Published in final edited form as:

Adv Exp Med Biol. 2016 ; 923: 113–120. doi:10.1007/978-3-319-38810-6_15.

Comparing the Effectiveness of Methods to Measure Oxygen in Tissues for Prognosis and Treatment of Cancer

Ann Barry Flood, Victoria A. Satinsky, and Harold M. Swartz

EPR Center for the Study of Viable Systems, Department of Radiology, HB 7785, Geisel School of Medicine at Dartmouth, Hanover, NH, USA

Abstract

Given the clinical evidence that hypoxic tumors are more resistant to standard therapy and that adjusting therapies can improve the outcomes for the subpopulation with hypoxic tumors, *in vivo* methods to measure oxygen in tissue have important clinical potential. This paper provides the rationale for and methodological strategies to use comparative effectiveness research to evaluate oximetry for cancer care. Nine oximetry methods that have been used *in vivo* to measure oxygen in human tumors are evaluated on several clinically relevant criteria to illustrate the value of applying comparative effectiveness to oximetry.

Keywords

Comparative effectiveness; Clinical oximetry; EPR (electron paramagnetic resonance); Tumor hypoxia; Cost effectiveness

1 Introduction

There is a considerable body of evidence to support the observation that solid tumors, especially malignant tumors, exhibit a very low oxygen concentration compared to normal tissue of a comparable type [1]. Hypoxia in tissues develops from a variety of factors, including the well-described poor vascularization and structural abnormalities in malignant tumors, which result in poor perfusion and diffusion of oxygen in the tumor [2, 3]. Likewise, the evidence is strong that hypoxia in cancer has very important clinical implications for predicting long-term patient outcomes, i.e., hypoxia is associated with poorer outcomes irrespective of type of treatment. Perhaps more importantly, for the subpopulation of patients whose tumors respond, there is strong evidence that the therapeutic ratio can be improved when radio- or chemotherapy is augmented by appropriately timed administration of hyperoxic therapy (such as breathing enriched oxygen or carbogen) and/or vasoactive agents (such as nicotinamide) or cytotoxic agents that function in the presence of hypoxia (such as tirapazamine) [2, 4–7].

The scientific evidence about the mechanisms leading to hypoxia, the biological consequences of its presence (such as increased insensitivity to radiotherapy or

chemotherapy, the proliferation of aggressive malignancies that can thrive in low oxygen), and the clinical importance of reducing hypoxia during therapy are all substantial and highly relevant to the importance of our thesis. However, these topics per se are not reviewed in depth here. Instead, the focus of this paper is how the clinical implications of these factors can be enhanced by greater application of comparative effectiveness methodology to meaningfully compare and contrast available methods to assess hypoxia for practical applications to improve cancer care.

Comparative effectiveness research is broadly defined as research into the relative effectiveness of two or more health care practices which can be used in a ‘real world’ clinical setting to treat a specific disease or condition [11]. Three primary strategies are used to compare effectiveness, based on assessing the risks and benefits of each treatment being evaluated: by synthesis of the literature and modeling, by observational studies, and by head-to-head trials [11, 12].

Comparative effectiveness methodology is an outgrowth of clinical studies designed to establish the effectiveness of a new clinical approach ranging from screening to diagnostics to therapeutics [8–10]. The term efficacy, in contrast to effectiveness, is generally used to differentiate outcomes of a treatment when performed under an idealized, highly controlled protocol for care in contrast to what can be achieved in conditions of usual clinical practice. Since their introduction in the 1940s, randomized controlled trials (RCTs), which typically compare a new therapeutic approach to a standard therapy or placebo, have been the ‘gold standard’ for determining the efficacy of a new strategy. Because RCT designs randomly assign eligible subjects to treatment arms, follow strict protocols, and have the potential for blinding the analysis, they eliminate selection bias and have strong internal validity.

However, while RCTs are well suited for assessing whether an individual method represents an improvement over random chance or the standard therapy, they do not address several important pragmatic issues regarding the generalizability of the results when used in usual care. In other words, because RCTs use strictly defined and controlled protocols for care, the results may only apply to idealized (i.e., ultra-controlled) clinical practices and may only be feasible to carry out in large medical centers with strict oversight of trials rather than in usual care settings such as in community hospitals or clinics. Moreover, intrinsic to the statistical designs governing RCTs, the analyses of outcomes in RCTs focus on assessing the average effect on patients. As a result, they cannot examine whether only certain subpopulations may have benefitted, without changing the underlying statistical assumptions. In addition, RCTs are not well suited to compare several widely disparate approaches to treatment of the same disease such as comparing prostate cancer treatments as varied as surgical intervention, radiation therapy, hormonal therapy or watchful waiting. They also ignore such practical issues as cost differences in the expenditures on resources to achieve any improvements in outcomes and ignore aspects that might lead patients or their doctors to prefer one strategy over another. Indeed, the rise of patient-centered care, with its emphasis on including patient preferences and/or in individualizing care based on biological variations in the patient’s tumor or its response to treatment, is arguably not well-suited to RCTs because they use a priori assignment of all eligible patients to treatment arms and do not permit deviation related to patient responses. Finally, RCTs are not designed to evaluate

multiple methods, all of which are still at an investigational development stage, i.e., they are not designed to assess the relative effectiveness or to compare the practical feasibility of methods still being developed for use in standard clinical care.

For all these reasons, alternative methodological approaches have emerged as important tools for clinical research to be able to answer these more complex and practical problems when assessing the clinical feasibility and other practical implications of alternative strategies. While several approaches have been proposed to assess the clinical implications, politically motivated concerns have led to the recent focus on *comparative* effectiveness to address these issues, especially in contrast to *cost* effectiveness research. For example, in the US, the federal stimulus bill of 2009 explicitly targeted that clinical studies use comparative effectiveness methodology. Subsequently, as a part of the 2010 major healthcare reform law informally called “Obamacare”, a new federal initiative called the Patient-Centered Outcomes Research Institute (PCORI) was established to conduct comparative effectiveness research. Included in this legislative mandate was an explicit prohibition that PCORI **not** include cost effectiveness thresholds in its studies. For example, policy-makers raised concerns that third party payers of health care might emphasize research that supported paying ‘only’ for less costly alternatives.

While some have argued that comparative effectiveness research, like RCTs, neglects patient-centered care, others argue that, because comparative effectiveness studies typically enroll a heterogeneous population, they are well-positioned to uncover subpopulations that can benefit most from particular treatments [13]. Likewise, effectiveness research can help identify treatments or services with essentially equivalent effectiveness, e.g., the likelihood of achieving 5-year disease-free status or survival. Where there is such equipoise in outcomes, i.e., where there is no dominant reason related to outcomes to choose one treatment over another, patient preferences for choosing one treatment over the other can be incorporated, thereby usually improving patient satisfaction with care, compliance with the treatment regimen and sometimes improving outcomes as well [14].

Yet, while there has been an important shift in emphasis to promote pragmatic clinical studies, the scientific body of clinical studies of outcomes remains overwhelmingly focused on RCTs. For example, Holve and Pittman [11] examined all of the clinical studies reported in a US federally mandated registry of clinical trials, [Clinicaltrials.gov](https://clinicaltrials.gov) [15]. Out of approximately 53,000 records on [Clinicaltrials.gov](https://clinicaltrials.gov) (with ~13,000 added annually), the overwhelming percent (~85 %) are RCTs, most of which are early Phase 1 or 2 studies, i.e., are preclinical studies or early safety trials. Of the approximately 1700 active Phase 3 or 4 studies (the majority involving cancer), fewer than one-third were designated as comparative effectiveness research. This provides evidence that, despite attempts to promote pragmatic and comparative studies of the clinical implications of alternative strategies, there is a general need for greater application of these methods to evaluate clinical care across the board.

2 Methods

More specifically on the topic of assessing tumor hypoxia and attempts to alter it during treatment, we found only one clinical study that attempted to directly compare the effectiveness of different methods for assessing hypoxia [21]

Nonetheless, we argue that the current status of technology and research supports turning to comparative effectiveness methods to help address some very important questions about when and whether to use oximetry, and about which oximetry method to use. Namely, there is evidence that methods to assess hypoxia in tumors have the potential to be clinically meaningful in at least three scenarios: (1) because hypoxia is associated with important long-term patient outcomes including death, it has prognostic value for physicians and patients and their families; (2) if treating physicians could identify the subset of patients whose tumor is resistant to radio- or chemotherapy, they could adapt their treatment to be most effective; and (3) if physicians knew both that a patient's tumor was hypoxic and that it responds to hyperoxic treatment, oximetry could help with the timing to administer care to improve outcomes. Secondly, there is an important need to compile evidence about the comparative effectiveness of oximetry methods. While about two dozen oximetry methods to assess hypoxia have been proposed, many have not been successfully applied to humans *in vivo* and most, if not all, have yet to establish their effectiveness in clinical settings [2].

Another reason that comparative effectiveness methods need to be applied to oximetry is because they are better suited than RCTs to address whether and when subpopulations of patients may experience better outcomes. Many have argued that hypoxia and individualized responsiveness to hyperoxic therapy are the major explanation as to why standard therapies or new variants appear to fail when their effectiveness is evaluated for the 'average' patient. It is because the subpopulation with hypoxic tumors require a different approach [2–5, 7, 16]. Oximetry could help establish when hypoxia is present, in order to validate these hypotheses in subpopulations.

Finally, oximetry methods need to be compared on the basis of practical considerations, such as the capability of the oximetry method to measure tumors at various depths, the sensitivity of the method to assess hypoxia at the most clinically sensitive levels of partial pressure of oxygen (pO_2), and the practicality of using the method during cancer therapy or during application of concurrent hyperoxic therapy (such as the rapidity with which measurements can be repeated or the degree to which specialized facilities and equipment would be needed to carry out oximetry). Comparative effectiveness can address all these important pragmatic questions as well as considering patient oriented preferences—such as the relative discomfort and invasiveness of oximetry methods.

3 Results

The first comparative effectiveness strategy to compare oximetry methods consists of a synthesis of publications, including prior reviews. While there are several excellent reviews related to oximetry methods, most focus on the evidence and mechanisms of how tumors become hypoxic, why hypoxia leads to complex responsiveness to primary therapy, and

evidence from cancer treatment studies [2, 3, 17, 18]. A few reviews include or focus on the potential of various types of oximetry to be used in clinical care [2, 4, 19].

In synthesizing the literature two guiding principles are key: (1) determine the methods that are most appropriate to compare for the same type of disease or condition and (2) decide on key criteria to evaluate. For purposes of this paper, we began with the list of 23 oximetry methods in Walsh and colleagues' "Table 4: existing techniques for *in vivo* assessment in tissue oxygenation" ([2], p 1525)). We selected the nine methods that have been or are being currently used in humans with tumors. Table 15.1 presents these nine methods, organized by whether the method directly assesses hypoxia or uses qualitative or indirect methods, such as those that evaluate the expression of endogenous markers in the presence of hypoxia. (See Swartz et al. in this same volume for further description of these methods [20].)

Table 15.1 compares these methods based on four criteria. Two criteria describe the capacity of the method to assess hypoxia (how deep below the surface the method can measure a tumor and how well the method can quantify the extent of hypoxia—the latter is especially needed to assess change in hypoxia in the presence of hyperoxic therapy). The third criterion, invasiveness of the method, is both a patient-centered indicator as well as a potential source for side effects. The last criterion is whether the measurement can be repeated, with the assumption that this capacity is more conducive to monitoring the effects of treatments.

Table 15.1 shows that three of the nine methods currently are limited as to the depth of the tumor they can assess and, while five assess the amount of hypoxia, only three are precise. The Eppendorf electrode probe is the only method labeled as significantly invasive in this table. Others, however, are minimally invasive. For example, the carbon-based EPR particulates leave a permanent tattoo, and all EPR particulates require being injected or implanted into the tumor initially but thereafter are noninvasive. Likewise, several imaging technologies require administering a tracer, usually intravenously, for each measurement session.

Repeatability in this table applies to all but one method. However, more subtle descriptions of repeatability, such as the ease and rapidity with which measurements can be repeated for monitoring during treatment, are not noted here. A practical consideration, such as how specialized the facility is or how cumbersome the device would be to co-locate during radiation therapy, may also preclude making rapidly repeated measurements in usual clinical care.

While observational studies, the second strategy of comparative research, are not detailed here, the practical considerations as well as strengths and limitations of a given oximetry method are not easy to discern in the literature to date. The third strategy, head-to-head comparisons of different oximetry methods to assess the same set of patients with the same condition, has seldom been carried out. A notable exception is a head-to-head comparison of five marker assays regarding their ability to predict local-regional control [21]. The authors reported diversity and little correlation among the markers within individual tumors of 67 head and neck patients in their study.

4 Conclusions and Implications

There is strong evidence that hypoxia in tumors is prognostic of poorer outcomes and can influence clinical treatment decisions, including the decision to use hyperoxic therapy concurrently with standard treatment to improve responsiveness to radio- and chemotherapy. Indeed tumor hypoxia is a powerful and independent prognostic factor, even when stage, tumor size and treatment type are considered [22]. More importantly for this paper, not all tumors of the same type exhibit hypoxia and not all respond to hyperoxic therapy. Consequently, there is an important need for practical, effective oximetry methods. While about two dozen oximetry methods have been proposed, not all are equally available or well suited for all types and depths of tumors, and some may not be practical in standard clinical practice. Comparative effectiveness strategies and methodologies offer an important means to rigorously compare and contrast oximetry methods, even at the stage of early clinical phases of research.

Besides the obvious implications of guiding treatment of cancer, a systematic evaluation of oximetry methods could result in some methods being invaluable in preclinical studies to understand the mechanisms of how hypoxia influences tumors even though not well suited for clinical studies or standard care. If one type of oximetry method is clearly more efficient and effective in predicting success in treatment, then there would be a ‘dominated’ oximetry winner—at least for some types of cancer. If instead there are several oximetry methods comparable in their effectiveness, other reasons to choose among them may begin to predominate, e.g., based on patient-based preferences such as the convenience or time to undergo oximetry measurements or based on physician or health care system related reasons, including having sufficient patient volume to justify the purchase of equipment or appropriate expertise available to administer the method.

Head-to-head clinical studies for comparative effectiveness research, like RCTs, can be very expensive to carry out [11] and so there is an imperative to use systematic and rigorous methods to evaluate the evidence already available in the literature wherever feasible. These other strategies developed for comparative effectiveness research can help provide the needed insights while avoiding the biases in observational studies carried out independently. Especially in complex situations where the underlying scientific and medical mechanisms are themselves evolving and there is no real ‘gold standard’ for assessing hypoxia, comparative effectiveness research needs to include the insights and perspectives of researchers who are cognizant of the problems with measuring oximetry as well as with the complexity of intervening on biological mechanisms.

So the challenges to apply comparative effectiveness to evaluate oximetry methods in cancer care are great, but so is the importance of the task to identify subpopulations with hypoxia in order to improve their outcomes.

Acknowledgments

This work was funded in part by a grant from the US National Cancer Institute, National Institutes of Health (PO1CA190193).

References

1. Vaupel P, Höckel M, Mayer A. Detection and characterization of tumor hypoxia using pO₂ histography. *Antioxid Redox Signal*. 2007; 9:1221–1236. DOI: 10.1089/ars.2007.1628 [PubMed: 17536958]
2. Walsh JC, Lebedev A, Aten E, et al. The clinical importance of assessing tumor hypoxia: relationship of tumor hypoxia to prognosis and therapeutic opportunities. *Antioxid Redox Signal*. 2014; 21:1516–1554. DOI: 10.1089/ars.2013.5378 [PubMed: 24512032]
3. Vaupel P, Mayer A. Hypoxia in cancer: significance and impact on clinical outcome. *Cancer Metastasis Rev*. 2007; 26:225–239. DOI: 10.1007/s10555-007-9055-1 [PubMed: 17440684]
4. Wilson WR, Hay MP. Targeting hypoxia in cancer therapy. *Nat Rev Cancer*. 2011; 11:393–410. DOI: 10.1038/nrc3064 [PubMed: 21606941]
5. Menon C, Fraker DL. Tumor oxygenation status as a prognostic marker. *Cancer Lett*. 2005; 221:225–235. DOI: 10.1016/j.canlet.2004.06.029 [PubMed: 15808408]
6. Bussink J, Kaanders JH, von der Kogel AJ. Tumor hypoxia at the micro-regional level: clinical relevance and predictive value of exogenous and endogenous hypoxic cell markers. *Radiother Oncol*. 2002; 67:3–15. DOI: 10.1016/S0167-8140(03)00011-2
7. Janssens GO, Rademakers SE, Terhaard CH, et al. Accelerated radiotherapy with carbogen and nicotinamide for laryngeal cancer: results of a phase III randomized trial. *J Clin Oncol*. 2012; 30:1777–1783. DOI: 10.1200/JCO.2011.35.9315 [PubMed: 22508814]
8. Gatsonis C. The promise and realities of comparative effectiveness research. *Stat Med*. 2010; 29:1977–1981. DOI: 10.1002/sim.3936 [PubMed: 20683886]
9. Goss CH, Tefft N. Comparative effectiveness research—what is it and how does one do it? *Paediatr Respir Rev*. 2013; 14:152–156. DOI: 10.1016/j.prrv.2013.06.006 [PubMed: 23890806]
10. Armstrong K. Methods in comparative effectiveness research. *J Clin Oncol*. 2012; 30:4208–4214. DOI: 10.1200/JCO.2012.42.2659 [PubMed: 23071240]
11. Holve, E., Pittman, P. A first look at the volume and cost of comparative effectiveness research in the United States. *Academy health*. 2009. http://www.academyhealth.org/files/filedownloads/ah_monograph_09final7.pdf. Accessed 30 Aug 2015
12. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA*. 2003; 290:1624–1632. DOI: 10.1001/jama.290.12.1624 [PubMed: 14506122]
13. Epstein R, Teagarden JR. Comparative effectiveness and personalized medicine: evolving together or apart? *Health Aff (Millwood)*. 2010; 29:1783–1787. DOI: 10.1377/hlthaff.2010.0642 [PubMed: 20921476]
14. O'Connor AM, Llewellyn-Thomas HA, Flood AB. Modifying unwarranted variations in health care: shared decision making using patient decision aids. *Health Aff (Millwood)*. 2004; Suppl Variation: VAR63-72. doi: 10.1377/hlthaff.var.63
15. Clinicaltrials.gov. U.S. National Institutes of Health; Rockville: 2015. <http://www.clinicaltrials.gov>. Accessed 25 Aug 2015
16. Dhani N, Fyles A, Hedley D, et al. The clinical significance of hypoxia in human cancers. *Semin Nucl Med*. 2015; 45:110–121. DOI: 10.1053/j.semnuclmed.2014.11.002 [PubMed: 25704384]
17. Koch CJ, Evans SM. Optimizing hypoxia detection and treatment strategies. *Semin Nucl Med*. 2015; 45:163–176. DOI: 10.1053/j.semnuclmed.2014.10.004 [PubMed: 25704388]
18. Overgaard J. Hypoxic modification of radiotherapy in squamous cell carcinoma of the head and neck—a systematic review and meta-analysis. *Radiother Oncol*. 2011; 100:22–32. DOI: 10.1016/j.radonc.2011.03.004 [PubMed: 21511351]
19. Springett R, Swartz HM. Measurements of oxygen in vivo: overview and perspectives on methods to measure oxygen within cells and tissues. *Antioxid Redox Signal*. 2007; 9:1295–1301. DOI: 10.1089/ars.2007.1620 [PubMed: 17576162]
20. Swartz, HM., Williams, BB., Hou, H., et al. Direct and repeated clinical measurements of pO₂ for enhancing cancer therapy and other applications [submitted same issue]. *Advances in experimental medicine and biology: oxygen transport to tissues XXXVIII*. 43rd annual meeting of the

- International Society on Oxygen Transport to Tissue; Wuhan, China. 11–16 July 2015; 2015. (in press)
21. Nordsmark M, Eriksen JG, Gebiski V, et al. Differential risk assessments from five hypoxia specific assays: the basis for biologically adapted individualized radiotherapy in advanced head and neck cancer patients. *Radiother Oncol.* 2007; 83:389–397. DOI: 10.1016/j.radonc.2007.04.021 [PubMed: 17499868]
 22. Vaupel P. Tumor microenvironmental physiology and its implications for radiation oncology. *Semin Radiat Oncol.* 2004; 14(3):198–206. doi: <http://dx.doi.org/10.1016/j.semradonc.2004.04.008>. [PubMed: 15254862]

Table 15.1

Methods of oximetry used in vivo in humans and criteria for effectiveness research

Criteria: Oximetry method:	Depth	Quantitative/Qualitative	Invasive?	Repeatable?
A. Direct assessment of hypoxia (O ₂ concentration or pO ₂)				
Polarographic oxygen electrode (Eppendorf)	3–4 cm	Quantitative	Yes	No
EPR oximetry (spectroscopy with particulates):				
Carbon- or phthalocyanine-based paramagnetic sensors	1–2 cm ^a	Quantitative	No ^b	Yes
Implantable wires with phthalocyanine-based sensors	Any	Quantitative	No ^b	Yes
B. Other assessment of hypoxia				
MRI/MRS techniques:				
DCE-MRI	Any	Indirect: not applicable	No	Yes
BOLD	Any	Indirect: not applicable	No	Yes
Diffusion and perfusion assessment	Any	Indirect: not applicable	No	Yes
Near-infrared (NIR) spectroscopy (multiple modalities)	1–6 cm	Indirect: not applicable	No ^b	Yes
PET:				
¹⁸ F tracers: FMISO, etc.	Any	Qualitative	No ^b	Yes
⁶⁴ Cu-ATSM	Any	Qualitative	No ^b	Yes

Acronyms: *BOLD* blood oxygen level dependent, *DCE* dynamic contrast enhanced, *EPR* electron paramagnetic resonance, *PET* positron emission tomography, *MRI/MRS* magnetic resonance imaging/¹⁹F-magnetic resonance spectroscopy, *FMISO* fluoro-misonidazole, *Cu-ATSM* Cu(II)-diacetyl-bis(*N*⁴-methylthiosemicarbazone)

^a EPR Probes can be placed intra-cavity so that depth can apply to oral cancer, uterine or cervical cancer, or colorectal cancers as well as at same depth below skin

^b Method is minimally invasive. EPR carbon-based particulates produce a permanent tattoo. EPR particulates need to be injected or implanted initially but thereafter are measured noninvasively. Tracers need to be injected intravenously prior to each measurement session