Author’s preprint

Submitted to European Journal of Medical Physics May 2022; accepted November 2022

Published as: European Journal of Medical Physics Correspondence| Volume 104, P145-148, December 01, 2022

DOI:https://doi.org/10.1016/j.ejmp.2022.10.013

The Perfect qMR machine: measurement variance much less than biological variance

Paul S Tofts, Brighton and Sussex Medical School, University of Sussex BN1 9PX, UK

1. Summary

Implementing quantitative MR (qMR) methodology can be a time-consuming task, sometimes seemingly without an end. The concept of the Perfect qMR Machine offers the possibility that the implementation is complete and that no further improvements are needed. This is achieved by making the *measurement repeatability variance much less than the biological variance*. Thus the proposal is:

***A Perfect Quantitative MR machine is one that, in making a measurement, contributes no significant extra variation to that which already exists from biological variation*.**

A medal system (platinum, gold, silver and bronze) recognises different sources of biological variance, depending on the type of measurement being carried out (whether a serial study or a group comparison, and different kinds of measurement variance (single machine or multi-centre). A perfect machine can in principle be demonstrated for each quantitative measure (T_1, ADC etc).

2. Introduction

The concept of quantitative MR is perhaps more than 50 years old; in 1971 Damadian measured excised rat tissue and showed abnormal tissue had altered values of relaxation time [1]. In-vivo qMR has been present for over 20 years, first in spectroscopy and then in imaging. In 1985 absolute concentrations of brain metabolites were measured [2]. A book setting out the principles was published in 2003 [3], with a recent new edition [4]. There are books on qMRI in cancer and spinal cord [5,6]. There is a ISMRM study group [7]. Recently the role of qMR measures as biomarkers has been recognised [8,9]. A recent large study shows multi-centre effects [10], and work in qMR continues [10,11,12]. Yet progress seems slow. QMR is not routinely supplied on an MRI machine[1], nor is it used in routine clinical practice. In research centres, MR physicists labour to continually re-implement methods as upgrades and new MRI machines arrive.

Repeatability of measurements within a centre is key in enabling clinical trials to be compact **(figure 1)** and for the method to be clinically useful **(figure 2**). Good multi-centre performance (i.e. reproducibility across centres) is also vital if a method to measure a quantity is to be accepted as

---

[1] There can be complex factors such as vendor interest and concern that the FDA might regulate a quantitative method.

valid. This note addresses how good these parameters need to be. The concepts have already been described in a preliminary form [3, 13].

3. Proposal

Perfect machine: The most important performance characteristic of a qMR method is its repeatability; this determines the smallest biological difference that can be detected. Total repeatability variance is the sum of the intrinsic machine variance, and the biological variance. Thus a qMR machine could be described as *perfect* when the variance produced by the machine (as assessed by repeatability measurements in an unchanging subject) is insignificant compared to the relevant biological variance. For example this could be caused by the subject naturally changing from moment to moment, or by natural between-subject differences when defining a normal range. A further reduction in machine variance would produce no change in its ability to characterise the biology of the subject under investigation.

Medal system: Different kinds of biological variation might be relevant, depending on the kind of measurement or study being carried out (**table 1**). The bronze, silver, gold and platinum progression is intended to denote an increasingly demanding specification on the machine performance. An idealised simple model for the total measurement variance TMV is:

$$\mathrm{TMV} = \mathrm{ISD}^2 + \mathrm{BCV} + \mathrm{WSV} + \mathrm{BSV} \quad (1)$$

Measurement variance is characterised by the ISD (Instrumental Standard Deviation); it can be estimated from pairs of repeated scans in a group of healthy controls (3); BCV = between-centre variance, WSV = within-subject variance and BSV = between-subject variance. Currently BCV is usually much larger than ISD i.e. between-centre differences dominate in a multi-centre study. From this equation can be inferred a progression of hard to relatively easy perfect machine criteria. Four possible situations are revealed, depending on whether BCV and BSV are present (i.e. non-zero) or not, and a complete analysis shows that four medals could in principle be awarded (see table 1).

Platinum medal: The hardest situation would be BCV>0, BSV=0 (i.e. a single subject serial or follow-up study carried out at multiple centres); then $\mathrm{TMV} = \mathrm{ISD}^2 + \mathrm{BCV} + \mathrm{WSV}$ and a group of perfect platinum machines has to achieve:

$$\mathrm{ISD}^2 + \mathrm{BCV} \ll \mathrm{WSV} \quad \text{(platinum)} \quad (2)$$

i.e. the combined instrumental and multicentre variance has to be much less than the within-subject variance. This is currently almost impossible to achieve (see discussion section (ii)).

Gold medal: in a single-subject measurement carried out at a single centre BCV=0, BSV=0, and $\mathrm{TMV} = \mathrm{ISD}^2 + \mathrm{WSV}$. Thus a perfect gold machine would have to meet the criterion of

$$\mathrm{ISD}^2 \ll \mathrm{WSV} \quad \text{(gold)} \quad (3)$$

i.e. the repeatability has to be much better than the within-subject variation.

Silver medal: In a multi-centre group comparison study, all the terms in eqn 1 are non-zero, and a perfect machine would require

$$\mathrm{ISD}^2 + \mathrm{BCV} \ll \mathrm{WSV} + \mathrm{BSV} \quad \text{(silver)} \quad (4)$$

Achieving this depends primarily on reducing BCV; a group of machines could be awarded a silver medal. This is probably less demanding than the gold criterion eq 3, since the between subject variance BSV has now been included. The gold criterion is more desirable, since it allows sensitive single-subject measurements.

Bronze medal: In a single centre group comparison (BCV=0), a perfect machine only has to achieve

$ISD^2 << WSV + BSV$ (bronze) (5)

which is less demanding than the silver criterion. Examples include when a comparison is made with a group of normal ('healthy control') subjects e.g. to assess whether a subject is outside a normal range, and the comparison of groups in a single-centre treatment trial. The repeatability only has to be better than the between subject variation.

4. Discussion

i) Three kinds of normal variation

Here three kinds of normal variation are discussed; these have to be understood and characterised in order to know when a qMR machine can be certified as Perfect. For each qMR parameter there is a normal range and normal variation, measured in healthy controls.

A. Within-Subject Variation WSV. Within-subject variation can be short-term (possibly at time scales of a few seconds, related to cardiac or respiratory factors), medium term (related to diurnal or weekly factors), or long-term (e.g. related to health or body mass changes over months or years). QMR measurements typically take a few minutes, thus they could be used to measure medium- or long-term variation (provided they are sufficiently precise), but probably not short-term variation. Conversely, a MRI machine could be established as perfectly golden in the context of medium- or long-term variation, but not short-term variation (of the order of seconds).

WSV can be measured using healthy controls in many situations; however for some pathologies, repeated studies in patients are needed. WSV will vary from subject to subject, and also depend on factors such as the kind of population and whether measures have been taken to control it (see discussion below on BSV).

Estimation of the Instrumental Standard Deviation (ISD) is affected by WSV; an immediate repeat will only include short term WSV, and is often dominated by image noise; conversely a delayed repeat could include more sources of WSV (depending on the subject), and also more sources of machine instability.

A sophisticated analysis of WSV and ISD has been carried out by Wenger et al [14] using intraclass correlation effect decomposition (ICED) [15], and this may point the way forward to appropriate analysis methods for the future.

B. Between-subject variation BSV. There have been several reports of normal ranges and variation, primarily in T_1, ADC and MTR [16,3], with typical values of 1-5%. These often vary considerably between centres. Any measurements to estimate between-subject variation will necessarily include a contribution from ISD and WSV, which should be subtracted to obtain the intrinsic BSV (see eq 1).

A reported normal range can be inflated in two ways. Firstly a large value of ISD (i.e. poor measurement repeatability) can increase the measured BSV (**figure 3**). Thus credible studies have to use a low and measured ISD (ideally a perfect bronze machine).

Secondly, the measured BSV has a contribution from confounding biological factors (e.g. sex, age and maybe body-mass). By controlling for these factors, the apparent normal range (i.e. BSV) will be reduced, and thus the criterion for achieving a bronze machine will be made more demanding. The confounding factors can be controlled in two ways: firstly, by only selecting subjects with a single value of the factor (e.g. only females aged 43), or secondly by modelling (e.g. with a linear model) the dependence of the MR quantity on the factor and correcting all measurements to a single value of the factor. Large cohort studies such as UK Biobank may help in disentangling these factors and estimating some of the variance components [17].

Further thought and data collection on normal ranges are probably needed, and the characterisation may need larger datasets than have been used up to now.

C. Between-centre variation BCV. BCV in uncontrolled multi-centre studies can be large. This can arise from intended, yet uncontrolled, systematic differences in implementation. This can prevent standard values being used internationally for treatment guidance. Pooling data from centres with poor technique reduces the power of the pooled data, so often there is an incentive to improve the between-centre agreement. BCV can be estimated using a few travelling subjects, or measuring normal values at each centre. With appropriate control of data acquisition and analysis methodology BCV can be small [18]. A silver certificate would be awarded to a group of machines at several centres (not an individual one), and for the group to be useful, each machine should achieve bronze status. The silver group would produce data where there was no detectable difference between the centres. Note that any measurements to estimate between-centre variation will necessarily include contributions from ISD, WSV and possibly BSV, which should be removed to obtain intrinsic BCV (see eq 1).

ii) Implementation of medal system for perfect qMR machines

A perfect machine can in principle be created for each biological or qMR quantity. Where there is large natural biological variation this is relatively easy. Some perfect machines probably already exist. ASL measurements have large within-subject variation [19] so gold medal performance is feasible. Lesion load in multiple sclerosis also shows large WSV over the medium term, and again machines already exist which are performing to a gold standard. However most quantities probably have a low WSV, and achieving a gold machine will be challenging. Silver performance has been achieved for normal MTR histograms [18].

With the magnitude of current multi-centre effects , for most MR parameters the platinum machine is well outside of what might be possible; however it is conceivable that future machines might be so well controlled that this did become possible. A subject in a sensitive serial study could then be measured on any machine that was available, instead of being restricted to always being measured on the same machine.

When an upgrade or new MRI machine arrives at a centre, and a qMR method is to be re-implemented, the Perfect Machine concept could guide this process. The criterion of how much less the measurement variance should be than the biological variance (eqs 2-5 and table 1) would have to be clarified, e.g. the measurement variance should perhaps be $\leq$10% of the biological variance.

Awards could be made in several possible ways. Investigators could claim this by submission of a paper for review in a scientific journal. Applications might come from individuals in research laboratories, or from MRI manufacturers (vendors). Medals could be awarded by radiological or MR groups e.g. ACR (American College of Radiology), RSNA (Radiological Society of North America), ESR (European Society of Radiology) or ISMRM. National metrology institutes could also play a part e.g.

NIST (National Institute of Standards and Technology, USA), NPL (National Physical Laboratory, UK) and PTB (German national metrology institute). Philanthropists might provide support in order to move the field forwards (as for the longitude problem see section iii).

iii) Inspiration and genesis.

The inspiration to move attention away from the activity of the implementation process to a higher-level (meta-) view came from Thomas Mann's book Death in Venice. *"To rest in the arms of perfection is the desire of any man intent upon creating excellence".* The concept of the Perfect Machine comes from the building of the 200 inch Mount Palomar telescope, which was the most perfect telescope that could be built at the time (1933-48). The idea of a medal to recognise and celebrate a machine that would be important for mankind came from the Longitude problem in the 18th century. In 1714 the British parliament offered a prize of £20,000 to anyone who could solve this problem. The concept of a *transparent machine* is attractive and relevant. Our most developed machines operate reliably and without their way of functioning being apparent e.g. voltmeters, weighing machines and automatic cars; a transparent MRI machine has minimal observer involvement and hence maybe less random measurement error. For more discussion see reference [20].

iv) Other considerations

The whole Perfect Machine concept is focussed on short-term random errors (which could arise principally from image noise). Systematic sources of error can also be present (e.g. model over-simplification or varying TE values in a multi-exponential T_2 system, slice selection effects, or machine upgrades). These could contribute to measurement variance if several different machines are used (either within a centre or in a multi-centre study). Thus changes over space (i.e. multi-centre effects) or over time (machine upgrades or replacements) can alter the effects so that they become time-varying and make a contribution to measurement variance.

The partitioning of variance into ISD and BCV (eq 1) is a simplification. They represent two extremes: ISD describes short-term behaviour on a single machine, BCV describes differences between centres. A long-term disease may be studied at a single centre, but as a machine becomes upgraded, or new machines are used at a centre, then between-machine effects come into play, possibly as large as between-centre effects.

Many pathological effects are currently studied using 'imperfect machines'. The biological effect size in disease is often usually greater than the machine variance characterised by ISD and BCV. Thus depending on the application of the qMR measurement, current machines may be quite adequate, and even 'perfect' in the context of the disease as currently understood and treated. The biological effect that is being searched for can be found (e.g. has a tumour substantially reduced its volume?). Nonetheless, an improvement in machine performance would probably open up new opportunities to characterise more subtle aspects of a disease.

Additionally, in some patient groups (e.g. movement disorders) the machine repeatability might be larger than the ISD measured in healthy controls. Thus repeatability does have to be measured in patient groups (as well as in healthy controls). For some diseases measures (e.g. tumour transfer constant K^{trans} value) there is no normal imaging option.

Patient outcomes would improve as measurement variance is reduced. Clinical trials would be smaller, shorter and less expensive (fig 1), giving more rapid access to new effective treatments. In individual patients abnormalities could be seen more easily (fig 2), and more subtle responses to treatment could be apparent.

5 Conclusion: the concept of the Perfect MR machine, and the associated medals, could provide context and give perspective to our work in qMR. It guides us on when further work in improving precision would be beneficial, and when there is no need for this. Clearly a knowledge of the variability from the scanner and other sources increases our ability to measure small changes in quantities associated with disease. Biological variance needs a deeper understanding and characterisation in order to make the most of the possibilities offered by qMR technology.

References

1 Damadian R. Tumor detection by nuclear magnetic resonance. Science 1971;171(3976):1151-3. doi: 10.1126/science.171.3976.1151

2. Tofts, P. S. & Wray, S. 1988, "Noninvasive measurement of molar concentrations of 31P metabolites in vivo, using surface coil NMR spectroscopy", *Magn Reson.Med.*, vol. 6,no. 1, pp. 84-86.

3. Tofts PS. *Quantitative MRI of the brain: measuring changes caused by disease.* New York: Wiley; 2003.

4. Tofts PS . Quality Assurance: accuracy, precision, controls and phantoms. Chapters 1-3 in *Quantitative MRI of the Brain*. Cercignani, Dowell and Tofts eds CRC press 2018

5. Thomas E. Yankeelov (Editor), David R. Pickens (Editor), Ronald R. Price (Editor) Quantitative MRI in Cancer 2011 CRC Press.

6. Julien Cohen-Adad (Editor), Claudia Wheeler-Kingshott (Editor) . Quantitative MRI of the Spinal Cord 2014 Academic Press.

7. Weingartner S, Desmond KL, Obuchowski NA, et al; the ISMRM Quantitative MR Study Group. Development, validation, qualification, and dissemination of quantitative MR methods: Overview and recommendations by the ISMRM quantitative MR study group. *Magn Reson Med*. 2021;00:1–23. doi:10.1002/mrm.29084

8. O'Connor, JPB et al . Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2017 14(3):169-186. doi: 10.1038/nrclinonc.2016.162. Epub 2016 Oct 11.

9. Amita Shukla-Dave et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. J Magn Reson Imaging 2019 Jun;49(7):e101-e121. doi: 10.1002/jmri.26518. Epub 2018 Nov 19.

10 Luca Fedeli et al.. On the dependence of quantitative diffusion-weighted imaging on scanner system characteristics and acquisition parameters: A large multicenter and multiparametric phantom study with unsupervised clustering analysis. *Physica Medica 85 (2021) 98–106*

11. Serena Monti et al. RESUMEN: A flexible class of multi-parameter qMRI protocols. *Physica Medica 88 (2021) 23-36*

12. Mazzoni et al. New developments in MRI: System characterization, technical advances and radiotherapy applications. *Physica Medica 90 (2021) 50–52*

13. Tofts PS British Chapter of ISMRM September 24-25 2018. *Towards the Perfect Quantitative MRI Machine* poster pp11

14. Wenger, E., Polk, S. E., Kleemeyer, M.M., Weiskopf, N., Bodammer, N. C., Lindenberger, U., & Brandmaier, A. M. (2022). Reliability of quantitative multiparameter maps is high for magnetization transfer and proton density but attenuated for R1 and R2* in healthy young adults. Human Brain Mapping, 1–19. https://doi.org/10.1002/hbm.25870

15. Brandmaier, A. M., Wenger, E., Bodammer, N. C., Kühn, S., Raz, N., & Lindenberger, U. (2018). Assessing reliability in neuroimaging research through intra-class effect decomposition (ICED). eLife, 7, 1–19. https://doi.org/10.7554/eLife.35718

16. Tofts PS, Collins DJ. Multicentre imaging measurements for oncology and in the brain. Br J Radiol 2011; 84 Spec No 2: S213-26.

17. https://www.ukbiobank.ac.uk/

18. Tofts PS, Steens SC, Cercignani M*, et al.* Sources of variation in multi-centre brain MTR histogram studies: body-coil transmission eliminates inter-centre differences. Magma 2006; 19(4): 209-22.

19. Parkes LM, Rashid W, Chard DT, Tofts PS. Normal cerebral perfusion measurements using arterial spin labeling: reproducibility, stability, and age and gender effects. Magn Reson Med 2004; 51(4): 736-43.

20. http://qmri.org/resources/perfect-machine

Figure captions

Figure 1

The effect of instrumental precision (ISD) on the power of a study, and the required sample size. By reducing the ISD, the sample size required is dramatically reduced, with a consequent saving in the cost and duration of the study (a simulation from reference [4]).

Figure 2

Simulation showing how the magnitude of ISD affects the ability to use an MR parameter to separate groups and individuals. Group separation is 10 units. With ISD=10 (right hand image), the groups overlap, and considerable statistical power would be needed to separate them (see fig 1). A reduced ISD=3 (centre) gives a good group separation. A further reduction to ISD = 0.5 (left hand) enables individuals to be accurately classified into their group. (from reference [4].

Figure 3

The influence of ISD on measured normal variation, for white matter MTR. Blue circles are published values of SD (units for MTR are pu; mean was 38-40 pu) from 8 centres. *Before* is authors' first value, almost the highest value of 9 centres. After solving a scanner instability problem ISD was low (≃0.2 pu) and the remeasured normal variation (*after)* dropped to the lowest value of 9 centres (from ref [4]).

Table 1 medal system

Medal	Target study	Criterion	BCV BSV values
platinum	serial study (multicentre) (awarded to group of machines)	$ISD^2 + BCV << WSV$	BCV>0; BSV=0
gold	serial study at one centre	$ISD^2 << WSV$	BCV=0; BSV=0
silver	group comparison (multicentre) (awarded to group of machines)	$ISD^2 + BCV << WSV + BSV$	BCV>0; BSV>0
bronze	group comparison at one centre	$ISD^2 << WSV + BSV$	BCV=0; BSV>0

Table notes:

ISD = Instrumental standard deviation

BCV = between-centre variance

BSV = between-subject variance

WSV = within-subject variance

Figure 1:

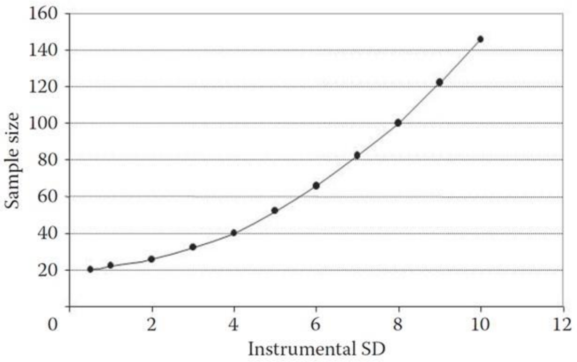


Figure 2:

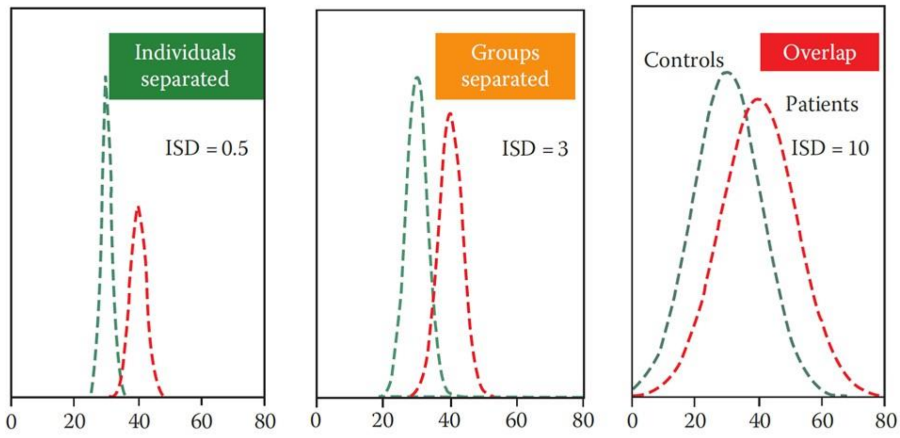
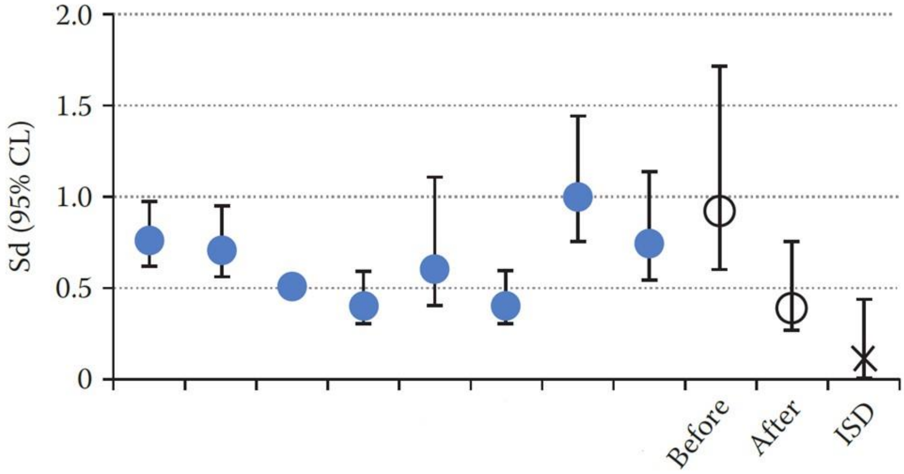


Figure 3:



END of note