

Journal of Participatory Medicine

Podcast

Can We Trust Traditional Peer Review?

[Peter Frishauf](#)

[Richard Smith](#)

[Liz Wager](#)

[Alex Jadad](#)

[Thomas \(Bo\) Adler](#)

Defining the Problems and Issues with Peer Review Today

Peter Frishauf:

Welcome to all of you. You are eminently qualified to participate in a discussion of peer review and quality in medical journals. I am going to start off the discussion by asking you, just in general, what the problems and issues you see with peer review, as being the most common in scientific publishing today. Liz, why don't you kick that off?

Liz Wager:

Okay. I think that one of the problems is we don't know what we are expecting peer review to do, and we expect it to do too much. I think most people who've been accepted agree that the comments quite often help to improve the article. But the process of filtering, of deciding which journal it's going to go to, is really not much better than random. So, we need to decide what we want peer reviews to do before we can decide if it's working or not.

Peter Frishauf:

In your publications Liz, you are actually quite specific about what some of the expectations are in terms of improving objectivity, in terms of eliminating fraud, in terms of being a quality filter for improving an article in general, and also doing it in a rather fast manner so that the information actually can get out there in less than years, which clearly does not serve the public or scientific interest. You come down pretty hard, present a lot of data that peer review has really failed in major ways in all of those areas. Could you expand on that a little?

Liz Wager:

Yes. Sadly, it's lousy as picking out really extreme fraud, fabricated data, and really extreme misconduct; it has a very poor track record. I think you might argue that it's not set up to do that, but there are still expectations that somehow the public, the reader, will be protected, because something has been peer-reviewed. Sadly, that's just not the case. I think in terms of the quality filter, there is some nice work by Bob and Suzanne Fletcher. They pull together various studies that look at how often reviewers have actually agreed whether to accept or reject an article.

Statistically speaking, you'd actually need six reviewers all agreeing to accept or reject an article to make it statistically robust. We never do that; we use two or three. We are really applying double standards. We are pretending that it's an accurate diagnostic tool whereas in fact, it's not. I really wish, in some way, that we had more courage and said it's not really about selecting what goes in my journal; I choose what goes in the journal, that's what being an editor is about.

It's about picking up little problems and improving the article slightly on top of getting advice from some people with different opinions. But, it's not an accurate way of deciding where things should be published. And then, there is the argument about should we be filtering stuff anyway, isn't it simply better with the web to simply have the information out there so that the reader can decide if it's any good or not, and let them be the final judge?

Peter Frishauf: Richard Smith, you were from 1979 to 2004 working for the British Medical Journal, and writing a book called *The Trouble with Medical Journals*. Your perspective?

Richard Smith: I think, unfortunately, peer review is an empty gun. I think it's largely useless, and it's taken us a very long time to recognize that. I think one of the things that has helped us is the research that's going on in the past twenty years, which never really happened before as far as peer review being as central to the process it sounds. I think that research has come up with endless problems, and has really shown very little benefit. So, let me quickly run through the problems as I see it. It's horribly slow and everybody knows that. Even in this age of the internet, a lot of them take more than a year for something to be published which is kind of crazy and might be tolerable if we thought there were other great benefits, but it's very hard to find those benefits.

Secondly, it's very expensive; most people spend time peer reviewing for free, but if you actually think of the amount that you might need to pay those people, it's a hugely expensive process and wasteful process, particularly when we are thinking of journal articles. Many journals, like the BMJ, like the New England Journal of Medicine have rejection rights well over 90%. There are lots of papers being reviewed that they are not going to publish. Those papers then get submitted somewhere else; they are reviewed again and that can go on several times. As Liz said, the process is something of a lottery that in many, many studies it is showing that reviewers often don't agree much more than we expect by chance.

I think the process is anti-innovatory; it's a lower common denominator process. You come up with a really brilliant idea; it's quite likely to be rejected. We know that it's very poor at detecting errors. We did a lot of studies at the BMJ where we inserted deliberate errors into articles, and then sent them to people. Many reviewers didn't spot any of the errors at all and

very few people spotted more than a few of them. We know, as Liz said, that it's hopeless at diagnosing trolls; there is good evidence of bias, particularly bias against the provincial, possibly bias against women. People in the developing world feel that the system is biased against them. When it's time to abuse, it's rather easy to abuse the peer review system, particularly if it's conducted as it usually is—confidentially. So, I think we have huge numbers of problems and we have very solid evidence of the problems; and we don't really have much evidence of benefit.

Certainly, people will tell you stories about papers they've submitted which have been improved, they think, but I worry that for every one that's improved, there are probably others that are made worse. This is a system that we should think very seriously about abandoning. A lot of people recognize the problems with peer review, but the usual excuse is that it's the least bad system, like democracy, but I wonder if we can't come up with something better.

Peter Frishauf:

Okay. It's slow, expensive, relies on the lowest common denominator that destroys innovative ideas. It doesn't identify errors, fraud or bias, and it's full of gamesmanship. That sounds like a wonderful system for our most prestigious STM (scientific, technical, and medical) journals to adopt! So yes, I think most of us, probably all of us, agree that it's broken and we need to fix it. So Alex, you've done a lot of work in medicine and public health on creating, with your current focus, virtual tools that you say will support the encounter between the public and the health system—to promote knowledge translation and mentorship of health professionals. But of course, none of that works if the underlying information is invalid and untrustworthy. So, what's your take on peer review as we have it today?

Alex Jadad:

It is ridiculous and difficult to justify. A lot of my work these days has to do with the role that young people play in the world. When I share with them how we decide what is valid and how we go through this big process to end that in what we still call a paper, they laugh! They find it absolutely ridiculous, and they are wondering who would be brave enough to get rid of the current system. So, one of my key questions for today is whether the Journal of Participatory Medicine would be that bold group that would lead the way into a future in which the notion of peer review is transformed completely? If you allow me, I would like to emphasize the fact that we have it and we still claim that it's the best of all the available options. We are still very much mentally stuck in a previous era, in an era in which printing capacity was limited, when we had limited amounts of space to present the information to an audience, and when we had limited access to sources of knowledge. All those limitations have disappeared.

Peter Frishauf:

We are definitely going to be talking about some of the proposed solutions and our hope is very much that the Journal of Participatory Medicine can be an innovator, both in creating new forms of peer review—which the journal itself follows—and perhaps also calling together a conference to critically evaluate what's going on. That's been done lots of times, to bring people from many disciplines of science together to discuss new forms and different forms of peer review that may be relevant to medical and STM publishing.

We are going to be talking about that in just a few minutes, but, before we do that I want to introduce Bo Adler, who does not come from the biomedical field. He is a PhD candidate at the University of California in Santa Cruz. But, Bo is no student; he is forty years old and has lots and lots of credentials in terms of studying text evolution in different publications. He recently came up with a methodology to try and assess trust of Wikipedia contributors and was featured in Wired Magazine recently on wired.com. He is the co-inventor on twelve patents, and also does work in healthcare technology in the area of creating sensors—biomedical sensors at Fujitsu Labs. Bo, you are from outside of medicine. What's your take on these views?

Bo Adler:

I pretty much agree with what everyone else said. Of course, in computer science, we have peer review too. In my own experience with trying to get my papers published is that sometimes the reviewers are helpful and have useful feedback. But, sometimes the reviewers just don't 'get' the paper; they either haven't spent the time with it or they just skim through it really quickly. The feedback that they give makes it really clear that just didn't understand what the paper was about or they bring up criticisms that were directly addressed in the paper that they skipped over. And, sometimes that's because the paper is dense. There are lots of potential problems there.

All the points that you bring up are really good, but I have this question about what's the—taking a step back—what's the purpose of the journal. For readers I think that the purpose of a journal is for, or for scientist readers, is for bringing together a bunch, a collective work that represents the leading edge of where science is going for that particular field right now. In that respect, I think that peer review serves as a filter to help the editors figure out what are the most relevant things for the target audience of the journal. That's a good way to think about the problem, that it's trying to figure out what is the good stuff that we should be looking at right now. But, I agree with all the points you brought up about that. It will definitely help detect fraud.

In my early graduate work, one of the things that they have us do a lot is recreating earlier works from previous papers, and then extending them just a little to give us practice on how science has typically been done. One of the things that I learned is pretty much every single paper where I try to recreate the work, there is either a mistake in the paper or some big gap that they left out that. Something that should have been obvious when you were reading the paper, but, when you actually try to go and do it, you are scratching your head and saying 'I don't actually see what they did,' and you have to take a guess. It's supposed to be reproducible, but it's difficult to reproduce actually, with the way that we have it right now. Partly, that's because of the limitations in the size for what you publish, but also just communication issues.

Light Versus Heavy Peer Review

Peter Frishauf:

Should peer review be light, lightweight really? Considering perhaps not whether a paper is original or important, but if the conclusions don't run ahead of the methods and results, or heavy, and just let them get out there and rely on a lot of post-publication peer review when an article is published online. Also, should authors know the name of reviewers and readers know the names of reviewers? Let's start with you, Liz.

Liz Wager:

Yes. I love the idea of post-publication peer review, but sadly up until now, it really has just been an idea. Various journals have experimented with it. The Medical Journal of Australia tried a sort of social peer review and found people didn't comment very much, and what they did have to say wasn't very useful. Nature did an experiment in really open peer review, i.e. just flinging an article up there, and seeing if people would comment and it didn't work. I think the idea is great; I like the idea. Wikipedia has been a fantastic model. People have contributed and they've contributed something useful. I don't quite know why, but we don't seem to have that working in peer review yet, and I'll be interested to know whether other people have theories as to why that hasn't worked.

Peter Frishauf:

Okay, go ahead.

Liz Wager:

Should peer review be light or heavy? I think you have to say *well, why are we doing it at all, you know what is the purpose?* Is it actually to improve the language, the readability of the paper? Oddly, although I am not sure that peer review works, I think technical editing and subediting can sometimes be very helpful. And, I might almost come down the side of

saying it really helps people to get their message across. But, in terms of spending a lot of time and lot of money, asking lots of different people about their views on a paper, then, no, I am not sure that's really useful at all.

Peter Frishauf:

Okay. Alex?

Alex Jadad:

I couldn't agree more. If that didn't work as we have it, and in fact we may be doing more harm than good due to the speed and the biases that have been incorporated time and time and time again, why don't we get rid of it?

Peter Frishauf:

Okay. So, you really think these other questions are irrelevant?

Alex Jadad:

Yes.

Peter Frishauf:

But, if we do, reviews—should they be lightweight or heavy and rely on post-publication peer review? Should authors and readers know the identity of reviewers?

Alex Jadad:

I think if what we are doing is part of a so called democracy, it's down to the reader to decide whether they should be named or not. Who are we to decide that? I think we are still displaying a lot of the paternalism that motivated the creation of the system in the first place. I think we are still treating the readers with disrespect and I think we are overvaluing what we are doing within journals. When we look at the citation impact factor, if we believe in those metrics for our top journals and consider them, that's an amazing achievement, then we are missing the point altogether. I think it's time for boldness. It's time for us to match the expectations and needs of the public. I cannot see any reason why we couldn't do something dramatically different and probably more consistent with what we already know.

Peter Frishauf:

Okay. Bo?

Bo Adler:

One of the things I realized while listening to you guys talk some more is that when we say peer review, right now we are all talking about the process that happens when papers are selected to be published in a journal. When I think back to my group and how we do our papers, there is this step of peer review that happens afterwards. Once you are published, everybody else gets to see your papers. And, when we read papers that are in our area, if

we disagree with the ideas we propose a different idea. We go and do the research and write what we call an answer. It's not formalized; you don't directly say 'oh, this is in response to,' but usually in your related works area you'll refer to the other paper and describe what they did, and why your idea is better. That's also peer review. I think that that's part of what science is about.

Peter Frishauf:

I think one of the reasons why post-publication curation of an article—or a work—works so well in Wikipedia is because of the form of the article. That the article is not perceived as a final work, but a beginning, and that it can be curated and improved over time. That provides a real incentive for others to contribute knowledge and corrections to the work.

I would also say that if you look outside of medicine in terms of the way knowledge and information is developed, it is frequently in a much more Wikipedia-like fashion. I would suggest a very good example of that is OpenWetWare in Synthetic Biology, and genetics, and systems of that nature which are really built on the same platform as Wikipedia with additional control. You have to be a member of the cohorts and develop some level of trust before you can contribute to it, and those are very robust productive methods, very trusted in terms of developing science in those fields.

So, I really see no reason, personally, why we can't adopt systems of publishing which are more like that than what we have in most medical journals today, including most online medical journals. I would certainly include PLoS in that group who publishes things as fixed PDF forms. That's my take. Richard?

Richard Smith:

I think the form of the article is one of the barriers that can stand in the way of effective post-publication review. In fact, I should lay my cards on the table straightaway and say that my instinct would be the same as Alex's. I think we should just get rid of this whole elaborate, expensive, cumbersome system and simply let the world decide.

I agree with Alex that back in the days when you could only print so much and it was expensive, maybe that made sense, but it doesn't make sense anymore. I also think it's what happens anyway. We know from many, many studies, there is endless amounts of rubbish published, much of which just disappears into the void. There are only a handful of things that really have an impact, and they have an impact because people pay a lot of attention to them after they are published. I think they are important—discuss them with each other.

I'd absolutely agree with Bo that peer review, in the sense of people looking at the work of others and deciding what they think of it, will always happen.

In the modern world, that doesn't need to be a handful of selected people before publication; that can be the whole world after publication, and publication can be very rapid. I do know why it can't be, and they are all bad reasons in my mind. One of them is the conservatism of the whole medical academic scientific system, which really surprises me by its extent. But, it makes a lot of sense if you think about it. If you are someone who has done well at this game, at the existing game, you are very reluctant to change to another game where you might not do so well.

So, power authorities on the whole are very much inside their peer review, and the way, ludicrously, that the whole academic credit system tends to be built off the back of where you publish gives absurd power, I think, to a handful of prestigious journals. So, we know that post-publication reviewers, maybe not the best way to describe it, but just review after publication happens in lots of other disciplines. It happens in high energy physics, it happens in mathematics; it happens in astronomy.

One of the reasons that people say it can't happen in medicine is because potentially the materials published in medicine might have a much more direct effect on ordinary people. People are not likely to be disturbed by some strange paper in high energy physics, but some paper being published suggesting that marmalade causes cancer can create a whole lot of furore. I have two answers to that. My main answer is, well, massively that kind of stuff is published already. We know that many of the scientific papers that are published in the medical press are just plain wrong. They are drawing associations between A and B, marmalade and cancer, and suggesting that they are causative when, almost always, they are not. So, there is a lot of that stuff out there already. Plus it tends often to appear not only in journals, but in newspapers. It's presented at conferences. And, the terrible thing there is that people who want to say well, should I believe this or not, you don't have a full paper and a full backup to go to, which, in the kind of world that I am proposing, you would have.

I think if we don't have the courage to go for full post-publication review, then I think light review is at least a step in that direction. And one of the reasons I asked that question is because light review is beginning to happen. I am on the board of the Public Library of Science, and it's a much lighter form of peer review that happens with PLoS One which is now publishing five hundred papers, six hundred papers a month. Nature has just announced that it's going to do something similar. So, actually the world, whereas as Liz said, the medical world has been very slow to adopt the possibility of post-publication review, it does seem to accept light peer review. And then, I would also say that, in my mind, if you are going to persist with some kind of pre-publication peer review, then in my view the more open it is (the authors know the identity of the reviewers, the readers know the identity of

the reviewers), I think the more fair and transparent it can be. So, I would favour that, but like I said, I really favour what we are now calling post-publication review.

Transparency in Peer Review

Peter Frishauf:

Okay, Liz?

Liz Wager:

Yes. I think the question about open peer review is also a very interesting one. I think, within a big community, open peer review absolutely makes a huge amount of sense. It feels more transparent, it feels more just and sensible. But personally, as a reviewer, if I've been asked to review a paper and I know the person quite well, I do wonder if perhaps I pull my punches, if I am a little bit more cautious if I know my name is going to be on that review. And again, there has been a lot of caution in journals; I was surprised initially that journals didn't just say *okay fine, let's all have open peer review* so, the reviewers know who the authors are, the authors know who the reviewers are, but it's a surprisingly small number that did adopt that. And, I think I have to say, that could be the reason.

I think it's a good principle, and it certainly works where you've got a very big research community. But, if you are a local publication or a very small community where people know people, maybe it's not such a good idea, but maybe that's counterbalanced by the fact that it's easier to just spot biases. It's fair on either side so the reader knows who the author is, the author knows who the reviewer is, and the reviewer knows who the author is, and so on. I like the idea, also, of readers knowing who the reviewers are, because I do think one problem is that if you expect quality review, whatever we mean by that, you want some reward, you want some recognition. If you do it in secret, then we are not going to get that. I think certainly, if you are doing post publication review, you should be prepared to put your name on something. You don't want anonymous comments, because they could just get personal or trivial and so on.

Alex Jadad:

I would like to emphasize what Liz said in terms of why she would judge the strength of her punches when she knows that her name will appear in the review. We need to acknowledge that we are operating in a very imperfect world where there are a lot of tensions across groups where careers depend on how you present yourself in front of more senior people and all those things. So, my suggestion about enabling the reviewer, and even the author at some point, to keep their name anonymous-may be a very important element of future peer review, because what if I am a student who has witnessed something inappropriate in a piece, and I want to disclose it. What if I disagree with somebody who can have a lot to do with my career

path? So, we might be able to put some safeguards behind that anonymity, and invite people to disclose their reasons why they may not want to share their names. And also, go all the way to even allow anonymous contributions by authors whose identity may be disclosable at a later date if they choose to do that.

Peter Frishauf:

So, if it's my boss' work or his or her friend's, it's the greatest.

Alex Jadad:

Yes.

Bo Adler:

I agree completely with Liz and Alex as well, and I think that the example of Nature that somebody brought up, where they opened up comments afterwards and nobody commented, is a really good one.

I have some personal experience too, because I started my own website, sort of as a regional group idea, because in grad school it's hard to find very many people that are interested in exactly the same topic as you are. So, I had this idea of posting summaries of papers on a Wiki essentially, and then allowing anybody to come and edit what they thought the paper was about so it could be different. Then I encouraged friends to post to it and everything, and I only had one friend who would post to it. I had several people that came and looked at it and said it was a great idea, but no one wanted to go on the record and put their name on it, because they were afraid of what it would do to their reputation. I think that that's a really big concern.

Richard Smith:

I disagree completely. I think that we should have the courage to stand by what we say, and I think if you are writing letters and comments anonymously, I think that's a shameful way to be. I can see that because of the power relations that perhaps people are nervous to criticize their seniors. But, what I am talking about here is not criticizing them as people, but criticizing their idea, their data, their controls, their methods. I mean, this is the absolute essence of science-that you put up some kind of thought processes, you gather some data. I think that we are living in a world, to some extent, where people are nervous to say well, I don't think your methods are right, I don't know that your controls are adequate; I think you could have done a different kind of analysis. I think you failed to recognize an important piece of work that's gone before. I mean, that's a terrible way to be, and yes, it reflects perhaps the sociology of the scientific system.

But, I think they are actually building into peer review systems, to build in, I mean the people would post publication comments to just weigh in without having to say who they are; I am totally against that. I find that really

unacceptable, rather than adapting a system to, a peer review system, to a corrupt scientific system. I would much rather work hard for a world where that's actually the fact that I, as a professor, say something and the student says 'I am sorry, I disagree. I think you've not thought about this properly and here is some evidence that points the opposite way.' I think that student should go straight to the top of the class.

Peter Frishauf:

One would hope that in academia, debate would be considered healthy. But, we all know that in many instances it's not, but I certainly agree that's no reason—we really need to stand up to that and confront it, because it's academic fraud when discussion and debate and people who challenge each other's ideas are discouraged, rather than encouraged.

Alex Jadad:

Okay. Richard, I agree fully with your vision of an ideal system. I have paid for being candid, pretty much throughout my career. What I was proposing was to look at opportunities for experimentation. Could we go gradual, not just to give a huge dose to people and say if you think about something that this is inappropriate, just disclose it and put your name, be courageous, okay? From what Bo said, from all of our experiences, we know that humans would protect themselves, and they know that there are consequences associated with candor and we are not in an ideal world.

My question had to do with whether we have an opportunity here to start exploring the best way to get to that idea. So, we leave it open, for example to people to decide if they do it anonymously or not, and then we add in a feature that enables that person to disclose their reason why that contribution was made anonymously. So, we start gathering information as to how we can facilitate the process to reach the level of openness that we all wish we had.

Peter Frishauf: Okay. I think I actually have a solution for the group on the anonymity question, and we are going to come back to that in a minute unless somebody...

Richard Smith:

Could I respond specifically to do that question, because I...

Peter Frishauf:

Okay.

Richard Smith:

I absolutely agree with experimentation, and I think actually it's the experimentation and gathering data that has led us to be very sceptical about peer review—the research that's been done in the past twenty years—certainly, when I was at the BMJ, we had a philosophy that before we

introduce a change in our peer review system, we should do some kind of research and gather some data. So actually Alex, in a way we did what you are proposing in reverse, because we started with a system of anonymous peer review which is really the known, the routine. We then did a randomized trial where we randomized people to give their reviews openly, to sign their reviews. We didn't ask them whether they wanted to. We said if you are in this arm, you must, or you don't review at all. We used as an outcome measure, the quality of the review, and our conclusion was that the quality of the review wasn't really changed by asking people to sign the review.

There was another study done in the British Journal of Psychiatry that suggested some improvement in the quality of the review. When we initially said to people, we are thinking of introducing open peer review, we went to a sample of reviewers and said would you be willing to review, and sign your reviews, 50 percent said no. When we actually did the trial, it was a very small percentage that said no; when we introduced the system, and it was maybe 5,000 reviewers that we had at the BMJ, it was about ten that refused to review, openly signing their review. So, I agree very much about experimentation, and I think we do have some data that are helpful to us.

Alex Jadad:

Yes. And, let me add that the anonymity you are talking about may be different from the anonymity we could achieve at this stage. The anonymity that the British Medical Journal used for those experiments was anonymity of reporting, you knew who their reviewers were as an editor, you selected them, and then...

Peter Frishauf:

Yeah, we knew as the editors, but the office didn't know, but...

Alex Jadad:

Exactly.

Peter Frishauf:

But, with the new system we did.

Alex Jadad:

Exactly. My point had to do with experimentation about the notion of anonymity in all its forms, you know anonymity of authorship.

Wikipedia-Style Peer Review...and Rating/Reputation Systems

Peter Frishauf:

We are going to, I think, talk about this a bit more in a minute. We haven't combined it with our conversation about what we might do in the future.

So, I would just like to review an excellent summary of some of the characteristics of what I think we all believe should be ideal in peer review. It's in an article that Liz co-authored with Tom Jefferson, Dr. Tom Jefferson, and Frank Davidoff, which was published in the June 5th, 2002 issue of JAMA that's called Measuring the Quality of Editorial Peer Review. And, she proposes seven things to strive for, the seven ideal indicators of quality that would come from an ideal peer review system; and that would be to assess the importance, usefulness, relevance, methodology, ethics, the completeness and accuracy of an article. I actually would add and maybe it's part of ethically sound or an eighth separate attribute, which would be accountable. And, what I mean by that is auditable, and one of the virtues really of Wikipedia, and one of the ironies. The irony is that academics dismiss it as being completely untrustworthy. The irony is that because of the technology and methodology, everything that happens on Wikipedia is auditable and measurable. And, what I mean by that is every word of every revision to an article on Wikipedia is recorded and viewable by any reader who goes to look it up. So, you can tell how an article looked at any point in time, and the identity, at least by IP address, and often by username. In some foreign language Wikipedia, such as German language Wikipedia, you always know the username of the revisions. But, the point is that it's a 100% accountable process, and the irony is that Wikipedia is probably the most robust Petri dish we have for actually studying the process of words and contributions, because it is auditable. T

his has led people like Bo, and other computer scientists, to download entire versions and renditions on Wikipedia and try and associate the trustworthiness of contributors and articles with the content that we as readers see. So, like it or detest Wikipedia, it is ironic to me that some of the most vociferous voices against it really hide behind a process which is not auditable at all. As Richard pointed out, the editor knows who the reviewers are, but most often, nobody else does. So, if these systems are to be so trusted, if what I see in the New England Journal is to be so trusted, why can't I actually look behind the curtain and see how the sausage is being prepared like we can in Wikipedia? I am a big fan of openness in process, and the kind of platform that Wikipedia is created on at least gives us the opportunity to do that. And, I would like Bo to briefly describe a little bit about his work in creating trust marks for articles in Wikipedia, his research in creating trust for individual contributors, and what this could mean for the future of other medical publishing.

Bo Adler:

Sure thing. Our problem really started a few years ago when there was a bunch of press about Wikipedia and the errors that it had, like John Seigenthaler was a newspaper editor who had a Wikipedia article about him, and somebody as a joke went and edited it and said that he was a participant

in the Kennedy assassination in some way. And so, there is a big uproar about Wikipedia and whether you could trust it at all, and if you could, how often could you trust it.

I believe it was Nature that did a study comparing articles between Britannica and Wikipedia, and what they found—that on the whole, when you averaged out over lots of articles, the number of errors seem to be about the same, or slightly better, in Wikipedia. But then, the question becomes, if you have all this information, how do you evaluate it and how do you know which is true and which is not? So, we started taking a look at the problem and what we decided is that the way—since Wikipedia works through crowd sourcing, the more that somebody looks at an article, the more likely that someone would notice that there is an error and go and fix it.

And so, our initial idea is if we could figure out how many times a page was viewed, then we'd be able to come up with some reputation measure that way, which is essentially what Google is doing too, or they do a variation of that. The problem there is that Wikipedia itself doesn't actually make that information available. We don't know which pages are the most popular pages that people are looking at. And, in Peter's article, he actually quotes a different paper from Minnesota where, using other data sources, they try to estimate that number and modify reputation system to make use of that. But, since we were stuck without that, we spent a lot of time thinking *well, what could you do*, and we realized that with the entire evolution of each and every article, we could figure out where every word came from and how long they last. We decided that that's essentially what we are saying that if someone sees an error, they are going to delete it right away, and hopefully they'll delete it pretty soon. In fact, my very first study was like, when there are mistakes and things are deleted, how long do they last in general? I don't remember the exact number, but it was something like 70% of text that's going to be deleted is deleted within seven minutes. It's surprisingly quick. Partly that's because there are people that sit and watch and look at every single change that's made to Wikipedia; they look for curse words and things that are obviously wrong. When they see that, they'll go and fix it. Over the years, people have made robots and stuff that do this automatically so that there is less work for the people to do. We just build on that idea that if something is deleted right away, it must have been bad. Whichever author was responsible for that did that work. If it lasts for a longer time, very likely it's true or at least people agree with it and so, it must be good. That's the basis of our reputation system.

If you think about that for just a little bit longer, you come to realize that all that we are really measuring then—we are not measuring truth, because that's not possible, because only, you have to leave that up to people. But, what we are measuring is consensus that people are willing to agree that this is good enough, that there is not something wrong with it that compels them to change it.

Peter Frishauf:

Okay. Liz what's your view on a system like that and does it have any relevance even to original research?

Liz Wager:

Yes. I think it's a very interesting possibility. I think one problem, though, is that Wikipedia should be dealing with established fact, things that can be checked, basically 'old ideas'; it's an encyclopaedia. When you are judging new work, as Richard says, one of the criticisms of peer review for a long time has been that it's been anti-innovation. Somebody said that the truly original has no peers. So, if you are going with consensus, I have a concern that we will stick with the old ways and we won't recognize the brilliance of the new. There is a difference, I think, between assessing established or old facts and 'new' facts. What the journals are doing is trying to present the new stuff. I think we ought to focus the peer review much more on methodology, and that's where it should lie. It's rather boring, it doesn't sound as sexy and as exciting, but that's really how to judge whether a paper is any good. Not by whether you like the results, but by whether the methods were sound.

Certainly, the idea of permanence is an interesting one. When Tom Jefferson and Frank Davidoff and I wrote that article about quality, one of the ideas we batted around—I can't remember if we actually put it in there—was that we would measure the quality of a randomized trial at least in clinical medicine by what—by how large an impact it would have on our matter analysis. So, we had the idea that you do your systematic review before you do your research; you do your research, and then if you haven't changed much, you haven't really made a big impact, whereas if you've actually shifted things one way or the other and made it more precise then you have. That has some big limitations. Obviously, it only applies to randomized trials; it applies to clinical stuff rather than sort of the early animal research and so on and preclinical work. But, it was an idea we were trying to get around for an objective measure of what meant—what was really good research.

Bo Adler:

I think that's exactly right, that peer review in any kind of concise measurement stifles innovation. And, you can ask yourself, did they laugh at Christopher Columbus when he said that the world was round, and did they try Galileo for heresy when he supported Copernicus? And, if the answer is yes, it's really hard to get new scientific ideas out there. So, consensus isn't necessarily the right method to use for a reputation system for a journal when you are trying to promote new science.

But, I think that there is an opportunity here for you guys to do something different. If you allowed people to comment, and the community actually took that up and started participating, there are lots of things that you could measure. You could measure page views, how many people are reading the

article, or reading the summary of the article, or at least how many people are commenting. If you allowed people to rate comments that came before them, so you could figure out which threads of comments were relevant and which ones were just personal attacks and whatever, there are lots of ways that you could help weed it out. And then, all of a sudden, maybe what a journal is supposed to be presenting to the public is not necessarily the best science—as in like, we know that this is absolutely accurate—but which is the most controversial, which has the biggest impact in terms of it's getting people talking whether for good or for bad.

Richard Smith:

I think that's very interesting Bo, and of course that actually is happening. It's from a couple of weeks ago that all of the journals, all of the articles published in the PLoS Journals, that's PLoS One, PLoS Medicine, PLoS Biology, and a few other journals now have attached to them something called article metrics. It tells you how many people have accessed that article and how many people have taken the trouble to download the PDF, which suggests a higher level of involvement. It shows you how many people have commented and you can see the comments. It shows you how many times those articles have been cited, how many times they've been mentioned in blogs. So, actually we begin to have that kind of thing. I think in a way that what happens in the real world is a kind of reputation system, anyway. I mean it is one of the innovative scientific publishers. Vitek Tracz started the thing called Faculty of a 1000. It is predicated on the idea that there is massive information out there within medicine science. Most of it not very interesting, not adding much value. Wouldn't it be nice if somebody you respect, somebody who has a high reputation said 'I looked through all this stuff and here are a few things I see, that looked to me very significant and important.' That whole innovation has been very successful.

I would also argue that the process whereby papers actually get to be thought important, get incorporated into the scientific system where this actual 'change thing' happens, is not just words on paper. That process, surely in some ways, is something fairly close to what might be called a reputation system. But, what I wonder is whether we can do it in some more methodical kind of way. Bo has looked at the Wikipedia, because there is a downside to reputation. And, I know that I mentioned in the article; I am always using Mark Twain's quote that once you have a reputation for being an early riser you can sleep until noon every day. Unfortunately, a lot of people have grand reputations and are kind of crazy, demented; I am probably one of those people myself. So, I think some system of reputation that wasn't based on how many times you published your name in a journal of medicine, or whether you had a Noble Prize—with some other more objective measure of reputation, might be very useful.

Bo Adler:

Well, maybe you don't mean objective. Maybe you mean time-based? Because, what you are really saying is that somebody might have been great once upon a time, but once you hit ninety years old, you are not going to be having the same kinds of ideas. Your reputation is a little different, and you can add that if you have mathematical formulas for calculating a reputation number.

Richard Smith:

I said that, but maybe I used a poor example, because, I think also there are people who have high reputations that are based on very, very little. There are people who've got nice warm hands and strong perfume, and they are very charming and charismatic, but perhaps absolute frauds and rogues. So, some kind of measure that is based on some kind of systematically collected data rather than just *oh he is a good chap*.

Bo Adler:

Right. Well, okay.

Richard Smith:

That you see clearly in my mind.

Bo Adler:

Related to that—and maybe this is less on the objective side—but one of the things that I see happen in papers is that reputations are accumulated over time. When a paper first comes out, people don't appreciate the full implications of it, what it could possibly mean. It's only a year or two later that the reputation really peaks for a paper and people understand better how it applies to everything else.

An idea that I had, if you had something like a Facebook for scientists, which I assume would be the Journal of Participatory Medicine's website where everyone has an account, there are a couple of other things you could do. If you let everyone rank the top twenty papers from the journal that they think really represent the best of science, you could look at it; not make it publicly available, but behind the scenes, the journal could keep track of everyone's top twenty. That data could be aggregated to figure out which articles really are the pinnacle of science for their particular specialty in whatever subfield of participatory medicine they are about. You can also have, like LinkedIn or Facebook, friends in a social network and give ratings to each of your friends, how you think of it, and again not share that. But, all of a sudden, once you have these personal reputations of people in your social network, the computer can go ahead and calculate reputations of articles. What do your friends think of articles, because if they rank articles highly, it's very likely that you'll also think the same kinds of articles are 'high.'

You can get more information about what might be relevant to you and have a journal that's customized to your particular interests.

Alex Jadad:

And, Bo and Peter if I may.

Peter Frishauf:

Please.

Alex Jadad:

Bo, I think you said several things that perhaps were not that explicit. And, please let me know if I mix them. You place a lot of value on quantitative, computer-driven assessments of reputation, for example. Throughout your comments, you also emphasized the importance of other more qualitative ways of judging importance.

Perhaps there is an opportunity here to blend methods to do it, to triangulate more, to 'quadrangulate' more, to ensure that we learn, how to get as close as possible to that ideal that we know we will not be able to reach. So, ranking, for example, the paper's value, and that can complement reputation of assessment that may be prone to manipulation as well. By comparing and contrasting, be able to get a much better picture of the value of a particular contribution.

Crowdsourcing Research/Peer Review

Peter Frishauf:

I would like to comment on something Richard said about how the reputation of an article changes over time anyway, and the value of crowdsourcing information post-publication. There was a very dramatic example of this last year, in February, when the proceedings of a National Academy of Sciences published a paper which was entitled Lithium Delays Progression of ALS. It was very promising to patients, a very promising report to the ALS Community, but it was based on a very small number of patients.

There is a site for ALS patients called PatientsLikeMe where the community was very, very interested—obviously—in this finding. And, they had 1,600 patients from the get-go who were actually on lithium. They basically created their own informal trial with these 1,600 patients to track the progression of the disease—those that were on lithium, really the largest ALS trial in history, so to speak. They followed the data provided by these patients—self-reported but highly structured—over a year's period of time, and really found that, unfortunately, the results were quite disappointing. It got a lot of attention.

Although it's not a vigorously strictly controlled clinical trial, this could be a very promising way of quickly validating, or at least providing evidence of,

the validity of findings published in the traditional literature—through the analysis of extensive amounts of broad data that are available on the site. So, maybe if our panel could comment on that, isn't that a form of crowdsourcing where you are actually providing data, clinical data about initial findings for an article, and the impact of work like that. Why don't we start with Liz here, and then go to Alex and Richard?

Liz Wager:

Okay. I'd like to pick up another point actually as well, which is, I think Bo was saying, that articles take a while to get established, especially the really innovatory ones. And, that is just such a great argument against the impact factor, which only looks at citations for a two-year period. It's heavily biased towards things that make an instant impact, but maybe were froth or fashion or whatever.

Coming back to the question—you do a sort of initial study and you follow it up. I think again, the idea is attractive, but my methodological hat warns me a bit, that actually even though you've got a big study, observational studies do have some awful problems, and there could be all sorts of confounding. And, it's not always the best way to follow-up; it might have been, it would have been great to get those hundreds or thousands of people into a properly randomized study. But, to get them involved in an observational study, I do have a bit of a problem with. I mean, the classic one was only stage 1 HRT, Hormone Replacement Therapy, which suggested it was wonderful. It lowered your cholesterol, it lowered your risk of having a heart-attack, it prevented Alzheimer's, it made you live longer—until we actually did a randomized trial and found that actually, it raised your cholesterol, it increased your risk of breast cancer, and so on and so forth. It was simply a confounding effect that richer, healthier women were using HRT. So, I do think you've got to be a little bit careful there. Sorry to sound like a party pooper on that one.

Alex Jadad:

Okay, I will explain. I agree with you Liz, and in fact we have been studying PatientsLikeMe since they started. In fact, I am registered as a member of the community, and I offer support to patients online who are also members of that community. I have the fortune to interact with leaders of that effort which is entirely driven by patients. They are prepared to do formal trials; in fact they would welcome that.

Liz Wager:

Fantastic.

Alex Jadad:

So, it's down to asking the establishment to recognize the power that the public has these days, to enable everybody to lead forward. So, it's possible; they did their durational study, because that's what they could do at that time.

Liz Wager:

Sure.

Alex Jadad:

I see, really, an opportunity for the Journal of Participatory Medicine to look at these kinds of communities that are already thriving outside their traditional academic environment, to agree with them and to figure out how we together can achieve the level of excellence that we all hope to get, and not consider it an either/or.

Peter Frishauf:

Richard.

Richard Smith:

Well, maybe. I've recently read Clay Shirky's book, *Here Comes Everybody*, which is full of examples of where crowds, people who can come very rapidly together and really give a different view of things, and can make things happen that were impossible back in the world before the internet.

Before the internet, it wasn't possible to come together in quite the same way or at least to do it on such a scale and as fast, and that book is full of examples of where they've made radical differences. When I reviewed the book I said, well, we don't really have many examples like that yet within medicine. But, I think it's only a matter of time, and the story you tell Peter, I think is quite an interesting example. I think back in the past when I was firstly with BMJ and many, many people in Britain were on tranquilizers, and a lot of people said 'well, we just can't come off these things, it's making our lives a type of misery when we try and stop taking them.' The conventional magical wisdom was 'that's because the symptoms that you were put on them for are returning', but that eventually turned out to be utter nonsense. That was an example of where the world at large was a lot smarter than the experts. I think we are going to see many more examples like that as people really begin to understand the ways in which you can make things happen very fast with the internet.

I would go back to one of the points Alex just made. It's very difficult for us to see how the world can be fundamentally different. But, I think it will be. Just as the peer review system will look archaic not many decades from now, I think many things will be very different in a way that those of us stuck firmly in the old paradigm have a lot of difficulty imagining.

Peter Frishauf:

I'd like to finish up by throwing out one more idea related to our conversation today. And, that is, if it were possible—and I believe that it is; it wouldn't be easy and it would take a lot of work—to create a trust mark for reviewers, readers, and authors of articles.

If we really thought about it, if we got the best minds together, I think we might be able to do this and really come up with a reliable, open, auditable way of scoring papers over time, perhaps even evolving them over time much like the Wikipedia article. I think fundamental to that idea would be to separate out credentials, that is, academic degrees as the only marker of expertise and recognize that expertise can be established without academic credentials. Of course, there are many, many notable examples of that. But, in order to have a system that would reliably credential, in fact, someone's expertise rather than solely their degrees, how could we do this and would this gain the trust of science and the public which now really relies on academia to be the arbitrator of trust and expertise?

So, I'd like to start off again with Richard, and then go to Bo and Liz and Alex on that very question.

Richard Smith:

That's a very hard question, Peter. I am thinking rapidly on my feet here; I certainly think academic credentials are a very rough and ready mechanism of deciding who can we really trust, and who's got something really interesting to say. I think we recognize that.

There are many people in the world who we pay a lot of attention to who don't have many degrees, and there are lots of people with heaps of degrees that nobody is much interested in. How could we systematize that? I find it hard to think on my feet how we could do that. But, what I would say is that I'd be very interested in experimenting to try and do something along those lines. Maybe I could end by just pointing at—Bo was talking about wouldn't it be good if people could score papers rather, as happens on eBay, Amazon, etcetera. Actually, that does happen on PLoS. You can go in and you can score an article and they put together the scores, but at the moment there are very, very few people doing that. I think we've got to build a system that not only will do what we want it to do, but will bring people to it—that they'll find attractive, and fun, and rewarding to interact with it in the way that clearly the people who've created Wikipedia do. Exactly how we do that, I don't know.

Peter Frishauf:

Maybe it should be a quid pro quo, so that you can't access the article unless you agree to participate in it. Bo?

Bo Adler:

Okay. I absolutely think that there are things that you could do to quantify reputation systems, because there is lots of work done on Hacker News, eBay, and Amazon. Richard brings up the point of scoring.

One of the things we wrestle with when we build reputation system is you are trying to create a community and you are trying to inform the community what the rules are by creating a scoring system. It's essentially a game that people play, and you know when you can say thumbs up and thumbs down. Sometimes thumbs down is a negative thing that encourages competition or encourages negative behaviour. I think science should be something where we are all colleagues and we are trying to work together to improve human knowledge. So, that's one of the things that I wrestle with when I think of reputation systems.

The idea of ranking—what are the top twenty papers that you think are important—is one of those ideas that gets around this whole negative thing. You are not actually saying that there are any papers that are bad; you are merely saying which papers you think are the best ones. It's in aggregate that we can get some information about what's useful and what's not so useful. Of course, everyone is going to say that their own paper is in the top twenty, right? So, every paper will get at least one vote. I think that there is lots of stuff you can do, and I guess a question I have for Richard is, he's mentioned the PLoS article metrics a couple of times. But then, he also said that he didn't have any idea for what a reputation system would look like.

One question I have is, is there anything that he doesn't like about the PLoS system? I think that what they have looks pretty impressive. It seems like that starts to form a basis for what a reputation system would look like; it's useful information for people to have.

Richard Smith:

Yes, I should quickly answer. It just started here a couple of weeks ago; it's the early days. I mean, to me it looks very impressive, and I think it is a beginning. I think the expectation is that slowly but surely, there will be other kind of metrics added. It does get us away from the tyranny of the impact factor that Liz was talking about. Slowly but surely, the whole world is becoming bloody fed up and impatient with the impact factor, which applies to whole journals. Having individual article metrics makes a lot more sense. Of course, most articles are written by many people and exactly how do you attribute the scores of those articles to individual people? We've got a lot of things to think about.

Building a Community

Bo Adler:

Right, and your final point from your earlier was that building a community is a hard thing, and that's where the real challenge is. You could make a game and no one might come; that's what I found with my own little reading group website. That's what Nature might have experienced as well. That's the really big challenge here. I think Wikipedia and DMOZ—they are a couple of projects that have really shown that you can build a community. When talking to friends about it, what they suggest to me is that the way DMOZ started—DMOZ is like Yahoo; it's an open source version of it. The way that Wikipedia and DMOZ started is that they went and found other sources of information and seeded the data. Wikipedia has the advantage that it's trying to represent truth, and everybody has an opinion of truth. So, when they see something wrong, they just jump in and edit it. Making comments about articles is a different thing. I think that you guys have a big challenge there on how to coax the community into participating.

Richard Smith:

Liz.

Liz Wager:

Yes, I have one question and one comment. One thing we seem to have skirted around and not addressed is, we seem to be assuming that the scientific literature is actually all made out about data and reports of original data, and that actually that sort of drives behaviour. Well, in fact if you are talking about clinicians, there are much more influenced by review articles, by editorials, by opinion pieces. And, I think if you are going to rate influence, you've actually got to think up a way of not only rating, you know how good methodologically was the clinical trial, but actually also look at the pieces that, you know the critics, the opinions, the background and so on as well.

And, don't forget that, you know they constitute a large part of the literature, so that's sort of challenge really. My *other question I suppose is PCU says* rating reviewers. Well, I thought we could have come around to thinking well, we don't really need these peer reviewers. So, are we talking about commentators; we are talking about post-publication review here?

Richard Smith:

Yes, we are. I think all, both pre-publication and post-publication reviews, should be published. They may have to be anonymized—hopefully they don't—but, the identity should still be known to a computer or some trusted system so that they could be random reviewed.

Liz Wager:

I certainly would like transparency in terms of people's financial interests, their competing interests. I want to know who these commentators are. It's funny, you mentioned a Facebook for science, and I actually seriously was wondering about whether you could generate a sort of Facebook, in terms of understanding who people worked for, who they worked with, who their friends were. That would actually give you a lot more information than just that sort of classical little disclosure and disclaimer that says what grants they've received and so on. I am afraid I am a bit cynical and skeptical; there are some big interests, big social interests out there. I've got to make sure that the system is not open to abuse. I do want to know the person's interests and where they are coming from. I am not sure I am happy with the idea of anonymity there.

Peter Frishauf:

Yes, I think that's a very valid comment, and any trusted system would have to be able to address that. Alex?

Alex Jadad:

Yes. Thank you very much; I had a wonderful time participating in this conversation.

I would like to make a few points to conclude. The first one is that, from research on quality, we know that it's a very tough construct to define, even, or to quantify. Quality is in the eye of the beholder. And, we have tried to use trust marks for information on the internet, for clinical trials, for quality of life, you name it. It's been—always—a very interesting, never ending process. One of the main reasons behind a lot of the challenges we have been discussing here today is that humans are very clever; we are master manipulators. So, it doesn't matter what system we put in place; it's always open to abuse. There is no way in which we can protect anything that involves competing interest. We can improve the transparency and we can make it more difficult to be manipulated, but we need to be fully aware that it's almost an impossible task. There are new options opening in front of us, and I think we need to acknowledge that we know very little as to how valuable they are going to be—how accepted, how successful. And, we need to remember one of my heroes, Maimonides, eight hundred years ago said *teach thy tongue to say I don't know, and thou shalt progress*. We need to acknowledge that we know very little about what these new possibilities offer to us. We know that the traditional system doesn't work. We need to consider it as obsolete and potentially harmful, and hopefully start from fresh.

So, I would like to suggest, to end, four issues. One is to idealize, idealization of what we would like to achieve. This comes back to the first question that Liz posed to us. What are we hoping peer review to achieve, and set an ideal. And then, be prepared to experiment, acknowledging that

we don't know, and not experiment on our, own which is what most journals are doing now. We should experiment in collaboration with other groups, especially those groups that are using social networking tools like PatientsLikeMe that may offer great opportunity for us to learn together and to progress together. And then, the final comment is to have reasonable expectations to understand that we are dealing with human nature here. Despite our best efforts to try to reach the ideal, we will never be able to get there. Competing interests will continue to creep up to restart the process.

Peter Frishauf:

Well, I thank you very much for that comment.

I agree with you entirely; it really would be very desirable if we could structure the nature of various rating systems that are used so that at least, many online publications put them into place. We could have some basis of comparing one against the other and we are not left with a complete mess. That was the idea behind my proposal that a conference on reputation systems—bringing out a lot of these competing methodologies that are being proposed, and in some cases, used—could air a lot of these different techniques and create some common ground so that we could compare the results of these experiments.

I'd like to thank all of you for participating in this call. If you have any final comments, please do make them now. Let's start again with Richard.

Richard Smith:

Well, just to say that I enjoyed this call. I was a little skeptical that we could talk in an interesting way about this for almost an hour-and-a-half. It certainly fascinates me, and it's been very pleasurable talking to you all. It's left me with a lot of ideas. I think we are at a beginning, not in an end; I'd very much like to be part of the continuing journey.

Peter Frishauf:

Thank you, and Liz?

Liz Wager:

Yes. It's been a good discussion. I am very much looking forward to seeing some participatory readers as well in this Journal of Participatory Medicine. That's the whole idea of the experiment, really developing a community and getting some new models off the ground. I think it should be exciting.

Peter Frishauf:

Bo?

Bo Adler:

Thank you all for inviting me to join your conversation. I mean since I started my own community study, I really believe in this whole idea that

science is an open discussion. So, I am looking forward to seeing you guys change the world.