Dear Dr Yang,

we thank you and the reviewers for the careful evaluation of our manuscript and for the helpful comments and suggestions. We think that the in-depth revision we conducted has substantially improved the manuscript. For your convenience, revised passages are tracked and marked in colour in the attached file "manuscript_changes marked.doc", while the file "manuscript_revised.doc" contains the final revised vision. Our detailed reply to the reviewers is attached below.

There is but one suggestion, raised by the reviewers 3 and 4, that we found difficult to follow and to implement. The reviewers recommend contrast-to-noise ratio (CNR) as the metric to measure lesion conspicuity. To calculate CNR, one would have to measure the standard deviation of "image noise" on the MR image in some location outside of the patient body.

Our MR physicists now argue that, using modern scanner hardware with multi-coil arrays and software with built-in image post-processing (filtering), image noise is no longer evenly distributed in MR images but depends to the site chosen for measurement. The traditional straight-forward approach of measuring image noise with ROI technique on the images can thus be affected in an unpredictable manner and extent under such circumstances. The concept of signal intensity ratios as used in our study avoids this problem. Even though it may have been less widely adopted to far, it is correct in terms of physics.

For the present revision, we thought it preferable to keep the SI ratios and to address this issue in "Discussion". In case that you and the reviewers consider switching to CNR as crucial and as a mandatory precondition for publication, we are prepared to provide these data.

Thanking you for taking care of our manuscript,

Sincerely,

Henning Neubauer, MD, MBA
Reviewer 1:

Introduction p. 3: If a patient has bilateral Wilms tumor, what difference does it make if NBS is detected in addition. This required further discussion.

Nephron-sparing surgery is the favoured approach (not only) in patients with bilateral Wilms tumour. If additional NBS foci cannot distinguished from multifocal Wilms tumour on pre-operative imaging, the surgeons would have to perform more extensive resection removing all potential Wilms tumour manifestations. Although, to our knowledge, there is no such data from a controlled clinical study available, it seems prudent to assume that accurate imaging is pre-condition for targeted surgery and may even result in a higher proportion of preserved renal tissue.

As our manuscript does not provide data to support (or contradict) this hypothesis, we chose to not elaborate on this issue in "Introduction", but added a short note to this effect in "Discussion" (page 9).

Methods p.4: The authors should discuss their choice of b-values. Would biexponential DWI lead to better results?

Biexponential modelling certainly is superior to the monoexponential approach. ADC values of the same lesion do differ to some extent, depending on the applied model. To our experience, these differences, however, are small compared to intra- and inter-observer variability of measuring ADC in a clinical setting. When we introduced DWI to our oncological MR imaging protocols in 2008, we evaluated several combinations of b-values, but ended up with 50/800 as the best practical choice. It provides acceptable image quality within not more than 5 minutes scanning time. Time is a crucial factor in paediatric imaging. The benefit of measuring additional b-values or higher b-values (requiring more averages because of less signal) does not outweigh the setback of additional scanning time in our young patients.

We added some discussion on this point in "Limitations".

Methods p.4: more details on the DWI sequence are required; did they authors use respiratory triggering? Did they use coregistration?

No and no. We have been using the standard free-breathing technique with multiple averages that comes with our SIEMENS scanners. We tried respiratory triggering, but it requires extra acquisition time and may not work properly in some patients, while the gain in image quality was found to be limited in paediatric patients.

On our new scanner (1.5 T Siemens Magnetom Aera) we have been testing alternative scanning techniques, including co-registration and multi-shot EPI DWI ("RESOLVE") with some success. Recently, our MR physicists provided a new high-resolution HASTE-DWI that we are now testing. DWI on the new scanner (open bore, short magnet) is more susceptible to B0 field inhomogenity, as compared to the old Symphony and the Avanto scanners, and we may even give up SS-EPI and move on the MS-EPI or HASTE-DWI for routine scanning.

As for the present manuscript, we now provide the requested details in M&M.
Methods p.4: How did the authors calculate ADC values? Where did they draw the ROIs?
We used the ADC maps that were calculated automatically by the scanner software. As also requested by another reviewer, we now provide a separate Figure (new Figure 1) to show an example of how the large ROI and the small ROI were placed for measurement.

Results, p.8: The authors should provide the mean ADC value for the NBS foci
The mean ADC value for all NBS foci is now included in the text as well as in Table 1. In the text, we provide data on mean ADC of small and large NBS foci. Partial volume effects as the most likely explanation for the observed difference in ADC are discussed in "Discussion."

Discussion, p.9: "DWI signal is related to cellularity". This should be discussed more in detail. The authors should also discuss T2-shine through effects.
We have added a passage elaborating on the hypothesized relation between tumour cellularity and diffusivity in "Discussion". Our strategy to prevent mis-readings caused by "T2 shine-through" artefacts is now stated in "M&M".

Discussion, p.9: as both cancer and NBS were detected by DWI, I do not think that the results of this study confirm the high sensitivity (or at least specificity) of DWI for malignant disease; the authors should moderate their claims.
We re-phrased the passage to state that DWI is sensitive for renal tumour lesions, yet incapable of distinguishing malignant and non-malignant focus, as based on ADC quantification.

Discussion, p.10: The authors should also provide changes in size (not only in ADC) of tumors during chemotherapy. This is essential in order to understand the meaning of DWI for treatment monitoring. Furthermore, the conclusions on the value of ADC for response assessment should be more moderate due to the small number of patients.
The chemotherapy-induced changes in size and ADC of Wilms tumours are given in Table 1. The ADC value and size of the two lesions without response to chemotherapy (NCC and WT recurrence) have been added to "Results". Our conclusions are now stated as preliminary experience based on a small number of patients.

General remark: English editing is required (e.g. p.6 "hepatic metastasis at initial diagnosis" or "the fourth patient had three ipsilateral....").
We performed English language editing and double-checked for typos.
Reviewer 2:

Page 5: It would be nice the explanation/justification of the additional ROI to be accompanied by a figure demonstrating the different ROI methods (showing the restricted diffusion area, cystic areas etc).

We now provide an additional figure (new Figure 1) presenting our evaluation strategy in an heterogeneous Wilms tumor.

Page 7: spelling mistake “…spread at nine months”.

We performed English language editing and double-checked for typos.

It is not a big surprise that such a simple model (ADC) cannot perform better in distinguishing difficult pathologies. It would be something worth mentioning in the limitations and that the field is now looking into alternative models for investigating cancer such as the IVIM, the Kurtosis and the VERDICT models.

At the request of another reviewer, we now discuss the issue of monoexponential vs. biexponential modelling for ADC. The "fast and simple" approach of ADC-based DW imaging is very valuable, especially for paediatric imaging. We have been evaluating IVIM in abdominal imaging for some time, but have not yet collected enough data to draw conclusions on it usefulness for tumour differentiation.

As requested we added some discussion and references on emerging techniques (IVIM, kurtosis, VERDICT) that look beyond ADC.

References: I find the reference list a little bit out of date, and I would suggest including a few more recent papers on renal tumour investigation with DWI.

We have updated our reference list. Quite some work has been done and published on renal masses in adults, but virtually no new studies are available from paediatric populations.
Reviewer 3:

1. The signal intensity ratio (tumor vs. adjacent tissue) is not a standard method to assess the detectability. I would suggest that authors provide contrast to noise ratio (CNR) which is often employed as an index for contrast.

We have repeatedly consulted our MR physicists on that topic, as the same issue had been raised in the review of previous work. The physicists say that CNR measurement in the way it is commonly performed (the image noise measured as standard deviation on the image) is a valid approach only in a very basic setting. With modern scanner equipment employing multiple coil arrays, multi-channel acquisition and image post-processing (filtering), image noise is no longer evenly distributed in the image and simple CNR measurements are thus inaccurate. Valid CNR measurements would require the scanning of field maps along with the diagnostic sequence in order to correctly quantify noise. The physicists therefore recommend the use of signal intensity ratios as we did in this study.

My personal "non-physicist" point of view is such that I do think measuring CNR in the traditional way is a good approximation, as far as the purpose of any clinical study is concerned. But probably it is the same as with statistical methods - everybody uses them, usually for a good end, but many - or so say the statisticians - do it the wrong way.

We could certainly provide CNR values, if this is considered mandatory by the reviewer and the editor. For the present revision, however, we keep the SI ratios, demonstrate this concept more in detail in a new Figure 1 and discuss the matter in "Discussion".

2. Please add more information about "two experienced observers", e.g. radiologists or physicists with # years experiences.

The requested information is now available in M&M. The readers were a resident with five years training in radiology and special training in paediatric imaging (I. P.) and the other a board-certified paediatric radiologist with seven years experience in extra-cranial DWI (H. N.).

3. Scan parameters of ce-T1w should be provided for the SI ratio comparison.

This information has been added to "M&M".

4. Please provide the ROI definition (tumor and adjacent tissue; large and heterogeneous tumors) in figures as examples of the SI ratio measurements.

As stated under 1., this information is now provided in the text and in our new Figure 1.
Reviewer 4:

Abstract:
Patients and Methods section: I think the description of the DWI sequence is too detailed and a little redundant for the abstract (length). The abstract already stated that images were acquired on a 1.5 T, so I would remove this statement. I would consider removing the b-value information (from the abstract). This is important, but can be stated with greater detail and at leisure in the body of the manuscript.

The abstract has been revised accordingly.

As per journal requirements and formatting, it may be helpful to spell out abbreviations in the abstract.
We now spell out all abbreviations in the abstract at first use.

BODY OF THE MANUSCRIPT:
Patients and Methods:
Please spell out all abbreviations upon first time use. Examples include: Magnetic Resonance (MR), Diffusion-weighted images (DWI), Turbo Spine-echo (TSE), T2-weighted (T2w).

We now spell out all abbreviations at first use.

• The second sentence of the second paragraph of the Patients and Methods section could be edited for clarity. It states: “seven patients below aged 0 to 6...”. I would simplify the wording to state that n (number of; age range) patients were sedated.
Corrected.

• The manuscript states that the DWI was acquired in the transverse plane. Is this the axial plane? This is minor linguistics but would be helpful for clarity.
Yes, transverse = axial. We routinely scan all DWI in the axial plane. Additional coronal DWI scans can be useful (as shown in Figure 4.), but only our AVANTO scanner produces coronal EPI-DWI with acceptable image quality. Severe distortion artefacts are very common with the new AERA scanner (large bore, short magnet architecture), and things are still worse at 3 Tesla.

• If possible, please include the type of fat suppression used.
We used standard spectral fat saturation, as now stated in M&M.

• I understand and support the rational that the authors have used, comparing signal intensity (SI) in a lesion to SI in the background parenchyma. However, I would STRONGLY encourage the use of a more widespread metric such as contrast to noise ratio (CNR). This
metric is well known in the abdominal MRI literature, and incorporates pure image noise into the calculation.

This point was also raised by another reviewer. We have again discussed this issue with our MR physicists - and not for the first time. It is the physicists’ view that, using modern scanner hardware and software, image noise is no longer evenly distributed in MR images when acquired with multi-channel receiver coils and built-in image filtering. The traditional straight-forward approach of measuring image noise with ROI technique on the images may thus be affected in an unpredictable manner and extent under such circumstances. The SI ratio avoids such issues, even though it is a less handy and perhaps less intuitive metric. One may quantify image noise by the means of field maps, but this was not done in our patients and does not seem viable in a clinical setting. For the present revision, we thought it preferable to keep the SI ratios and to address this issue in "Discussion".

In case that the editors consider switching to CNR as crucial and as a mandatory precondition for publication, we are prepared to provide these data.

- Please provide additional details about the interpretation of these studies. Were the DWI and the conventional images interpreted simultaneously? Was there an interval (to avoid recall bias) between interpretation of DWI and conventional images?

The readings of DWI and contrast-enhanced T1w imaging were performed one week apart to avoid recall bias to the extent possible in our small group of patients, as now stated in M&M.

- Also please state what images of the DWI acquisition were analyzed (b=800 only, b=50 and b=800, trace diffusion map). Please state if these were reviewed in default window setting, or “inverted gray scale” which the authors have chosen for the figures.

When reading DWI, it is important to look at all image information (low b, high b, ADC map). We use standard window settings as the first choice in routine readings. Inverted gray scale is sometimes useful as it may make small foci stand out more clearly from the background. We have got no data to support this statement, though. There are studies on pulmonary nodules on chest films where gray scale inversion was found helpful for the trained eye, and it might be the same with DWI. We also use inversion for reading whole-body DWI studies.

Our choice to present inverted gray scale images in this manuscript, however, is primarily based on the experience that image quality in print is much better with inversion. Furthermore, our clinicians are used to seeing scintigraphic studies and PET images, but rarely trouble to understand DW images. As DWI with inverted gray scale somewhat resembles bone scintigraphy or PET scans, we found it easier to convey our imaging findings with inverted DWI when communicating with the referring clinicians.

We added some explanation of this point to M&M.

- Results:

- Typographical error: “Juvenil” renal cell carcinoma, as opposed to “Juvenile”.

Corrected.
• In the section “Patients with primary diagnosis of Wilms Tumor”: I understand the observation that NBS was better seen in DWI, and I sympathize with the statement. However, this remains subjective. Could the authors be more explicit and say how many lesions were seen only on DWI, and how many lesions were seen on DWI and conventional images, but were better seen on DWI.

We now provide the requested data.

Discussion:
The third paragraph is long, and should be shortened. I would suggest shortening information on ADC values and their usefulness in other tissues. I believe all the references the authors cite are pertinent and establish a precedent for the use of ADC to differentiate benign vs. malignant lesions. I’m not opposed to mentioning the quantitative thresholds (from literature) they reported to distinguish benign vs. malignant lesions, as they provide a general “framework” of magnitudes observed in different tissues. However, I would refrain from comparing the results of these references between each other (references 17, 19, 23). There are several sentences discussing how reported ADC in musculoskeletal neoplasms compares to regenerative nodules in cirrhotic liver. These sentences should be deleted, as they are discussing results from work that is not part of this submission. Also, comparing diffusivity between musculoskeletal neoplasms and HCC adds little to the discussion of diffusion in renal neoplasms.

The respective paragraph has been re-written. The cross-comparison between the different entities was deleted, as suggested.

I would make a little more emphasis on great quality of images on free breathing DWI.

We now again highlight this finding in "Discussion" and in our conclusion.

Limitations:
In general, its well drafted. There is a sentence that states, “The scanning protocols were designed with a particular emphasis on comparability and our results in previous studies did not show any substantial inconsistencies”. I find this sentence confusing, as the authors are referring to prior studies, and not data from this study. I don’t think prior results increase validity or reliability of the observations in this manuscript, unless the other studies pertain directly to renal masses with this protocol (ref would need to be included). Overall I would favor deleting it.

The small number of patients in our study was indeed not sufficient for any between-scanners analysis. In earlier studies, we compared ADC of reference tissues between different scanners at 1.5 and 3 Tesla (references 23, 27 and 29). As it is, these findings cannot really support the findings in this manuscript, so we deleted this passage, as suggested.

Tables:
• The units of ADC and the exponents (mentioned in the abstract), need to be mentioned again in the caption.

Corrected.
Figures and legends:

• Figure 1. Image quality is good. I’m a little bit surprised by the decision to depict the DWI images on inverted gray scale. While I consider this a valid approach that can be helpful (different window settings have proven useful in adult stroke literature), I’d be curious to hear why the authors chose to do so, and if they do it in routine clinical practice. While only a minor concern, I think most people look at the DWI/ADC in the standard display settings, and interpretation of the images provided in this manuscript may be less intuitive. We discussed the matter above (M&M). After all, it may just be a matter of what one is used to looking at. If thought necessary, we could switch back to non-inverted figures.

• Figure 4. I would slightly edit the caption, which currently states “MRI showed a small focal signal increase”. This goes back to the use of inverted gray scale. Although I believe that the lesion has decreased diffusivity the display gets confusing since the caption says high signal intensity and the image (arrow) demonstrates a dark focus (inverted scale). The focus of high signal on DWI is as a matter of fact not displayed. If the authors choose to maintain this display, they probably should refer to these foci as areas of decreased/restricted diffusion.

The caption has been re-written, as requested, to emphasise the restricted diffusivity rather than the high/low signal.

(end)