Patient Matching Power Team
Draft Transcript
August 10, 2011

Presentation

Judy Sparrow – Office of the National Coordinator – Executive Director
Thank you, operator. Good morning, everybody, and welcome to the Standards Committee's Patient Matching Team. This is a Federal Advisory Call so there will be opportunity at the end of the call for the public to make comment.

A quick roll call – Marc Overhage?

Marc Overhage – Regenstrief Institute
Present.

Judy Sparrow – Office of the National Coordinator – Executive Director
David McCallie?

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Present.

Judy Sparrow – Office of the National Coordinator – Executive Director
Nancy Orvis? Cris Ross?

Cris Ross – LabHub – CIO
Present

Judy Sparrow – Office of the National Coordinator – Executive Director
Walter Suarez?

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
Present.

Judy Sparrow – Office of the National Coordinator – Executive Director
Shaun Grannis? Lisa Gallagher? Judy Murphy? Deven McGraw

Deven McGraw – Center for Democracy & Technology – Director
Here.

Judy Sparrow – Office of the National Coordinator – Executive Director
Did I leave anyone off? Alright, with that I'll turn it over to Marc Overhage.

Marc Overhage – Regenstrief Institute
Super. Thanks, everybody, for joining this morning. Since our last meeting, we had the chance to present our current state of thinking at the HIT Standards Committee and received some just wonderful and helpful feedback and subsequent discussion with a variety of folks that I think helped crystallize some of the questions that people have, and some of the potential concerns and ways to improve things that folks had.

I've been working towards, and thought I would have by today but didn't get there, a sort of draft letter building on Walter Suarez's outline. I'm still hoping to get that out to all of you today as a first draft; very much a draft to look at. But there were a couple of issues that at least I couldn't readily sort of come to peace with based on some of the feedback that we got from the Standards Committee presentation. And one of those, in particular, was the area I hoped we could spend the bulk of our time on today and then I could share with you the sort of approach I was taking to the question of what patient attributes do we think about, and then whatever else the group would like to cover. Is there anything else that folks would like to make sure we touch on today?

So the issue that I was having a discussion with a couple of people about and would like to get this group's sort of … guidance about is while we certainly want to stimulate innovation and thoughtful and creative ways of patient matching, David has guided us several times – we don't want to stifle innovation. On the other hand, and I wish Shaun Grannis were on the call today because he and I spent a lot of time trying to think through this and describe this. It seems that, to me anyway and I think Shaun is on the same boat but there may be others too—that we need, that if there's not some level of commonality of the patient attributes that I would expect to receive in a query, and be using to help match patients, then we're going to get into some very awkward places.

So let me describe if in the world where we sort of say, okay it's the Wild West; use whatever you want for matching; have the ability to ask for additional stuff but recognizing that people may well not have captured that additional stuff in a high-quality way, if there's no commonality. So the scenario is that Entity A asks Entity B to match a patient and Entity A provides three patient attributes, A, B and C. And Entity B, however, uses C, D and E for matching; so there's only the common element C; the kind of situations that I think—I fear you run into is that either in B's matching process there's only one patient, if you will, that has attribute C and you assume it's a match because you have no other data to validate that when, in fact, the chances are it's well not a match. The other scenario is that you have lots of people that have that value for attribute C and obviously this is kind of a simplification exaggeration but I think it applies when you have twelve elements as well as when you have one. You know, you have a hundred patients that have attribute C's value that was sent in the query and there's no way to tell which one of them is the correct match, so you return none.

So I think some people would say well, of course that's okay because what you do is if, you would always say, "Well, gee I can't find your patient with only C; you need to send me D and E." But again, if there's no commonality the chance that the querying group is going to have D and E available in a quality form, who knows what that is. So that's one dimension of why it seems to me there needs to be some level of commonality, at least as a foundation.

The other question that I have that perhaps some of you can help me think through is, is there—given that one of our goals is to have a level of sensitivity and specificity that the match will deliver, if you don't know what attributes you're potentially going to get I don't know that you can estimate the sensitivity and specificity. In other words, you can't know for any sort of random set of patient attributes that you get to actually do the work to identify the sensitivity and specificity for that is significant, and is it reasonable to expect that every organization, every EHR that's going to do matching, has done the sensitivity and specificity work for every possible combination of patient attributes that they might receive.

So I guess those were the two things that led me to be thinking that we need to suggest that there is a—needs to be a level of, a foundational level of commonality. While accommodating the opportunity to include new and innovative things, it seems like you have to start with a foundation but maybe there's a way out of that conundrum that some of you could suggest.

Deven McGraw – Center for Democracy & Technology – Director
Well, so—Mark, this is Deven. You know, while we did say on the Policy Committee side that we didn't think that use of any particular data field ought to be required and we can go into a little more detail about why we reached that conclusion because there might be some flaws in our logic; we also recognize that there is a standard set of demographic data fields that people typically use when they match. And we got to that conclusion through the testimony that we had in our all-day hearing which included a lot of people who are in the business of matching patients today, including Shaun Grannis and … SureScripts, including Catholic Healthcare West as an example of a provider. So name and address and gender, I can't off the top of my head rattle off exactly what those were, but they were—

Marc Overhage – Regenstrief Institute
Date of birth is the other one that commonly comes up.

Deven McGraw – Center for Democracy & Technology – Director
Right, so they're typically used and everyone said without hesitation that consistent presentation of the data in these fields would greatly enhance the accuracy in matching. Because that's really what we're aiming for here, right? It's not that everybody does this—uses the same fields but that everybody achieves a consistent level of matching. Recognizing that if people are wildly inconsistent in the data that they chose to present in a matching circumstance, that won't work either. But I didn't get the sense that it was an—that people were using A, B and C and other people were using D, E and F. It's more that there are common fields that people use to match but the data is a mess in those fields.

Marc Overhage – Regenstrief Institute
Right. And that's clearly—that's clearly heard and is one of the core recommendations with some specifics about ways to approach that. So yes.

Deven McGraw – Center for Democracy & Technology – Director
So I honestly think that what we had said at the Policy Committee and the direction you guys are heading in is very closely aligned, except for the fact that we did not specifically recommend the use of any particular number. Whether it's development of a universal identifier, whether it's a voluntary identifier or whether it's a social security number. And that—you all seem to be headed in the direction of not necessarily requiring the use of a social security number, but creating a standard field for it to use that raise some concerns in my mind because of the huge identify theft problems that we have with respect to use of the social security number and the fact that there's a number of efforts going on at the state level, as well as within the federal government, to try to reduce the dependence on the social security number. So we seem to be swimming in the opposite direction.

Marc Overhage – Regenstrief Institute
And just to be fair I don't think that this group was necessarily saying that social security number was—I think it's a demographic or a descriptor field that we should have a common way to represent and if somebody is using it, clearly they should use whatever quality they need to do. I do understand your— is this something that, like you say, a typical or common so the—one of those typical or common attributes—I don't know that—yes, that that's something that this group has said.

Deven McGraw – Center for Democracy & Technology – Director
Well and I—don't get me wrong. I get asked for my social security number all the time, including every time I walk into a healthcare provider. But it makes me very nervous.

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
So Mark, this is Walter – just wanted to jump in and take this from a little bit of a different perspective. I was thinking of the response part and so if an entity receives a request for a match, whether it's a – I'm

sending you this single set of data elements about one person I'm looking for, or whether there is a whole host of here are all these elements about all these people. Some of them only have three demographic elements; others have five. The way I was looking at it was the response that that request would generate. So the responses could be not match, single exact match, or multiple matches which is more likely than anything and in the multiple matches the concern, of course, is what do I respond? Do I send back, "Well based on your data these are all the possible people we have." Or whether I send back, "Well based on this data we do have multiple matches and if you have more data we can refine those matches." And then provide some level of the sensitivity specificity of, "If you provide this additional data element we will be able to match with 99.9% certainty." So I'm trying to see whether we had a core or base or common set of parameters, what do we see the responses of the—from an entity that is being inquiried—or that is receiving an inquiry about a person or a group of people that are being looked for. So you know, and I'm not trying to deviate from the question about whether there is a need for a base or common set of parameters but I'm trying to understand, from the other end, the response—what is it the response—what do we expect the response should be to guide the decision as to whether there should be a base or not?

M
My understanding of where we were on that was that we—if we could specifically identify the person, of course, return the person; if you could not make a match, then the option would be to, as you suggested Walter, provide additional information that if you had available you might be able to match the person. Although having to be thoughtful there about helping people narrow down or fish for things. But—

David McCallie – Cerner Corporation – Vice President of Medical Informatics
You there's—this is David, one conundrum here is that we are trying to uniquely identify a person but we're worried that we might be uniquely identifying people. At some point you just have to say in a healthcare setting, for purposes of healthcare, assuming appropriate access to the system, what's it going to take to identify a person for safe health care? And then worry a little bit less about some of these other things because you can't both be incredibly precise and also prevent people from identifying people. You know, the whole point is to identify people. We just want to do it with the minimum risk at side effects. So I want to take the question back to Marc; you opened with putting—sounded a little bit like putting me in opposition to the notion of common elements and I want to make sure that that's not what I feel. I just want to make sure we don't put an upper bound of … that could be in use because I think we'll discover as systems evolve and society's expectations changes that we may find additional useful identifier capabilities emerging, just we don't want to rule those out. But I would certainly support some notion of a common set—although I would refer to it more as baseline, and sort of what's the minimum expected.

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
And you know, the reason I was asking my question is because depending on the response that people will give to this, there is no need to establish a minimum required set of the elements. There is a need to establish a list of possible elements that people can use whether we call it base or common is the kind of data that people will use and so whenever you have that data, send it. And if you have only three out of seven, send those three. And so we concentrate more, in my mind, on the parameters of the response rather than on the parameters of the request, of the inquiry. In other words, sure we can list the seven or eight or whatever list of common elements that entities use to do queries for matches but aside from listing them, there's no sense of us will be the direction of suggesting or requiring those in all cases or—so if we go, if we pass the point of saying okay well let's just agree on the seven or five, and that's—that's not just we're going to say but then let's concentrate more on the response part.

David McCallie – Cerner Corporation – Vice President of Medical Informatics

I think that's an important consideration in the general rubric of you don't respond with stuff that you weren't given is a pretty good starting point because, you know, somebody could obviously be fishing if you're sending back stuff that the person doesn't proffer in the first place. If they can't give you a candidate in a field, don't respond to that field.

Deven McGraw – Center for Democracy & Technology – Director
Yes.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
That makes—I mean, that's a … recommendation from way back. That's makes a lot of—

M
Right. Yes.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
But I think the expectation of this group is that we would give some advice on what are the useful fields that we think, in general, systems should be managing for the purposes of patient matching. Meaning not that we would give an exhaustive or a binding list but we've had discussions, for example, that address is not as useful as people think because the amount of moving. So are we going to put address on the list of what we would consider to be common or baseline, or not? We had this discussion about social security number, are we going to say, "Well no social security number but last four digits is okay." Are we going to recommend that or not? Isn't that what they expect of us, to kind of nail down on what we think is backed up by evidence as being useful?

Cris Ross – LabHub – CIO
Sure. This is Cris. Let me just ask a question and make a comment, David. I think what you're saying makes a lot of sense. I guess the question that I've got is there's probably two reasons why people, why an institution would want to maintain data about an individual patient. One would be kind of for their own record which in that instance, keeping the address, you know billing address and billing information and insurance and demographics, would be very useful. And then the clinical data associated with that patient, right? That's sort of one step.

It feels to me as if the idea here is that we want to support innovative methods for patient identification. That that may be a separate way of storing information about that patient that's not related to the core patient record. I think that's relevant because if we're going to try and be determinative and say, "We recommend using social security number as the only identifier and every system must maintain it." That allows us to store that data in the patient record and also use it for matching. But if we're saying that there's a variety of ways that institutions may want to use to do matching, that suggests that you probably want to separate the way that you store data for patient matching, in a somewhat separable way than what you're doing in the core patient record. So upshot of it is from a direction to software developers and those who use software may be that the data and methods and the algorithms that we use for patient identification need to be loosely coupled, or in some ways decoupled, from the core patient record. So I just want to make that distinction.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
This is David. I'm a skeptic of the notion of something that can be tightly coupled when you need it to be tightly coupled but is otherwise loosely coupled; because if it's tightly couple-able, and it would need to be to find the record, then it's available. I mean if it's a database join, it's a join. If it's a service call, it's a service call. Somebody can get it if they want it. I don't see the technical advantage of an ... notion that we should separate these.

M
Well unfortunately I don't know that we have to worry about—I mean that becomes an implementation issue that I think it's probably beyond our—

Cris Ross – LabHub – CIO
Well it might. I'll drop the point except to say I think if we are not clear about what we're trying to do we could end up setting a standard that says, "You have to store all these things in a patient record system, and oh by the way because we're fostering innovation we expect that this is going to require that you have a bunch of different algorithms you can support." Meaning that an implementer might find themselves believing, or actually being true, that we're going to put a lot of burden on core patient systems. And that's my concern.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
This is David again. I agree. I'm not—I wouldn't disagree with that point. I was just saying that the data's in the database and it's joinable, it's in the database and it's joinable. Meaning that—

Cris Ross – LabHub – CIO
Hey, I totally agree with you and obviously we want to exploit data that's in a core patient record, but I don't think we want to impose on core patient records that they support absolutely every single possible algorithm or method for patient matching that they might run into in correspondence with another organization, right?

David McCallie – Cerner Corporation – Vice President of Medical Informatics
So, Marc take us back to what you think the question would be—

Cris Ross – LabHub – CIO
There you go.

Marc Overhage – Regenstrief Institute
I think this is very helpful. What I'm hearing in this discussion is that folks do believe there is a—the term is a baseline, common, whatever set of attributes and I think the sorts of things that people are thinking about are name, date of birth, gender, a handful of things that are both useful to validate that whatever match you came up with kind of made sense. It's enough to, probably coupled with something else, because we know that those elements alone are not enough to achieve the levels of sensitivity and specificity that we think that the Policy Committee will eventually recommend. And I think I heard that from the group. Is that a fair impression?

Deven McGraw – Center for Democracy & Technology – Director
Well I think the only caveat Mark – and this is Deven again – that I would say is that we probably should be careful not to keep looking for a data field, another data field, to solve our matching problems. Because, again, one of the things that came through very clearly from the hearing that we had was that this is both a technical and mostly a people problem. And you don't solve the people problem just by asking people to populate yet another data field. They have to—you know, organizations—healthcare organizations have to start dedicating some time to cleaning up their data.

Marc Overhage – Regenstrief Institute
Absolutely.

Deven McGraw – Center for Democracy & Technology – Director

And if we give them a set of consistent representations and data fields and EMRs that require them to sort of, at least, enter it accurately or at least accurately as far as the data standard is concerned, we can make some strides and then we'll be in a lot better position to assess whether, in fact, there's another data field that will, in fact, get us to the level. But we're in such a mess right now that I'm really reluctant to say yes, all of the consistent representation and high-quality population of the demographic data fields that people typically use won't be enough and we're going to need something else for people to populate in a data field.

Marc Overhage – Regenstrief Institute
Can you say that last bit again; I think I was with you until I lost—

Deven McGraw – Center for Democracy & Technology – Director
Again, maybe it was not—I might have been not—it was meant to underscore the point that I think it's a mistake, or at least it's premature, to say we won't get—to say that what we need is another data field to make matching accuracy work better in healthcare versus cleaning up the quality of the data that we currently use to match. It's a mess.

Marc Overhage – Regenstrief Institute
So let me—and I think the—your statement makes sense; but it depends on what fields you're talking about using, today. So in other words let me use a specific example to make sure I understand where you're going, Deven. So if we have perfect name, date of birth, gender data. I think there—and it was perfect and clean everywhere. You still could not match at the level that we would expect people to need.

Deven McGraw – Center for Democracy & Technology – Director
Well that's a conclusion you're making. That's not a conclusion that based on the efforts of the Policy Committee that I feel like we … on.

Marc Overhage – Regenstrief Institute
So, now I think—this is a good issue and something that—because I think that's actually demonstrable.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Yes, I agree. That's just statistics of the U.S. population and there's good studies to show it's what you expect. The RAND Study has those numbers nailed.

Marc Overhage – Regenstrief Institute
And even if the perfect data—I think that's true.

Deven McGraw – Center for Democracy & Technology – Director
Right.

Marc Overhage – Regenstrief Institute
So that doesn't invalidate your point that you still need better—higher quality date of birth data and higher quality, you know—those will help. But they'll be marginal. They'll be one or two percent improvement in matching versus getting it from 80% to 96% or something.

Deven McGraw – Center for Democracy & Technology – Director
Right. So—alright, well I mean … our point but I guess I think that if you're saying we want another data field I actually think you may be straying into some important policy questions about what that data field ought to be.

Marc Overhage – Regenstrief Institute

There is—so let me try this. So the thing that feels like solid ground is we need more than name, date of birth, gender.

Deven McGraw – Center for Democracy & Technology – Director
Okay.

Marc Overhage – Regenstrief Institute
And as you said that begs the question of, well what else?

Deven McGraw – Center for Democracy & Technology – Director
Right.

Marc Overhage – Regenstrief Institute
You know, I think we all believe that we need to leave the answer to that open in many ways because there could be a variety of—there could be local solutions.

Deven McGraw – Center for Democracy & Technology – Director
Right.

Marc Overhage – Regenstrief Institute
You know, a dominant health plan identifier might work perfectly in a given market, for example.

Deven McGraw – Center for Democracy & Technology – Director
Right, or a system using a voluntary identifier or—you know, I haven't dug into this in any detail but I understand that now that there's cell phone number portability that the cell phone is becoming actually an increasingly reliable data point.

Marc Overhage – Regenstrief Institute
Sure. So there's a lot of possibilities and obviously with an answer to question of what else do you use is going to be driven by a variety of factors including how historical do you want to go, because even if cell number portability were perfect today and everybody had a cell phone, which of course they don't—

Deven McGraw – Center for Democracy & Technology – Director
Right. A lot of people do.

Marc Overhage – Regenstrief Institute
Well they do, but it's only about 79% in the U.S. So I mean there's still a huge chunk of people who don't. Will that change? Yes. But you do have to look to a degree, historically, as well and say how do you deal with the data from last week?

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
And kids under 13, or maybe it's now under 12, won't have cell phones.

Deven McGraw – Center for Democracy & Technology – Director
Might be under 10.

Marc Overhage – Regenstrief Institute
Or—obviously it's an increasing thing but I guess you do have to look, to a degree, where—there's data today being generated that needs to be matched tomorrow and so while there's some sort of forward looking 20-year down the road or 10-year down the road thing, I think you have to, at least, weight in the nearer term, how do I do it next year.

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
So is this like a tiered listing of data, where we say well the base is those three, name, gender, date of birth; then the next tier, let's call it that way for lack of a better word, but next tier would be address, cell phone, and any unique identifier whether it's a unique identifier assigned by health plan, by provider, by—could be even a social security number if that's the case. And then the next tier is anything else that can be provided; maybe a—

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Facial image.

Deven McGraw – Center for Democracy & Technology – Director
Biometric.

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
Biometric.

Marc Overhage – Regenstrief Institute
I guess my only reluctance, Walter, and again I welcome the group's thoughts. I would be reluctant to have more than two tiers in many ways. I think that sort of baseline things that you describe, and then sort of others—

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
Okay. I'm fine with that. I think that—

Marc Overhage – Regenstrief Institute
Does that make sense to folks?

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
Yes. That makes sense to me. Yes, absolutely.

Deven McGraw – Center for Democracy & Technology – Director
Yes.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
So what's that baseline? Is the baseline just those three?

M
I don't know else you'd put on there that's, frankly—

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
Well the question is, for example, whether address should be really part of base.

Marc Overhage – Regenstrief Institute
I think so. I—let me give you my pitch again. I've been trying to come up with a way to sort of look at and think about these things and so I put together a little graph that had sort of the number of times in a lifetime that an attribute changes. And then the percentage of the population that changes. So last name, for example, typically changes four to five times for about 35% or 40% of the population. So you could, sort of, on a two dimensional graph think about that. Address changes a lot—

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy

Yes, no absolutely.

Marc Overhage – Regenstrief Institute
—over the lifetime. And it's hard to get statistics on some of this stuff, but 15% of population moves every year. So you might conservatively think about a number of somebody changing 20 times in their lifetime address. And that's everybody. That's the whole world—changes.

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
And then the concern, of course, is to what extent having address for example, or city, might actually confound the other three base elements because there might be actually someone in that city that matches more likely those elements. So to what extent, because of those multiple changes that people have through their lifetime on some of these elements, having that element becomes a confounding factor rather than a discriminating factor.

Marc Overhage – Regenstrief Institute
That is an interesting point; I don't know if anybody's analyzed that. That's a really interesting point, Walter, because there's a not independence of last name and location, for example. That's fascinating.

So I guess I would have some reluctance to throw address on it just because of the number of changes over a life; now having said that, clearly SureScripts has, because of their temporal constraints we're only looking back a year—you know the chance that somebody has changed address in the last year, and you have their current address to query with which you presumably if you're actively involved in care is reasonable, maybe that's okay. Or maybe that's useful if you've got a time window like that. That seems like a very specific scenario.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
But are we going to be in a position of recommending those minimum three, and then having to say that that only gets you X% accuracy?

Marc Overhage – Regenstrief Institute
I think that's what I'm kind of hearing, Deven. I hope this is congruent with what you were saying, that—and if those don't get you there you need something else.

Deven McGraw – Center for Democracy & Technology – Director
Well I mean I think that's right although as I mentioned to you in an email, we really struggled in the Tiger Team discussions at the policy level and David, you can correct me if I'm wrong or misinterpreting this in some way, but we didn't feel like we had enough evidence, even after a full day hearing, to set an accuracy threshold.

Marc Overhage – Regenstrief Institute
Right.

Deven McGraw – Center for Democracy & Technology – Director
But we specifically recognized that one should be set.

Marc Overhage – Regenstrief Institute
Right. And I think we're not attempting to usurp that and agree there needs to be one set. But for the purposes of thinking about these things, one of the ways, I guess, I was trying to craft this letter and it's why it's been so darn hard to write, is sort of saying okay if you make for the purpose of thinking about this assumption that you want at least a 95% sensitivity and specificity level – you may want higher, you may want a little lower, but as a post that you can sort of work from which doesn't seem crazy that the, for

example, these three attributes as David said, the literature would say this doesn't get you there; it gets you to 70-some percent. So we at least know that unless you're going to drop a threshold to 80% you're probably not going to get there with those measures. So that's how I was kind of thinking about it, sort of a bracketed—without having the benefit of a specific threshold to reach, I think there's still some guardrails we can put around it in the sense of saying okay, best evidence, best thinking is those three things would not get you to the kinds of levels of sensitivity and specificity you need, even if they're cleaned up and perfect.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
And if we make that as our only conclusion, I think we'll have a number of people saying, "Well thanks for the common sense."

Marc Overhage – Regenstrief Institute
Well I think the other part of it is, as Deven and others have highlighted, we're not just saying as Walter kind of had put together sort of the specific areas of kind of one of the key things I think about and this is consistent with the Tiger Team's recommendations, is some thoughts about the specific formats for the data, some thoughts about specific things to do to improve the quality of the data. So it's not like—I agree with you to a degree, David. This part of it may feel like a, okay you're not giving me any answers. And in part, we are hard pressed to give an answer—a specific answer, because we don't have a specific threshold.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
No, but we know that the threshold that those three will get you is not enough. We know that—

Marc Overhage – Regenstrief Institute
Correct.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
—no healthcare system on the planet that would stop with just those three. So if people want advice from the Power Team about what to do to remedy the situation, we give them strong advice that we think is free from policy confusion or we come out with a set of policy questions that would be necessary to be answered before the advice could be taken. I mean if we're going to do the separation between policy and technology, which we're kind of forced into by the structure of the … we need to supply, if you added last four digits of social security this is the accuracy that you would get; however, we recognize this is a policy question and we're done. I mean, not done. But you know what I'm saying, we've done—

Marc Overhage – Regenstrief Institute
Right, as much as we can do.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
As much as we can do until somebody says, "Okay, last four is an appropriate tradeoff as long as you do the following constraints. The querier has to submit the last four and the return can only say matched, didn't match, they can't—

Marc Overhage – Regenstrief Institute
Right.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
--you know, say, "Did you mean" or "This was close," or stuff like that.

Marc Overhage – Regenstrief Institute

Right.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
I mean that seems to me that if we're going to justify our mornings on these calls that we need to either turf a question back to policy or admit that it's just not a problem that can be addressed by a federal committee.

Deven McGraw – Center for Democracy & Technology – Director
Well right, although we've said—I'm actually acknowledging that that's an approach that makes sense for you all to follow; I guess I'd ask the question of whether social security number or last four of social security is the only other data field option that you would put on the table as potentially moving the needle higher or to a better place in terms of levels of sensitivity.

Marc Overhage – Regenstrief Institute
So I think it is—and so let me make it some less provocative—

Deven McGraw – Center for Democracy & Technology – Director
Yes, I mean only because the social security number raises a lot of policy—if you can get an answer back on the policy side, no guarantees on this but says "no we don't want you to use it."

David McCallie – Cerner Corporation – Vice President of Medical Informatics
This is David – I didn't mean to imply that was our only choice; I think that's the choice that's been studied the most. That's the number the RAND Study did and they get all these permutations and combinations and study it against a what? An 80 million record database. So yes, we have pretty darn good statistics on what that particular field brings to the table. We can put concrete choices down; you go from this percent accuracy to this percent accuracy by the addition of these four digits. We could do that with other choices too; I didn't mean to imply that was the only one. It's just the one that's been studied well.

Deven McGraw – Center for Democracy & Technology – Director
Okay.

Marc Overhage – Regenstrief Institute
That was—kind of the approach I was imaging, David, and I appreciate your kind of crystallizing that we need to have some worthwhile conclusions. And I sort of … if our thresholds need to be of this order of magnitude and then we've got these core—the core's a bad word, these common set of attributes, whatever we want to call it, that then we can comment on what is known is these things and part of the conclusion, it feels like, maybe because there is not solid data and part because these are very difficult studies to do about what would be the value, for example, of the mother's maiden name as a matching attribute. You know, that's certainly one that's been used, has some stability characteristics; is that a—but it hasn't been studied, I don't think, at a scale and a degree of certainty that would let us answer the question, would it get us there. So maybe part of this conclusion is, I think you were suggesting, is there's a menu of some possibilities that at least would warrant further evaluation. Does that make sense to folks?

Deven McGraw – Center for Democracy & Technology – Director
Yes.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Yes. I agree. I mean this is—

Marc Overhage – Regenstrief Institute

And for the more forward looking, you know cell phones for example; cell phone numbers—although of course there's privacy issues there as well. And it's amazing how hard it is to find data about this. I've actually spent a lot of time trying to dig into data about this. Like for example there's a segment of the population, about 30% seems to be the number that comes up a lot, that changes their email every six months. And there seems to be a similar population and harder to get numbers on that changes their cell phone number every six months or so. I mean, intentionally and consciously, because they don't want anybody to know their cell phone number. So—

David McCallie – Cerner Corporation – Vice President of Medical Informatics
People that want to evade detection will falsely supply any of these fields; they'll change their date of birth. So that's—I think that's a mute point.

Marc Overhage – Regenstrief Institute
Yes. I was just surprised that it was that high. The 30% changing every six months surprised me.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
That surprises me as well; that's contrary to stuff I've read but anybody who wants to evade detection is going to do so by any of these volunteered fields. So I think you just take that off the table.

Marc Overhage – Regenstrief Institute
So I get that feeling like I've got a—I mean this sort of gels and helps me with some of the areas of sort of pulling this together that I was struggling with. So if I can just sort of summarize what I think—where my head is in terms of what we turn into recommendation is essentially the set of principles that we arrived at and have been—you know things like, you don't spit anything back that you didn't get. But we provide a—here's a set of things that have been used; here's a set of things that maybe haven't been used but sound promising; things like voluntary identifiers or future cyber identifies. Here is what we know about what this sort of common, relatively uncontroversial set of fields; what levels of sensitivity and specificity those can achieve. Here's what we know about what some of those other fields can contribute to sensitivity and specificity and here are the things—the fields that we don't have information about. The recommendation that comes out of that is in essence to say we need something more than this base set. There probably needs additional evaluation and studies to assess the contribution some of these other things can make, but that the sort of common set should be common.

On the data quality front, we have some specific ideas and suggestions for things and processes that should be put in place, but in terms of in systems and in terms of processes to improve data quality. Then we had some specific recommendations that are consistent with the metadata recommendations that Walter's group put together regarding data formats and so on. And then—I think that's it. Is that generally feel like—is that a, I know it's a very high level prospective and the devil's in the details, but does that flow of things and those sets of recommendations seem like as far as we can get.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
This is David. I mean, that sounds good to me but I'm—I just wonder if we should try to push back, maybe we shouldn't, but push back in front this question of last four digits of social security number. I mean I was just re-reading the RAND Study here while we were talking. Improvement in accuracy in going from the three that we've talked about, plus the addition of the last four digits is astonishing. It's like four orders of magnitude improvement. I mean four or five orders of magnitude. How could you not recommend that if you care about patient safety? The only reason is if there's some policy objection that insurmountable and it just seems to me that with the caveat that we're putting on here that you don't respond with anything, you merely acknowledge that that number matched and you only use the last four for matching, etc., I would be hard pressed to say anybody—why wouldn't you do that and could we give

people that implement systems some comfort if the Policy Committee said this seems safe to us. I don't know, Deven, maybe you think it just wouldn't stand a chance of going anywhere but I think if people actually looked at the statistics, it's dramatic.

Deven McGraw – Center for Democracy & Technology – Director
I can't say that it would have no chance of going anywhere because I've given—I've long given up predicting how the Policy Committee will come down on anything that has privacy and security implications to it. But the issue with the four numbers if you're going to have to trade off that level of accuracy and have a robust policy discussion on the tradeoff with the identify theft problems that are also popping up with the use of the last four numbers. Because people have been migrating to the use of just the four numbers in order to try to ameliorate the identify theft problems associated with the collection of whole numbers, and while certainly collecting just four is better than collecting all, Marc started this phone call and you stepped out David, so you might night have heard it, about a Carnegie Mellon study where they were able to predict the first five numbers of people's social security numbers using publicly available software and data, and so that's—it just underscores the issue that, notwithstanding the value that it might bring from a matching perspective, it carries some baggage with it on the other end. And do we really want to tie ourselves to a solution that may become less valuable over time because the negatives associated with it might start to overcome the positive matching impact of it. So, I think you just have acknowledge that the use of the last four numbers does not necessarily get you out of the hot water associated with the use of the whole—all of the digits.

Marc Overhage – Regenstrief Institute
But I guess, David where I thought we were sort of reflecting that was if we—in our recommendation say last four social security number whatever, driver's license number, here's what's known about the levels of increases that you can see. Part of our recommendation is obviously you've got to—somebody has to decide on the level of sensitivity and specificity that they're willing to tolerate and trade that off, as Deven said, with the privacy implications of that. It seems like as far as we can go is to say you need something more than these common elements; here's what's known. As Deven said there's a tradeoff to be made. And that does become the policy question that we can beg, based on the data and the technology possibilities.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
I think, yes, I mean I'm not sure we will have contributed much to the field with that conclusion but yes, that's certainly unarguable. But I keep coming back to this, sort of the tautological impossibility here. We're trying to come up with fields that will uniquely identify someone but will not reveal their identify. I mean how—you can't do that, right?

Deven McGraw – Center for Democracy & Technology – Director
Well although David though, I wouldn't even have suggested that that's the goal that we were trying to achieve. The problem with the use of the social security number is not that it identifies people; it does. It's that it's the key to a whole lot of other—it's the key to opening up accounts; it's the key to establishing credit; it's—and that's the reason why that number, in particular, has become the gateway to identify theft.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Right, but it's all nine digits necessary to open up an account or go establish credit, not—

Deven McGraw – Center for Democracy & Technology – Director
Right, but what I said to you before is that it's becoming increasingly more obvious that the use of the last four numbers, in collection with other data that allows you to reasonably predict the first five, is

increasingly becoming a problem. It is, admittedly, not the same problem as the use of all of the digits but it's—

David McCallie – Cerner Corporation – Vice President of Medical Informatics
So are you suggesting—?

Deven McGraw – Center for Democracy & Technology – Director
--to the extent that it's become the new, it's becoming the new trend in identify theft I think we need to acknowledge it. It's—the choice of that number is more problematic, potentially, than the choice of other numbers like a health plan ID which doesn't get you anything other than who the person is in the health plan. It's not the typology of, oh – it proves somebody's identify and therefore we can't use it. It's the fact that we are tinkering with recommending a number that, in fact, we're trying to reduce the use of because of identify theft problems.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
I understand and it's the joinability of that number to other databases that is the issue. But everything we can collect about the patient is joinable against the other databases, cell phone number—

Deven McGraw – Center for Democracy & Technology – Director
Oh, that's right but not in exactly the same way.

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
Deven this is Walter; just one quick question, just to clarify the concern because the exchange here is between someone that is looking for someone else and so they are providing name and address and date of birth—name and gender, date of birth, and then other things. I mean they provide the last four digits—you know it's on the receiving end of that inquiry that they will decide whether they disclose the data to the requestor so the way I see it this isn't being used out of the context of a healthcare information inquiry and so I'm trying to understand, where do you see the implication for like identify theft and—?

Deven McGraw – Center for Democracy & Technology – Director
Oh well Walter, every time that number gets collected, it gets collected initially for a legitimate reason, unless—but the more frequent collection of it and the more places where it's stored, you're just multiplying the possibility that somebody will take it and while it was initially collected for good purposes, use it for not good purposes. So whether it's somebody on the staff of a healthcare facility that takes a number and does something inappropriate with it; whether you have a breach of data which oh, we know never happens in the healthcare industry, or whether it gets end-received by a hacker. I mean it's not that the initial use of the data is problematic; it's that its collection and storage just creates the—it creates more possibility for it to be used for nefarious purposes which is the reason why federal government agencies are trying to reduce the use of it—reduce the collection and use of it for identify purposes. It starts out as being for completely benign and beneficial purposes but nevertheless, you've just increased the opportunities for theft; that's the problem.

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
But are you arguing that because of the recommendations of this Power Team there will be more collection of social security? Because what we're saying is mainly if you have the four digits—I mean if you have the social security, provide the last four digits. So we're not saying you should have collected the social security number. I mean that clearly is not what we are recommending.

Deven McGraw – Center for Democracy & Technology – Director

I guess I read your recommendation to be to require there to be a data field and certified EMRs for population of the last four digits of a social security number. That doesn't require its collection, I'll admit. But it's a strong incentive to do so and I think it has policy implications that need to be discussed.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Right. And there's no question about the policy implications. I'm just—I'm trying to push the envelope a little bit here to say do we want to turn it back into a policy question based on the statistical issues that the studies, which have evaluated to increase resolution and power of adding those four digits, bring to the table which is a dramatic increase in power. Is that sufficiently balanced against the policy risk to make it a recommendation—to the identify theft risk to make a recommendation. I'm trying to force the question—

Marc Overhage – Regenstrief Institute
And I appreciate that David because I think what you've done is help me, anyway, we're going to come back to—like you say, we have to have some value here and really challenging to say have we done enough. So let me—this has actually been an incredibly clarifying for me and maybe we'll fix my writer's block because I think I kept cycling back to this. But it feels like at least a way that we can draft a letter that I can sort of finish, get out to the committee to an asynchronously review and provide feedback on, which was sort of my original plan but then got stuck on this topic. I do think that the feedback and discussion that we've had from the presentation of these ideas at the Standards Committee has been actually very fruitful and helpful in terms of refining and focusing this.

So I will finish my draft, and draft is maybe even too strong a word - it's sort of a trying to get something down on paper with the specifics to your points, David and Deven, where we're sort of crystallizing into words something that we can then argue about, and get that out to folks to review and comment on asynchronously. And now turn this back to Judy for any public comment that we might have.

Judy Sparrow – Office of the National Coordinator – Executive Director
Great. Thanks, Marc. And operator, could you check and see if anybody from the public wishes to comment please?

Operator
I do not have any comments at this time.

Judy Sparrow – Office of the National Coordinator – Executive Director
Okay, thank you operator. Thanks, Marc, and everybody.

Marc Overhage – Regenstrief Institute
Thanks, everyone, and stay cool and we'll talk to you all soon.

Deven McGraw – Center for Democracy & Technology – Director
Thank you.

Walter Suarez – Kaiser Permanente – Director, Health IT Strategy
Thanks, Marc. Bye-bye.

M
Thanks, Marc.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Bye-bye.