scan parameters other than those required in **Table 7**. The imaging protocol had to follow the

standard-of-care at each site, including the use of enteric contrast material, intravenous contrast material injection method including the use of saline flush, tube potential, and additional image reconstruction such as coronal images.

CT technology has been rapidly evolving, particularly toward using lower radiation dose. Because LOCAT participant recruitment took several years, sites were allowed to adopt the advances during the study period so that CT imaging parameters were as up to date as possible. Within the LOCAT protocol, changes in the imaging protocol, even including the radiation dose level (target DLP) in the CDCT group, were allowed during the study period. Any such changes (or any installation of a new CT machine during the study period) had to be reported to the site lead investigator and the LOCAT office.

Table 7. Required imaging parameters

Required scan protocol	
Intravenous contrast enhancement	
Intravenous access	Antecubital, not lower extremity
Contrast material	Iodine amount, 400−800 mg/kg
Scan timing	Portal venous phase
Scan	
Range	From 4 cm above the liver dome to 1 cm below the ischial tuberosity
Collimation	Use all detector rows available
Automatic exposure control	Use all techniques available
Reconstruction*	
Thin-client image distribution solution is available	
Thick transverse images	Section thickness, 3−5 mm; overlap, 20% or more
Thin transverse images	Section thickness ≤ 2 mm; Reconstruction interval ≤ 1 mm
Thin-client image distribution solution is not available	
Thick transverse images	Section thickness, 3−5 mm; Reconstruction interval ≤ 2 mm
Thick coronal images	Section thickness, 3−5 mm; Reconstruction interval ≤ 2 mm

Note.—*Minor deviation of the suggested section thickness or reconstruction interval was not regarded as a protocol non-adherence.

Radiation Doses

LOCAT was aimed to compare 2-mSv CT vs. CDCT (standard-of-practice at each site). The unit of radiation dose that was primarily referred in the LOCAT protocol was effective dose (in mSv). In general, effective dose provides a general idea of detrimental effect from ionizing radiation in comparing different imaging techniques or in justifying an imaging study (60). Importantly, the effective dose is estimate of generic risk to a generic individual (both sexes, all ages, standard-sized patient) and does not represent the actual risk for any individual participant. A minute

difference of decimals in mSv was unimportant in the range of 1−10 mSv.

There was considerable variation across the sites in the radiation dose conventionally used for the diagnosis of appendicitis (4). Also, the standard-of-care radiation doses in sites could gradually decrease to some extent during the study period, with advances in CT technology and with greater awareness of the associated carcinogenic risk (59). Taking into consideration these variations and changes, LOCAT had a unidirectional standpoint in determining and adjusting the radiation doses: being flexible towards dose decreases in either group while being strict against dose increases. It was noted that lowering the radiation dose in the CDCT group would affect the study results toward the noninferiority. Nevertheless, the unidirectional policy was in line with the ultimate goal of LOCAT Group, which was disseminating the 2-mSv CT technique throughout the sites and other hospitals over the course of LOCAT.

It was mandatory to use all automatic exposure control techniques (163) that were available for each CT machine. In general, the purpose of automatic exposure control techniques is the consistency of image quality, which is similar to photo-timer in a camera. With the use of the automatic exposure control techniques, the X-ray tube current is decreased for slender patients or low-attenuating body parts and boosted for obese patients or high-attenuating body parts. The failure to use the automatic exposure control techniques would result in unnecessarily high doses to slender patients and poor image quality in obese patients. In LOCAT, the actual radiation dose varied substantially from the preset reference values, because the tube current and/or tube potential had been automatically adjusted according to the individual participant’s body size and shape.

Record of Modulated Radiation Dose

To ensure the accuracy and completeness in the data collection about the radiation dose, it was mandatory to use CT machines (or scanning programs) capable of capturing the dose information. For each participant, the modulated radiation dose was recorded in a DICOM image containing a screen capture of a text table that summarized overall X-ray exposure. The collected dose data included $CTDI_{vol}$ (based on the use of 32-cm [not 16-cm] diameter reference phantom) and DLP. The radiation dose reported in LOCAT referred to that used for the single-phase helical scan and did not include the dose used to obtain CT radiographs (scout images) or other preparation scans such as bolus tracking. When an additional scan was performed for any reason (e.g., in the case of protocol non-adherence rescanning for the non-diagnostic initial study, scanner failure, or extravasation), then the total amount of DLP for the multiple helical scans as well as the DLP for each of the helical scans were reported.

The following data were collected: $CTDI_{vol}$, DLP, anteroposterior, and lateral diameters of participant abdomen, and size-specific dose estimate (59).

Target Median DLP Values for the 2-mSv CT and CDCT groups

In each CT scanner, reference tube-current–time products (or noise levels) were adjusted aiming at target DLP values for an average-size patient (142) in the 2-mSv CT and CDCT groups, respectively.

For the low-dose group, the target DLP was set as 130 mGy·cm for all CT machines, which corresponded to effective doses of 2 mSv with a conversion factor of 0.015 $mSv \cdot mGy^{-1} \cdot cm^{-1}$ (62). This “low” dose was empirically determined based on experience in depicting the appendix using reduced tube currents (27, 28, 101)

and then employed in the previous single-center trial (29).

In the CDCT group, the target DLP was determined for each CT machine at the discretion of the lead radiologist in line with the standard-of-practice at each site. The target DLP was typically 530 mGy·cm corresponding to 8 mSv but did not exceed the typical dose used in standard-of-practice at each site. In comparison, reference values often quoted at the time the LOCAT protocol was written ranges from 7 to 10 mSv (164-166). It should be noted that changing the target DLP in each CT machine for the CDCT group at the discretion of the lead radiologist was allowed during the study period and that no lower limit was defined for the target DLP for the CDCT group. This flexibility was needed because the "standard" dose used at each site could be decreased gradually during the study period (59).

Calibration of Radiation Doses

The calibration (142) was performed by the lead CT technologists at each site in conjunction with the CT technique coordinator and technicians from the CT manufacturers. Owing to rapid technologic advances, CT manufacturers had developed diverse automatic exposure control techniques, use proprietary nomenclature (163), but often did not clarify the technical details. Furthermore, the techniques varied with the CT machine model and software version even within the same CT manufacturer. To avoid confusion and to facilitate the calibration process, the CT technique coordinator had developed and appropriately updated Radiation Dose Calibration Sheet/Manual which covered all individual CT machines used in LOCAT. The CT technique coordinator was also responsible for keeping thorough record (e.g., target DLP values) of each calibration step before registering the first participant or during the study period for each CT machine by using Radiation Dose

Calibration Sheet/Manual. The outline of the calibration process for each CT machine (**Fig. 1**) was as follows. The details were available in Radiation Dose Calibration Sheet/Manual.

- *Before* registering the first participant at each site, the automated scan program for CDCT in each CT machine was tested in regular abdomen and pelvis CT examinations which were performed for various purposes in patients not enrolled in LOCAT. For each CT machine, it was mandatory to draw a box-and-whisker plot of DLP distribution in 100 or more consecutive patients to ensure the appropriate calibration. When the measured median DLP value was greater than 110% of the predefined target DLP (i.e., out of the error range of +10% from the target DLP), it was mandatory to adjust the reference tube-current–time product (or noise level) as appropriate, and then to draw a new box-and-whisker plot of DLP distribution in the next 100 (or more) consecutive patients. This calibration process was iterated until the median DLP value reaches the target DLP within the error margin of +10%. If the median DLP value was less than 90% of the target DLP (i.e., out of the error range of -10% from the target DLP), the lead radiologist could decide whether to decrease the target median DLP at his or her discretion. The determined parameters for the automated scan program in each CT machine was recorded in Radiation Dose Calibration Sheet/Manual. Written informed consent was not required for the patients involved in this calibration which should be regarded as a standard-of-practice quality assurance program at each site.
- *Before* registering the first participant at each site, the determined parameters for the automated scan program for 2-mSv CT in each CT machine were recorded in Radiation Dose Calibration Sheet/Manual. The technical details for

individual CT machines are available in Radiation Dose Calibration Sheet/Manual.

- *During* the study period, the box-and-whisker plot was drawn for every 50 participants for each of 2-mSv CT and CDCT groups for each CT machine. If the median DLP value exceeded the target DLP by 10%, then the dose adjustment for the next 50 patients was mandatory, while it was left to the discretion of the lead radiologist whether to adjust for the reverse error. The adjusted scan parameters were recorded in Radiation Dose Calibration Sheet/Manual. As long as the dose calibration process conformed to the LOCAT protocol, a minor deviation of the median DLP value out of the error range was not regarded as a protocol non-adherence, and the participants remained for the analysis in the group to which they had been originally assigned.

Estimation of Carcinogenic Risk Associated with CT Examination

The risk of cancer associated with CT radiation is a highly debatable topic (167). The carcinogenic risk in each group of LOCAT was estimated as follows. Organ-specific radiation doses were calculated using Monte Carlo radiation transport from the measured DLP and other CT scan parameters (84). Sex- and organ-specific lifetime excess incidence of radiation-induced cancer was estimated using radiation risk models developed by the Biological Effects of Ionizing Radiation VII Committee (168). This approach has been commonly used to estimate radiation-related cancer risks from CT scans (22, 23, 169). As the cancer risk was not a true measurement but estimate based on unverified assumptions, it was not included as a secondary endpoint and will be reported separately from the main LOCAT results.

In the previous single-center trial (29), the risk of cancer associated with CT

radiation has been calculated at two dose levels similar to those used in LOCAT. For the 2-mSv CT (median effective dose, 2 mSv) group, an exposure at the age of 30 years was estimated to result in a lifetime excess risk of 14 and 16 cancers per 100,000 male and female patients, respectively. For CDCT (median effective dose, 8 mSv) group, the estimated risks were 63 and 72 cancers per 100,000 male and female patients, respectively. These point estimates imply that using CDCT instead of 2-mSv CT at the age of 30 years in estimated 2,000 male or 1,800 female patients would result in one additional cancer. With an exposure at the age of 30 years, the largest proportion of total cancer incidence was attributable to colon cancer (28% for males and 17% for females), followed by bladder cancer (18% for males and 17% for females). Additional data for different ages can be found elsewhere (29).

Image Interpretation

In making CT reports, radiologists at each site reviewed the thick-section images on a picture archiving and communication system workstation at each site.

In sites where a thin-client image distribution solution such as AquariusNET (TeraRecon, San Mateo, CA) or equivalent was available, it was recommended to reconstruct the additional thin-section images and to review the thin-section images using the thin-client image distribution solution as follows. In addition to the thick-section images, the thin-section images were reviewed as needed for a more confident diagnosis, using the multiplanar sliding slab averaging technique, a real-time image post-processing technique widely used to efficiently review large thin-section CT datasets. To make the thin-section image datasets timely available 24 hours per day, 7 days per week, in reporting areas, it was recommended the sites to use a thin-client image distribution solution.

The technical details of the sliding slab averaging technique have been described elsewhere (170-172). Importantly, this visualization technique reduces noise in the final displayed image by averaging the pixel values within the slab, particularly in the grainy 2-mSv CT images. This technique has been used to depict the appendix (27-29, 101, 133, 173), enhancing radiologists' confidence in making the diagnosis of appendicitis as compared with conventional thick-section CT reviewing (117, 133) which is likely to have been employed in the previous studies measuring the effect of preoperative CT on NAR (7, 9-13). A recent study (117) reported that sliding slab averaging review of thin sections is helpful particularly when the diagnosis of appendicitis is difficult at 2-mSv CT.

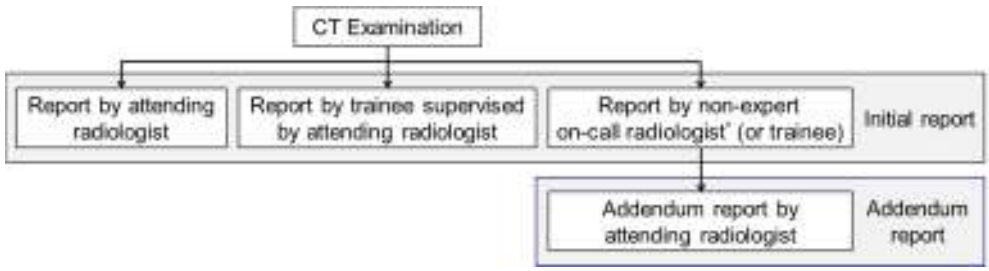
It should be noted that the importance of the two-tier (thick and thin) image reconstruction and multiplanar sliding slab averaging technique is often overlooked by radiologists, although many average hospitals now have sufficient hardware and network resources to implement the techniques.

Radiologists and CT Reports

The CT images were interpreted as a part of daily clinical practice. During the working hours, the initial reports were made by attending radiologists or trainees supervised by attending radiologists. CT examinations performed after hours were given initial reports by on-call radiologists (or trainees) with various levels of experience in abdominal radiology (**Fig. 3**). In all sites except one where the previous single-center trial (29) had been conducted, most radiologists had limited experience in interpreting low-dose abdomen CT.

Figure 3. Reporting procedure.

*Including radiology trainees and board-certified radiologists with various levels of experience in abdominal radiology.



All initial reports made by the non-expert on-call radiologists were reviewed in the morning of the next business day by the attending radiologists and then, they made an addendum report and gave educational feedback to the on-call radiologists. Any important changes in the initial report had to be immediately notified to the referring physician so that the patient management could be changed. The addendum reports were not included in the main LOCAT results for two reasons. First, due to the limited availability of the attending radiologists around the clock, some addendum reports could be made after patient disposition regarding surgery. Second, even in cases in which the addendum report was apparently made before patient disposition, it was difficult to determine objectively how the addendum report alters or consolidates the clinical decision regarding the patient disposition. As the study analyses did not include the addendum reports that obviously changed patient disposition in some cases but included the initial reports by on-call radiologists in such cases, the measured diagnostic accuracy of the CT reports in LOCAT were underestimated to some extent.

In making the reports, the radiologists were allowed to access medical records, including clinical and laboratory findings, and to contact the referring physician for

a consultative discussion. The CT reports were made in a predefined structured format to meet the minimum requirement in **Table 5**. The likelihood of appendicitis was rated on a 5-point scale, and visualization of the appendix was rated on a 3-point scale. The presence of appendiceal perforation was determined. Alternative diagnosis to explain the abdominal pain was proposed whenever possible.

Radiologist Training

With degraded image quality, 2-mSv CT is likely to be less straightforward to interpret than the CDCT, especially for inexperienced radiologists such as the non-expert on-call radiologists. At the time of writing of the initial version of the LOCAT protocol, no published data existed on the learning curve for interpretation of the 2-mSv CT images for the diagnosis of appendicitis. The LOCAT Group later analyzed the learning curves of 46 attending radiologists and 153 radiology residents from 22 sites and reported that the learning curve was prolonged and formed gradually over the years by overall radiology training and clinical experience in general rather than by experience with low-dose appendiceal CT specifically (141).

Before the initiation of LOCAT, the LOCAT coordinating committee had, therefore, planned a self-learning course, LOCAT-Training, to train the radiologists at all sites. LOCAT-Training was directed by Dr. Min Hee Lee (141). The development of the e-learning course was led by Dr. Hyoun Sik Woo (174). LOCAT-Training is publicly available (www.locat.org). The training materials included introductory PowerPoint slideshows and 2-mSv CT cases with direct feedback regarding the appendix location and final diagnosis. Any questions on LOCAT-Training should be directed to the LOCAT office.

LOCAT-Training had several purposes. First, to ensure the safety of the

participants in LOCAT, it was mandatory for each site that above 80% of all site radiologists potentially involved in LOCAT, who were the attending radiologists or on-call radiologists, should complete LOCAT-Training before registering the first participant at the site. When a site recruited new radiologist(s) potentially involved in LOCAT during the study period, the new radiologist(s) were also recommended to complete LOCAT-Training. Second, during the development of LOCAT-Training, the learning curve was measured for the radiologists with different experience levels, and the results were used to eventually improve LOCAT-Training itself. Third, the radiologists could become familiar with the sliding slab averaging technique and predefined structured CT report form used in LOCAT as the same image reviewing technique and report form were used in the LOCAT-Training. This could help the collection of high-quality data across the sites in LOCAT. Fourth, the promotion of the 2-mSv CT technique was intended through the course of LOCAT-Training as well as LOCAT, to disseminate the 2-mSv CT technique in domestic and international radiology communities.

Considerations Regarding Technical Advantages over Previous Studies

There have been remarkable advances in CT technology over the last decade, including improved spatial resolution, higher signal-to-noise ratio, faster scanning, increased use of multiplanar images, and the introduction of the sliding slab averaging technique. Therefore, with the same radiation dose used, the CT imaging protocol in LOCAT is considered advantageous in depicting the appendix compared with the CT protocols employed in the previous studies that have measured the effect of preoperative CT on NAR (7, 9-13).

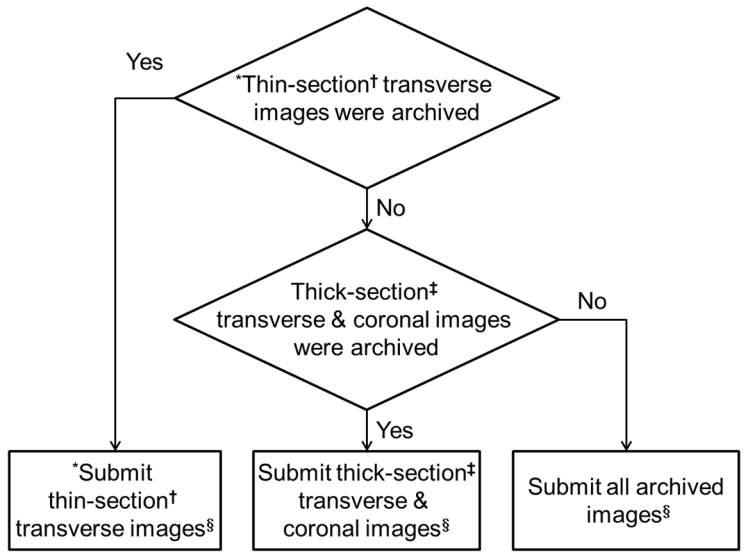
Image Submission

The CT images in DICOM format, along with an Image Submission Worksheet, were submitted to the data center according to the timeline and instruction specified in Image Submission Worksheet and LOCAT protocol. The thin-section transverse image datasets, whenever they were available, have been submitted; otherwise, both transverse and coronal thick section images were submitted (**Fig. 4**). If an iterative reconstruction had been used in addition to filtered back projection, all the images from both reconstructions were submitted.

For technical support in image data handling, sites could contact the image transfer coordinator at the LOCAT office. All DICOM header records containing confidential participant information were scrubbed at each site before the images are transferred. This anonymization involves replacing both Participant Name (tag number: 0010,0010) and Participant ID (tag number: 0010,0020) with the Enrollment Number (e.g. SNUBH1001).

Figure 4. Image submission algorithm.

If iterative reconstruction had been used in addition to filtered back projection, both images were submitted. *Most recommended. †Section thickness ≤ 2 mm, reconstruction interval ≤ 1 mm. ‡Section thickness, 3−5 mm, reconstruction interval ≤ 2 mm. §According to image submission worksheet.



Co-intervention

Emergency physicians initially assessed the patients. As allowed by the trial protocol (39), 37 patients in the 2-mSv CT group and 39 in the CDCT groups underwent additional abdominal ultrasonography or CDCT at the discretion of the care providers as the diagnosis of appendicitis remained undetermined after the initial CT and clinical observation.

Appendectomy was the treatment of choice for appendicitis at all sites. The 2-mSv CT and CDCT groups were comparable for the time required for patient disposition. The median interval between CT and appendectomy was 5.2 (interquartile range [IQR] 3.1–9.5) vs. 5.4 (3.3–10.4) hours. The median interval between CT and discharge without surgery was 1.6 (IQR 0.9–3.8) vs. 1.7 (1.0–4.2)

hours. The median length of hospital stay associated with appendectomy was 3.0 (IQR 2.4–4.0) vs. 2.9 (2.2–3.9) days.

Additional Imaging

If the diagnosis of appendicitis remained undetermined after initial CT examination and clinical observation, then additional abdominal imaging test(s), including abdominal ultrasonography (134, 175) or CDCT, could be performed at the discretion of the emergency department physician or surgeon. An additional imaging test was defined as one that was performed within 7 days of the initial CT to diagnose or rule out appendicitis.

General Treatment Guidelines

Except for the CT radiation dose in the 2-mSv CT group, all diagnostic and treatment processes in both groups followed the standard-of-care at each site. As most of the medical staff members involved in the participant care were not LOCAT investigators, they were expected to provide care as they did in their normal clinical practices. Before the registration of the first participant at each site, the site lead investigator submitted to the LOCAT office a completed Site Staff and Investigator List containing information on the numbers and experience levels of medical staff involved in the patient care. At the beginning of every academic year during the study period, the site lead investigator revised the Site Staff and Investigator List and submitted it to the LOCAT office.

Typically, the participants were initially evaluated by one of the emergency department physicians, including attending physicians, fellows, and residents. Each of these physicians, under the supervision of one of the attending physicians,

recorded patient history (using the Patient Questionnaire), performed a physical examination, determined the need and timing of diagnostic tests including imaging studies, contacted radiologists or surgeons for a consultative discussion, and determined the timing of hospital discharge. It was recommended, although not mandatory, that all potentially eligible patients were given the Patient Questionnaire to solicit chief complaint and past medical history.

Radiology service at each site was provided immediately upon request, 24 hours per day, 7 days per week, including the acquisition and interpretation of CT and ultrasonography, as well as the timely availability of thin-section CT images in the reporting areas (162).

All participants who potentially were required surgical exploration were referred to the surgical departments. A surgical resident was typically the first surgeon to evaluate the participant, and the final decision to operate was approved by one of the attending surgeons who performed or supervised the surgery. While appendectomy was the procedure of choice in patients with a preoperative diagnosis of appendicitis, the surgical plan was individualized for each participant as appropriate, including the need for preoperative percutaneous drainage of periappendiceal abscess and the use of the laparoscopic approach.

Criteria for hospital discharge following appendectomy are tolerance of a soft-blend meal, safe ambulation, and afebrile status without major complications. The short-term follow-up was typically scheduled around 7 to 14 days after an appendectomy at the outpatient department for the survey of complication and stitch-out. Pathologic examinations of the surgical specimens were performed by site pathologists following the standard-of-practice at each site.

Follow-up

Data collected during the follow-up include events and measurements regarding endpoints and adverse events (AEs). Every attempt was made to collect complete follow-up information according to study calendar (**Table 8**), whether the participants complete CT examination as they had been originally assigned, except for those who specifically withdraw consent for release of such information. If the participant had visited another hospital or had been transferred to another hospital, all possible efforts were made to obtain source documents from that hospital.

Table 8. Study calendar

Task	Initial visit	3 months
Initial clinical evaluation	O	
Laboratory tests	O	
Request for CT examination	O	
Screen for eligibility	O	
Obtain written informed consent	O	
Enrollment Number	O	
Contact phone number	O	
Randomization	O	
CT examination	O	
Radiation dose record of CT scan	O	
CT report	O	
Additional imaging test(s)	O	
Percutaneous abscess drainage	A	
Surgery	S	
Discharge	O	
Pathology report	S	
Telephone follow-up		O
Adjudication of final diagnosis	S	O

Note.—O, all participants; A, participants undergoing preoperative percutaneous abscess drainage due to perforated appendicitis; S, participants undergoing surgery.

For all participants, a structured telephone interview of the participant or proxy was conducted by the site research coordinators 3 months after the randomization following the instruction in the Case Report Forms (CRF). The elements of the interview were as follows.

- Persistence, development, or worsening of any symptom, sign, illness, or discomfort in the abdomen or any other body parts after the hospital discharge.
- Onset, duration, severity*, and resolution of the symptom, sign, illness, or discomfort.
- Any hospital visit, hospitalization, or surgery.
- An open-ended question, "how are you feeling?"

*Specifically, if self-care Activities of Daily Living have been limited (89).

The telephone contact occurred after 90 days from the randomization. Ideally, contact should have occurred no later than 14 days after the 90 days from the initial presentation. No fewer than 3 attempts should have been made by the site research coordinator to contact the participant or proxy. Ideally, efforts were continued until contact is made. Attempts were logged in the CRF. Participants should not be considered "lost to follow-up" until the final follow-up for the last participant in the study.

If the participant or proxy responded that any hospital visit, admission, or surgery related to abdominal symptoms had occurred, the site research coordinator then conducted a further review of the participant's medical records.

Endpoints in LOCAT

The endpoints were as follows.

Primary Endpoint

The primary endpoint was NAR. Negative appendectomy indicated the removal of the uninflamed appendix. NAR was defined as the percentage of negative appendectomies (unnecessary removal of uninflamed appendix) out of all nonincidental appendectomies (7, 9-13). As a secondary analysis, NAR in an alternative definition was calculated by excluding cases with appendiceal neoplasms without superimposed appendicitis from the numerator, as appendectomy would be clinically necessary for such patients (4). Any surgery performed for the treatment of presumed appendicitis was counted as nonincidental appendectomy, even though the surgical procedures could be more extensive than simple appendectomy (e.g., ileocecectomy).

Secondary Endpoints

- Clinical outcomes
  - APR: the percentage of perforated appendicitis[*] of all confirmed appendicitis cases (9-12)
  - The percentage of negative appendectomies out of all randomized cases[†]
  - The percentage (i.e., prevalence) of perforated appendicitis[*] out of all randomized cases[†]
  - The percentage (i.e., prevalence) of non-perforated appendicitis[*] out of all randomized cases[‡]
  - The proportion of participants requiring additional imaging test(s) to diagnose or rule out appendicitis
  - Delay in patient disposition

- ✓ Interval from CT acquisition[§] to appendectomy[‖] in participants undergoing nonincidental appendectomy
  - ✧ Delayed appendectomies following percutaneous abscess drainage and/or medical treatment are not included in this analysis.
- ✓ Interval from CT acquisition[§] to hospital discharge[¶] in participants not undergoing surgery

➢ Hospital stay associated with nonincidental appendectomy: the interval from CT acquisition[§] to hospital discharge[¶] after nonincidental appendectomy

- ✓ Delayed appendectomies following percutaneous abscess drainage and/or medical treatment are not included in this analysis.

➢ The percentage of nonincidental appendectomies out of all randomized cases[†]

- Diagnostic performance of CT reports
  - ➢ Diagnosis of appendicitis
    - ✓ Area under receiver-operating-characteristic curve
    - ✓ Diagnostic sensitivity and specificity: for calculation of the sensitivity and specificity, the 5-grade likelihood scores for appendicitis were collapsed into binary responses with a decision threshold of a score ≥ 3 as positive for the diagnosis. This decision threshold was based on the fact that appendicitis is actually present in 13-73% of patients with CT scans interpreted as equivocal (176) and that appendiceal perforation as a consequence of a false-negative diagnosis was considered more harmful to patients than a negative appendectomy as a consequence of a false-positive diagnosis.
- Diagnostic confidence in diagnosing and ruling out appendicitis

- Likelihood score for appendicitis in participants confirmed as having appendicitis
- Likelihood score for appendicitis in participants confirmed as not having appendicitis
- The frequency of indeterminate CT interpretation (grade 3)
- The frequency of normal appendix visualization at CT
    - The visualization of the normal appendix is a paramount sign in ruling out appendicitis (101, 135, 136).

- Diagnosis of appendiceal perforation at CT
    - Diagnostic sensitivity: the number of correct detections of the perforation divided by the number of cases of perforated appendicitis
    - Diagnostic specificity: the number of correct ruling out the perforation divided by the number of cases of appendicitis without perforation

[*]By using both broader and narrower definitions of appendiceal perforation.

[†]Added in Jan 2016.

[‡]Added in Nov 2016.

[§]Defined as DICOM tags, Study Date [tag number: 0008, 0020] and Study Time [tag number: 0008, 0030]).

[‖]Defined as the induction of anesthesia.

[¶]In participants discharged directly from emergency department without inpatient hospitalization at a ward, the hospital discharge was defined as the time of leaving the emergency department. For participants admitted to wards, the hospital discharge was defined as midday of the discharge date from the wards.

Considerations for NAR and APR

NAR and APR are the two reciprocal, established measures of quality of care (40) which represent the consequences of false-positive and delayed diagnoses, respectively. While NAR explicitly indicates false-positive diagnosis, there has been no practical outcome measure suggested to directly indicate the false-negative diagnosis. Therefore, APR has been frequently used as a surrogate index indirectly representing delayed diagnosis associated with diagnostic uncertainty in predicting appendicitis, although it is also associated with many other factors, including disease severity at the time of presentation and non-medical factors that delay the treatment (177). Since some appendiceal perforation might result from a separate clinical process than the one at work in non-perforated appendicitis (48), it is increasingly recognized that APR is not an ideal indicator of the quality of care. It should be noted that the definition of appendiceal perforation is missing or inconsistent in many of the previous studies addressing APR (43, 178).

It has been traditionally believed that an inverse relationship exists between NAR and APR if the overall performance of a diagnostic system is stably maintained. It has been asserted that a certain level of NAR (up to 20% before the introduction of CT) is an appropriate index of management and that the failure to maintain such a surgical threshold is an indication of insufficient surgical aggressiveness and an excessive rate of delayed diagnosis.

Changes in Endpoints

The primary endpoint was NAR, which was defined as the percentage of negative appendectomies out of all nonincidental appendectomies. The assumption was that the number of appendectomies would be similar between the 2-mSv CT and

CDCT groups. In Jan 2016, when 2,299 participants were enrolled, the Data and Safety Monitoring Board (DSMB) suggested that this assumption may not be the case and recommended to change the endpoints. The difference in the number of appendectomies may have been attributable to the potential difference in the diagnostic sensitivity, particularly for milder diseases (179) that may resolve without appendectomy (41, 180). It has been recently suggested that NAR is not an ideal indicator of the quality of care, as NAR does not take into account the appendicitis cases that resolve without appendectomy (41).

The coordinating committee, protocol development and revision committee, DSMB, and invited experts intensively discussed the need for changing the primary endpoint to the percentage of negative appendectomies out of all randomized cases. The coordinating committee and protocol development and revision committee concluded to keep the primary endpoint of NAR unchanged and to instead add three secondary endpoints as follows. First, the percentage of negative appendectomies out of all randomized cases, second, the percentage (i.e., prevalence) of perforated appendicitis out of all randomized cases, and third, the percentage of appendectomies out of all randomized cases]. Had the primary endpoint been changed to the percentage of negative appendectomies out of all randomized cases, the final sample size (in terms of enrolled participants) would increase by less than 100 from the original sample size (when other assumptions in the sample size calculation were unchanged). This increase was regarded as an unimportant change.

In the same context, the coordinating committee and protocol development and revision committee decided to add another endpoint, the percentage (i.e., prevalence) of non-perforated appendicitis out of all randomized cases, in Nov 2016 when more than 3,000 participants were enrolled.

Reference Standards

According to the trial protocol (39), independent assessors (five radiologists and two emergency physicians with 2- or 3-year experience) adjudicated the final diagnosis based on the trial data including pathologic findings and standardized 3-month follow-up telephone interview. The assessors were blinded to the index test results.

Site pathologists examined 1,069 appendectomy specimens as a part of daily practice. The pathologists were instructed to adhere to the definition of acute appendicitis: mural neutrophil infiltration or mucosal neutrophils with ulcerations (39). Appendicitis was confirmed primarily based on the pathological criteria. Additionally, a small number of patients who underwent interval appendectomy following percutaneous abscess drainage or antibiotic treatment were regarded as having appendicitis according to the study protocol (39) even if the pathologic findings did not meet the criteria.

Overview of Reference Standards

The independent outcome assessors determined the final diagnosis based on all available surgical findings, pathologic findings, other medical records, and the standardized telephone interviews. Pathologic examinations of the surgical specimens were performed by site pathologists following the standard-of-practice (181) at each site. The outcome assessors were emergency department physicians or radiologists blinded to the allocation. If an outcome assessor had not been confident in establishing a final diagnosis, the case was submitted to a multidisciplinary expert panel at each site to determine a working diagnosis.

Definition of Acute Appendicitis

The presence of acute appendicitis was determined based on the histopathologic findings. Histopathologic diagnosis of acute appendicitis was defined as neutrophil infiltration in the appendiceal wall, including the mucosa, submucosa, and muscularis propria (181). If neutrophilic collections had been confined to the mucosa, the diagnosis of acute appendicitis was based on the presence of mucosal ulcerations (182). It should be noted that the pathologic diagnosis of appendicitis is not always straightforward and that specific definition of appendiceal inflammation lack consensus agreement (41, 183). A systematic review (149) has found that the pathologic criteria for appendicitis were missing or inconsistent in many previous studies. To ensure diagnostic reproducibility across pathologists and sites in LOCAT, all pathologists involved in the patient care had to review Brief Guideline for Pathologic Diagnosis and make a pathologic diagnosis according to the guideline.

The absence of acute appendicitis (i.e., not having appendicitis) was confirmed based on the negative histopathologic findings from appendectomy specimen, gross surgical findings, and/or clinical follow-up, including telephone interview.

Mild or Early Acute Appendicitis

The histopathologic diagnosis of established appendicitis showing transmural infiltration of neutrophils is straightforward. However, diagnostic ambiguity may arise when an appendix, removed from a patient with the clinical presentation of acute appendicitis, shows only mild acute inflammation confined to the mucosa. The clinical significance of pure mucosal inflammation is uncertain, particularly when ulcer is absent (184).

Previous studies have reported that up to 35% of appendices removed incidentally show small collections of neutrophil polymorphs in the lumen, focal ulceration of the surface epithelium with pus cells in the adjacent lamina propria, and even a few crypt abscesses (185). On the other hand, studies of experimentally induced appendicitis have shown that identical mucosal lesions can progress rapidly to established acute appendicitis with gangrene and perforation (186). Confounding the issue further, one study (187) has demonstrated that a substantial proportion of histologically normal appendices from patients with a clinical diagnosis of acute appendicitis showed increased amounts of cytokines, sensitive markers of inflammation.

Appendiceal Diverticulitis

Cases of appendiceal diverticulitis were also counted as appendicitis since the distinction between the two diseases was not considered clinically important, and the two diseases commonly coexist (173, 188).

Cases of Delayed Appendectomy

As an exception, it was considered that appendicitis was present in LOCAT even if typical histopathologic findings were not present in participants who underwent delayed appendectomies, which was followed by percutaneous abscess drainage and/or medical treatment. In this case, the independent outcome assessors verified appendiceal perforation based on a medical record review.

Periappendicitis

Periappendicitis was defined as appendiceal serosal inflammation without

mucosal involvement. Common causes of periappendicitis are mechanical manipulation of the appendix during surgery, pelvic inflammatory disease, urologic disorders, and gastrointestinal perforation (189-191). Periappendicitis, without acute inflammation in the appendix wall, was differentiated from acute appendicitis.

Definition of Appendiceal Perforation

In LOCAT, the presence of appendiceal perforation was based on the spillage of the appendiceal contents, peritonitis, or abscess observed during the surgery (192), or pathologically confirmed appendiceal wall defect due to transmural necrosis. This definition of appendiceal perforation has been used in the previous single-center trial (29), and later further supported by an extensive literature review (43). Again, in cases undergoing delayed appendectomies following percutaneous abscess drainage and/or medical treatment, appendiceal perforation could exceptionally be considered present in LOCAT even if the typical surgical or pathologic findings were not present, only if the perforation was verified based on a medical record review by the outcome assessors.

Surprisingly, the definition of appendiceal perforation has been missing or inconsistent (43, 178) in many previous studies (9-12, 58, 193) addressing APR. Importantly, in the previous studies, it has been particularly unclear whether appendiceal perforation indicates only gross periappendiceal abscess or generalized peritonitis, or also includes micro-perforation with localized peritonitis of minimal extent which can be identifiable only by microscopic examination of the appendectomy specimen. While the former is associated with surgical approach or with patient prognosis, the latter is unlikely so (43). In LOCAT, the aforementioned broader definition was primarily used to cover both extreme types of perforations. A

narrower alternative definition of appendiceal perforation will be additionally used by including surgically-identified perforations but excluding pathologically-identified perforations (43). The results of this alternative analysis will be reported separately from the report of the main LOCAT results.

It is unlikely that any of surgeons' inspection, gross pathologic examination of appendectomy specimen, or microscopic examination can serve as a sole reference standard for the presence of appendiceal perforation. Surgeons' inspection or gross pathologic examination is sometimes limited (182) in cases with extensive inflammation (194). Pus can be observed on the serosal surface, even in cases without appendiceal perforation (194). On the contrary, microscopic examination may miss tiny perforation if the area of perforation fails to be included in the tissue preparation.

Reporting AEs

AE reporting policy in the LOCAT protocol was modified from that in the American College of Radiology Imaging Network AE Reporting Manual for Commercial (Non-IND [Investigational New Drug]) Imaging Agent Trials (195). The policy defined in the LOCAT protocol covered only AE reporting to the LOCAT office. Each site lead investigator had to be aware of and abide by the policy imposed by the site institutional review board (IRB), in addition to the policy defined in the LOCAT protocol. Individual site IRBs could have different AE reporting policies. Any site lead investigators or research coordinators having questions on AE reporting in LOCAT could be able to contact the LOCAT office, which forwarded the issues to the AE adjudication committee.

As virtually all diagnostic and therapeutic procedures in LOCAT, except for the

use of the 2-mSv CT technique, followed the standard-of-care at each site, clinically and scientifically meaningful AEs were anticipated to be few and minor. On the contrary, participants with various abdominal diseases were enrolled, and therefore, a great diversity of AEs was expected during many downstream diagnostic and therapeutic pathways following initial CT examinations. Therefore, AEs reportable to the LOCAT office were limited to unexpected serious adverse events (SAEs), so that the AE reporting could be truly meaningful and feasible.

As LOCAT was a diagnostic trial, the eligibility criteria were inevitably broad, and a great variety of events could occur during many downstream diagnostic and therapeutic pathways following the initial CT examinations. No published guideline or consensus statement was available at the time of the study as to how AEs should be reported in a diagnostic trial. It was unavoidable that some subjectivity is involved in the definition or characterization of AEs in some cases in LOCAT. To minimize the subjectivity and to ensure the consistency in reporting and characterizing AEs, the independent AE adjudication committee was established to help the site lead investigators' judgments. The AE adjudication committee was blinded to assigned group while the site lead investigators could not be.

Definition of AE

An AE is the development or worsening of any symptom, sign, illness, or experience that is temporally associated with study procedures, regardless of causality. These include events that occur as a result of study procedures (including the use of iodinated contrast materials) or preexisting medical conditions (including appendicitis), and that are judged by the Investigators to have emerged or worsened in severity or frequency during the AE reporting period. The AE reporting period for

a participant is defined as the period from the initiation of any study procedures and up to the last follow-up. The term of AE was recorded according to the Common Terminology Criteria for AEs (CTCAE) 4.0 (89).

Definition of SAE

SAE was defined as any untoward medical occurrence that:

- Resulted in death
- Was life-threatening (at the time of the event)
- Required inpatient hospitalization or prolongation of an existing hospitalization[*†]
- Resulted in persistent or significant disability or incapacity, or
- Resulted in congenital anomalies/birth defects.

[*]Hospitalization was defined as medically or surgically required one so that the term could be reserved for situations in which the AE truly fitted the definition. For example, a hospitalization is not reportable when it is for diagnostic or elective surgical procedures for a preexisting condition and the outcome is uneventful (e.g., uneventful negative appendectomy). On the contrary, if abdominal pain persisted after the first hospital discharge and led to another hospitalization for appendectomy, the event was regarded as a reportable AE.

[†]A hospital stay over 7 days following nonincidental appendectomy was regarded as a prolongation of hospitalization.

AE Characteristics

Site lead investigators were responsible for the characterization of each AE according to the following guideline.

Grade

Grade was used to denote the severity of the AE, according to the CTCAE 4.0 (89).

Expected/Unexpected AEs

In LOCAT, participants with various abdominal diseases were enrolled, and therefore, a great variety of AEs was expected during many downstream diagnostic and therapeutic pathways following initial CT examinations. Therefore, it was inevitable to use a broad definition for expected AE.

- An expected AE was defined as one that was consistent with the natural course of management of a participant with a given suspected or established diagnosis.
- An unexpected AE was defined as one that was considered to occur rarely given the medical context.

The judgment was made by the site lead investigator in conjunction with the AE adjudication committee on a sound medical and scientific basis, assuming the best treatment in the absence of co-morbidities. For example, if a final diagnosis of colitis had been established in a participant during the treatment course, the diagnosis of colitis was regarded as an expected event. If vomiting developed during the hospital course in a participant who had been eventually diagnosed to have adhesive ileus, vomiting was considered as an expected event.

Attribution

- Attributable – The AE is clearly or probably related to the study procedure.

- Not attributable – The AE is clearly or probably not related to the study procedure.

Individual Symptoms vs. Single Diagnosis

If a disease was known or suspected at the time an AE was reported, this diagnosis had to be recorded on the Reportable AE Form rather than listing individual symptoms. However, if a cluster of symptoms could not be identified as a single diagnosis, each individual event was reported separately. If a diagnosis had been subsequently known, it was reported as follow-up information.

Who Should Report AEs

Complete, accurate, and prompt reporting of AEs is the legal and ethical responsibility of the site lead investigator, assisted by the research coordinators as appropriate. Active surveillance of reportable AEs was performed through the structured telephone interview of the participants or proxies by the site research coordinators 3 months after the randomization. Also, at each contact with the participants, including site visits, information on AEs were elicited, and if indicated, the participant was evaluated clinically. The site lead investigator, with the help of the AE adjudication committee, assigned the grade, attribution, and expectedness for each AE.

How to Report AEs

For all reportable AEs, the site lead investigator reported the AE within 7 calendar days of learning of the event. The reporting was defined as the submission of a completed Reportable AE Form to the LOCAT office according to the

instructions specified in the Reportable AE Form. The LOCAT office forwarded the information to the safety officer in the DSMB.

Follow-up for AEs

The clinical course of each reportable AE was followed by the site lead investigator until resolution or stabilization, or until it had been determined that the study procedure or participation not be the cause. AEs that had been still ongoing at the end of the AE reporting period were further followed up to determine their outcome.

Ethical Considerations

Through the course of the design, conduct, data analysis, and results reporting in LOCAT, the LOCAT investigators executed special rigorous efforts to justify the randomized controlled trial. The considerations included but were not limited to; what is the most appropriate questions to ask, who are ethically eligible to be randomized, what are the most ethical comparisons to make, and how and when the participants should be randomized.

Ethics and Responsibility

LOCAT was conducted in compliance with the LOCAT protocol, the site's standard operating procedures, the site IRB regulations, the International Conference on Harmonisation (ICH) Good Clinical Practice (GCP) guidelines, the Declaration of Helsinki, and applicable government regulations. The LOCAT protocol, Informed Consent Form, and any amendments of them were submitted to the site IRBs for formal approval of the study conduct. Site lead investigator submitted a copy of the

initial IRB approval letter to the LOCAT office before the registration of the first participant. Site lead investigator also provided copies of the IRB approval letters for any amendments and annual renewals.

Informed Consent Form

All participants were given an IRB-approved Informed Consent Form describing LOCAT and providing sufficient information for participants to make informed decisions about their participation in LOCAT. The Informed Consent Form was signed and dated by the participant or a legally acceptable representative. The site lead investigator or a designated research staff obtained the consent before the participant was subjected to any study procedures. Any revisions to the Informed Consent Form at any time during the trial was submitted to each site IRB for approval.

Data Security and Participant Confidentiality

Enrolled participants were assigned an Enrollment Number. The link to confidential participant information was stored securely and separately from the study data. Only the independent outcome assessors who adjudicated the final diagnosis were allowed to access the link. All personal information and study data were kept confidential. A password was required to access computer data file to prevent unauthorized access to confidential participant information.

Early Stopping Rules in LOCAT

The LOCAT protocol did not have any stopping rule for futility in proving the noninferiority of 2-mSv CT to CDCT. Instead, the stopping rule of LOCAT was defined in terms of AEs. Monitoring of AEs was conducted by the data center

research associates. If two or more SAEs with a severity of CTCAE 4.0 (89) Grade 4 (life-threatening) or 5 (death) were reported in any of the two groups, and then the DSMB urgently investigated if the events were attributable to study procedures and recommended whether LOCAT should be terminated early or not. Participant enrollment had been suspended until a formal recommendation from the DSMB Chair was received.

Data Management

Site lead investigators were obliged to provide the data center with complete study data in accordance with the LOCAT protocol. Data collection at each site occurred through the submission of completed case report forms (CRFs) and images to the data center according to the timelines and instructions specified in the CRFs and LOCAT protocol. The site lead investigator was responsible for the completeness, accuracy, and timeliness of the data submission. The data center periodically prompted site lead investigators and research coordinators for the submission of necessary data. After a CRF had been submitted to the data center, no direct revision of the data was allowed. In the unlikely event that correction of data was needed, the site lead investigator was required to submit a revised CRF to the data center and a written explanation. It was site lead investigator's responsibility to confirm the timely receipt of the completed CRFs by the data center. The site lead investigator had retained all LOCAT-related documents until the LOCAT office informed the site lead investigator that the documents were no longer needed.

All data received from each site were electronically stamped with the date and time of receipt. Trained research associates at the data center performed extensive data checks for accuracy and completeness. If missing or problematic data were

detected, the data center research associates contacted the site lead investigator and research coordinators to specify the problem, and to request clarification.

The collected data were kept in the central data archive of the data center. The data archive had a built-in security feature preventing unauthorized access to confidential participant information. Access to the system was controlled by a sequence of identification codes and passwords. The data was not available to the Investigators until the completion or early termination of LOCAT.

The database was locked on April 1, 2017.

Case Report Forms

LOCAT used electronic CRFs which could be accessed at www.locat.org. All data requested on the CRFs were recorded, and any missing data was explained. Trial statistician (Yousun Ko) generated Data Dictionary (**Appendix 1**) containing more than 200 variables and their descriptions.

Monitoring Participant Accrual

The data center monitored participant accrual and regularly reported the accrual status, participant characteristics, and protocol compliance to the LOCAT office and DSMB. The accrual goal was 20 participants per site for every month. If the target had not been reached, the data center, coordinating committee, and site lead investigator would conduct a review with the intention of discovering and resolving any recruitment barriers.

Monitoring Data Quality

Site lead investigators permitted monitoring, auditing and inspection by

providing direct access to source data and documents. Data quality was monitored according to the LOCAT protocol and ICH GCP Guidelines. The data center performed periodic site monitoring through site visits. The primary goal of this monitoring was to assess overall compliance with the LOCAT protocol and to detect an unforeseen difference in study procedures among sites. If patterns had been discovered in the data that appeared to arise from causes specific to a site, the data center apprised the site lead investigator and coordinating committee and worked with the site lead investigator and coordinating committee until the problem was resolved.

The monitoring was implemented after the registration of the first participant at each site. The data center informed the site lead investigator when the monitoring was implemented. To help sites prepare for the monitoring, the data center sent monitoring instructions to the site lead investigator beforehand and offered continuous training to sites.

Each session of the site monitoring required the presence of the site lead investigator. The site lead investigator allocated adequate time and space for the monitoring activities and allowed access to all study-related documents and facilities. The research associates from the data center reviewed source documents (e.g., medical record, Patient Questionnaire, DICOM image dataset, and telephone interview sheet) against the submitted CRFs for all cases accrued. If an item had not been mentioned in the source document, it was assumed as not present. Major discrepancies were reported to the coordinating committee, site lead investigator, DSMB, and site IRB. Initial and revised regulatory documents were also monitored.

Data and Safety Monitoring Board

The DSMB was responsible for the review of data and identification of any potential safety issues. All members of the DSMB were entirely independent of LOCAT and had no financial, scientific, or other conflict of interest with LOCAT. Collaborators or associates of the coordinating committee members were not eligible to serve on the DSMB. All members of DSMB declared any conflicts of interest should they arise. In regard to the continuation or termination of the study along with any concerns of the DSMB on the participant safety, each DSMB meeting created a recommendation report that was to be decided by the formal DSMB majority or unanimous vote. Once approved by the DSMB members, the DSMB chair forwarded the formal recommendation to the coordinating committee. It was the responsibility of the coordinating committee to distribute the recommendations to all site lead investigators and to ensure the copies were submitted to all the IRBs associated with the study. Details are available in Data and Safety Monitoring Board Charter.

Statistical Analysis

The primary investigator of the trial and the trial statistician (Yousun Ko) planned all analyses and the statistician conducted all analyses.

Considerations for Primary Endpoint

The primary endpoint was NAR. Secondary endpoints included APR, the proportion of the participants who required the additional imaging test(s) to diagnose or rule out appendicitis, delay in patient disposition, and diagnostic performance of the CT reports.

Of these endpoints, NAR and APR were two most important established

reciprocal measures of quality of care (40) in the diagnosis of appendicitis by representing consequences of false-positive diagnosis and delayed diagnosis, respectively. If LOCAT could prove the noninferiority of 2-mSv CT to CDCT regarding APR as well as NAR, the LOCAT results were more conclusive than only proving the noninferiority for NAR, in establishing 2-mSv CT as the first-line imaging test. Therefore, the protocol development and revision committee extensively discussed the need for APR in addition to NAR as a co-primary endpoint, however, have finally decided not to include APR in co-primary endpoints for several reasons. First, ambiguity exists in defining appendiceal perforation (43, 178), which may partly explain the wide variation in reported APRs in previous studies (134) and across the sites in a retrospective study (4). Second, in contrast to NAR explicitly indicating the clinical consequence of false-positive diagnosis, APR is not so directly linked with false-negative diagnosis, as APR is also affected by many other factors (48, 177). Third, the presence of appendiceal perforation, especially in a mild form, would not always affect clinical outcomes (43). Therefore, LOCAT had a single primary endpoint of NAR.

Nevertheless, noninferiority testing was performed for APR as well as NAR. While testing two different hypotheses simultaneously (one for NAR and the other for APR) generally required the control of the statistical false positive rate (or Type I error, α), we used a hierarchical approach that enabled the testing of ordered hypotheses without the need for the α adjustment (196). By performing the two statistical tests according to a pre-specified hierarchical strategy, the noninferiority for APR provided an additional basis supporting the use of 2-mSv CT as the first-line imaging test. Therefore, as we set overall study α as 0.05, each of the two noninferiority hypotheses was tested at the full α level of 0.05. This fixed-sequence

testing allowed the noninferiority hypothesis for APR to be tested only if the noninferiority for NAR was established first.

Analysis Plans

Data analyses were performed by the data center led by Yousun Ko.

All participants undergoing randomization were included in the analysis in the groups to which they were originally assigned (intention-to-treat). Additional per-protocol analysis could be used in case it was needed. Although per-protocol analysis was generally preferred in a noninferiority trial due to the possibility that protocol non-adherence could bias study results toward noninferiority (197), we chose to use the intention-to-treat analysis primarily for the following reason. The motivation behind LOCAT was to hopefully replace CDCT with 2-mSv CT as the first-line imaging test. In other words, LOCAT was intended to compare the two diagnostic pathways including each physicians' clinical assessment and final clinical judgment based on the integration of all available diagnostic information such as the additional imaging test results as well as the initial CT results. This comparison was different from comparing 2-mSv CT and CDCT in a simple test-to-test manner.

According to the intention-to-treat principle, participants not undergoing the CT examination originally assigned (protocol non-adherence) were included in the analysis in the groups to which they were originally assigned. Presumably, drop-outs could occur more in participants not undergoing surgery than in participants undergoing surgery. Participants not undergoing surgery and then lost to follow-up were not counted in either numerator or denominator in calculating the NAR, according to the definition. Otherwise, missing data for the primary endpoint due to non-retention or protocol non-adherence were expected to be very rare.

The NARs in both groups and the two-sided 95% confidence interval (CI) for the differences were calculated. The noninferiority of 2-mSv CT to the CDCT was accepted if the upper bound of the two-sided 95% CI lied below the prespecified noninferiority margin, 4.5 percentage points.

The same noninferiority analysis was performed for APR with a prespecified noninferiority margin of 10.0 percentage points. Chi-square tests or Fisher's exact tests, Mann–Whitney U tests, and receiver-operating-characteristic analysis (nonparametric Wilcoxon statistic) were used in comparing the other secondary endpoints. A two-sided *P* value less than 0.05 indicated statistical significance.

For the NAR and APR, numbers-needed-to-treat was calculated as the reciprocals of the measured absolute differences between the 2-mSv CT and CDCT groups.

If the study results showed considerable variations across the sites, generalized estimating equations could be used to account for clustering effect by site.

Reportable AEs were tabulated with intention-to-treat and per-protocol manners.

Sample Size

The rationale for sample size determination is as follows.

Sample Size Considerations

Study sample size was determined to provide 90% power for the noninferiority test with respect to NAR. Sample size was computed using PASS version 11.0 (PASS; NCSS, Kaysville, Utah).

We assumed 4% NAR following CDCT based on the previous data from one or more of the sites (4, 27, 29, 198). The same NAR was assumed following 2-mSv CT.

We judged 8.5% NAR to be clinically acceptable following 2-mSv CT, which corresponded to a noninferiority margin of 4.5 percentage points, considering the potential reduction in carcinogenic risk associated with CT. With these assumptions, 399 nonincidental appendectomies per group were needed to obtain 90% statistical power with a two-sided α equal to 0.05 **(Table 9**) according to the following equation:

NARLD: NAR in the 2-mSv CT group, NARCD: is NAR in the CDCT group, nLD: number of nonincidental appendectomies in the 2-mSv CT group, nCD: number of nonincidental appendectomies in the conventional-dose CT group

$\Delta = NAR_{LD} - NAR_{CD}$; $\Delta > 0$ implies CDCT is better than 2-mSv CT.

H_0: $\Delta \geq \Delta_0$ (CDCT is better by at least Δ_0) vs. H_1: $\Delta < \Delta_0$ (2-mSv CT is not worse by as much as Δ_0)

$$n_{LD} = n_{CD} = \left(NAR_{LD}(1 - NAR_{LD}) + NAR_{CD}(1 - NAR_{CD})\right) \frac{\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{(\Delta_0 - (NAR_{LD} - NAR_{CD})^2}$$

$$= 2 * 0.04 * 0.96 \frac{\left(z_{1-\frac{0.05}{2}} + z_{0.9}\right)^2}{0.045^2}$$

Table 9. Sample size simulation for NAR

NAR	Noninferiority margin	Nonincidental appendectomies per group
3%	4.0%	383
	4.5%	302
	5.0%	245
3.5%	4.0%	444
	4.5%	351
	5.0%	284
4%	4.0%	505
	4.5%	**399**
	5.0%	323

Note.—The number of nonincidental appendectomies per group required to obtain 90% statistical power with a two-sided Type I error equal to 0.05.

Final Sample Size

To achieve a power of 90% for NAR, 399 nonincidental appendectomies per group were required. The sample size was inflated to 444 nonincidental appendectomies per group to account for a 10% drop-out rate. Therefore, participant recruitment was continued until the number of nonincidental appendectomies per group exceeds 444.

It should be noted that participants who did not undergo appendectomy during the study period were also included in LOCAT, although the required sample size was determined in terms of the number of appendectomies. Given that appendectomy was eventually performed in 40–44% of the patients undergoing appendiceal CT in one of the sites (27-29), we assumed that appendectomy would be finally performed in at least 30% of all enrolled participants, considering the variability across the sites. With this assumption, the expected total number of participants enrolled in LOCAT approximated 3,000.

Rationale for the Noninferiority Margin

We judged 8.5% NAR to be clinically acceptable following 2-mSv CT, which corresponded to a noninferiority margin of 4.5%. To justify the noninferiority margin, we summarized the previously reported NARs as follows.

Reported NARs Following Preoperative CT

Several previous studies have estimated a decrease in NAR from 12%–29% to 3%–11% with the introduction of preoperative CT as reviewed by Coursey *et al*. (7) According to more recent large studies, which have likely used modern CT scanners with radiation doses presumably similar to the conventional dose in LOCAT, the reported NAR was:

- 6% of 3,540 appendectomies in 15 hospitals in Washington State in 2006–2007 (86% of the patients underwent preoperative CT or ultrasonography) (12)
- 8.2% of 232 appendectomies in an urban university hospital in North Carolina in 2006–2007 (93% of the patients underwent preoperative CT) (7)
- 3.0% of 233 appendectomies in a tertiary center in Massachusetts in 2006-2007 (97% of the patients underwent preoperative CT) (6)
- 7.5% of 716 appendectomies following CT examinations in a tertiary center in Wisconsin in 2000-2009 (58)
- 3.2% of 186 appendectomies in a tertiary center (one of the sites) in metropolitan Seoul in Korea in 2009-2011 (29)
- 5.4% of 19,327 appendectomies in 55 hospitals in Washington State in 2006–2011 (91.3% of the patients underwent preoperative CT, ultrasonography, or magnetic resonance imaging) (5)

  - 4.1% in a subgroup (the number of the patients is unclear from the report) who underwent preoperative CT
  - 4.6% in a subgroup (the number of the patients is unclear from the report) of adolescents and young adults (15-30 years in age) who underwent preoperative CT
- 3.9% of 2,320 appendectomies in 11 hospitals (including sites) in metropolitan Seoul in 2011 (99.7% of the patients underwent preoperative CT or ultrasonography) (4)
  - 3.9% in a subgroup (n = 1,395) of adolescents and young adults (15–44 years in age) who underwent preoperative CT as the first-line imaging test

Reported NARs in Patients Without Preoperative CT

To our knowledge, there has been no randomized controlled trial demonstrating the efficacy or effectiveness of the CDCT over placebo (no CT), which can be used as the basis for the statistical reasoning of the noninferiority margin (199). Instead, we summarized the reported NARs in patients without preoperative CT as follows.

- According to a meta-analysis (200),
  - 21.5% during pre-CT era as compared to 10.0% during CT era, from 10 original studies compared NARs between pre-CT (n = 4,485) and CT eras (n = 1,629).
  - 16.7% without preoperative CT as compared to 8.6% with preoperative CT, from 20 original studies that reported the NARs in patients who underwent clinical evaluation alone (n = 3,125) and those who underwent preoperative CT (n = 2,491).
- In a study including 19,327 appendectomies in 55 hospitals in Washington State

in 2006–2011 (5)

- 15.4% in a subgroup (approximately 168 patients) who did not undergo preoperative imaging
- 10.4% in a subgroup (the number of the patients is unclear from the report) who underwent preoperative ultrasonography
- 12% in a subgroup (the number of the patients is unclear from the report) of adolescents and young adults (15–30 years in age) who underwent preoperative ultrasonography

- In a study including 2,320 appendectomies in 11 hospitals (including sites) in metropolitan Seoul in 2011 (4)
  - 8.5% in a subgroup (n = 152) who underwent preoperative ultrasonography as the first-line imaging test

Sample Size Considerations on APR

According to a review by Birnbaum *et al*. (134), previously reported APR ranges 16%–39% with a median of 20%. For the noninferiority test regarding APR in LOCAT, we assumed 25% APR following CDCT based on the previous data (ranging from 23% to 31%) from one or more of the sites (4, 27-29). The same APR was assumed following 2-mSv CT. We judged 35% APR to be clinically acceptable following 2-mSv CT, which corresponded to a noninferiority margin of 10 percentage points, considering the potential reduction in carcinogenic risk associated with CT. With these assumptions, to obtain 90% statistical power with a two-sided α equal to 0.05, 395 cases of confirmed appendicitis per group were needed (**Table 10**). This corresponds to 412 nonincidental appendectomies per group by assuming 4% NAR.

The sample size determined to conclude the noninferiority in NAR, 444 nonincidental appendectomies per group, corresponded to 384 cases of confirmed appendicitis per group, with the assumption of a 10% drop-out rate and 4% NAR. With this sample size, the power to conclude noninferiority in APR was 89%.

Table 10. Sample size simulation for APR

APR	Noninferiority margin	Confirmed appendicitis per group
20%	10%	337
	15%	150
	20%	84
25%	**10%**	**395**
	15%	176
	20%	98
30%	10%	442
	15%	197
	20%	111

Note.—The number of confirmed appendicitis per group required to obtain 90% statistical power with a two-sided Type I error equal to 0.05.

Subgroup Analyses for APR and NAR

Subgroup analyses were performed to explore whether estimated NAR difference (and also APR difference) between the 2-mSv CT and CDCT groups vary significantly between subcategories of trial participants. The subgroup categories were defined in **Table 11**. Forest plots were generated to display the NAR (and APR) differences across the subgroups. Due to multiple comparison issues, no further formal hypothesis testing was done for the subgroup analyses.

Table 11. Definition of subgroup categories

Subgroups	2-mSv CT group	CDCT group	Difference in percentage points (95% CI)
Sex			
Female			
Male			
Body size			
Body mass index (kg/m^2)*			
< 18.5 (underweight)			
18.5–24.9 (normal)			
25.0–29.9 (overweight)			
30.0–34.9 (class I obesity)			
35.0–39.9 (class II obesity)			
≥ 40.0 (class III obesity)			
Effective diameter (cm)†			
< 20.0			
20.0–24.9			
25.0–29.9			
≥ 30.0			
Clinical risk score for appendicitis			
Alvarado score‡			
Low risk (0–4)			
Indeterminate risk (5–6)			
High risk (7–10)			
Appendicitis inflammatory response score§			
Low risk (0–4)			
Indeterminate risk (5–8)			
High risk (9–12)			
Time of CT examination			
Working hours‖			
After hours			
CT machine¶			
Radiologist who made initial CT report			
Attending radiologist			
Non-expert on-call radiologist (or trainee)			

Table 11. Definition of subgroup categories (continued)

Subgroups	2-mSv CT group	CDCT group	Difference in percentage points (95% CI)
Site[¶]			
LDCT experience in the previous trial			
Number of beds			
Annual number of appendectomies			

Note.—[*]The body-mass index is the weight in kilograms divided by the square of the height in meters. [†]The square root of the product of the anteroposterior diameter and lateral diameter of the abdomen, as measured on the transverse CT image at the umbilicus level (108). [‡]Categorized according to Alvarado (106). [§]Categorized according to Andersson *et al*. (107). [‖]8:00 AM to 5:00 PM on working days. [¶]Categories depending on participating sites.

Subgroup Analyses for Diagnostic Performance

In this dissertation research, we performed the subgroup analyses for diagnostic performance in the diagnosis of appendicitis. The primary investigator of the trial and the trial statistician (Yousun Ko) planned all analyses and the statistician conducted all analyses in line with published guidelines (35, 37).

We compared the baseline characteristics between the 2-mSv CT and CDCT groups using Fisher's exact tests and Mann-Whitney U tests. We measured the diagnostic performance for appendicitis in terms of sensitivity and specificity as well as AUC. In calculating sensitivities and specificities, we used a predefined decision threshold of the likelihood of appendicitis ≥ 3 as positive for appendicitis (39, 176). We have already reported elsewhere (30) the results of overall diagnostic performance using Fisher's exact test or Chi-square statistics.

In general, subgroup analysis is prone to multiple comparisons issue. Particularly *post hoc* analysis observations are regarded as unreliable unless they can be replicated (35). Therefore, we limited the tested subgroups to those predefined in

the trial protocol (39): patient sex, body size (i.e., body mass index [BMI] and effective diameter), clinical risk scores for appendicitis (i.e., Alvarado score and appendicitis inflammatory response score) (106, 107), time of CT examination (i.e., working hours vs. after hours), number of channels in the CT machines, radiologists' experience (i.e., attending vs. on-call radiologists), site experience in 2-mSv CT from the previous single-center trial (29), and site practice volume (i.e., number of beds and annual number of appendectomy).

We drew forest plots for the between-group differences across the subgroups. If the number of patients in any subgroup (combined for the 2-mSv CT and CDCT groups) was less than 200 in comparing either sensitivity or specificity, the results from the subgroup were considered not meaningful. In comparison, the denominators of previously reported sensitivities of LDCT for appendicitis in adults in retrospective studies (31-33, 74) have rarely exceeded 100 patients (**Table 1**).

We tested for additive and multiplicative interactions (37, 201) between each subgroup attribute and radiation dose (i.e., 2 mSv and conventional dose) on each of sensitivity and specificity. For additive interaction, we calculated the relative excess risk due to interaction (RERI). If the 95% CI of a RERI included 0, we considered there was no significant additive interaction. For multiplicative interaction, the null hypothesis was that the logistic regression coefficient for each treatment-by-subgroup product term was 0. If a *P* value could not be calculated due to a very small event rate, we used Firth's logistic regression (202). For testing interaction on sensitivity, we included the disease-positive patients (i.e., patients with confirmed appendicitis), and outcome of interest was true-positive diagnosis (i.e., the likelihood of appendicitis ≥ 3). For testing the interaction on specificity, we included the disease-negative patients (i.e., patients confirmed as not having appendicitis), and

outcome of interest was true-negative diagnosis (i.e., the likelihood of appendicitis < 3).

Unlike intention-to-treat analyses for clinical outcomes reported elsewhere (30), we opted for per-protocol principle because we were interested in comparing diagnostic performance between competently performed 2-mSv CT and CDCT. Intention-to-treat analyses may have led to an overestimation of diagnostic performance of 2-mSv CT relative to CDCT, particularly because there were patients assigned to the 2-mSv CT group but mistakenly underwent CDCT. Therefore, as we mentioned earlier, we excluded the patients who did not adhere to the trial protocol regarding radiation dose or eligibility. The intraclass correlation coefficients for site clustering were minimal (the intraclass correlation coefficient was 0.02 for sensitivity and < 0.01 for specificity) for the two groups combined, and therefore we did not consider the clustering effect. Missing data were rare and not included in the analyses. All statistical analyses were performed with Stata version 15.1 (StataCorp, College Station, TX, USA). A two-sided *P* value less than 0.05 indicated statistical significance. Since we repeated the statistical tests 44 times for additive interaction (for 22 subgroups for sensitivity and specificity, respectively) and 22 for multiplicative interaction (for 11 subgroup attributes for sensitivity and specificity, respectively), up to three additive interaction tests and two multiplicative interaction tests could show statistical significance on the basis of chance alone.

RESULTS

Patient Characteristics

The two groups were well balanced for most baseline characteristics (**Table 12**) and involved nearly the same care providers (**Table 6**). Among the 2,773 patients, 493 in the 2-mSv CT group and 540 in the CDCT group were determined to have appendicitis. The remaining patients were considered as not having appendicitis.

Table 12. Patient characteristics

Characteristic	2-mSv CT group (N = 1,392)	CDCT group (N = 1,381)	*P* value
Age (years)			0.26
15–24	516 (37.1%)	554 (40.1%)	
25–34	493 (35.4%)	464 (33.6%)	
35–44	383 (27.5%)	363 (26.3%)	
Sex			0.68
Female	767 (55.1%)	749 (54.2%)	
Male	625 (44.9%)	632 (45.8%)	
Ethnicity			0.004
Korean	1,389 (99.8%)	1,366 (98.9%)	
Non-Korean	3 (0.2%)	15 (1.1%)	
Body size			
Body mass index (kg/m^2)			0.56
< 18.5 (underweight)	138 (9.9%)	133 (9.6%)	
18.5–24.9 (normal)	958 (68.8%)	922 (66.8%)	
25.0–29.9 (overweight)	243 (17.5%)	265 (19.2%)	
≥ 30.0 (obese)	44 (3.2%)	54 (3.9%)	
Effective diameter (cm)*			0.55
< 20.0	209 (15.0%)	220 (15.9%)	
20.0–24.9	815 (58.5%)	773 (56.0%)	
25.0–29.9	326 (23.4%)	348 (25.2%)	
≥ 30.0	42 (3.0%)	40 (2.9%)	

Table 12. Patient characteristics (continued)

Characteristic	2-mSv CT group (N = 1,392)	CDCT group (N = 1,381)	*P* value
Chief complaint			0.75
Abdominal pain	1,313 (94.3%)	1,307 (94.6%)	
Nausea/vomiting	35 (2.5%)	31 (2.2%)	
Fever	25 (1.8%)	22 (1.6%)	
Others	19 (1.4%)	21 (1.5%)	
Duration of symptoms			0.30
≤ 12 hr	554 (39.8%)	560 (40.6%)	
13–24 hr	363 (26.1%)	392 (28.4%)	
2–3 days	350 (25.1%)	322 (23.3%)	
≥ 4 days	125 (9.0%)	107 (7.7%)	
Location of abdominal pain†			
Right lower quadrant	1,225 (88.0%)	1,217 (88.1%)	0.95
Suprapubic	203 (14.6%)	184 (13.3%)	0.35
Right flank	187 (13.4%)	165 (11.9%)	0.25
Periumbilical	161 (11.6%)	157 (11.4%)	0.91
Epigastric	143 (10.3%)	111 (8.0%)	0.048
Other area(s)	161 (11.6%)	123 (8.9%)	0.024
No pain	19 (1.4%)	31 (2.2%)	0.088
Migration of pain‡			0.62
Yes	425 (30.5%)	409 (29.6%)	
No	967 (69.5%)	972 (70.4%)	
Abdominal tenderness†			
Right lower quadrant	1,192 (85.6%)	1,182 (85.6%)	> 0.99
Epigastric	128 (9.2%)	134 (9.7%)	0.65
Left lower quadrant	114 (8.2%)	87 (6.3%)	0.057
Suprapubic	104 (7.5%)	103 (7.5%)	> 0.99
Periumbilical	96 (6.9%)	118 (8.5%)	0.117
Other area(s)	95 (6.8%)	81 (5.9%)	0.31
No tenderness	129 (9.3%)	127 (9.2%)	> 0.99
Rebound tenderness			0.048
Yes	576 (41.4%)	520 (37.7%)	
No	816 (58.6%)	861 (62.3%)	
Body temperature (°C)	36.8 (36.5–37.2)	36.8 (36.5–37.2)	0.99
Blood test results			
White blood cell (10^3/mm^3)	10.6 (7.8–13.6)	10.7 (8.1–14.0)	0.191
Segmented neutrophil (%)	74 (64–82)	75 (64–82)	0.66
C-reactive protein (mg/dL)	0.7 (0.2–3.3)	0.7 (0.1–3.4)	0.39

Table 12. Patient characteristics (continued)

Characteristic	2-mSv CT group (N = 1,392)	CDCT group (N = 1,381)	*P* value
Clinical risk scores for appendicitis			
Alvarado score			0.87
Low risk (0–4)	512 (36.8%)	528 (38.2%)	
Indeterminate risk (5–6)	443 (31.8%)	430 (31.1%)	
High risk (7–10)	430 (30.9%)	415 (30.1%)	
Appendicitis inflammatory response score			0.46
Low risk (0–4)	772 (55.5%)	763 (55.2%)	
Indeterminate risk (5–8)	585 (42.0%)	575 (41.6%)	
High risk (9–12)	16 (1.1%)	26 (1.9%)	
Time of CT examination			0.88
Working hours	602 (43.2%)	602 (43.6%)	
After hours	790 (56.8%)	779 (56.4%)	
CT machine			> 0.99
16-channel	286 (20.5%)	288 (20.9%)	
64-channel	362 (26.0%)	356 (25.8%)	
128-channel	484 (34.8%)	480 (34.8%)	
256- or 640-channel	260 (18.7%)	257 (18.6%)	
Target effective dose[§]			> 0.99
2 mSv vs. 3 mSv	18 (1.3%)	18 (1.3%)	
2 mSv vs. 5 mSv	32 (2.3%)	34 (2.5%)	
2 mSv vs. 6 mSv	335 (24.1%)	331 (24.0%)	
2 mSv vs. 7 mSv	491 (35.3%)	491 (35.6%)	
2 mSv vs. 8 mSv	516 (37.1%)	507 (36.7%)	
Individual radiation dose			
Dose-length product (mGy·cm)	131 (117–147)	481 (389–554)	NA
Volume CT dose index (mGy)	2.6 (2.2–2.7)	9.4 (7.6–10.4)	NA
Size-specific dose estimate (mGy)	4.0 (3.7–4.5)	14.4 (12.9–16.2)	NA
Iterative reconstruction			< 0.001
Used	557 (40.0%)	149 (10.8%)	
Not used	835 (60.0%)	1232 (89.2%)	
Radiologist who made initial CT report			0.91
Attending radiologist	781 (56.1%)	771 (55.8%)	
On-call radiologist or trainees	611 (43.9%)	610 (44.2%)	

Table 12. Patient characteristics (continued)

Characteristic	2-mSv CT group (N = 1,392)	CDCT group (N = 1,381)	*P* value
Site			
LDCT experience in the previous trial			0.86
Yes	153 (11.0%)	148 (10.7%)	
No	1,239 (89.0%)	1,233 (89.3%)	
Number of beds			0.95
< 650	344 (24.7%)	347 (25.1%)	
650–949	503 (36.1%)	492 (35.6%)	
≥ 950	545 (39.2%)	542 (39.2%)	
Annual number of appendectomies			0.96
< 150	51 (3.7%)	51 (3.7%)	
150–299	311 (22.3%)	299 (21.7%)	
300–449	440 (31.6%)	447 (32.4%)	
≥ 450	590 (42.4%)	584 (42.3%)	

Note.—Data are numbers (and percentages) or median numbers (and interquartile ranges). For each characteristic, there were missing data in less than 1.4% of the included patients. CDCT = conventional-dose CT, NA = not applicable. *The square root of the product of the anteroposterior diameter and lateral diameter of the abdomen, as measured on the transverse CT image at the umbilicus level. †Patients could fit into more than one category. ‡Defined as pain starting in the epigastrium or periumbilical area and migrating to the right lower quadrant in a few hours. §Although the effective dose was about 2 mSv for all CT machines, the target effective dose was individualized for each CT machine following the institutional normal dose.

Overall Diagnostic Performance

We have reported the overall between-group differences elsewhere (30). The sensitivity was 97.2% (479/493) in the 2-mSv CT group and 98.0% (529/540) in the CDCT group, showing the difference of -0.8 percentage points (95% CI, -2.7 to 1.1). The specificity was 95.8% (861/899) in the 2-mSv CT group and 94.2% (792/841) in the CDCT group, showing the difference of 1.6 percentage points (95% CI, -0.5 to 3.7). The AUC was 0.982 in the 2-mSv CT group and 0.986 in the CDCT group, showing the difference of -0.003 (95% CI, -0.013 to 0.006).

Subgroups of Limited Comparison

Despite the use of the large trial data, the 95% CIs particularly for sensitivity were wide due to small sizes (< 200) for the following subgroups: BMI less than 18.5 kg/m^2, BMI of 30.0 kg/m^2 or greater, effective diameter less than 20.0 cm, effective diameter of 30.0 cm or greater, high risk of appendicitis inflammatory response score, 256- or 640-channel CT machine, hospitals with 2-mSv CT experience in the previous trial, and hospitals with annual number of appendectomy less than 150. Any results from these subgroups were considered not meaningful and are not detailed hereinafter.

Between-group Differences for Subgroups

Otherwise, most of the subgroups showed trends similar to the overall results. For sensitivity and specificity as well as AUC, the 95% CIs for the between-group differences in most subgroups contained the overall between-group differences as well as null hypothesis value (i.e., 0) (**Figs. 5–7**). Exceptionally, the 95% CIs for the specificity difference skewed favoring 2-mSv CT, not covering the null hypothesis

value, for the subgroups with low risk of Alvarado score (between-group difference [95% CI], 3.5 [0.9 to 6.0] percentage points) and CT reports made by attending radiologist (2.7 [0.4 to 5.1] percentage points). For AUC, the 95% CIs for the between-group differences for all subgroups contained the null hypothesis value.

Figure 5. Forest plots for sensitivity.

RERI = relative excess risk due to interaction. *The 95% CI was considered wide as the sample size combined for the two groups was smaller than 200.

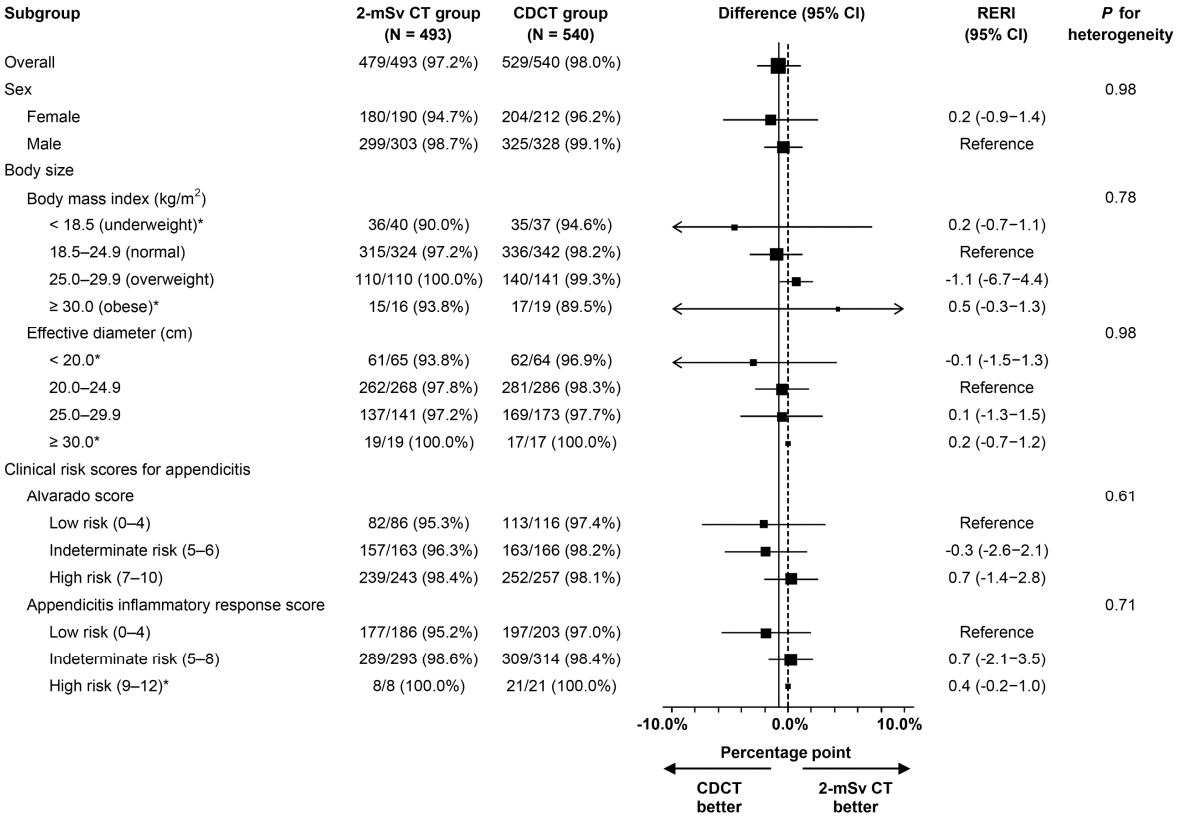


Figure 5. Forest plots for sensitivity (continued).

RERI = relative excess risk due to interaction. *The 95% CI was considered wide as the sample size combined for the two groups was smaller than 200.

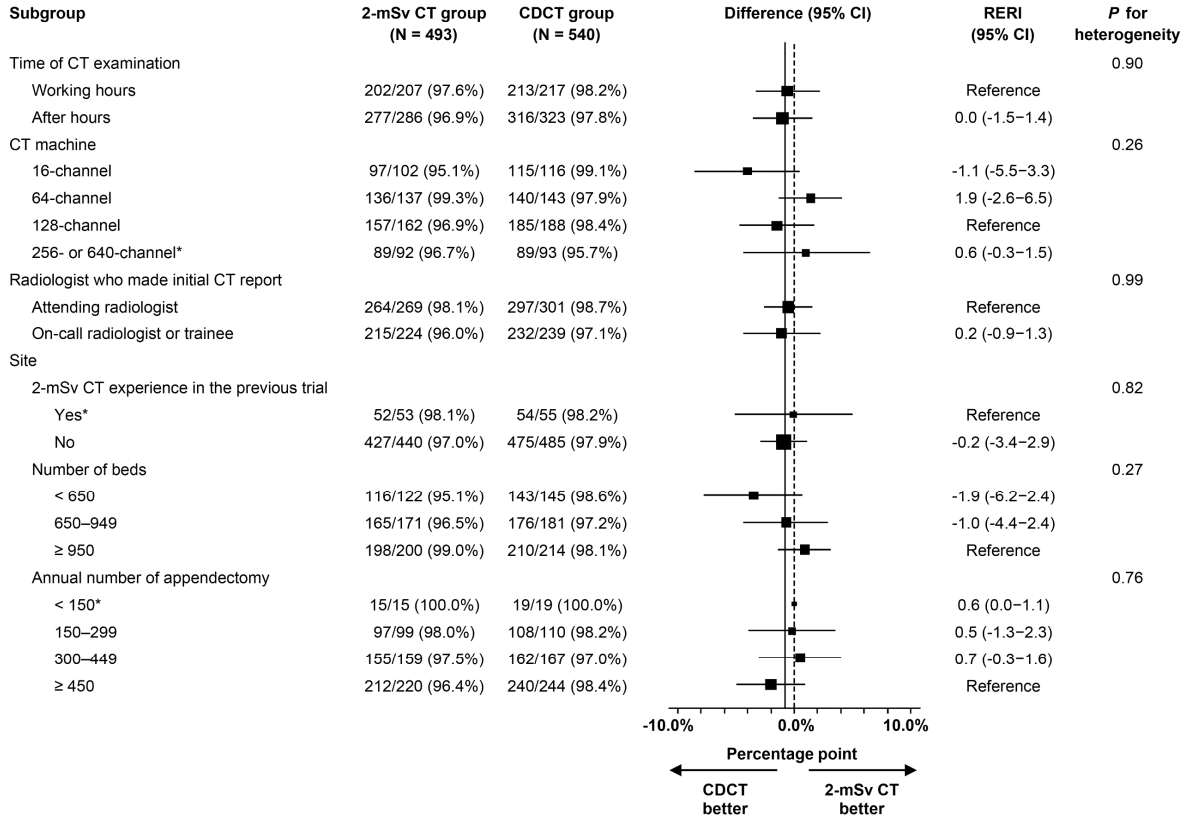


Figure 6. Forest plots for specificity.

RERI = relative excess risk due to interaction. *The 95% CI was considered wide as the sample size combined for the two groups was smaller than 200.

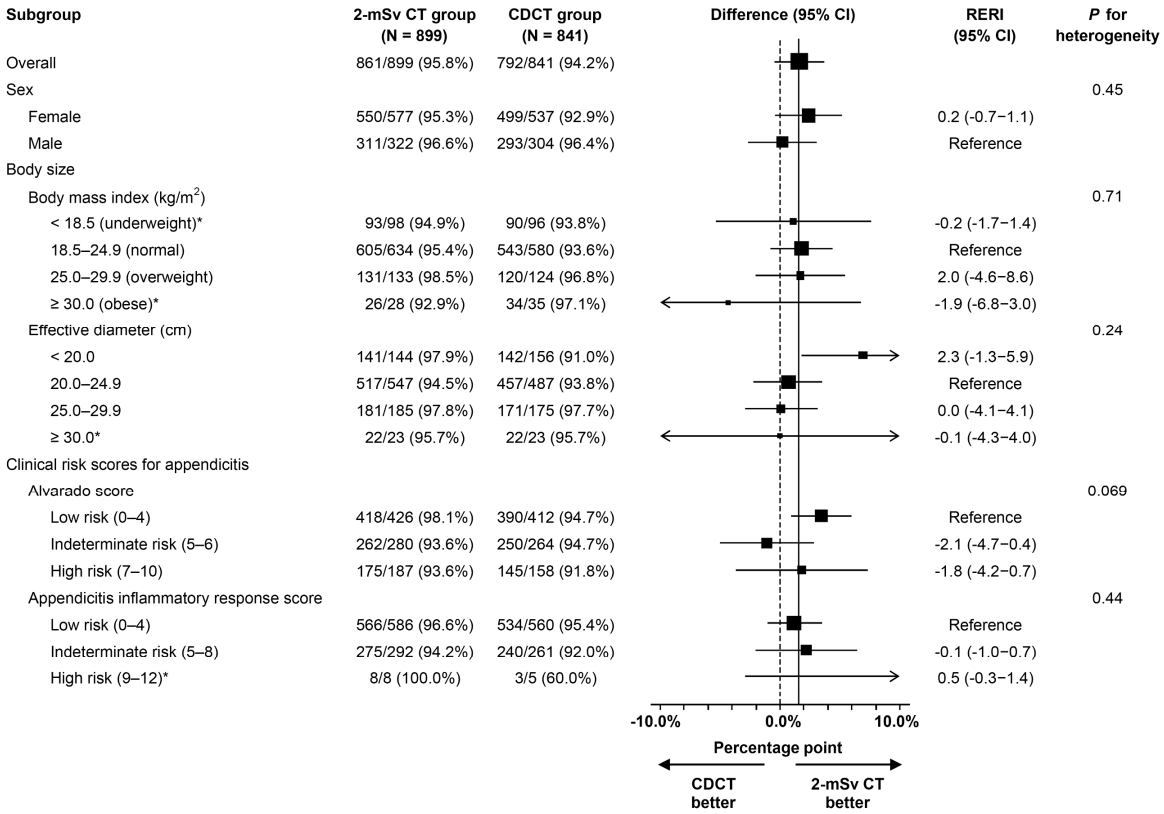


Figure 6. Forest plots for specificity (continued).

RERI = relative excess risk due to interaction. *The 95% CI was considered wide as the sample size combined for the two groups was smaller than 200.

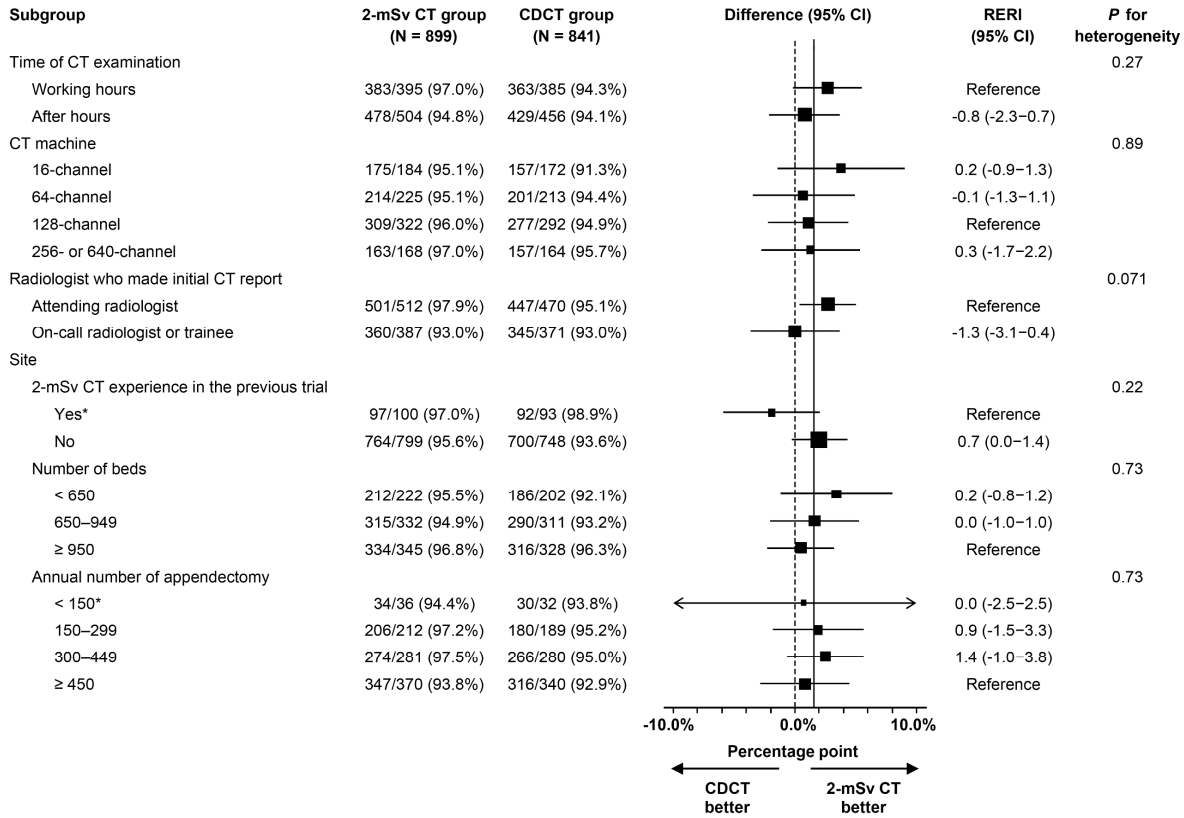


Figure 7. Forest plots for AUC.

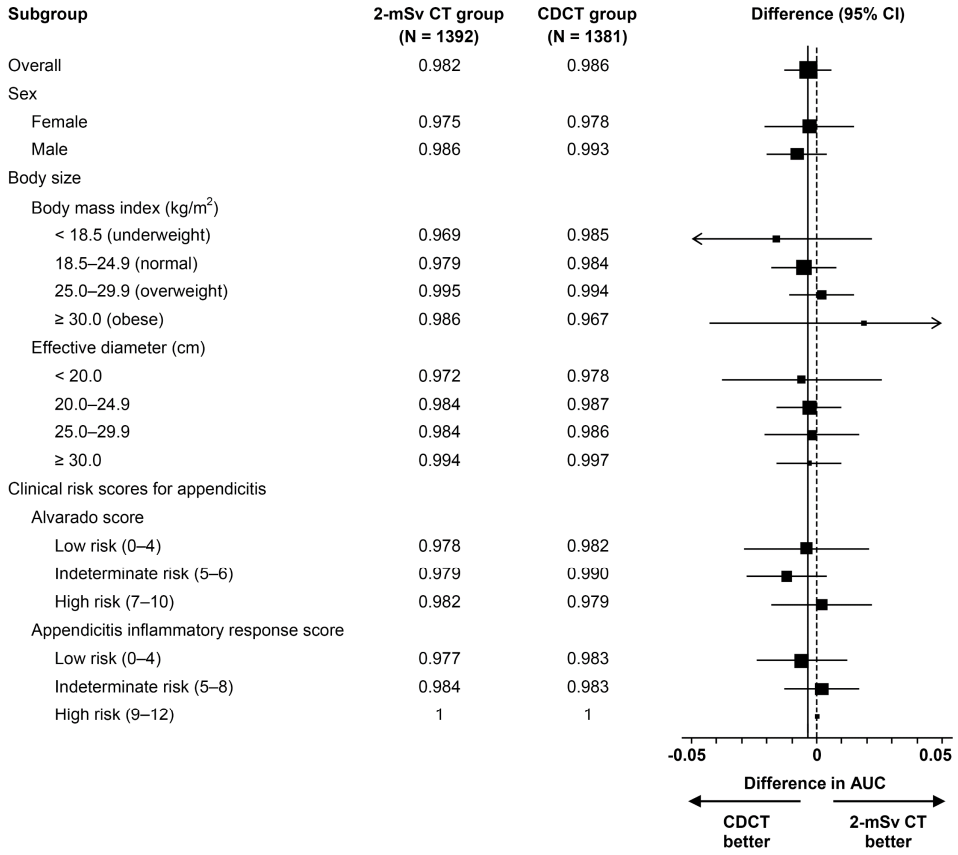
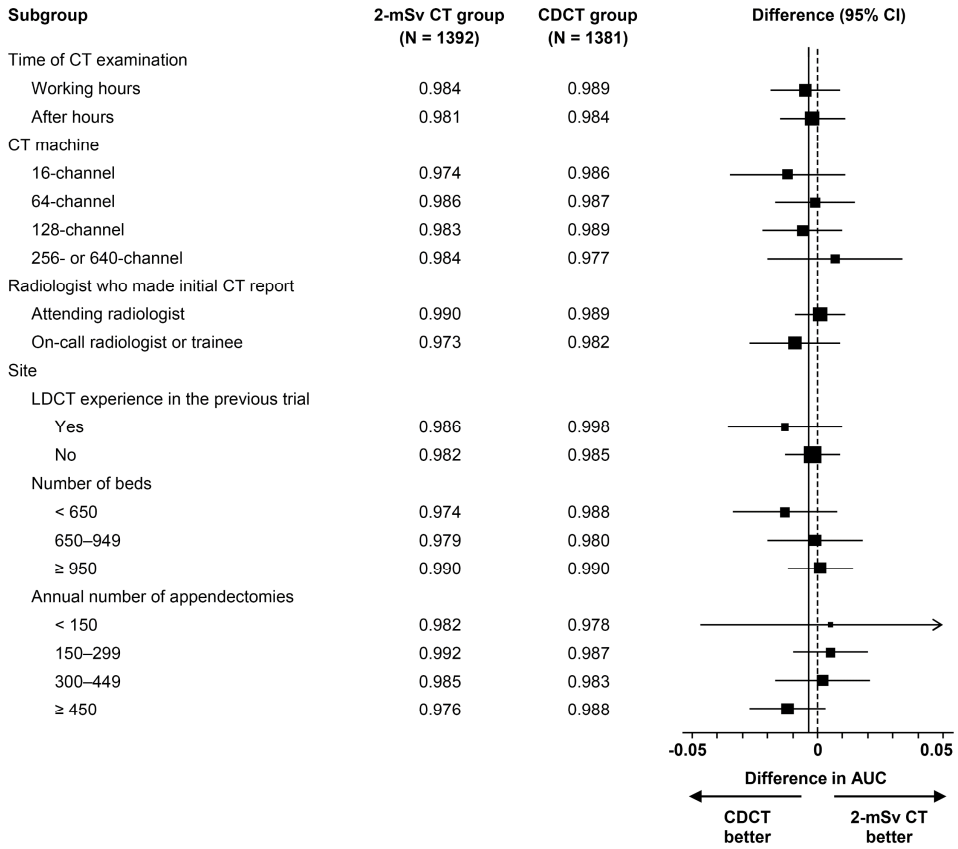


Figure 7. Forest plots for AUC (continued).



Heterogeneity

We did not find any significant additive or multiplicative treatment-by-subgroup interaction on each of sensitivity and specificity. For the tested subgroup attributes, the *P* value for multiplicative interaction was 0.26 or greater for sensitivity and 0.069 or greater for specificity. For all the subgroup attributes, the 95% CIs of RERIs included 0 for both sensitivity and specificity.

DISCUSSION

Summary of Results

We explored any heterogeneity across predefined subgroups in the diagnostic performance of 2-mSv CT relative to CDCT in the dataset of large the randomized controlled trial involving 2,773 patients and 160 radiologists from 20 hospitals. The 95% CIs for the differences between the 2-mSv CT and CDCT groups, particularly for sensitivity, were wide due to small sizes (< 200) for the subgroups of extreme body sizes, high risk of appendicitis inflammatory response score, newer CT machines, hospital with prior experience in 2-mSv CT, and hospitals with small appendectomy volume. Otherwise, sensitivity and specificity point estimates as well as their between-group differences were consistent across the subgroups. The 95% CIs for the between-group differences in most subgroups contained the previously reported overall between-group differences as well as null hypothesis value. There was no significant additive or multiplicative interaction on either sensitivity or specificity.

Clinical Implications of Study Results

As emphasized in guidelines on subgroup analysis (35, 146), caution is needed in interpreting our data. Scientific mistakes have been common in subgroup analysis of randomized control trial data (203). As Rothwell stated, unfounded clinical concerns about possible heterogeneity or inappropriately narrow indications for a new treatment would reduce the use of the effective treatment in routine practice (204). Importantly, it is inappropriate to test for any significance of between-group

difference in an individual subgroup. Instead, subgroup analysis has to focus on if the between-group difference differs significantly across the subgroups, which can be measured with the interaction tests (35).

Our results clearly show that the overall results of comparable diagnostic performance between the 2-mSv CT and CDCT groups were consistent across many subgroups. These results imply that 2-mSv CT can replace CDCT in diverse populations, mitigating unfounded concern that the use of 2-mSv CT may increase misdiagnosis in some particular patient subgroups (34). Importantly, the consistent results across the subgroups imply that the overall results of comparable diagnostic performance between the two groups are very unlikely attributable to a qualitative interaction.

Our results reinforce the rationale for adopting 2-mSv CT as a routine practice in most adolescents and young adults requiring appendiceal CT, who are our target population. Acute appendicitis is a very common illness. The use of CT is very popular in diagnosing appendicitis. Adolescents and young adults undergoing appendiceal CT typically have otherwise normal life expectancies. Previous large epidemiological studies (20, 55) showed that CT radiation is carcinogenic particularly to children and adolescents. In this target population, it is uncertain if the diagnostic advantage of appendiceal CT truly outweighs the potential radiation-associated carcinogenic risk (29). Based on linear-no-threshold assumption, Kim *et al*. (29) estimated that using 2-mSv instead of 8-mSv in appendiceal CT in estimated 2,000 patients of 30 years of age would prevent one case of cancer. If this is projected to 120,000 annual appendiceal CT examinations in a total population of 21 million people aged 15–44 years in Korea (17), using 2 mSv instead of 8 mSv would prevent 60 cancers each year (30).

Strengths of Our Data

Importantly, we used the data of the large multi-center pragmatic randomized controlled trial.

RCT Data

First, our data had enhanced between-group comparability owing to the randomization. To our knowledge, there have been ten published studies of head-to-head comparison between low-dose CT (1–4 mSv) and CDCT in adolescents and adults with suspected appendicitis (26-29, 72-74, 87, 88, 158). Only two of them were randomized trials conducted by our group, while the remaining eight studies were retrospective studies.

Multi-center Pragmatic Trial

Second, our data had enhanced generalizability as the trial setting was multi-center pragmatic. This is important because appendicitis is a very common disease encountered in nearly every hospital, often involving less experienced care providers. Most of our trial procedures including co-interventions followed daily practice in individual sites. All the sites but the lead site had little prior 2-mSv CT experience. Approximately 40% of the CT reports were made by radiology residents.

Large Size

Third, owing to the unprecedentedly large data size, most of the subgroups we tested could have reasonably large size. Our data size accounted for 1% of the estimated 320,000 appendiceal CT examinations from the total population aged 15–

44 years, and a quarter of the 74 teaching hospitals in Korea during the study period (30). The number of our patients is more than twice greater the total number of patients included in the eight previous retrospective studies. As we mentioned earlier, the denominators of the reported sensitivities in those studies (31-33, 74) have rarely exceeded 100 patients.

Subgroup Attributes

It would be worthwhile to discuss specific considerations on several subgroup attributes. Several studies have addressed that patient sex (7, 205) and body size (108, 206) may affect CT diagnosis of appendicitis.

Body Size

For patients with extreme body sizes, we were unable to draw a firm conclusion if 2-mSv CT can replace CDCT because the subgroups were small. We recognize the concern that the diagnostic performance of 2-mSv CT may be compromised in very large patients. However, the physical principle that image noise increases with increasing body size does not directly project to modern CT machines equipped with automatic exposure control that can keep consistent image quality across different body sizes. Large patients tend to have more intra-abdominal fat, which in fact helps to visualize the appendix on CT images (108). We are unaware of any published data suggesting that larger body size limits the performance of LDCT in the diagnosis of appendicitis or alternative diagnoses. Unfortunately, none of the previous studies on LDCT included a sufficient number of obese patients to answer the question.

On the other hand, smaller body size can arguably limit the performance of LDCT. Two small studies (73, 76) have suggested that 1–2-mSv CT may have

limited diagnostic sensitivity in slender patients. This finding was attributed to the fact that it is difficult to identify diseased or normal appendix at CT in patients with sparse pericecal fat. However, many other studies (26-29, 76) showed no notable effect of body mass index or pericecal fat amount (100) on the diagnosis of appendicitis at 1–4 mSv CT. Again, each of these studies had very small number of slender patients. Regardless of these data, ultrasonography instead of CT should be used primarily for such slender patients, as ultrasonography would be accurate and technically easy in these patients who generally have good sonic window.

Radiologist Experience

We found no notable heterogeneity across subgroups by radiologist experience (i.e., attending radiologists vs. on-call radiologists who were mostly residents), site experience in 2-mSv CT, or site practice volume (i.e., number of beds and annual number of appendectomy). Our results are in line with previous learning-curve (141) study showing little effect of radiologist experience level on the diagnostic performance at LDCT. Overall, these findings imply that 2-mSv CT can replace CDCT without specific education for radiologists inexperienced in 2-mSv CT.

Clinical Score for Appendicitis

We found no notable heterogeneity across subgroups by clinical scores for appendicitis. Importantly, in the subgroups with indeterminate risk scores, the 95% CIs for the between-group differences were reasonably narrow and still contained the null hypothesis value. These findings would justify the incorporation of 2-mSv CT instead of CDCT in a stepwise selective CT utilization approach (110-112) of limiting CT utilization to subgroups having intermediate clinical risk for appendicitis

while avoiding CT in low- or high-risk subgroups.

Study Limitations

Our study has several limitations.

Subgroups of Limited Comparison

First, our analyses were limited for the following subgroups of small sizes. For the subgroups of extreme body sizes, high clinical risk score for appendicitis, or hospitals with small appendectomy volume, further studies may be needed to confirm if 2 mSv CT can replace CDCT. However, for the subgroups of newer CT machines or prior site experience in 2-mSv CT, it would be reasonable to assume comparable diagnostic performance between 2-mSv CT and CDCT.

post hoc Analyses

Second, our study was *post hoc* subgroup analyses for secondary endpoints in the trial. Although most of the individual subgroups were larger than the sample sizes in the previous retrospective studies, Type II error in detecting across-subgroup heterogeneity may have occurred because LOCAT sample size was not determined for the purpose of our subgroup analyses. As mentioned earlier, we considered the results of a subgroups as not meaningful if the subgroup was smaller than 200 patients (combined for the 2-mSv CT and CDCT groups). On the other hand, type I error may also have occurred due to multiple testing (146), although we observed no significant subgroup heterogeneity.

Verification Bias

Third, our study may have been prone to verification bias, since appendectomy was performed selectively in patients with positive CT results for appendicitis. As we discussed elsewhere (30), we observed potential between-group imbalance in the number of appendectomies or confirmed appendicitis cases which may be attributable small number of spontaneously-resolving incipient appendicitis that were under-diagnosed at 2-mSv CT (or over-diagnosed at CDCT) (30). This imbalance may indicate that the verification biases might have occurred differently in the 2-mSv CT and CDCT groups.

Generalizability

Fourth, despite the large scale, the representativeness of our data was compromised to some extent, and therefore, whether our results can be generalized to hospitals worldwide remains uncertain. All the participating sites were teaching hospitals, mostly large, which could have had better resources or been motivated to use 2-mSv CT than non-participating hospitals. Only a third of the eligible patients were randomly assigned because of logistical reasons in the sites. The catchment area was limited to Korea, where extremely large body habitus is rare and CT is highly used for diagnosing appendicitis.

CONCLUSION

In conclusion, we found no notable subgroup heterogeneity in the comparable diagnostic performance between 2-mSv and CDCT in adolescents and young adults with suspected appendicitis. Our results mitigate unfounded concerns of misdiagnosis at 2-mSv CT in some patient subgroups and therefore reinforce the rationale for broadening indications for 2-mSv CT in routine practice. Further studies, however, are needed for the populations for which our subgroups were small.

Given the vast number of appendiceal CT examinations done worldwide, the use of 2-mSv CT instead of CDCT could prevent a sizeable number of radiation-associated cancers in the future. Based on linear-no-threshold assumption (29), it is estimated that using 2-mSv instead of 8-mSv in appendiceal CT in roughly 2,000 young adults (aged 15–44 years) will prevent one case of cancer. If this is projected to 120,000 annual appendiceal CT examinations in a total population of 21 million people aged 15–44 years in Korea (17), using 2 mSv instead of 8 mSv will prevent 60 cancers each year (30).

REFERENCES

1. Addiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. Am J Epidemiol 1990;132:910-925.

2. Terasawa T, Blackmore CC, Bent S, Kohlwes RJ. Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. Ann Intern Med 2004;141:537-546.

3. van Randen A, Bipat S, Zwinderman AH, Ubbink DT, Stoker J, Boermeester MA. Acute appendicitis: meta-analysis of diagnostic performance of CT and graded compression US related to prevalence of disease. Radiology 2008;249:97-106.

4. Park JH, for LOCAT Group. Diagnostic imaging utilization in cases of acute appendicitis: multi-center experience. J Korean Med Sci 2014;29:1308-1316.

5. Drake FT, Florence MG, Johnson MG, et al. Progress in the diagnosis of appendicitis: a report from Washington State's Surgical Care and Outcomes Assessment Program. Ann Surg 2012;256:586-594.

6. Raja AS, Wright C, Sodickson AD, et al. Negative appendectomy rate in the era of CT: an 18-year perspective. Radiology 2010;256:460-465.

7. Coursey CA, Nelson RC, Patel MB, et al. Making the diagnosis of acute appendicitis: do more preoperative CT scans mean fewer negative appendectomies? A 10-year study. Radiology 2010;254:460-468.

8. Paulson EK, Kalady MF, Pappas TN. Clinical practice. Suspected appendicitis. N Engl J Med 2003;348:236-242.

9. Flum DR, Morris A, Koepsell T, Dellinger EP. Has misdiagnosis of appendicitis decreased over time? A population-based analysis. JAMA 2001;286:1748-1753.

10. Raman SS, Osuagwu FC, Kadell B, Cryer H, Sayre J, Lu DS. Effect of CT on false positive diagnosis of appendicitis and perforation. N Engl J Med 2008;358:972-973.

11. Rao PM, Rhea JT, Rattner DW, Venus LG, Novelline RA. Introduction of appendiceal CT: impact on negative appendectomy and appendiceal perforation rates. Ann Surg 1999;229:344-349.

12. Cuschieri J, Florence M, Flum DR, et al. Negative appendectomy and imaging accuracy in the Washington State Surgical Care and Outcomes Assessment Program. Ann Surg 2008;248:557-563.

13. Rhea JT, Halpern EF, Ptak T, Lawrason JN, Sacknoff R, Novelline RA. The status of appendiceal CT in an urban medical center 5 years after its introduction: experience with 753 patients. Am J Roentgenol 2005;184:1802-1808.

14. Rao PM, Rhea JT, Novelline RA, Mostafavi AA, McCabe CJ. Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. N Engl J Med 1998;338:141-146.

15. Lahaye MJ, Lambregts DM, Mutsaers E, et al. Mandatory imaging cuts costs and reduces the rate of unnecessary surgeries in the diagnostic work-up of patients suspected of having appendicitis. Eur Radiol 2015;25:1464-1470.

16. National Hospital Discharge Survey, 2010. Number of all-listed procedures for discharges from short-stay hospitals, by procedure category and age: United States, 2010. CDC/NCHS. http://www.cdc.gov/nchs/data/nhds/4procedures/2010pro4_numberprocedureage. pdf. Published 2012. Accessed Mar 11, 2019.

17. 2017 Statistics on major surgeries in Korea. National Health Insurance Service. http://www.nhis.or.kr/menu/boardRetriveMenuSet.xx?menuId=F3326. Published 2018. Accessed Jun 09, 2019.

18. Budhraja V. Radiation exposure from medical imaging procedures. N Engl J Med 2009;361:2290-2291; author reply 2291-2292.

19. Karnath BM, Luh JY. Suspected appendicitis. N Engl J Med 2003;349:305-306; author reply 305-306.

20. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. Lancet 2012;380:499-505.

21. Smith-Bindman R, Lipson J, Marcus R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med 2009;169:2078-2086.

22. Brenner DJ, Hall EJ. Computed tomography-an increasing source of radiation exposure. N Engl J Med 2007;357:2277-2284.

23. Berrington de Gonzalez A, Mahesh M, Kim KP, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Arch Intern Med 2009;169:2071-2077.

24. Tsapaki V, Rehani M, Saini S. Radiation safety in abdominal computed tomography. Semin Ultrasound CT MR 2010;31:29-38.

25. Johnson PT, Horton KM, Mahesh M, Fishman EK. Multidetector computed tomography for suspected appendicitis: multi-institutional survey of 16-MDCT data acquisition protocols and review of pertinent literature. J Comput Assist Tomogr 2006;30:758-764.

26. Keyzer C, Tack D, de Maertelaer V, Bohy P, Gevenois PA, Van Gansbeke D. Acute appendicitis: comparison of low-dose and standard-dose unenhanced multi-detector row CT. Radiology 2004;232:164-172.

27. Kim SY, Lee KH, Kim K, et al. Acute appendicitis in young adults: low- versus standard-radiation-dose contrast-enhanced abdominal CT for diagnosis. Radiology 2011;260:437-445.

28. Seo H, Lee KH, Kim HJ, et al. Diagnosis of acute appendicitis with sliding slab ray-sum interpretation of low-dose unenhanced CT and standard-dose intravenous contrast-enhanced CT scans. Am J Roentgenol 2009;193:96-105.

29. Kim K, Kim YH, Kim SY, et al. Low-dose abdominal CT for evaluating suspected appendicitis. N Engl J Med 2012;366:1596-1605.

30. LOCAT Group. Low-dose CT for the diagnosis of appendicitis in adolescents and young adults (LOCAT): a pragmatic, multicentre, randomised controlled non-inferiority trial. Lancet Gastroenterol Hepatol 2017;2:793-804.

31. Aly NE, McAteer D, Aly EH. Low vs. standard dose computed tomography in suspected acute appendicitis: Is it time for a change? Int J Surg 2016;31:71-79.

32. Yoon HM, Suh CH, Cho YA, et al. The diagnostic performance of reduced-dose CT for suspected appendicitis in paediatric and adult patients: a systematic review and diagnostic meta-analysis. Eur Radiol 2018;28:2537-2548.

33. Yun SJ, Ryu CW, Choi NY, Kim HC, Oh JY, Yang DM. Comparison of low- and standard-dose CT for the diagnosis of acute appendicitis: a meta-analysis. Am J Roentgenol 2017;208:W198-W207.

34. Kim HJ, Lee KH, Kim MJ, Park SB, Ko Y. Changing standard-of-practice to low-dose CT in adolescents and young adults with suspected appendicitis: willingness survey for 579 referring physicians, surgeons, and radiologists from 20 hospitals. In submission.

35. Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet 2005;365:176-186.

36. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Ann Intern Med 2007;147:W163-194.

37. VanderWeele TJ, Knol MJ. A tutorial on interaction. Epidemiol Methods 2014;3:33-72.

38. Rothwell PM. Can overall results of clinical trials be applied to all patients? Lancet 1995;345:1616-1619.

39. Ahn S, LOCAT Group. LOCAT (low-dose computed tomography for appendicitis trial) comparing clinical outcomes following low- vs standard-dose computed tomography as the first-line imaging test in adolescents and young adults with suspected acute appendicitis: study protocol for a randomized controlled trial. Trials 2014;15:28.

40. Velanovich V, Satava R. Balancing the normal appendectomy rate with the perforated appendicitis rate: implications for quality assurance. Am Surg 1992;58:264-269.

41. Bhangu A, Søreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. Lancet 2015;386:1278-1287.

42. Andersson R, Hugander A, Thulin A, Nystrom PO, Olaison G. Indications for operation in suspected appendicitis and incidence of perforation. BMJ 1994;308:107-110.

43. Kim HJ, Kim MS, Park JH, et al. Meaningful standard of reference for appendiceal perforation: pathology, surgery, or both? Ann Surg Treat Res 2017;93:88-97.

44. Buckius MT, McGrath B, Monk J, Grim R, Bell T, Ahuja V. Changing epidemiology of acute appendicitis in the United States: study period 1993-2008. J Surg Res 2012;175:185-190.

45. Annual mid-year population estimates, 2013. Office for National Statistics. http://www.ons.gov.uk/ons/dcp171778_367167.pdf. Published 2014. Accessed Jun 11, 2019.

46. Anderson JE, Bickler SW, Chang DC, Talamini MA. Examining a common disease with unknown etiology: trends in epidemiology and surgical management of appendicitis in California, 1995-2009. World J Surg 2012;36:2787-2794.

47. Garcia EM, Camacho MA, Karolyi DR, et al. ACR Appropriateness Criteria® right lower quadrant pain-suspected appendicitis. J Am Coll Radiol 2018;15:S373-S387.

48. Livingston EH, Woodward WA, Sarosi GA, Haley RW. Disconnect between incidence of nonperforated and perforated appendicitis: implications for pathophysiology and management. Ann Surg 2007;245:886-892.

49. Doria AS, Moineddin R, Kellenberger CJ, et al. US or CT for diagnosis of appendicitis in children and adults? A meta-analysis. Radiology 2006;241:83-94.

50. Lee J, Ko Y, Ahn S, et al. Comparison of US and CT on the effect on negative appendectomy and appendiceal perforation in adolescents and adults: a post-hoc analysis using propensity-score methods. J Clin Ultrasound 2016;44:401-410.

51. Toorenvliet BR, Wiersma F, Bakker RF, Merkus JW, Breslau PJ, Hamming JF. Routine ultrasound and limited computed tomography for the diagnosis of acute appendicitis. World J Surg 2010;34:2278-2285.

52. National Surgical Research Collaborative. Multicentre observational study of performance variation in provision and outcome of emergency appendicectomy. Br J Surg 2013;100:1240-1252.

53. Jaunoo SS, Hale AL, Masters JP, Jaunoo SR. An international survey of opinion regarding investigation of possible appendicitis and laparoscopic management of a macroscopically normal appendix. Ann R Coll Surg Engl 2012;94:476-480.

54. Duke E, Kalb B, Arif-Tiwari H, et al. A systematic review and meta-analysis of diagnostic performance of MRI for evaluation of acute appendicitis. Am J Roentgenol 2016;206:508-517.

55. Mathews JD, Forsythe AV, Brady Z, et al. Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. BMJ 2013;346:f2360.

56. Krille L, Dreger S, Schindel R, et al. Risk of cancer incidence before the age of 15 years after exposure to ionising radiation from computed tomography: results from a German cohort study. Radiat Environ Biophys 2015;54:1-12.

57. Huang WY, Muo CH, Lin CY, et al. Paediatric head CT scan and subsequent risk of malignancy and benign brain tumour: a nation-wide population-based cohort study. Br J Cancer 2014;110:2354-2360.

58. Pickhardt PJ, Lawrence EM, Pooler BD, Bruce RJ. Diagnostic performance of multidetector computed tomography for suspected acute appendicitis. Ann Intern Med 2011;154:789-796.

59. Bankier AA, Kressel HY. Through the looking glass revisited: the need for more meaning and less drama in the reporting of dose and dose reduction in CT. Radiology 2012;265:4-8.

60. Mettler FA, Jr., Huda W, Yoshizumi TT, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: a catalog. Radiology 2008;248:254-263.

61. ICRP. Diagnostic reference levels in medical imaging: review and additional advice. Ann ICRP 2001;31:33-52.

62. Deak PD, Smal Y, Kalender WA. Multisection CT protocols: sex- and age-specific conversion factors used to determine effective dose from dose-length product. Radiology 2010;257:158-166.

63. Tack D, Jahnen A, Kohler S, et al. Multidetector CT radiation dose optimisation in adults: short- and long-term effects of a clinical audit. Eur Radiol 2014;24:169-175.

64. Demb J, Chu P, Nelson T, et al. Optimizing radiation doses for computed tomography across institutions: dose auditing and best practices. JAMA Intern Med 2017;177:810-817.

65. European Commission. European guidelines for quality criteria for computed tomography. EUR 16262 EN. Luxembourg: European Commission, 2000.

66. American College of Radiology (ACR) practice guidelines for diagnostic reference levels in medical X-ray imaging. American College of Radiology. http://www.acr.org/Quality-Safety/Standards-Guidelines/Practice-Guidelines-by-Modality/General-Diagnostic. Published 2008. Accessed Mar 11, 2019.

67. National Diagnostic Reference Level Service (NDRLS). Australian Radiation Protection and Nuclear Safety Agency. https://www.arpansa.gov.au/research-and-expertise/surveys/national-diagnostic-reference-level-service. Published 2018. Accessed Jun 11, 2019.

68. Jang J, Jung SE, Jeong WK, et al. Radiation doses of various CT protocols: a multicenter longitudinal observation study. J Korean Med Sci 2016;31 Suppl 1:S24-31.

69. McCollough C, Branham T, Herlihy V, et al. Diagnostic reference levels from the ACR CT Accreditation Program. J Am Coll Radiol 2011;8:795-803.

70. Fukushima Y, Tsushima Y, Takei H, Taketomi-Takahashi A, Otake H, Endo K. Diagnostic reference level of computed tomography (CT) in Japan. Radiat Prot Dosimetry 2012;151:51-57.

71. Tsai HY, Tung CJ, Yu CC, Tyan YS. Survey of computed tomography scanners in Taiwan: dose descriptors, dose guidance levels, and effective doses. Med Phys 2007;34:1234-1243.

72. Keyzer C, Cullus P, Tack D, De Maertelaer V, Bohy P, Gevenois PA. MDCT for suspected acute appendicitis in adults: impact of oral and IV contrast media at standard-dose and simulated low-dose techniques. Am J Roentgenol 2009;193:1272-1281.

73. Platon A, Jlassi H, Rutschmann OT, et al. Evaluation of a low-dose CT protocol with oral contrast for assessment of acute appendicitis. Eur Radiol 2009;19:446-454.

74. Kim SH, Yoon JH, Lee JH, et al. Low-dose CT for patients with clinically suspected acute appendicitis: optimal strength of sinogram affirmed iterative reconstruction for image quality and diagnostic performance. Acta Radiol 2015;56:899-907.

75. Park JH, Kim B, Kim MS, et al. Comparison of filtered back projection and iterative reconstruction in diagnosing appendicitis at 2-mSv CT. Abdom Radiol 2016;41:1227-1236.

76. Karabulut N, Kiroglu Y, Herek D, Kocak TB, Erdur B. Feasibility of low-dose unenhanced multi-detector CT in patients with suspected acute appendicitis: comparison with sonography. Clin Imaging 2014;38:296-301.

77. Poletti PA, Platon A, De Perrot T, et al. Acute appendicitis: prospective evaluation of a diagnostic algorithm integrating ultrasound and low-dose CT to reduce the need of standard CT. Eur Radiol 2011;21:2558-2566.

78. United Nations Scientific Committee on the Effects of Atomic Radiation. Sources and effects of ionizing radiation. 2008 Report to the General Assembly Scientific Annexex. New York: United Nations, 2010.

79. Park JH, Jeon JJ, Lee SS, et al. Can we perform CT of the appendix with less than 1 mSv? A de-escalating dose-simulation study. Eur Radiol 2018;28:1826-1834.

80. Walsh L, Shore R, Auvinen A, Jung T, Wakeford R. Risks from CT scans--what do recent studies tell us? J Radiol Prot 2014;34:E1-5.

81. Journy N, Rehel JL, Ducou Le Pointe H, et al. Are the studies on cancer risk from CT scans biased by indication? Elements of answer from a large-scale cohort study in France. Br J Cancer 2015;112:185-193.

82. Bosch de Basea M, Pearce MS, Kesminiene A, et al. EPI-CT: design, challenges and epidemiological methods of an international study on cancer risk after paediatric and young adult CT. J Radiol Prot 2015;35:611-628.

83. Boone JM, Hendee WR, McNitt-Gray MF, Seltzer SE. Radiation exposure from CT scans: how to close our knowledge gaps, monitor and safeguard exposure--proceedings and recommendations of the Radiation Dose Summit, sponsored by NIBIB, February 24-25, 2011. Radiology 2012;265:544-554.

84. Kim KP, Berrington de Gonzalez A, Pearce MS, et al. Development of a database of organ doses for paediatric and young adult CT scans in the United Kingdom. Radiat Prot Dosimetry 2012;150:415-426.

85. National Research Council. Secondary Health risks from exposure to low levels of ionizing radiation: BEIR VII phase 2. In. Health risks from exposure to low levels of ionizing radiation: BEIR VII phase 2. Washington, DC, The National Academies Press, 2006.

86. Kiatpongsan S, Meng L, Eisenberg JD, et al. Imaging for appendicitis: should radiation-induced cancer risks affect modality selection? Radiology 2014;273:472-482.

87. Chang CC, Wong YC, Wu CH, et al. Diagnostic performance on low dose computed tomography for acute appendicitis among attending and resident radiologists. Iran J Radiol 2016;13:e33222.

88. Yun SJ, Kim HC, Yang DM, et al. Diagnostic usefulness of low-dose nonenhanced computed tomography with coronal reformations in patients with suspected acute appendicitis: a comparison with standard-dose computed tomography. J Comput Assist Tomogr 2016;40:485-492.

89. Douglas PS. The theory and practice of imaging outcomes research. Lancet 2015;385:2334-2335.

90. Braveman P, Schaaf VM, Egerter S, Bennett T, Schecter W. Insurance-related differences in the risk of ruptured appendix. N Engl J Med 1994;331:444-449.

91. Varadhan KK, Neal KR, Lobo DN. Safety and efficacy of antibiotics compared with appendicectomy for treatment of uncomplicated acute appendicitis: meta-analysis of randomised controlled trials. BMJ 2012;344:e2156.

92. Kim HY, Park JH, Lee YJ, Lee SS, Jeon JJ, Lee KH. Systematic review and meta-analysis of CT features for differentiating complicated and uncomplicated appendicitis. Radiology 2018;287:104-115.

93. Foley WD. CT features for complicated versus uncomplicated appendicitis: what is the evidence? Radiology 2018;287:116-118.

94. Kim HY, Park JH, Lee SS, et al. CT in differentiating complicated from uncomplicated appendicitis: any-of-10 CT features vs. radiologists' gestalt assessment. Am J Roentgenol in press.

95. Lee KH, Lee JM, Moon SK, et al. Attenuation-based automatic tube voltage selection and tube current modulation for dose reduction at contrast-enhanced liver CT. Radiology 2012;265:437-447.

96. Marin D, Nelson RC, Schindera ST, et al. Low-tube-voltage, high-tube-current multidetector abdominal CT: improved image quality and decreased radiation dose with adaptive statistical iterative reconstruction algorithm-initial clinical experience. Radiology 2010;254:145-153.

97. Schindera ST, Odedra D, Mercer D, et al. Hybrid iterative reconstruction technique for abdominal CT protocols in obese patients: assessment of image quality, radiation dose, and low-contrast detectability in a phantom. Am J Roentgenol 2014;202:W146-152.

98. Vaishnav JY, Jung WC, Popescu LM, Zeng R, Myers KJ. Objective assessment of image quality and dose reduction in CT iterative reconstruction. Med Phys 2014;41:071904.

99. Bongartz G, Golding SJ, Jurik AG, et al. Secondary European guidelines on quality criteria for computed tomography. In. European guidelines on quality criteria for computed tomography. Office for Official Publications of the European Communities, 2000.

100. Karabulut N, Boyaci N, Yagci B, Herek D, Kiroglu Y. Computed tomography evaluation of the normal appendix: comparison of low-dose and standard-dose unenhanced helical computed tomography. J Comput Assist Tomogr 2007;31:732-740.

101. Joo SM, Lee KH, Kim YH, et al. Detection of the normal appendix with low-dose unenhanced CT: use of the sliding slab averaging technique. Radiology 2009;251:780-787.

102. Pooler BD, Lawrence EM, Pickhardt PJ. Alternative diagnoses to suspected appendicitis at CT. Radiology 2012;265:733-742.

103. Weisenthal K, Karthik P, Shaw M, et al. Evaluation of kidney stones with reduced-radiation dose CT: progress from 2011-2012 to 2015-2016—Not there yet. Radiology 2018;286:581-589.

104. Tack D, Bohy P, Perlot I, et al. Suspected acute colon diverticulitis: imaging with low-dose unenhanced multi-detector row CT. Radiology 2005;237:189-196.

105. O'Neill SB, Mc Laughlin PD, Crush L, et al. A prospective feasibility study of sub-millisievert abdominopelvic CT using iterative reconstruction in Crohn's disease. Eur Radiol 2013;23:2503-2512.

106. Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med 1986;15:557-564.

107. Andersson M, Andersson RE. The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. World J Surg 2008;32:1843-1849.

108. Benjaminov O, Atri M, Hamilton P, Rappaport D. Frequency of visualization and thickness of normal appendix at nonenhanced helical CT. Radiology 2002;225:400-406.

109. Lee CC, Golub R, Singer AJ, Cantu R, Jr., Levinson H. Routine versus selective abdominal computed tomography scan in the evaluation of right lower quadrant pain: a randomized controlled trial. Acad Emerg Med 2007;14:117-122.

110. Kollar D, McCartan DP, Bourke M, Cross KS, Dowdall J. Predicting acute appendicitis? A comparison of the Alvarado score, the appendicitis inflammatory response score and clinical assessment. World J Surg 2015;39:104-109.

111. Scott AJ, Mason SE, Arunakirinathan M, Reissis Y, Kinross JM, Smith JJ. Risk stratification by the Appendicitis Inflammatory Response score to guide decision-making in patients with suspected appendicitis. Br J Surg 2015;102:563-572.

112. Tan WJ, Acharyya S, Goh YC, et al. Prospective comparison of the Alvarado score and CT scan in the evaluation of suspected appendicitis: a proposed algorithm to guide CT use. J Am Coll Surg 2015;220:218-224.

113. Poortman P, Oostvogel HJ, Bosma E, et al. Improving diagnosis of acute appendicitis: results of a diagnostic pathway with standard use of ultrasonography followed by selective use of CT. J Am Coll Surg 2009;208:434-441.

114. Lameris W, van Randen A, van Es HW, et al. Imaging strategies for detection of urgent conditions in patients with acute abdominal pain: diagnostic accuracy study. BMJ 2009;338:b2431.

115. Kaza RK, Platt JF, Goodsitt MM, et al. Emerging techniques for dose optimization in abdominal CT. Radiographics 2014;34:4-17.

116. Mayo-Smith WW, Hara AK, Mahesh M, Sahani DV, Pavlicek W. How I do it: managing radiation dose in CT. Radiology 2014;273:657-672.

117. Lee YJ, Kim B, Ko Y, et al. 2-mSv CT in adolescents and young adults with suspected appendicitis: advantages of additional review of thin sections using multiplanar sliding-slab averaging technique. Am J Roentgenol 2015;205:W485-491.

118. Dym RJ, Duncan DR, Spektor M, Cohen HW, Scheinfeld MH. Renal stones on portal venous phase contrast-enhanced CT: does intravenous contrast interfere with detection? Abdom Imaging 2014;39:526-532.

119. Preminger GM, Tiselius HG, Assimos DG, et al. 2007 guideline for the management of ureteral calculi. J Urol 2007;178:2418-2434.

120. Drake FT, Alfonso R, Bhargava P, et al. Enteral contrast in the computed tomography diagnosis of appendicitis: comparative effectiveness in a prospective surgical cohort. Ann Surg 2014;260:311-316.

121. Fefferman NR, Roche KJ, Pinkney LP, Ambrosino MM, Genieser NB. Suspected appendicitis in children: focused CT technique for evaluation. Radiology 2001;220:691-695.

122. Rao PM, Rhea JT, Novelline RA, Mostafavi AA, Lawrason JN, McCabe CJ. Helical CT combined with contrast material administered only through the colon for imaging of suspected appendicitis. Am J Roentgenol 1997;169:1275-1280.

123. Brassart N, Winant C, Tack D, Gevenois PA, De Maertelaer V, Keyzer C. Optimised z-axis coverage at multidetector-row CT in adults suspected of acute appendicitis. Br J Radiol 2013;86:20130115.

124. Kamel IR, Goldberg SN, Keogan MT, Rosen MP, Raptopoulos V. Right lower quadrant pain and suspected appendicitis: nonfocused appendiceal CT--review of 100 cases. Radiology 2000;217:159-163.

125. Jacobs JE, Birnbaum BA, Macari M, et al. Acute appendicitis: comparison of helical CT diagnosis focused technique with oral contrast material versus nonfocused technique with oral and intravenous contrast material. Radiology 2001;220:683-690.

126. Zanca F, Demeter M, Oyen R, Bosmans H. Excess radiation and organ dose in chest and abdominal CT due to CT acquisition beyond expected anatomical boundaries. Eur Radiol 2012;22:779-788.

127. Lee CH, Goo JM, Ye HJ, et al. Radiation dose modulation techniques in the multidetector CT era: from basics to practice. Radiographics 2008;28:1451-1459.

128. Coakley FV, Gould R, Yeh BM, Arenson RL. CT radiation dose: what can you do right now in your practice? Am J Roentgenol 2011;196:619-625.

129. Willemink MJ, Leiner T, de Jong PA, et al. Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality. Eur Radiol 2013;23:1632-1642.

130. Johnson PT, Horton KM, Kawamoto S, et al. MDCT for suspected appendicitis: effect of reconstruction section thickness on diagnostic accuracy, rate of appendiceal visualization, and reader confidence using axial images. Am J Roentgenol 2009;192:893-901.

131. Sprawls P. AAPM tutorial. CT image detail and noise. Radiographics 1992;12:1041-1046.

132. Paulson EK, Harris JP, Jaffe TA, Haugan PA, Nelson RC. Acute appendicitis: added diagnostic value of coronal reformations from isotropic voxels at multi-detector row CT. Radiology 2005;235:879-885.

133. Lee KH, Kim YH, Hahn S, et al. Computed tomography diagnosis of acute appendicitis: advantages of reviewing thin-section datasets using sliding slab average intensity projection technique. Invest Radiol 2006;41:579-585.

134. Birnbaum BA, Wilson SR. Appendicitis at the millennium. Radiology 2000;215:337-348.

135. Nikolaidis P, Hwang CM, Miller FH, Papanicolaou N. The nonvisualized appendix: incidence of acute appendicitis when secondary inflammatory changes are absent. Am J Roentgenol 2004;183:889-892.

136. Ganguli S, Raptopoulos V, Komlos F, Siewert B, Kruskal JB. Right lower quadrant pain: value of the nonvisualized appendix in patients at multidetector CT. Radiology 2006;241:175-180.

137. Godwin BD, Drake FT, Simianu VV, et al. A novel reporting system to improve accuracy in appendicitis imaging. Am J Roentgenol 2015;204:1212-1219.

138. Park SB, Kim MJ, Ko Y, Sim JY, Kim HJ, Lee KH. Structured reporting versus free-text reporting for appendiceal computed tomography in adolescents and young adults: preference survey of 594 referring physicians, surgeons, and radiologists from 20 hospitals. Korean J Radiol 2019;20:246-255.

139. Brown TW, McCarthy ML, Kelen GD, Levy F. An epidemiologic study of closed emergency department malpractice claims in a national database of physician malpractice insurers. Acad Emerg Med 2010;17:553-560.

140. Howell JM, Eddy OL, Lukens TW, Thiessen ME, Weingart SD, Decker WW. Clinical policy: critical issues in the evaluation and management of emergency department patients with suspected appendicitis. Ann Emerg Med 2010;55:71-116.

141. Yang HK, Ko Y, Lee MH, et al. Initial performance of radiologists and radiology residents in interpreting low-dose (2-mSv) appendiceal CT. Am J Roentgenol 2015;205:W594-611.

142. Huda W, Mettler FA. Volume CT dose index and dose-length product displayed during CT: what good are they? Radiology 2011;258:236-242.

143. Ford I, Norrie J. Pragmatic Trials. N Engl J Med 2016;375:454-463.

144. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370:1453-1457.

145. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Radiology 2015;277:826-832.

146. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—reporting of subgroup analyses in clinical trials. N Engl J Med 2007;357:2189-2194.

147. World Health Organization. Obesity: preventing and managing the globar epidemic-report of a WHO consultation of obesity. Geneva: World Health Organization, 1997.

148. Wagner JM, McKinney WP, Carpenter JL. Does this patient have appendicitis? JAMA 1996;276:1589-1594.

149. Andersson RE. Meta-analysis of the clinical and laboratory diagnosis of appendicitis. Br J Surg 2004;91:28-37.

150. Lameris W, van Randen A, Go PM, et al. Single and combined diagnostic value of clinical features and laboratory tests in acute appendicitis. Acad Emerg Med 2009;16:835-842.

151. Leeuwenburgh MMN. Appendicitis is easily missed or wrongly diagnosed - imaging is needed. BMJ 2011;343:d5976.

152. Weston AR, Jackson TJ, Blamey S. Diagnosis of appendicitis in adults by ultrasonography or computed tomography: a systematic review and meta-analysis. Int J Technol Assess Health Care 2005;21:368-379.

153. Gilmore T, Jordan C, Edelstein E. Right-sided diverticulitis mimics appendicitis. J Emerg Med 2013;44:e29-32.

154. Kim JH, Cheon JH, Park S, et al. Relationship between disease location and age, obesity, and complications in Korean patients with acute diverticulitis: a comparison of clinical patterns with those of Western populations. Hepatogastroenterology 2008;55:983-986.

155. Bachur RG, Levy JA, Callahan MJ, Rangel SJ, Monuteaux MC. Effect of reduction in the use of computed tomography on clinical outcomes of appendicitis. JAMA Pediatr 2015;169:755-760.

156. Humes DJ, Simpson J. Clinical presentation of acute appendicitis: clinical signs—laboratory findings—clinical scores, alvarado score and derivate scores. In: Keyzer C, ed. Imaging of acute appendicitis in adults and children. Berlin: Springer, 2012; 13-21.

157. Chan AW, Tetzlaff JM, Gotzsche PC, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ 2013;346:e7586.

158. MacGregor K, Li I, Dowdell T, Gray BG. Identifying institutional diagnostic reference levels for CT with radiation dose index monitoring software. Radiology 2015;276:507-517.

159. Pinto Leite N, Pereira JM, Cunha R, Pinto P, Sirlin C. CT evaluation of appendicitis and its complications: imaging techniques and key diagnostic findings. Am J Roentgenol 2005;185:406-417.

160. Bixby SD, Lucey BC, Soto JA, Theysohn JM, Ozonoff A, Varghese JC. Perforated versus nonperforated acute appendicitis: accuracy of multidetector CT detection. Radiology 2006;241:780-786.

161. Tsuboi M, Takase K, Kaneda I, et al. Perforated and nonperforated appendicitis: defect in enhancing appendiceal wall-depiction with multi-detector row CT. Radiology 2008;246:142-147.

162. Lee KH, Lee HJ, Kim JH, et al. Managing the CT data explosion: initial experiences of archiving volumetric datasets in a mini-PACS. J Digit Imaging 2005;18:188-195.

163. Kalra MK, Maher MM, Toth TL, et al. Techniques and applications of automatic tube current modulation for CT. Radiology 2004;233:649-657.

164. National Council on Radiation Protection and Measurements. Ionizing radiation exposure of the population of the United States. NCRP report No. 160. Bethesda, MD National Council on Radiation Protection and Measurements, 2009.

165. Brix G, Nagel HD, Stamm G, et al. Radiation exposure in multi-slice versus single-slice spiral CT: results of a nationwide survey. Eur Radiol 2003;13:1979-1991.

166. Hart D, Wall BF, Hiller MC, Shrimpton PC. Frequency and collective dose for medical and dental X-ray examinations in the UK, 2008. HPA-CRCE-012. Chilton, UK2010.

167. Smith-Bindman R, Kwan ML, Miglioretti DL. Who gets to decide? Radiology 2016;278:635-637.

168. National Academy of Sciences. Health risks from exposure to low levels of ionizing radiation: Biological Effects of Ionizing Radiation Committee VII phase 2. Washington D.C. 2005.

169. Kim KP, Einstein AJ, Berrington de Gonzalez A. Coronary artery calcification screening: estimated radiation dose and cancer risk. Arch Intern Med 2009;169:1188-1194.

170. Kim B, Lee KH, Kim KJ, Mantiuk R, Kim HR, Kim YH. Artifacts in slab average-intensity-projection images reformatted from JPEG 2000 compressed thin-section abdominal CT data sets. Am J Roentgenol 2008;190:W342-350.

171. Lee KH, Hong H, Hahn S, Kim B, Kim KJ, Kim YH. Summation or axial slab average intensity projection of abdominal thin-section CT datasets: can they substitute for the primary reconstruction from raw projection data? J Digit Imaging 2008;21:422-432.

172. von Falck C, Galanski M, Shin HO. Informatics in radiology: sliding-thin-slab averaging for improved depiction of low-contrast lesions with radiation dose savings at thin-section CT. Radiographics 2010;30:317-326.

173. Lee KH, Lee HS, Park SH, et al. Appendiceal diverticulitis: diagnosis and differentiation from usual acute appendicitis using computed tomography. J Comput Assist Tomogr 2007;31:763-769.

174. Yang HK, Woo H, Jo J, et al. Development of an online radiology case review system featuring interactive navigation of volumetric image datasets using advanced visualization Techniques. J Korean Soc Radiol 2015;73:312-322.

175. Sim JY, Kim HJ, Yeon JW, et al. Added value of ultrasound re-evaluation for patients with equivocal CT findings of acute appendicitis: a preliminary study. Eur Radiol 2013;23:1882-1890.

176. Daly CP, Cohan RH, Francis IR, Caoili EM, Ellis JH, Nan B. Incidence of acute appendicitis in patients with equivocal CT findings. Am J Roentgenol 2005;184:1813-1820.

177. Sicard N, Tousignant P, Pineault R, Dube S. Non-patient factors related to rates of ruptured appendicitis. Br J Surg 2007;94:214-221.

178. St. Peter SD, Sharp SW, Holcomb GW, 3rd, Ostlie DJ. An evidence-based definition for perforated appendicitis derived from a prospective randomized trial. J Pediatr Surg 2008;43:2242-2245.

179. Black WC. Randomized clinical trials for cancer screening: rationale and design considerations for imaging tests. J Clin Oncol 2006;24:3252-3260.

180. Flum DR. Clinical practice. Acute appendicitis–appendectomy or the "antibiotics first" strategy. N Engl J Med 2015;372:1937-1943.

181. Rosai J. Appendix. In: Rosai J, ed. Rosai and Ackerman's surgical pathology. St. Louis, MO: Mosby, 2011; 714-718.

182. Fenoglio-Preiser CM, Noffsinger AE, Stemmermann GN, Lantz PE, Isaacson PG. Nonneoplastic diseases of the appendix. In: Fenoglio-Preiser CM, ed. Gastrointestinal pathology. Philadelphia, PA: Lippincott Williams & Wilkins, 2008; 504-505.

183. Riber C, Tonnesen H, Aru A, Bjerregaard B. Observer variation in the assessment of the histopathologic diagnosis of acute appendicitis. Scand J Gastroenterol 1999;34:46-49.

184. Tang C-K. Disorders of the vermiform appendix. In: Ming S-C, ed. Pathology of the gastrointestinal tract: Williams & Wilkins 1998; 905.

185. Pieper R, Kager L, Nasman P. Clinical significance of mucosal inflammation of the vermiform appendix. Ann Surg 1983;197:368-374.

186. Campbell F. Inflammatory disorders of the appendix. In: Shepherd NA, ed. Morson and Dawson's gastrointestinal pathology: Wiley-Blackwell, 2012; 482-483.

187. Wang Y, Reen DJ, Puri P. Is a histologically normal appendix following emergency appendicectomy alway normal? Lancet 1996;347:1076-1079.

188. Phillips BJ, Perry CW. Appendiceal diverticulitis. Mayo Clin Proc 1999;74:890-892.

189. Bloch AV, Kock KF, Saxtoft Hansen L, Sandermann J. Periappendicitis and diagnostic consequences. Ann Chir Gynaecol 1988;77:151-154.

190. Fink AS, Kosakowski CA, Hiatt JR, Cochran AJ. Periappendicitis is a significant clinical finding. Am J Surg 1990;159:564-568.

191. Sandermann J, Glenthoj A, Nielsen KK. Peroperative mechanical manipulation of the appendix. A cause of periappendicitis? Ann Chir Gynaecol 1989;78:127-129.

192. Lally KP, Cox CS, Andrassy RJ. Appendix. In: Townsend CM, ed. Sabiston Textbook of Surgery: the biological basis of modern surgical practice. Philadelphia: Elsevier Saunders, 2004; 1382.

193. O'Leary DP, Redmond HP, Andrews EJ. Low-dose abdominal CT for diagnosing appendicitis. N Engl J Med 2012;367:478; author reply 478-479.

194. Park CS, Chang MS, Park I, Kim YI, Choe G. Pathologic analysis of 2159 cases of appendix. Korean J Pathol 2000;34:39-49.

195. American College of Radiology Imaging Network Adverse Event Reporting Manual. American College of Radiology. https://www.acrin.org/RESEARCHERS/REGULATORYRESOURCES.aspx. Published 2009. Accessed Mar 11, 2019.

196. Committee for Proprietary Medicinal Products. Points to consider on multiplicity issues in clinical trials. Committee for Proprietary Medicinal Products. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003640.pdf. Published 2002. Accessed Jun 09, 2019.

197. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG, Group C. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA 2012;308:2594-2604.

198. Kim K, Lee CC, Song KJ, Kim W, Suh G, Singer AJ. The impact of helical computed tomography on the negative appendectomy rate: a multi-center comparison. J Emerg Med 2008;34:3-6.

199. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. Ann Intern Med 2006;145:62-69.

200. Krajewski S, Brown J, Phang PT, Raval M, Brown CJ. Impact of computed tomography of the abdomen on clinical outcomes in patients with acute right lower quadrant pain: a meta-analysis. Can J Surg 2011;54:43-53.

201. Andersson T, Alfredsson L, Kallberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. Eur J Epidemiol 2005;20:575-579.

202. Firth D. Bias reduction of maximum likelihood estimates. Biometrika 1993;80:27-38.

203. Zhang S, Liang F, Li W, Hu X. Subgroup analyses in reporting of phase III clinical trials in solid tumors. J Clin Oncol 2015;33:1697-1702.

204. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365:82-93.

205. Antevil J, Rivera L, Langenberg B, Brown CV. The influence of age and gender on the utility of computed tomography to diagnose acute appendicitis. Am Surg 2004;70:850-853.

206. Levine CD, Aizenstein O, Lehavi O, Blachar A. Why we miss the diagnosis of appendicitis on abdominal CT: evaluation of imaging features of appendicitis incorrectly diagnosed on CT. AJR 2005;184:855-859.

APPENDIX

Appendix 1. Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Registration	Patient's unique identifier	UID	String		[Site number][Assign][Categorization of patient for surgery][Sequence disit 5]S########## - 01 LD - 01 Referred appendectomy - 02 SD - 02 Other abdominal surgery - 03 Did not undergo surgery
	Site number	SiteNo	String		[S][Sequence digit 2] S##
	Visit	Visit	String		[Baseline]
	Assign	assign	Integer		[1] Low-dose group [2] Standard-dose group
	Enrollment date	ENDAT	Date		[YYYY-MM-DD]
Demographics	Ethnicity	ETHNIC	Integer		[1] Korean [2] Non-Korean
	Sex	SEX	Integer		[M] Male [F] Female
	Date of birth	BRTHDAT	Date		[YYYY-MM-DD]
	Age	AGE	Integer	year(s)	[XX]
	Height	HEIGHT	Integer	cm	[XXX]
	Weight	WEIGHT	Integer	kg	[XXX]
	Body mass index	BMI	Float	kg/m^2	[XX.X]
	Body temperature	TEMP	Float	℃	[XX.X]
History	History of abdominal surgery	UDZ_ABDOP	String		[N] No [Y] Yes
	History of specification of abdominal surgery	UDZ_ABDOPHX	String		
	History of diverticulitis	UDZ_DIVER	String		[N] No [Y] Yes
	History of urinary stone	UDZ_UST	String		[N] No [Y] Yes
	History of OBGY disease	UDZ_OBGY	String		[N] No [Y] Yes
	History of specification of OBGY disease	UDZ_OBGYHX	String		
	History of acute pyelonephritis	UDZ_APN	String		[N] No [Y] Yes
	Other notable history	UDZ_OTH	String		[N] No [Y] Yes
	Specification of other notable history	UDZ_OTHTX	String		

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Initial pain (IP)	Absence of pain at the time of symptom onset	IP_NONE	String		[N] No [Y] Yes
	RLQ pain at the time of symptom onset	IP_RLQ	String		[N] No [Y] Yes
	Epigastric pain at the time of symptom onset	IP_EPIGA	String		[N] No [Y] Yes
	Periumbilical pain at the time of symptom onset	IP_PERIUM	String		[N] No [Y] Yes
	Right flank pain at the time of symptom onset	IP_RFLK	String		[N] No [Y] Yes
	LLQ pain at the time of symptom onset	IP_LLQ	String		[N] No [Y] Yes
	RUQ pain at the time of symptom onset	IP_RUQ	String		[N] No [Y] Yes
	LUQ pain at the time of symptom onset	IP_LUQ	String		[N] No [Y] Yes
	Left flank pain at the time of symptom onset	IP_LFLK	String		[N] No [Y] Yes
	Suprapubic pain at the time of symptom onset	IP_SUPPU	String		[N] No [Y] Yes
Current pain (CP)	Absence of current pain	CP_NONE	String		[N] No [Y] Yes
	Currently RLQ pain	CP_RLQ	String		[N] No [Y] Yes
	Currently epigastric pain	CP_EPIGA	String		[N] No [Y] Yes
	Currently periumbilical pain	CP_PERIUM	String		[N] No [Y] Yes
	Currently right flank pain	CP_RFLK	String		[N] No [Y] Yes
	Currently LLQ pain	CP_LLQ	String		[N] No [Y] Yes
	Currently RUQ pain	CP_RUQ	String		[N] No [Y] Yes
	Currently LUQ pain	CP_LUQ	String		[N] No [Y] Yes
	Currently left flank pain	CP_LFLK	String		[N] No [Y] Yes
	Currently suprapubic pain	CP_SUPPU	String		[N] No [Y] Yes

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Migration of pain	Migration of pain	MIG	String		[N] No [Y] Yes
Tenderness (TD)	Absence of tenderness	TD_NONE	String		[N] No [Y] Yes
	RLQ tenderness	TD_RLQ	String		[N] No [Y] Yes
	Epigastric tenderness	TD_EPIGA	String		[N] No [Y] Yes
	Periumbilical tenderness	TD_PERIUM	String		[N] No [Y] Yes
	Right flank tenderness	TD_RFLK	String		[N] No [Y] Yes
	LLQ tenderness	TD_LLQ	String		[N] No [Y] Yes
	RUQ tenderness	TD_RUQ	String		[N] No [Y] Yes
	LUQ tenderness	TD_LUQ	String		[N] No [Y] Yes
	Left flank tenderness	TD_LFLK	String		[N] No [Y] Yes
	Suprapubic tenderness	TD_SUPPU	String		[N] No [Y] Yes
Costovertebral angle tenderness (CVAT)	Right of costovertebral angle tenderness	CVAT_RT	String		[N] No [Y] Yes
	Left of costovertebral angle tenderness	CVAT_LT	String		[N] No [Y] Yes
Rebound tenderness (RT)	Absence of rebound tenderness	RT_NONE	String		[N] No [Y] Yes
	RLQ rebound tenderness	RT_RLQ	Integer		[0] No [1] Slight [2] Moderate [3] Severe
	Epigastric rebound tenderness	RT_EPIGA	Integer		[0] No [1] Slight [2] Moderate [3] Severe
	Periumbilical rebound tenderness	RT_PERIUM	Integer		[0] No [1] Slight [2] Moderate [3] Severe

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Rebound tenderness (RT)	Right flank rebound tenderness	RT_RFLK	Integer		[0] No [1] Slight [2] Moderate [3] Severe
	LLQ rebound tenderness	RT_LLQ	Integer		[0] No [1] Slight [2] Moderate [3] Severe
	RUQ rebound tenderness	RT_RUQ	Integer		[0] No [1] Slight [2] Moderate [3] Severe
	LUQ rebound tenderness	RT_LUQ	Integer		[0] No [1] Slight [2] Moderate [3] Severe
	Left flank rebound tenderness	RT_LFLK	Integer		[0] No [1] Slight [2] Moderate [3] Severe
	Suprapubic rebound tenderness	RT_SUPPU	Integer		[0] No [1] Slight [2] Moderate [3] Severe
Symptom	Nausea	NAUSEA	String		[N] No [Y] Yes
	Vomiting	VOMIT	String		[N] No [Y] Yes
	Anorexia	ANORX	String		[N] No [Y] Yes
	Chill	CHILL	String		[N] No [Y] Yes
	Presence of diarrhea	DIA	String		[N] No [Y] Yes
	Characteristic of diarrhea	DIA_C	Integer		[1] Watery [2] Loose stool
	Number of diarrhea	DIA_NO	Integer		XX
	Mucoid stool	STOOL_MC	String		[N] No [Y] Yes
	Bloody stool	STOOL_BL	String		[N] No [Y] Yes

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Chief complain (CC)	Chief symptom	CC_SYMP	Integer		[1] Abdominal pain [2] Nausea/Vomitting [3] Fever [4] Diarrhea [5] Other
	Specification of other chief symptom	CC_OTH	String		
	Duration of chief symptom	CC_DUR	Integer		[1] ≤ 12 hours [2] 13–24 hours [3] 2–3 days [4] ≥ 4 days
Laboratory results	White-cell count	WBC	Float	$\times 10^3/\mu\ell$	[XX.X]
	Segmented neutrophils	NEU	Integer	%	[XX]
	C-reactive protein	CRP	Float	mg/dl	[XX.X]
	Serum creatinine	CREAT	Float	mg/dl	[X.X]
CT examination	Date of CT	CT_CTDAT	Date		[YYYY-MM-DD]
	Time of CT	CT_CTTIM	Time		[HH:MM]
	Vendor and number of channels	VENDER	String		[Vender name][Channel]
	$CTDI_{vol}$	CTDIVOL	Float	mGy	[XX.XX]
	Dose-length product	DLP	Float	mGy*cm	[XXX.X]
	Anteroposterior dimension	AP	Float	cm	[XX.X]
	Lateral diameter	LD	Float	cm	[XX.X]
	Effective diameter	EFF_DIA	Float		[XX.X]
	Size-Specific Dose Estimates	SSDE	Float		[XX.X]
	Presence of additional scan	DLP_ADDNA	String		[N] No
	Dose-length product of additional scan	DLP_ADD	Float	mGy*cm	[XXX.X]
	Total of dose-length product	TOTDLP	Float	mGy*cm	[XXXX.X]
	Patient's unique identifier for submitted CT images	IMG_UID	String		[Site acronym][[Sequence disit 4]
	Thin-section (section thickness ≤ 2 mm, reconstruction interval ≤ 1 mm) transverse images	IMG_THIN	Integer		[N] Not archived [Y] Archived
	Thick-section (section thickness 3−5 mm, reconstruction interval ≤ 2 mm) transverse and coronal images	IMG_THICK	Integer		[N] Not archived [Y] Archived
	Image datset to submit to Data Center	PLAN	String		[A] Thin-section transverse images [B] Thick-section transverse and coronal images [C] Submit all archived images and report protocol non-adherence
	Iterative reconstructure	IR	Integer		[0] No [1] Yes
	Series description	IMG_SRDSC	String		

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
CT examination	Images orientation	IMG_TCS	String		[T] Transverse [C] Coronal [S] Sagittal
	Reconstruction technique	IMG_FBPIR	String		[F] FBP [I] IR
	Thickness	IMG_THICKMM	Float	mm	
	Interval	IMG_INTVMM	Float	mm	
	Number of images	IMG_COUNT	Integer		
Verification of submitted CT images	Were patient name and medical record number anonymized using the enrollment number?	IMGUP_ANNY	Integer		[0] Not anonymized [1] Anonymized
	Is the number of submitted images identical to that to the number of images archived in the locat pacs?	IMGUP_NUMEQ	Integer		[0] Not identical [1] Identical
	Were all images reviewed to confirm no corruption during the data transfer?	IMGUP_RESOK	Integer		[0] Not confirmed [1] Confirmed
	Verification completed?	IMGUP_UPCOMP	Integer		[0] Not completed [1] Completed
Initial CT report (INI)	Radiologist's unique identifier	RadID	String		[Site number][Department digit 2][Position digit 2][Sequence digit 2] S######## - 01 radiology - 01 abdominal attending - 02 surgery - 02 non-abdominal attending - 03 emergency - 03 fellow - 04 pathology - 04 resident
	Position of radiologist	INI_RADS	Integer		[1] Trainee [2] Non-abdominal attending [3] Abdominal attending
	Likelihood of appendicitis	INI_LKHAPP	Integer		[1] 1 [2] 2 [3] 3 [4] 4 [5] 5
	Perforation	INI_PERFO	Integer		[0] 0 [1] 1 [2] 2
	Drainable abscess	INI_DRNAB	Integer		[0] 0 [1] 1
	Appendix visualization	INI_APPVS	Integer		[0] 0 [1] 1 [2] 2
	Alternative diagnosis	INI_ALTDX	String		
	Presence of incidental clinically significant findings	INI_OTHNN	String		[N] No
	Incidental clinically significant findings	INI_OTH	String		

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Addendum CT report (AND)	Presence of addendum CT report	ADN_CT	Integer		[0] Absent [1] Present
	Radiologist's unique identifier of addendom CT report	AND_RadID	String		[Site number][Department digit 2][Position digit 2][Sequence digit 2] S######## - 01 radiology - 01 abdominal attending - 02 surgery - 02 non-abdominal attending - 03 emergency - 03 fellow - 04 pathology - 04 resident
	Position of radiologist	ADN_RADS	Integer		[1] Trainee [2] Non-abdominal attending [3] Abdominal attending
	Likelihood of appendicitis	ADN_LKHAPP	Integer		[1] 1 [2] 2 [3] 3 [4] 4 [5] 5
	Perforation	ADN_PERFO	Integer		[0] 0 [1] 1 [2] 2
	Drainable abscess	ADN_DRNAB	Integer		[0] 0 [1] 1
	Appendix visualization	ADN_APPVS	Integer		[0] 0 [1] 1 [2] 2
	Alternative diagnosis	ADN_ALTDG	String		
	Presence of incidental clinically significant findings	ADN_OTHNN	String		[N] No
	Incidental clinically significant findings	ADN_OTH	String		
Additional imaging (AI)	Presence of additional ultrasonography	AI_US	Integer		[0] No [1] Yes
	Date of additional ultrasonography	AI_USGDAT	Date		[YYYY-MM-DD]
	Presence of additional CT	AI_ADCT	Integer		[0] No [1] Yes
	Date of additional CT	AI_ADCTDAT	Date		[YYYY-MM-DD]
Admission	ER discharge or ward admission	RSTER	Integer		[1] Discharge from ER [2] Admission to ward [3] Transfer to another hospital
	Hospital to which the patient was transferred	HP_TRNSF	String		
	Time of discharge from ER or ward	DSCH_DAT	Date		[YYYY-MM-DD]
	Time of discharge from ER or ward	DSCH_TIM	Time		[HH:MM]

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Medical treatment and preoperative percutaneous abscess drainage (PCD)	Percutaneous abscess drainage for perforated appendicitis	PCD	Integer		[0] No PCD [1] PCD
	Date of percutaneous abscess drainage	PCD_DAT	Date		[YYYY-MM-DD]
	Presence of medical treatment	MEDITR	Integer		[0] Absent [1] Present
Operation record (OP)	Presence of operation record	OP_RPTNO	Integer		[0] Absent
	Surgeon's unique identifier	SurID	String		[Site number][Department digit 2][Position digit 2][Sequence digit 2] S######## - 01 radiology - 01 attending - 02 surgery - 03 fellow - 03 emergency - 04 pathology
	Date of operation (induction of the anesthesia)	OP_DAT	Date		[YYYY-MM-DD]
	Time of operation (induction of the anesthesia)	OP_TIM	Time		[HH:MM]
	Appendectomy	OP_APPY	Integer		[1] Appendectomy for suspected appendicitis [2] Incidental appendectomy [3] No appendectomy
	Mode of approach for appendectomy	OP_APPRO	Integer		[1] Laparoscopic [2] Open [3] Conversion to open
	Extent of appendectomy	OP_EXTENT	Integer		[1] Simple appendectomy [2] Ileocecectomy [3] Cecectomy [4] Right hemicolectomy [5] Other
	Other extent of surgery	OP_EXTENT_TX	String		
	Appearance of appendix	OP_APPENDIX	Integer		[1] Hyperemic [2] Suppurative or gangrenous [3] Normal
	Surgically-identified appendiceal perforation	OP_PERFO	Integer		[0] Not perforated [1] Perforated
	Name of surgery	OP_NAM	String		
	Categorization of patient for surgery	OP_CAT	Integer		[1] Referred appendectomy [2] Other abdominal surgery [3] Did not undergo surgery
Pathology report (PA)	Presence of pathology report	PA	Integer		[0] Absent [1] Present
	Pathologist's unique identifier	PathoID	String		[Site number][Department digit 2][Position digit 2][Sequence digit 2] S######## - 01 radiology - 01 attending - 02 surgery - 03 fellow - 03 emergency - 04 pathology

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Pathology report (PA)	Presence of appendicitis	PA_APPE	Integer		[1] Appendicitis [2] Not appendicitis [3] No appendectomy
	Pathologically-identified appendiceal perforation	PA_PERFO	Integer		[0] Not perforated [1] Perforated
	Presence of specific pathologic diagnosis other than appendicitis	PA_DXNN	String		[N] No
	Specific pathologic diagnosis other than appendicitis	PA_DX	String		
Final diagnosis (FD)	Final diagnosis	FD	String		
	Categorization of final diagnosis	FD_CAT	String		
	Data supporting the alternative diagnosis using CT	FD_CT	String		[N] No [Y] Yes
	Data supporting the alternative diagnosis using medical record	FD_MEDR	String		[N] No [Y] Yes
	Data supporting the alternative diagnosis using telephone interview	FD_TEL	String		[N] No [Y] Yes
Follow-up (FU)	Withdrawal from study	FU_WIDTH	String		[N] No [Y] Yes
	Success of telephone interview after three months	threefu	String		[N] No [Y] Yes
	Success of follow-up during study period	fu	String		[N] No [Y] Yes
	Specific follow-up loss	fuloss_ty	Integer		[1] Did not obtain the operation record [2] Did not obtain the pathology report [3] Fail of telephone interview after three months [4] Did not obtain both the operation record and the pathology report [5] Did not obtain the pathology report and fail of telephone interview after three months [6] Did not obtain both the operation record and the pathology report and fail of telephone interview after three months
Adverse event (AE)	Presence of serious adverse event	SAE	String		[N] No [Y] Yes
	Death	DEATH	Integer		[0] No Death [1] Death
	(Only patients who did not undergo appendectomy) After you returned home from our emergency department, did your abdominal pain (or other symptom(s)) persisted (or recurred), which finally turned out to be appendicitis which needed appendectomy in another hospital?	PI_EXTEND1	Integer		[0] No [1] Yes [2] Underwent appendectomy when I first visited emergency department

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Adverse event (AE)	(Only patients who underwent appendectomy) After you underwent appendectomy, were you hospitalized for more than 7 days? A hospital stay over 7 days following nonincidental appendectomy is regarded as a prolongation of hospitalization.	PI_EXTEND2	Integer		[0] No [1] Yes [2] Did not undergo appendectomy
	After you returned home from our emergency department, were you hospitalized for treatment other than appendectomy? A hospitalization is not reportable when it is for diagnostic or elective surgical procedures for a preexisting condition and the outcome is uneventful (e.g., uneventful negative appendectomy).	PI_EXTEND3	Integer		[0] No [1] Yes
	After you returned home from our emergency department, were you hospitalized for intensive care?	LIF_THREAT	Integer		[0] No [1] Yes
	Persistent or significant disability or incapacity, or congenital anomalies/birth defects	DISABILITY	Integer		[0] No [1] Yes
	Expectedness of adverse event	AE	Integer		[1] Expected AE [2] Unexpected AE [3] No AE
Reportable AE	Presence of reportable adverse event	AE_REPORT	Integer		[0] No Reportable AE [1] Reportable AE
	CTCAE term	AE_TERM	String	CTCAE 4.01	
	Onset date of reportable AE	ONSETDAT	Date		[YYYY-MM-DD]
	Stop date of reportable AE	STOPDAT	Date		[YYYY-MM-DD]
	Grade	GRADE	Integer		[1] 1 [2] 2 [3] 3 [4] 4 [5] 5
	Attribution to study procedures	ATTRBUT	Integer		[0] Not attributable [1] Attributable
	Association with missed appendicitis or other important diagnosis	MISSEDDX	Integer		[0] Not missed diagnosis [1] Missed diagnosis
	Action taken	ACTTAKEN	Integer		[0] None [1] Medication therapy [2] Procedure [3] Hospitalizaton
	Outcome	OUTCOME	Integer		[1] Resolved [2] Persistent/Sequela [3] Death [4] Unknown
Nonadherence	Nonadherence	Nonadherence	String		

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Generated variables					
Alvarado score	Presence of migration of pain according to a definition of Alvarado score	al_mig	Integer		[0] No [1] Yes
	Presence of vomiting according to a definition of Alvarado score	al_vom	Integer		[0] No [1] Yes
	Presence of RLQ pain according to a definition of Alvarado score	al_rlq	Integer		[0] No [2] Yes
	Presence of anorexia according to a definition of Alvarado score	al_ano	Integer		[0] No [1] Yes
	Presence of rebound tendeness according to a definition of Alvarado score	al_rt	Integer		[0] No [1] Yes
	Degree of fever according to a definition of Alvarado score	al_tem	Integer		[0] < 37.3°C [1] ≥ 37.3°C
	Degree of NEU according to a definition of Alvarado score	al_neu	Integer		[0] < 75% [1] ≥ 75%
	Degree of WBC according to a definition of Alvarado score	al_wbc	Integer		[0] < $10\times10^3/\mu\ell$ [2] ≥ $10\times10^3/\mu\ell$
	Alvarado score	Alvarado	Integer		[0–4] Low risk [5–6] Indeterminate risk [7–10] High risk
AIR score	Presence of vomiting according to a definition of AIR score	air_vom	Integer		[0] No [1] Yes
	Presence of RLQ pain according to a definition of AIR score	air_rlq	Integer		[0] No [1] Yes
	Presence of rebound tendeness according to a definition of AIR score	air_rt	Integer		[0] No [1] Slight [2] Moderate [3] Severe
	Degree of fever according to a definition of AIR score	air_tem	Integer		[0] < 38.5°C [1] ≥ 38.5°C
	Degree of NEU according to a definition of AIR score	air_neu	Integer		[0] < 70% [1] ≥ 70% and < 85% [2] ≥ 85%
	Degree of WBC according to a definition of AIR score	air_wbc	Integer		[0] < $10\times10^3/\mu\ell$ [1] ≥ $10\times10^3/\mu\ell$ and < $15\times10^3/\mu\ell$ [2] ≥ $15\times10^3/\mu\ell$
	Degree of CRP according to a definition of AIR score	air_crp	Integer		[0] < 1mg/dl [1] ≥ 1mg/dl and < 5mg/dl [2] ≥ 5mg/dl
	AIR score	AIR	Integer		[0–4] Low risk [5–8] Indeterminate risk [9–12] High risk

Data Dictionary

Group	Meaning	Variable name	Data type	Unit	Values
Others	Detailed categorization of patients whose final diagnosis is "Presumed acute appendicitis without pathologic confirmation."	presumedappe	Integer		[1] Lost to follow-up, but we presumed the patient undergo appendectomy for the treatment of presumed appendicitis [2] Did not obtain pathology report, but we presumed the patient undergo appendectomy for the treatment of presumed appendicitis [3] The patient had medical treatment
	Patients who underwent interval appendectomy due to medical treatment	PROLONG	Integer		[0] No [1] Yes
	Appendectomy as a denominator of a primary endpoint	appendectomy	Integer		[0] Did not undergo appendectomy [1] Underwent appendectomy
	Confirmed appendicitis through the pathology report	confirmedappy	Integer		[0] Confirmed the patient has not appendicitis [1] Confirmed the patient has appendicitis
	Reference standards for analyses of diagnostic results	RS	Integer		[0] Patients had incomplete reference standards for the presence or absence of appendicitis [1] Patients had complete reference standards for the presence or absence of appendicitis
	Negative appendectomy as a numerator of the primary endpoint	NA	Integer		[0] No [1] Yes
	Presence of appendiceal neoplasm	NEOPLASM	Integer		[0] No [1] Yes
	Presence of appendiceal perforation	PERFO	Integer		[0] No appendiceal perforation [1] Surgically- or pathologically-identified appendiceal perforation
	Patient visit ER in working hour (from 08:00 to 17:00)	workinghour	Integer		[0] No [1] Yes (from 08:00 to 17:00, Monday to Friday)
	Interval from CT acquisition to surgery	HR_CTtoOP	Float	hour(s)	
	Interval from CT acquisition to hospital discharge	HR_CTtoDSCH	Float	hour(s)	
	Interval from CT acquisition to hospital discharge	DAY_CTtoDSCH	Float	day(s)	

Abstract in Korean

충수염 의증 청소년 및 젊은 성인에서 2-mSv CT와 기존 선량 CT의 민감도 및 특이도: LOCAT의 사후 하위그룹 분석

서론: 본 연구는 충수염 의증 청소년 및 젊은 성인에서 기존 CT와 비교하여 2-mSv CT의 진단 민감도 및 특이도에서 환자 또는 병원의 특성에 따른 이질성이 있는지를 탐색하는 연구임.

방법: 본 연구는 2013년 12월에서 2016년 8월 사이에 15–44세의 환자에서 2-mSv CT와 기존 선량 CT (일반적으로 7 mSv)를 비교한 대규모 비열등성 무작위배정 임상시험의 프로토콜 별 분석세트를 사용함. 본 연구에는 20개 병원에서 2,773명의 환자 (중앙값 연령 [사분위수 범위], 28 [21–35]세)가 포함되었으며, 160명의 판독의가 참여함. 환자의 성별, 신체 크기, 충수염에 대한 임상 위험 점수, CT 검사시간 (일과시간 [근무일 기준 오전 8시부터 오후5시] 또는 일과시간 이후), CT 장비, 판독의의 경험정도, 2-mSv CT에 대한 이전 경험 여부, 그리고 병원의 임상규모 등의 사전 정의된 하위 그룹에서 충수염 진단을 위한 민감도 및 특이도의 이질성을 테스트함. 두 군의 차이를

숲그림으로 제시하고, 민감도와 특이도에 대한 덧셈 및 곱셈 상호작용을 테스트함.

결과: 많이 날씬하거나 뚱뚱한 경우, 충수염 염증 반응 점수가 높은 경우, 최신 CT 기기를 사용한 경우, 2-mSV CT 의 이전 경험이 있는 병원, 그리고 충수절제술 규모가 작은 병원의 경우 등 특정 하위 그룹은 작은 크기 (< 200)로 인해 민감도에 대한 95 % 신뢰구간이 넓었음. 그 외, 대부분의 하위 그룹에서 그룹 간 차이에 대한 95 % 신뢰구간은 이전 보고된 전체 그룹 간 차이 및 귀무가설 값 (즉, 0)을 포함하였음. 2-mSv CT 군과 기존 선량 CT 군 간에 민감도 및 특이도에서 덧셈 또는 곱셈 상호작용을 보이는 하위 그룹은 없었음.

결론: 충수염 의증 청소년과 젊은 성인에서 2-mSv CT와 기존 선량 CT 간에 민감도와 특이도에서 이질성을 보이는 하위그룹은 없었음. 이는 2-mSv CT가 다양한 집단에서 기존 선량 CT를 대체할 수 있음을 의미함. 다만, 본 연구에서 작은 크기를 가진 일부 하위 그룹에 대해서는 추가적인 연구가 필요함.

주요어: 진단능, 충수염, 전산화단층촬영, 하위 그룹 이질성

학번: 2016-34869