NPL REPORT MS 28

Statistical analysis of temperature rise in passive medical implants in a magnetic resonance imaging environment

MAURICE COX
KAVYA JAGAN
SUNDER RAJAN

APRIL 2021

Statistical analysis of temperature rise in passive medical implants in a magnetic resonance imaging environment

Maurice Cox, Kavya Jagan
Data Science Department, National Physical Laboratory, UK
Sunder Rajan
US Food and Drug Administration, US

ISSN: 1754-2960

https://doi.org/10.47120/npl.MS28

National Physical Laboratory
Hampton Road, Teddington, Middlesex, TW11 0LW

Approved on behalf of NPLML by Dr P M Harris
Science Area Leader for Data Analytics & Modelling,

EXECUTIVE SUMMARY

Purpose: Data records from test laboratories had been gathered for small passive medical implants immersed in gel phantoms in an MRI environment. The data had been analyzed by the Food and Drug Administration of the USA (FDA) in order to understand the extent to which temperature rise around the implant is correlated with implant primary length. This paper extends that work to other factors that may influence RF heating.

Methods: Parametric models are constructed to try to understand which factors have most influence on temperature rise. Explanatory models can be used to predict heating rise for implants whose characteristics lie within the span of the data on which the models are based.

Results: Models were analyzed that yield temperature rise as a function of various geometric parameters and test laboratory, taking measurement uncertainties into consideration. A main result is that there is dependence on the primary length parameter but secondary and tertiary length parameters have little effect. Also, the test laboratory appears not to be a significant influencing factor.

Conclusions: Since the data set is small, our conclusions are inevitably tentative. Our analyses should be regarded as preliminary to a more extensive treatment involving more data, which will be required to make more concrete statements. Those statements can be strengthened, and measurement traceability issues handled when there is access to uncertainties associated with all measured quantities. Currently quoted uncertainties are too small by a factor of about two.

CONTENTS

TABLES

1 INTRODUCTION

Data records from test laboratories had been gathered for a number of passive medical implants immersed in gel phantoms in a magnetic resonance imaging (MRI) environment and provided to the Food and Drug Administration of the USA (FDA) as part of pre-market submissions. For each device tested, the data included implant dimensions, the material, the MRI system, temperature rises and specific absorption rate (SAR) values.

The FDA carried out a retrospective analysis of these data records, especially concerning the observations of radio-frequency (RF) induced heating [13]. The analysis was primarily concerned with understanding the extent to which temperature rise around the implant is correlated with the primary length (longest reported dimension) of the implant.

This paper extends that work, using the same data records, to other factors that may influence RF heating. Those factors include secondary and tertiary implant length, the test laboratories themselves, and prior knowledge of the measurement uncertainties associated with the data. Parametric models are constructed to try to understand which factors are most important. Explanatory models can be used to predict heating rise for implants whose characteristics lie within the span of the observations on which the models are based. Attention is paid to the uncertainties associated with the development and use of such models.

2 METHODS

Data records were gathered for a number of passive medical devices, with primary length up to 185 mm from a number of test laboratories. Devices included cardiac plugs, aortic valves, stents, inferior vena cava filters, fracture fixation screws, inter-body spinal fusion devices, and dental bridges. 53 of the devices were stents, four vascular, five dental and 24 ‘other’.

The data records, 86 in all, each of which corresponded to one device tested and a scan duration of 15 min, were based on RF-heating test reports conforming to the ASTM F2182-11a standard [2]. The MRI type used was described in each case. Following that standard, the implant was submerged in gelled saline with electrical conductivity of $0.47\,\mathrm{S\,m^{-1}} \pm 0.05\,\mathrm{S\,m^{-1}}$ (interpreted as defining a 95 % confidence interval). The implant was secured in a plastic holder at locations at least 2 cm away from the gel surface, bottom and walls of the ASTM phantom. The fibre optic probes used for measuring temperature were sited in control positions as well as in locations expected to yield maximum changes. Temperature increases were obtained over a 15 min RF exposure window and scaled to 2 kg to 4 kg whole-body (WB) average SAR (determined calorimetrically). The increase in temperature at the same location in the absence of the implant was also measured to estimate the local background (LB) SAR. The LB SAR in the region of the implant was computed from the slope of the temperature-time graph using the expression

$$\mathrm{SAR} = c\frac{T}{t},$$

where $c = 4150\,\mathrm{J\,kg^{-1}}$ is the specific heat capacity of the gel; T is the change in temperature; and t is the duration of the exposure. The test report included values of WB SAR and LB SAR at the location of maximum heating on the implant.

Seven test laboratories submitted data that corresponded to most devices being tested once at a field of 1.5 T and once at 3 T. For each device, temperature change and SAR were recorded for these two fields. Data were recorded regarding the dimensions of the implants, the material, the MRI system and some of the other factors on which temperature rise and SAR depended.

Table 1 lists possible factors (independent variables) and possible responses (dependent variables). Other factors that could be considered include shape of implant and implant coating type, but information on such factors was either not available or only available for some devices.

Among the material properties in the table are coating type and the baseline viscosity of the gel used. The latter affects heat conduction during the 15 min RF exposure window, but no information on this quantity, which can vary with temperature, was provided in the reports.

Table 1: Possible factors (independent variables) and responses (dependent variables)

Possible factor	**Possible response**
Device name	Temperature change per unit mass
Manufacturer	SAR in the presence of implant
Test laboratory	Background SAR (no implant)
Device type	
Device model	
Device dimension 1 (primary dimension)	
Device dimension 2	
Device dimension 3	
Material properties (various)	

Our study has the objective of extending the statistical analysis in [13] in various ways: to examine the effect of all three length dimensions on temperature rise, to investigate the influence of test laboratory, and to account for prior knowledge of uncertainties associated with the data. Ideally, data for each category of implant would be analyzed separately. However, because of the smallness of the total data set, all data were considered in the same analysis as in [13], although there attention was also paid to a separate analysis of stent data.

The data points in figure 1 (also see [13]) portray temperature rise normalized with respect to local background SAR against implant length (termed the *primary* length L_1 below) for a field of 1.5 T and (right) 3 T. There is considerable scatter in the data presumably due to factors such as shape, thermal conductivity, electrical conductivity and measurement errors, and, in general, factors such as listed in table 1).

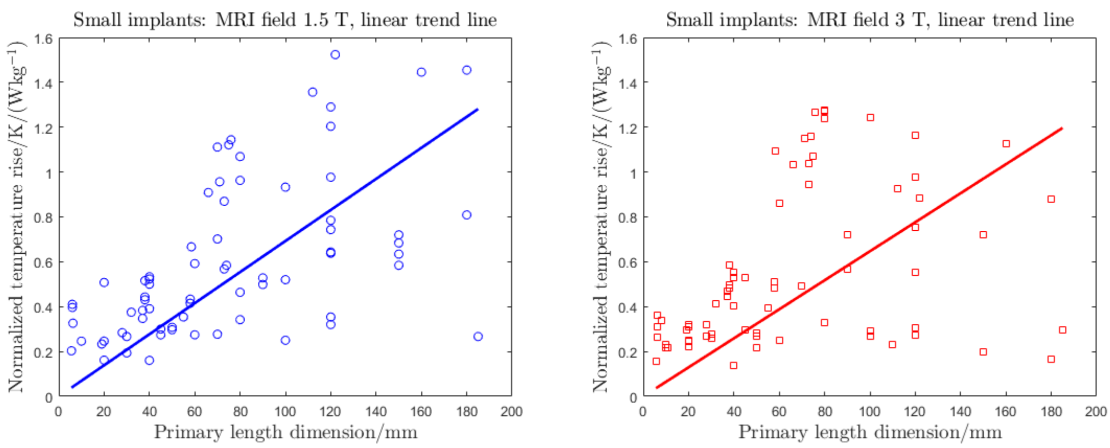

Figure 1: 1.5 T and (right) 3 T data and according linear trend lines for normalized temperature rise as a function of the primary dimensional parameter of the implant

3 Results

3.1 Temperature dependence on primary length

Figure 1 also shows the linear trend lines determined by ordinary least squares (OLS) [7], that is, with no weighting applied, for the normalized temperature rise as functions of the primary dimension; further details are given in section 3.2. Higher-order (quadratic, cubic, etc.) trend lines were also obtained, but gave negligible improvement (statistically insignificant) over the linear trend lines, as might be expected visually from figure 1.

For each of these devices the information in table 1 is given in the provided data records. For a few devices some of this information is absent.

According to [13] a linear univariate model relating T, temperature change per unit mass SAR in the presence of the implant, to primary dimension L_1 for this data has an R^2-value of 0.21, implying that 21 % of the variability in the data is explained by this model. This result applies to taking all data records into consideration, that is, for both 1.5 T and 3 T.

Analysis of data for the two fields separately gives the linear trend lines depicted in figure 1. Inspection of that figure indicates that the data for the 3 T field are slightly more scattered about their trend line than that for the 1.5 T field, as reflected in the R^2 statistics of 0.14 and 0.18, respectively, for the two cases: see section 3.2.

We prefer to use the root-mean-square residual (RMSR) rather than R^2 as a measure of the goodness of fit since the former can be generalized to regressions other than OLS: see section 4.3. RMSR is defined as the square root of the quotient of (a) the sum of the squares of the deviations of the T-data values from the corresponding model values, and (b) the degrees of freedom (number of data records used minus the number of model parameters). The RMSR values for the data for the 1.5 T and the 3 T fields are respectively 0.29 K/(W/kg^{-1}) and 0.33 K/(W/kg^{-1}).

It could rightly be argued that since there should be negligible heating rise for implants of near-zero primary length, the trend lines shown in figure 1 should pass through the origin. The trend line for the 3 T field passes very near the origin, but that is not the case for the 1.5 T field. Repeating the calculation for a trend line with no intercept term gave for the 1.5 T field an increased values of RMSR of 0.38 K/(W/kg^{-1}).

3.2 Temperature dependence on up to three length dimensions

Since the inclusion of additional variables in a regression model can improve its ability to represent the data, results for models relating T to L_1 and the two other dimensional parameters L_2 and L_3 were obtained. Three regression analyses were carried out:

- Model 1: T as a function of L_1: $T = b_0 + b_1 L_1$ (subsection 3.1),
- Model 2: T as a function of L_1 and L_2: $T = b_0 + b_1 L_1 + b_2 L_2$,
- Model 3: T as a function of L_1, L_2 and L_3: $T = b_0 + b_1 L_1 + b_2 L_2 + b_3 L_3$.

Some of the data records are incomplete. For instance, not all L_1,L_2 and L_3 were reported in every case, presumably because the relevant implants were characterized geometrically in terms of just one or two lengths. Accordingly, the data were pre-processed to include only that data for which all factors under consideration were present. Thus, for example, 65 of the 86 data records for the 1.5 T field were used for Model 2 for which data values for L_1, L_2 and T were all available, whereas 56 of the data records were available for Model 3. Thus, a direct comparison of the three models cannot be made although a comparison based on the 56 data records for which all three lengths were identified indicated similar results.

The results are summarized in table 2 for the 1.5 T and the 3 T fields, where estimated parameters are denoted as hatted. The data for both fields is no better explained, as judged by RMSR in the final column, by including in the regression model more than one length parameter.

Table 2: Regression results for normalized temperature rise T for a 1.5 T field as a function of $n = 1,2,3$ dimensional parameters

		Regression parameter estimates					
	n	$\widehat{b}_0$/ K	$\widehat{b}_1$/ $\mathrm{K\,mm^{-1}}$	$\widehat{b}_2$/ $\mathrm{K\,mm^{-1}}$	$\widehat{b}_3$/ $\mathrm{K\,mm^{-1}}$	No. of data records	RMSR/ K/(W/kg^{-1})
	1	0.262	0.00435			73	0.29
1.5 T	2	0.275	0.00437	0.00042		65	0.30
	3	0.271	0.00442	−0.01353	0.0116	56	0.31
	1	0.364	0.00274			74	0.33
3 T	2	0.335	0.00245	0.00722		64	0.35
	3	0.368	0.00189	−0.02097	0.02923	53	0.34

Further statistical analyses were carried out on the data. One analysis involved a comparison across linear models involving L_1 only, L_1 and L_2, and L_1, L_2 and L_3, but using only the data records for which all three lengths were reported. Statistically, proceeding this way avoids the bias involved in the results given in table 2, which had different numbers of records involved in the three cases. However, negligible change from the previous results was observed. Another analysis included a further term in the model, namely, $L_1L_2L_3$, representing a volumetric effect. Again negligible improvement over previous results was observed.

3.3 Temperature dependence on test laboratory

There would also be a variable relating to the RF exposure system used and therefore the test laboratory. Such a variable could be considered to be categorical. However, here we include a variable representing ‘laboratory effect’ for each test laboratory in the overall model. This variable would constitute an additive bias term for each test laboratory, which would in general take a different value for each laboratory (cf. [10]). It was reported in [13] that results differed ‘significantly across the test houses due to differences in RF coil design’.

The test laboratories providing data, p, say, in number, are assigned anonymous identifiers $1,2,\ldots,p$. Thus each point in figure 1 is associated with a known laboratory. For the data considered here, $p = 7$. In terms of the data from test laboratory ℓ, the counterpart of Model 1 used for laboratory ℓ is

$$T = b_0 + b_1L_1 + d_\ell,$$

where d_ℓ is a systematic error or offset for that laboratory. The d_ℓ are regarded as drawn from the same probability distribution. Similar extensions are used for Models 2 and 3. We only present results for Model 1 extended by laboratory effects: see section 4.5.

There are $p+2$ parameters b_0 and b_1, representing the underlying straight line, and the laboratory biases d_ℓ, $\ell = 1,\ldots,p$. Since it is impossible to solve uniquely for $p+2$ such constants, the constant term b_0 is arbitrarily set to zero, giving the $(p+1)$-term model

$$T = b_1L_1 + d_\ell, \qquad \ell = 1,\ldots,p.$$

We could equally well set any of these constants to an arbitrary value or impose some condition such as the sum of the constants is zero. It is straightforward to move from one such choice to another following the determination of a particular solution.

The corresponding statistical model (or observation equation) [11] is

$$T_i = b_1 L_{1,i} + d_{\ell(i)} + e_i, \qquad i = 1, \ldots, m, \tag{1}$$

where T_i is the ith recorded normalized temperature rise, $L_{1,i}$ is the primary length in the ith data record, $d_{\ell(i)}$ denotes the bias for the laboratory that contributed the ith data record, and e_i is the value of a random variable. The e_i are assumed to be drawn from the same probability distribution. Model (1) can be expressed in matrix-vector form as

$$\boldsymbol{T} = \boldsymbol{A}\boldsymbol{c} + \boldsymbol{e}, \tag{2}$$

where $\boldsymbol{T}$ is the vector of the T_i, $\boldsymbol{c}$ denotes the vector containing b_1 and the d_ℓ, $\boldsymbol{e}$ is the vector of the e_i, and $\boldsymbol{A}$ is a transformation matrix termed the *design matrix*. To illustrate, the design matrix for a situation in which there are three test laboratories, with the number of implants tested by these laboratories being respectively 3, 4 and 2, is

$$\boldsymbol{A} = \begin{bmatrix} \times & 1 & & \\ \times & 1 & & \\ \times & 1 & & \\ \times & & 1 & \\ \times & & 1 & \\ \times & & 1 & \\ \times & & 1 & \\ \times & & & 1 \\ \times & & & 1 \end{bmatrix},$$

where $\times$ denotes a length value and a space a zero element.

The least-squares solution of the observation equation (2) is given by minimizing $\boldsymbol{e}^\top \boldsymbol{e}$, that is, the OLS (unweighted) solution of

$$\boldsymbol{A}\boldsymbol{c} \approx \boldsymbol{T}. \tag{3}$$

In the current instance [13] the numbers of implants measured by the seven test laboratories are respectively 31, 12, 1, 19, 16, 3 and 4, that is, 86 in all.

Primary length and temperature values were reported in 70 of the 86 data records for the field 1.5 T and in 71 data records for 3 T. Table 3 shows for both fields the estimates obtained of the parameters related to primary length L_1 and test laboratory biases. It also shows the standard uncertainties associated with these estimates.

The corresponding trend lines for each laboratory are shown in figure 2 for 1.5 T and (right) 3 T. These trend lines are parallel since the only difference between laboratories in the regression model is the constant terms associated with them.

Compared with the case of Model 1, when no account was taken of laboratory biases, in section 3.2, the difference in the RMS residuals was negligible. For 1.5 T, RMSR changed from 0.29 K/(W/kg^{-1}) to 0.30 K/(W/kg^{-1}), and for 3 T, RMSR changed from 0.33 K/(W/kg^{-1}) to 0.36 K/(W/kg^{-1}).

We can conclude the following from an examination of table 3. If twice the standard uncertainty associated with an estimate of a quantity is no greater than the absolute value of that

Table 3: Estimates of parameters related to primary length L_1 and test laboratory biases and their associated standard uncertainties for 1.5 T and 3 T

		b_1/	Test laboratory bias/[K/(W/kg^{-1})]						
		[K/(W/kg^{-1})mm^{-1}]	1	2	3	4	5	6	7
1.5 T	Estimate	0.005 95	0.16	0.29	0.01	−0.10	0.14	0.48	0.06
	Std unc	0.002 07	0.26	0.34	1.00	0.38	0.45	0.59	0.74
3 T	Estimate	0.005 43	0.20	0.12	0.29	−0.10	0.41	0.33	0.03
	Std unc	0.002 09	0.27	0.33	1.00	0.39	0.42	0.58	0.71

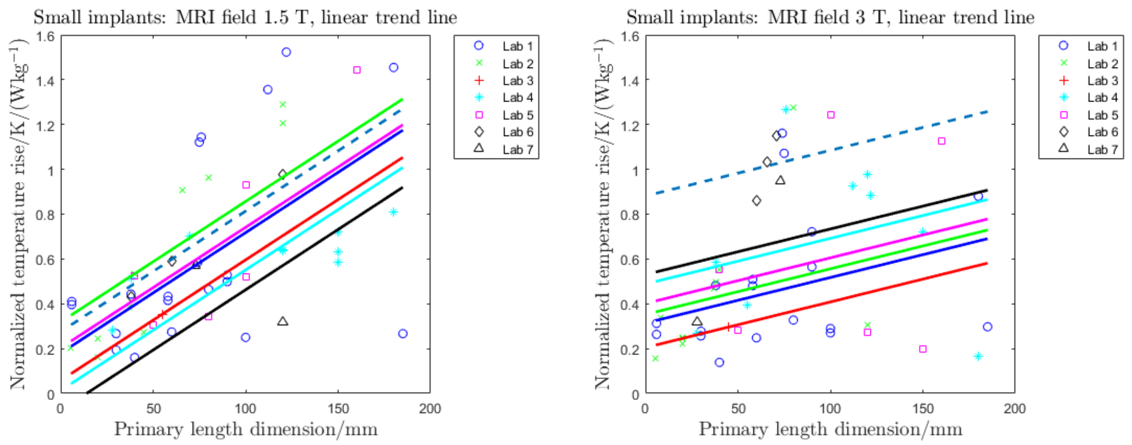

Figure 2: Linear trend lines for approximating the data by a model including a systematic effect for each test laboratory for 1.5 T and (right) 3 T

estimate, the estimate can be considered significant at an approximately 95 % level of confidence. Statistically, therefore, primary length has an effect on temperature rise: for 1.5 T, $2 \times 0.002\,07$ K/((W/kg^{-1})mm^{-1}) is less than the estimate 0.005 95 K/(W/kg^{-1})mm^{-1}, with a similar statement for 3 T. On the same basis and in contrast, not one of the estimates of laboratory bias was significantly different from zero for either field. It can generally be concluded that the results from this retrospective analysis do not indicate significant bias due to the test laboratory involved.

The above statement is reinforced by table 4, which shows the estimated biases arranged in decreasing order (top to bottom lines in figure 2) for 1.5 T and 3 T. It might be expected that because of individual test laboratory commonalities, similar biases would be present within each laboratory's results for 1.5 T and 3 T. That there is very little correlation with respect to order across the two fields is inconsistent with this expectation.

Table 4: Test laboratories ordered in terms of decreasing estimated bias

Field/T	Test laboratory order						
1.5	7	2	5	4	6	1	3
3	6	1	3	5	4	2	7

4 DISCUSSION

4.1 LIMITATIONS OF THE DATA SET USED FOR ANALYSIS

The number of data records available for analysis is limited. Hence, for each model studied, where the factors under consideration were (a) primary length only, (b) primary length and secondary length and (c) primary length, secondary and tertiary length, the maximum quantity of data according to that model was taken into consideration. The different numbers of data records in each category made direct comparison of results less than straightforward. The situation was ameliorated somewhat by the fact that these numbers were not too different. For the 1.5 T field, these numbers were 75, 66 and 56 in the three cases and for the 3 T field, they were 73, 66 and 56. See table 2.

The analysis could be repeated if necessary, using only the 56 data records for which all three implant lengths were reported for both the 1.5 T and the 3 T field, enabling a more rigorous statistical comparison. As reported in section 3.2, however, the outcome was similar.

It is said in [13] that a limitation of the study carried out there is that it is a retrospective analysis and therefore suffers from gaps in data, which makes it difficult to make interesting comparisons of subgroups (for instance, screws versus stents or coated versus un-coated). It is stated, however, that the data analyzed provides useful insights into the dependence on device length of the RF heating effect in MRI.

In this paper we have used the same data as that used in [13]. Because of the unavailability of a sufficient quantity of data, as in [13] testing results for a number of implant types have been aggregated in an attempt to draw conclusions for small implants. Since the data set is very small, our conclusions are tentative. Our analyses should be regarded as preliminary to a more extensive treatment involving much more data, which will be required to make more concrete statements. Those statements will be strengthened, and measurement traceability issues handled when there is access to uncertainties associated with all measured quantities.

4.2 OTHER FACTORS

Many other factors such as categorical factors device type, material and coating type have not been considered in this study. Another factor that has not been entertained is patient age, where correlation between age and temperature rise when undergoing MRI scans has been reported [1, 8, 12]. For instance, Kim [8] summarized results from 69 patients receiving clinical brain MRI. Patients' tympanic temperatures were recorded before and immediately after the MRI procedure. Among factors considered, age was found to be the most significant in the case of 3 T MRI. Interestingly, there was no correlation between temperature and age for 1.5 T MRI. Based on Kim's results it seems that age and other physiological factors should be recorded and taken into consideration when analyzing MRI data.

In general, there are three potential sources of data, from patients, phantoms and models. By its nature, the amount of patient data is extremely limited because of practical and ethical reasons. However, some patient data such as that considered by Kim [8] and others (section 4.2) has been acquired. Phantom data have been the consideration of this paper, but that too is limited because of commercial reasons. Model data is more prolific, many studies having reported data from simulations. Ideally, model data should be validated by phantom data from test laboratories.

Generally, estimating local SAR, and hence normalized temperature rise, by invasive measurements within the human body is not feasible [6]. The authors of that paper also report that discrepancies between simulations and experiments vary in a range between 15 % and 100 % depending on the quantity under analysis and the procedure adopted for the comparison.

4.3 Input uncertainties

One of the test laboratories quoted an uncertainty of 30 % in the measured heating values [13], with sources of error that include positioning of the fibre optic probe and the properties of the gel medium. Bottauscio et al. [6] have determined the uncertainty due to measurement covering sources such as implant positioning and probing inaccuracy. They also claim that a maximum of 30 % uncertainty can be attributed to measurements.

Under the assumptions that the above uncertainties are understood as relative standard uncertainties and applied for all the data under consideration, the calculations in sections 3.2 and 3.3 were repeated using identical relative standard uncertainties instead of absolute standard uncertainties. Those subsections used ordinary least squares (OLS). To take account of the above somewhat sparse knowledge of input uncertainties requires weighted least squares (WLS). In the absence of further knowledge, we make the assumption that a relative standard uncertainty $u_{rel}(T) = 0.3$ is applicable to *all* normalized heating rise data. In contrast to determining the least-squares solution of $\boldsymbol{A}\boldsymbol{c} \approx \boldsymbol{T}$ in formulation (3), we therefore seek the least-squares solution of

$$\boldsymbol{W}\boldsymbol{A}\boldsymbol{c} \approx \boldsymbol{W}\boldsymbol{T}. \tag{4}$$

Here $\boldsymbol{W}$ is a diagonal weighting matrix with diagonal elements equal to the product of $u_{rel}(T)$ and the reciprocals of the T_i. As a consequence, since $u_{rel}(T)$ is a common factor on both sides of formulation (4), we seek a vector $\boldsymbol{c}$ such that $\boldsymbol{W}\boldsymbol{A}\boldsymbol{c}$ is the closest vector to the vector of ones. The factor does not disappear completely, however, since it is required in the evaluation of the uncertainty associated with $\boldsymbol{c}$.

Qualitatively, results were comparable with the previous calculations, which assumed implicitly constant and unknown absolute uncertainties. As a figure of merit, the relative RMSR, defined by the ratio of RMSR and the root-mean-square value of normalized temperature rise, was used. For the 1.5 T field, the values of relative RMSR were respectively 56 %, 56 % and 57 % for one, two and three length parameters. For the 3 T field, the corresponding values were 67 %, 67 % and 67 %. For the model including test laboratory biases the relative RMSR values were 41 % for the 1.5 T field and 55 % for the 3 T field. We conclude that an input standard uncertainty of 30 % is inadequate to explain the dispersion in the data analyzed. An input standard uncertainty of some 40 % to 70 % would be more realistic.

A suggestion for future testing is that wherever possible test laboratories should provide (standard) uncertainties or some comparable quantification of the quality of the data they have gathered. In the absence of such knowledge, the starting assumption would be that all data carried comparable absolute or relative uncertainties, that is, the laboratories and the measuring systems they use have identical capability, which is unlikely to be the case.

A further reason for the importance of taking account of uncertainty associated with the data records is the following. The Guide to the expression of uncertainty in measurement (GUM) [4] is widely regarded as the 'bible' in the world of measurement uncertainty. It considers *measurement models*, that is, models that may be mathematical or algorithmic (such as the solution process for a regression problem as in this study), with a number of input quantities and an output quantity. (One of the supplements to the GUM [5] treats multivariate output quantities.) Given estimates of the input quantities and standard uncertainties associated with these estimates, the GUM gives guidance on the estimation of the output quantity and the evaluation of its associated standard uncertainty. An uncertainty evaluation carried out according to the GUM or its supplements has an internationally recognized degree of rigour and acceptability associated with it.

It should be stated that the GUM and its supplements can be applied to regression problems. A thermometry application example is given in the GUM in which formulæ are provided for the intercept and gradient of the OLS straight line that best fits a data set from that area [4, annex H.3]. Those formulæ constitute the measurement model through which uncertainties associated with the data are propagated. Relevant formulæ can be derived for the parameters in fitting functions other than a straight line and uncertainties accordingly propagated through them. Even if the parameters are defined algorithmically, the GUM and its supplements can still be applied [5, clause 3.9].

4.4 Dark uncertainty

It is said in section 4.3 that the calculations of sections 3.2 and 3.3 indicated strong inconsistencies with the assumption of relative input standard uncertainties of 30 %. The implication is that there is considerable unexplained uncertainty. In many measurement situations, measured values are substantially over-dispersed by comparison with their individual, stated uncertainties, thus suggesting the existence of yet unrecognized sources of uncertainty: 'dark uncertainty'. This dark uncertainty [14] can be attributed to unmodelled effects including influencing factors that have not been measured or recorded, but drawing on experience in other areas such as interlaboratory comparisons [9] suggests it could could be caused by many effects, which can be systematic, the overlooking of causes, mistakes, etc. [3]. Such speculation should ideally be tested if possible.

In general, the models used in this paper do not explain the considerable dispersion of the values of normalized temperature rise between test laboratories. Neither do they generally explain the dispersion of values *within* test laboratories. There are exceptions in the latter case: in [13], the data provided by one test laboratory was analyzed independently of all other data records. It was found that for that test laboratory the temperature rise data was explained very well in terms of a polynomial of degree two (quadratic) in primary length for a 1.5 T field, and as a polynomial of degree three (cubic) for a 3 T field. R^2 values of 0.985 and 0.973, respectively, were reported. These results appear to be atypical. For the other test laboratories involved, no comparably good polynomial models were found by the present authors.

4.5 Other possibilities

For the set of data records analyzed, the use of variables normalized in various ways, by primary length, for example, could be considered. Further, models that take account of more factors could be entertained, such as the 11-parameter model given by including the three length parameters and laboratory biases. However, on the basis of some analyses we have undertaken, we believe the results would be changed a little in a quantitative sense but not qualitatively.

5 Conclusions

The Food and Drug Administration of the USA had previously analyzed data records from test laboratories for passive medical implants immersed in gel phantoms in a magnetic resonance imaging (MRI) environment. The analysis was concerned with understanding the extent to which temperature rise around the implant due to radio-frequency (RF) induced heating is related to the primary length (longest reported dimension) of the implant.

The work described here extended that analysis to other factors that may influence RF heating. Those factors include secondary and tertiary implant length, the test laboratories themselves, and prior knowledge of the measurement uncertainties associated with the data.

Parametric models were constructed to study which factors are most important and to provide a capability to predict heating rise for implants whose characteristics lie within the span of the observations on which the models are based.

It was concluded that the limited amount of data prohibited strong statements to be made about the dependence of induced RF heating rise around an implant on its geometric properties and implant type. Considerably more data would need to be acquired for such statements to be possible. Importantly, although a number of test laboratories provided data, there was no evidence that any of the laboratories provided results containing significant bias. Thus, data for all test laboratories could be aggregated for analysis as was done here.

Acknowledgment

Contributions to this paper by the NPL authors was carried out under project JRP 17IND01 MIMAS — Procedures allowing medical implant manufacturers to demonstrate compliance with MRI safety regulations — and has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

This paper has benefited from detailed technical reviews at the authors' organizations.

Disclaimer

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services of the U.S. Government.

References

[1] Eleanor R. Adair, Kevin S. Mylacraine, and Stewart J. Allen. Thermophysiological consequences of whole body resonant RF exposure (100 MHz) in human volunteers. *Bioelectromagnetics*, 24(7):489–501, 2003.

[2] ASTM F2182-11a. Standard Test Method for Measurement of Radio Frequency Induced Heating On or Near Passive Implants During Magnetic Resonance Imaging, 2011.

[3] David C. Bailey. Not normal: the uncertainties of scientific measurements. *Royal Society Open Science*, 4(1):160600, 2017.

[4] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology, JCGM 100:2008.

[5] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Supplement 2 to the "Guide to the expression of uncertainty in measurement" — Models with any number of output quantities. Joint Committee for Guides in Metrology, JCGM 102:2011.

[6] O Bottauscio, A M Cassarà, J W Hand, D Giordano, L Zilberti, M Borsero, M Chiampi, and G Weidemann. Assessment of computational tools for MRI RF dosimetry by comparison with measurements on a laboratory phantom. *Physics in Medicine and Biology*, 60(14):5655–5680, 2015.

[7] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley, New York, 1998. Third edition.

[8] Myeong Seong Kim. Investigation of Factors Affecting Body Temperature Changes During Routine Clinical Head Magnetic Resonance Imaging. *Iranian Journal of Radiology*, 13(4), 2016.

[9] Amanda Koepke, Thomas Lafarge, Antonio Possolo, and Blaza Toman. Consensus building for interlaboratory studies, key comparisons, and meta-analysis. *Metrologia*, 54(3):S34–S62, 2017.

[10] Ignacio Lira, Anna Chunovkina, Clemens Elster, and Wolfgang Woger. Analysis of key comparisons incorporating knowledge about bias. *IEEE Transactions on Instrumentation and Measurement*, 61(8):2079–2084, 2012.

[11] Antonio Possolo. Statistical models and computation to evaluate measurement uncertainty. *Metrologia*, 51(4):S228–S236, 2014.

[12] Frank G. Shellock and John V. Crues. MR procedures: Biologic effects, safety, and patient care. *Radiology*, 232(3):635–652, 2004.

[13] Ting Song, Zhiheng Xu, Maria Ida Iacono, Leonardo M. Angelone, and Sunder Rajan. Retrospective analysis of RF heating measurements of passive medical implants. *Magnetic Resonance in Medicine*, 2018.

[14] M. Thompson and S. L. R. Ellison. Dark uncertainty. *Accred. Qual. Assur.*, 16:483–487, 2011.