# NATIONAL QUALITY FORUM

## Moderator: Sarah Fanta
## August 05, 2011
## 11:00 am CT

Operator: Welcome to the conference. Please note, today's call is being recorded.

At this time, I'd like to turn the call over to Sarah Fanta. Please go ahead.

Sarah Fanta: Good morning, everyone. Thanks for joining us today. We're going to be having our last call for the bone joint task. We're very excited about that. I'm joined here today by (Tarone Amine) and (Lorelei Dorian) and by phone Ashlie Wilbon and Sally Turbyville.

So right now, we're going to begin our discussion of Measure 1603, submitted by Ingenix. As you recall, it's their ETG hip fracture measure, originally it was submitted in conjunction with pelvic fracture; however, they took out the pelvic portion and just focused on hip fracture right now.

So if (Sherrie) or Tom want to begin and give a brief overview of the measure?

Tom Lynn: Yes, this is Tom Lynn. This is a measure that captures the cost of care for hip fracture based on the episode treatment grouper methodology, which creates episodes of disease from administrative claims data.

Once the episode is created, the administrative claims diagnostic data is used to gather severity markers and measure the severity of the hip fracture episode, and that's my overview.

Sarah Fanta: Okay, great. Thanks, Tom. Before we begin our discussion, I just wanted to do a brief roll call so your fellow TAP members know who's here right now and just so it's on the record. So, Jim Weinstein?

Jim Weinstein: Yes.

Sarah Fanta: Okay. Mary Kay O'Neill?

Mary Kay O'Neill: Yes, I'm here.

Sarah Fanta: Elizabeth Paxton? John Ratliff?

John Ratliff: I'm here.

Sarah Fanta: Okay. Catherine Roberts?

Catherine Roberts: Here.

Sarah Fanta: Craig Rubin?

Craig Rubin: Here.

Sarah Fanta: Okay. And (Patty)?

Patricia Sinnott: (Patsy)? Yes, I'm here.

Sarah Fanta: Okay, great. Thank you. Okay, thanks for that overview, Tom. I think that we're going to spend the majority of today, as I mentioned, just discussing basically the scientific acceptability. I know that the primary reviewers were Craig, Liz, and Catherine.

I know that you probably haven't had a chance to discuss it since Ingenix resubmitted their measure, but if you all feel comfortable and want to lead that part of the discussion, that would be great.

And then I did email out the SurveyMonkey link. As we're discussing the measure, if you want to just go through and fill in your ratings, if you do have access to a computer. If not, if you all could send in those ratings by next Friday, that would be great.

So we can go ahead and get started. Looks like Craig, do you want to start off?

Craig Rubin: Sure.

Sarah Fanta: ...(2A1). And I'll put it up on the slide.

Craig Rubin: So, let me...

Sarah Fanta: So (2A1) is that the measure is well defined and precisely specified so that it can be implemented consistently across organizations.

Craig Rubin: Yes, so just in general, this may not sit completely in the format, but hopefully I will touch upon the elements in the measure criteria. But as they point out, it's a very important problem - common problem. There's no question about the - I think the importance and the growing

importance that the cost of hip fractures and morbidity/mortality, especially over the coming decades.

The - my (major) comments or questions are with regard to the scientific validity and whether it captures, I think, some of the elements that are really, I think, critical to measure outcomes, especially comparing perhaps one site to another.

The data is generated primarily from a young population, those less than 65. This (steward) points out that the problem with hip fractures is predominantly one of an aging population and they - very few instances of capturing it - this is a clinical picture of most of this patient population.

So we can go over later in some of the details in terms of co-morbidities, severity, complexities, but they don't match with the patient population. That is (that) fractures, increase in frequency, incidents, prevalence, predominantly in the age 70-75 group and higher. And most of their information comes from a younger group.

And so they note that hip fractures will contain all clinically relevant information related to the condition, but I question whether that's true. And again, examples of conditions, status factors, (there's) diabetes and chronic bronchitis. There's a mismatch.

You know, in the clinical evaluation of these patients, and I'll also defer to the orthopedic surgeons, but, you know, the surgery is one issue and I think hip fractures is a perfect clinical condition to be evaluated in that it's, you know, very clear cut if you have a fracture.

But the course and cost in terms of management and comparing quality after the repair includes thing such as delirium, include typical post-operative complications, which they, I think, did address better in this. I didn't see previously pulmonary embolism or (DBT). It is in one of their co-morbidities and (now) I can find it.

But things such as delirium, you know, is not accounted for. When they asses severity of illness, and I'm sorry I have to click around here, you know, they use - so they do use age in the severity score. But, you know, they assign the same risk whether you're 65 or 90, and that doesn't make any inherent sense.

If I'm a provider of care and look at a cohort of people 65 to 70 with hip fracture, compared to a population of 80 plus, which is more likely to be the patient population you're going to be encountering, those risks are very different.

So, you know, why are they assigning the same weight is very discrepant, populations with different risks and mortalities - morbidities. I would think if I'm a surgeon I would much rather operate on a 65 year old than the 85 year old. Costs involved.

And they appropriately allude to the importance of these measures in terms of accountable care organizations and the medical home, but in particular this highlights the issues - these are system-based issues and they do capture nursing home, rehab and other costs involved in the care of these patients.

But don't drill down to those important elements in comparison, (the) descriptions of patient populations that would be relevant to accountable care organizations and the medical home systems of care.

So, you know, I'll stop here but (I can) answer any questions or what others weigh in.

Hello?

Patricia Sinnott: Yes. This is (Patsy) and I just have a couple of other questions. I'm, again, looking at the Excel spreadsheet of reliability and validity. I see the comparisons are across health care organizations, across peer groups and for cost and utilization.

And this is kind of, you know, just a technical issue. The results across peer groups for cost and utilization seem to include pelvic fracture still. And - so it's a little hard to tell - well it's very hard to tell what your comparison - comparing in the episodes.

I haven't gone back and compared it to the other spreadsheet but...

(Sherrie): Can you tell me what section you're referring to? This is (Sherrie) from Ingenix. Because we did alter all of our output results for the utilization factors to be just hip fractures...

Patricia Sinnott: Okay, so I'm on - I'm looking at the spreadsheet that's called reliability ((inaudible)) (valid), and if you look at results across - the first tab is reliability across HCOs. It calls it - for question essay answer Ingenix, reliability, the measures hip/pelvic fracture.

And then the - I'm sorry, that's cost. So the HCO one still is called hip/pelvic fracture, although it is now the ETG-based 713103 hip fracture is - the results are described in the tab reliability across HCOs.

But if you look at results across peer groups, it includes pelvic fracture, the ETG group pelvic fracture 70328, and if you look at results across peer group's utilization, it also includes peer group - pelvic fracture. Do you see what I mean?

Tom Lynn: Yes, I see it. This is Tom Lynn. Oh wait a minute, (I'm on mute).

Patricia Sinnott: And also the exclusion seems to describe hip/pelvic fracture also. But we don't really know. So I would like to go - my guess is that's, you know, something that was missed or maybe just a typo.

But if we go back to reliability across HCOs, what I understand you describing is you run through the episode grouper and define episodes separately for this condition, the hip fracture, for different health care organizations and compare them on cost and utilization or episode numbers - yes, visits.

So it's cost and visits. So what are we to understand from these tables that - I mean there's a lot of variation in the number of episode quantities, there's a lot of variation in the cost per episode, the primary care cost per episode. I mean what are we supposed to infer from this - these tables as to reliability?

Tom Lynn: I think you're supposed to infer from these tables that, you know, the variability that you see here could be - is, you know, what we'd expect amongst variabilities in health care organizations and is relatively a standard cost for the episode.

You know, I mean we're trying to show - that's what we're trying to show here.

Patricia Sinnott: But we don't know whether running data from different organizations is - through the episode grouper is actually including and excluding - we don't know what the reliability of the grouper is from this, correct?

Tom Lynn: Well, I think it's a measure reliability in that it's showing a relatively consistence cost across these health care organizations.

Patricia Sinnott: But we don't know that the cost - the relative consistency of the cost is due to the grouper function or the charge function, although my recollecting is (just) standardized cost, is that correct?

Tom Lynn: I believe this data is generated on using standardized cost.

Patricia Sinnott: Yes. But again, we're back to the question of the episode grouper, which is the basis for all of this.

Tom Lynn: Correct.

Patricia Sinnott: We don't know that the grouper is reliable from these data, from these results.

Tom Lynn: Well, I mean - I think this data alone maybe doesn't show it, but I think that it does show that it's a relatively consistent cost across health care organizations. I don't know if I'm adequately answering your question.

Patricia Sinnott: No, I mean you're answering it. But for example, I would want to know that - so I take the ETG grouper and I can change some of the criteria, for example, defining a low outlier or a high outlier based on cost or utilization. I can - I might change the pharmacy inclusion of exclusion. I might change the duration of the episode or the absence of care algorithm.

And then I would take one group of data or one HCO or all of them and change the criteria for the episode definition and the inclusion and exclusion definition. And I'd want to make sure that if I ended up with 1000 episodes with the first definition of inclusion and exclusion, that if I - let's say I changed the high outlier from $50,000 to $25,000.

I should know exactly how many episodes were excluded if I change that criteria.

Tom Lynn: Yes, we have that information. And in that particular example, it wouldn't change the number that were excluded, it would change the dollars that were excluded. But we cap the high cost, we don't throw those out.

But I hear what you're saying. We have that information about, you know, what's excluded and we certainly could provide information for what would be excluded.

The dollars that are excluded (in that case) is that different outlier threshold we could provide. We do exclude cases that are low in cost, we absolutely exclude them. We don't (half) them or whatever the bottom part of that is. We do exclude them, so we could vary the low outlier and talk about the number of cases that are excluded by varying a low outlier.

We have the data to do that, I guess you know that.

Patricia Sinnott: Okay. But it's not here in the report.

Tom Lynn: It is not.

Patricia Sinnott: Right, okay.

Sarah Fanta: Okay, so it sounds like we've pretty much covered the (2A1), the measures well defined and precisely specified as well as (2A2), the reliability testing. So if anyone - if there's nothing else to add, we can move along to (2B1), that the measure specifications are consistent with the evidence presented to support the focused measurements.

And that the measure is specified to capture the most inclusive target population indicated by the evidence and exclusions are supported by the evidence. It looks like that was also assigned to you, Craig.

Craig Rubin: Yes. So, again, I'm coming from the, you know, clinician's perspective in terms of validity - (2B1), right, we're talking about?

Sarah Fanta: Yes, you're right.

Craig Rubin: The validity and...

Sarah Fanta: Well, (2B1), yes, the specifications are consistent with the evidence presented.

Craig Rubin: Right. And, again, the problem that I have is, you know, the evidence presented is, you know, hasn't captured the patient population most likely to be affected by this problem. And no matter what numbers you generate, then how relevant they are, you know, I question the validity.

Male: Yes, I mean I, you know, I'm just going to make a comment here as a measure developer. Don't mean to interrupt or speak out of turn. But, you know, the fact of the matter is, you know, just to be fair and open. You guys probably already know this. This rule was developed with commercial populations in mind.

And, you know, most of the data that we have here is commercial data. So we don't have a ton of Medicare data. So, you know, I have to, you know, say that that's true.

Craig Rubin: So, then the - and I understand that, I sympathize with that. But then if you want to reconcile that with the issues of relevancy to, you know, broad populations to, you know, the national health care issues in terms of this problem, the relevance and validity may be limited.

It may not be any fault of the, you know, technique used, but the data may have limited broad applicability to those outside of the group studied, but those outside is the majority of the ((inaudible)), you know. And what was studied was the minority.

So that gets the question of validity and applicability. I don't know how to reconcile that.

Mary Kay O'Neill: This is Mary Kay and, you know, I was thinking about what you're saying.

In some ways, clinically, if you look at hip fractures in a younger population and in older population, they may actually represent quite different clinical situations, you know, other than major trauma if a 40 year old has a hip fracture, that's telling you something quite different than if a 70 year old has a hip fracture.

And so that incidence rates and utilization patterns and complications and co-morbidities, even as people are getting to the older age of the commercial age cohort, may be more Medicare (liked) from a clinical situation.

So it - as a population measure, it may actually be kind if confusing different clinical entities and might be hard to tease out actionable items because we're really talking about very different clinical conditions in patient populations.

Male: Right. I agree completely and you articulate if very effectively. But the reality is that the minority of - I mean it's rare for a 35, or the example given, a 57 year old, to have a hip fracture unless, you know, it's in the traumatic fracture, a motor vehicle accident.

But the world, in terms of - well, the patients, the latter situation, the older patient, and as the developer points out, falls, you know, is the typical factor, and that is the patient population that has the greatest relevancy for health care policy, for expenditures.

And they are very different. And so if you're trying to capture cost or savings, it's hard to identify with the input or logic that was provided for the risk factors and so forth, it's going to be hard to (tease) out, you know, the origin of cost in different populations if those weren't even specified.

The typical perioperative complications in the 80 year old, that is the real world in terms of numbers, not the 57 year old. So if anything, you know, I sympathize with the challenge because the database mostly had younger patients.

But the clinical problem as demonstrated and put forth, that this is a serious problem, growing numbers, you know, ten times - this is the developer's numbers - ten times higher risk of fracture.

In 80 year olds this is the largest growing segment of our population, and so, you know, I think the utility of this measure is going to be very limited because there just aren't that many hip fractures, you know, in a - non-traumatic hip fractures in those, you know, less than 65.

John Ratliff: This is John Ratliff. Just chiming in, not to (beat) on this horse further. But for the relative risk assessment that you have and the severity score you proffer, the explanation you give is quite good.

But all of the measures that you use by definition are coming from a patient population of less than age 65 since the database you're using is, again, defined as like a non-elderly patient population. And then for each of these individual co-morbidities, you're taking out the impact of their - you're taking out their impact on patient outcomes to like four decimal places.

I mean if you're pretty darn confident that what you're saying is accurate and, again, you're confident that the measure is meaningful. I don't know how valid co-morbidities in a 35 year old are going to be in comparison to a 75 year old. A 35 year old with diabetes is going to be different than a 75 year old with diabetes. Similarly with like COPD.

I mean it's a really different patient population. And I don't know if I can get that bridge from the database that this is based on to the patients in the real world where we're going to be capturing (capital) outlay for hip fractures.

Sarah Fanta: Okay, great. So looks like we adequately covered (2B1) and we can move along to (2B2) which speaks to the (Blood Z) testing, making sure that it's demonstrated across the measure data elements.

It looks like Liz Paxton was assigned this but I don't believe that she's here today, so if anyone wants to go ahead and jump in to begin that discussion, that would be great.

Female: Just confirming, it's (2B1) right?

Sarah Fanta: Right now we're on (2B2).

Female: ...(2B2).

Sarah Fanta: Yes.

Female: Well, I think we've talked about this.

Male: Right. I think just in looking at the statistician's comments, if I may, in (2B2) Item C, the data representative of the target population question mark, ((inaudible)). And then, you know,

furthermore, another question, validity, is - I'm not sure (if there's) a risk adjustment model for those in different age groups, 64 to 74, 75 to 84 and greater than 84.

And the statistician comments, whether this is adequately represented in the database. But my - the question I have is they're all assigned the same risk coefficients, whether you're 65 or 84, and I didn't, you know, understand how that ((inaudible)) validity is possible.

Male: You want the developer to answer that question?

Male: Yes, that - (I'd) be happy, yes.

Male: Yes, I mean you guys have pointed it out. I mean it's the limitation of the data, and we - it's a commercial data set, you know, there were - there was a relatively small percentage, certainly not representative of folks over 65 who, you know, happen to be on commercial programs for - I think there's some like Medicate Part C folks that were managed by various health insurers.

And we - so it was - we didn't have enough patients to sort of stratify them further between - from 65 and up. And, you know, it's a limitation of the data issue.

Male: Right. So your findings were that the 85 year old risk was the same as the 65 to 74 risk? And that's (the data you had).

Male: Yes, I wouldn't say those were our findings. They were...

Male: Well, they all have the same value in the coefficients. Is that - I'm trying to find the right...

Male: They do. They do have the same value because there's only one age range and that's over 65.

Male: But in one of the tables you separate them out - you separate those age groups out which I can't find which - in risk factors which table that was, but...

Male: ((inaudible)).

Male: Pardon?

Male: I think it's F-8, and there's a demographics tab on F-8, hip fracture clinical logic. I'm not sure it's exactly set up the same way with you guys. I hope so.

Male: Yes, right. Okay. So there's, you know, each (side) is minus 0.89 and it's the same for all those age groups. And that's what you may have found, is that right? Am I...

Male: No, no, no. That's not right. That - we were only - because of numbers, we were only able to really stratify based on 65 and above and then we applied those to all those age ranges.

Male: Okay. So it's unlikely - I don't have the data but I'm fairly confident to say that the risk of an 85 and 65 year old is not likely to be the same.

Male: Right.

Male: So, you know, in terms of (2B2) is the data representative of the target population? Again, I don't want to ((inaudible)) but I think, you know, this speaks to it again.

Sarah Fanta: All right, great. Now we'll move along to (2B3) which are the exclusions are supported by the clinical evidence, otherwise they're supported by evidence of sufficient frequency of occurrence that the results are distorted with the exclusion. ((inaudible)), go ahead.

Male: Yes, that's me. But I don't really have anything else to add.

Sarah Fanta: Anyone else? Then we can move along to (2B4) which is resource use measures and other measures when indicated in evidence based risk adjustment strategy as specified. And it's based on patient clinical factors that influence the measured outcome.

So if anyone wants to begin discussing that, that would be great.

Female: I wonder if the developer could talk a little bit more about the low dollar exclusion that you use for inclusion in the analysis per provider?

Male: Yes. The - that exclusion is based on group - well first we exclude some very low dollar amounts, but I think that's very low, like something that would be less than an office visit.

Female: Yes.

Male: And then we take - we use the (stress hold) for the bottom to be the bottom 2-1/2% of episodes. That's how we came up with that.

Female: And is that per provider or per all episodes?

Male: That's per - that's done by - it's done overall episodes, but there's a separate one for severity and I think treatment indicator.

Female: So you're saying - for example, I think you're saying that for all Severity 1 hip fractures combined, you exclude the bottom 2-1/2% of episodes from further analysis?

Male: That's correct.

Female: You do not do it by individual provider type - by individual provider so that Dr. (Jones) gets the bottom 2-1/2% and Dr. (Smith)...

Male: No.

Female: Okay.

Male: No, it's not by provider. It's - we could calculate separate (trim) points for each peer group, not by provider. I think in this particular case there's only one peer group, so it's not really an issue.

Female: Yes. And what is the hypothesis there?

Male: That we cut out the bottom 2-1/2%?

Female: Yes.

Male: The idea there is that, you know, the bottom 2-1/2% are probably most likely to be mistakes or miscodes or things like that. And so they're eliminated.

Female: Okay, so that they're probably incomplete.

Male: Correct.

Sarah Fanta: All right. If no one else has any other questions about that section we can move on to (2B5) which deals with the identification of statistically significant or practically or clinically meaningful differences in performance.

Patricia Sinnott: So I think - this is (Patsy) again, I'm looking at (S10_RAXLS) which is the risk adjustment method. It's a comparison of individual provider to peer group, right? And so you're looking at a relative cost of care ratio. Are there any other statistical tests that you recommend to define someone as different?

Are you using a 5% - a 0.05 or 0.10?

Male: We do recommend a statistical method. (Sherrie) is going to have to help me whether it's described in detail here because we do recommend that only statistically significant differences be reported or acted upon or that the measurement be reported with a confidence interval.

And we use a method that was described. It's often used by (RAND) when they do these episodes (for the) studies that measures the variance of the ratio.

Patricia Sinnott: Yes.

Male: So we wouldn't say that just because it's 5% different that it's different. It depends on what the confidence interval is for that measurement.

Patricia Sinnott: Yes.

Male: And of course that depends on the underlying variance of the episode cost as well as the number of cases that you have.

Patricia Sinnott: Yes.

(Tarone Amine): This is (Tarone). I would also urge the committee to look at Page 71 on the sample (edition) profile to see some of the output to see if it's meaningful clinically and statistically meaningful.

Female: And so that's in the submission?

(Tarone Amine): Yes.

Female: So my page is - I'm sorry. So what's your page number again?

Sarah Fanta: This is on Page 71 of...

Female: Okay.

Sarah Fanta: ...((inaudible)) for 1603 and it's also being projected on the screen - the webinar screen.

Female: Okay.

Sarah Fanta: So if we have no other comments was can move along. We generally are skipping (2B6) which is the criteria that deals with multiple data sources. That's typically just been skipped. So we can move along to 2-C which discusses disparities in care.

And it looks as though Catherine Roberts, you'd like to begin leading this discussion?

Catherine Roberts: Absolutely. Actually I had no idea that was assigned to me. That must be on a different one. I have 3-A, B, C and 4-C and D, no?

Sarah Fanta: Let me see. Oh you're right. Actually, it should be Craig.

Craig Rubin: Yes. So...

Sarah Fanta: ((inaudible)) if you want to.

Craig Rubin: They do address some disparities and incidents of hip fractures in different racial groups, but the numbers are very small. And they clarify that they are - age and sex are part of the risk adjustment models but, you know, the numbers are very small in terms of different racial groups.

Sarah Fanta: Okay, great. And if there's no other comments about scientific acceptability, that was the last criterion. Before you all go ahead and vote on the survey tool, if you want to use this time to ask any general questions to the developer about any issues with usability, feasibility or importance, you can use this time right now.

I don't know if anyone has any questions.

Ashlie Wilbon: Hey Sarah, this is Ashlie. Maybe have them discuss the first three - or maybe the middle two usability criteria just based on some of the discussion that came off the list. The applicability of the measure to the data that they have available, so 3-B and 3-C maybe.

Sarah Fanta: Okay, we can do that. So 3-B, is that the measure performance results are considered meaningful, understandable and useful to the intended audience for both public reporting and quality improvement? So I think I got this right this time, but I think Catherine, that was your criteria?

Catherine Roberts: Absolutely. And, you know, well I would like to first start and recognize Ingenix for their massive improvement in this measure in a very short period of time. So thank you to Ingenix.

Of course no measure is entirely perfect. I believe the only thing with usability preliminarily that we really wanted clarified was the data, whether that was, you know, both public and private. And I think they've pointed out that this is predominately private data.

The usability, at least specifically for 3-B, Craig and I kind of rated as moderate. And, you know, it's hard to be perfect, but within the limits of what they have, it seems adequate.

Sarah Fanta: All right, great. So if there aren't any other comments about 3-B, we can move along to 3-C, which is that the data and result detail are maintained such that the resource use measure, including the clinical and construction (logic) for defining a measurement can be decomposed to facilitate transparency and understanding.

Catherine Roberts: And again, sort of within the limits of what we've talked about and practicality, again, that's a moderate rating at least for me and preliminarily from Craig I believe.

Craig Rubin: Yes, I've just - I mean I know - I agree with the effort and work put into this. There's no question about that. But in terms of having a measure and endorsing something for the general public, you know, it's going to have very limited, I think, usability because the patient population is not what - it was not the focus.

And it's, you know, nobody's fault but as I said it is what it is. So I think the usability - it will be very focused.

Sarah Fanta: Okay, any other general usability or feasibility issues anyone wants to raise at this point?

Mary Kay O'Neill: This is Mary Kay, can you hear me?

Sarah Fanta: Yes.

Mary Kay O'Neill: Okay, sorry. I have a new phone and the mute thing is not intuitive.

Sarah Fanta: Okay.

Mary Kay O'Neill: So the, you know, I guess my concern, given the discussion about the age cohort for the database Ingenix has to work with, which is understood, and the natural history of this clinical condition.

And the fact that what we might be doing clinically is reporting on mixes of patients in very different clinical situations and looking at utilization data by individual physicians or systems around really a mismatch of clinical settings, you know, clinical storylines.

So I'm just always very hesitant to have measures go out there that appear to explain something they're not truly explaining.

And so I would have some reservation about sort of (blanketly) applying this across a commercial population base without being clear that somebody - a young healthy person with a traumatic, you know, sports or motor vehicle related hip fracture is not being really looked at the same way an osteoporotic older person is with their hip fracture.

And maybe the risk adjustment manages that, but it's not clear to me that that's actually the case. So we might be sort of dumping what are essentially different diagnoses in the same bucket. I'm done.

Sarah Fanta: Thank you. So in conclusion, Dr. Weinstein, I don't know if you have any comments that you just want to make overall about this measure?

Jim Weinstein: Well, no I didn't want to comment too much. I mean I think that the group has hit all the issues because the sample that was given or that the Ingenix folks used is - makes it hard to validate as was discussed and it's quite different to see a fracture in somebody less than 65 and somebody 80.

I mean the co-morbidities, the survival rate, the treatment, the - I mean just everything is different. And, you know, I'm not sure how valid you'd want to suggest this measure is. And it's - on the other hand the algorithms with the right population probably could work pretty well.

So it's unfortunate. Maybe we should test this on a Medicare database or if they wanted we could probably do that for them or with or however they want it. You know, we have those populations at Dartmouth. We do it all the time with the Dartmouth Atlas.

So I hate to throw away the good work, but I certainly think we should probably run this algorithm with a different population before I would want it to be going out to anybody as a hip fracture measure.

Male: We would, you know, certainly love to have that sort of opportunity to work with you. We're working with Medicare data right now that has the variations in age that, you know, addresses the population that's...

Jim Weinstein: We probably have, you know, we have years and years of data of hip fractures. It's a population we've studied a lot. And we have cohorts of populations, you know, we probably have several hundred thousand or a million hip fractures in over 65.

So we could even do segments by age which would be interesting. So if you wanted to work with us on that we'd be happy to be helpful. I've been talking to (Sam Hold) about some other work with you guys to try to help as well, but I leave that to you.

But I would worry just from this exercise to support this without that kind of effort because that's really the population that we need to study.

Craig Rubin: This is Craig. Can I add one other point?

Sarah Fanta: Go ahead, sure.

Craig Rubin: Again, a different day, but I would urge if there is an analysis done of the Medicare population, there are some very obvious co-morbidities and complications that are not being addresses above and beyond age.

And that is, you know, dementia which is a risk factor for the hip fracture in the first place, but also recovery and other expenses. Post-opt delirium, you know, which doesn't appear here as a co-morbidity or severity index score.

You know, wound healing, urinary tract infections post-op. You know, all the clinical challenges once a person's out of the OR that are very common that should also be evaluated. Thanks.

Male: (Thanks).

Sally Turbyville: This is Sally. I have a question for Jim, given your comment as well as everybody on the panel. So I believe, and Tom you can correct me, did you test this measure only in commercial data as presented to the TAP?

Male: Yes.

Sally Turbyville: Okay. So just as a reminder, that would mean this measure would be endorsed for the use in commercial populations only until you had the opportunity to come back with testing in Medicare, which I doubt there is sufficient time to do that under this project.

So, you know, and to your point, Jim, absolutely. As you know, in general, for all measures we're looking for the industry to move to testing it across more populations.

So that said, Jim, what I'm interested, because your comment struck me as to whether or not you would then consider this measure if it were to be endorsed for implementation in commercial populations only to then not meet the importance criteria that is for the commercial population because this type of medical occurrence, a hip fracture, is not sufficiently prevalent in the commercial population.

So just wanted to probe your question a little bit further.

Ashlie Wilbon: And, Sally, this is Ashlie, just to piggy-back on that. And whether or not I think this also touches on the usability criteria as well.

So not only in the importance but if the - your concerns about the applicability of this measure in the commercial population impacts the usability in such a way that the measure greatly limits the usability of the measure and how that might impact your recommendations for the measure, then obviously I would be interested in hearing about your thoughts on that as well.

Jim Weinstein: Yes. Well, as an orthopedic surgeon, you know, the number of hip fractures that we treat in people less than 65 is pretty small. And they're very - they're usually traumatic, very serious injuries and have very significant ramifications. But it's a different animal.

Maybe that's not the right word, but - so I think everybody's, you know, Craig and everybody's highlighted that and the issues surrounding these patients - once they - the hip fracture in itself is almost a lethal event. I mean 30% mortality in one year after hip fracture.

And then you throw in all the other issues with cognition and ambulation and many of them actually, you know, become necessary to move out of their homes into assisted living facilities. I mean it's a devastating event for most people.

And the other thing, the epidemiology of these fractures is changing. We're starting to see this in older people even as Craig alluded to another decade older, which then creates even more problems because they're more serious and more difficult and they're more complicated fractures in these people.

So I think it warrants an effort that, you know, again, we'd be happy to work with the group or Ingenix to try to be helpful if that would be useful. But I understood what the comment - the questions were, and I couldn't - I'm not sure I could support it in a different population because I'm not sure how relevant it is.

Craig Rubin: This is Craig. I just wanted from the staff a clarification in criteria 1-D in terms of importance to measuring report. It seems that a requirement is - I'm reading - a demonstrations resource use or cost problem and opportunity for improvement. Goes on - and delivery of care across providers and/or population groups.

So I don't see that justifying - I'm sorry, justifying is probably not the right word, but indicating that a sub-group, for example, commercial only, is appropriate. I thought by that definition it should be appropriate across population groups.

(Tarone Amine): It would - this is (Tarone) by the way. It would be relevant for the population of focus within the measure, and since the measure is specified for use in the commercial population, really should be looked at in that context.

Craig Rubin: Thanks.

Male: Seems to me that's kind of a bait and switch.

Craig Rubin: Yes.

Male: I think of a hip fracture and I don't think of like a commercially insured 45 year old patient. I think of a 75 year old. I think the measure of the foundation here is valid. This is an important public health issue. Maybe the measure of the tool we're studying isn't getting there.

So I kind of answered the first 1-A, B, and C from a framework of thinking about grandpa with a hip fracture, not a commercially insured patient with a hip fracture.

Mary Kay O'Neill: This is Mary Kay again. You know, I guess what I'm trying to sort out in my mind is there's these sort of like a clinical entity which might be called geriatric hip fracture.

And then there's probably two other entities, one of them is trauma fracture in otherwise healthy people and then there's, you know, those few cases for younger people with systemic diseases that give them osteoporosis or whatever.

But - which is too small to measure I'm sure. But in the commercial population, a couple, you know, we're having a little bit of the demographic shift in the (covered lives) getting into older ages as people extend their work life, (timing).

And so there is a sub population within the commercial database that is essentially behaving in the same way as Medicare folks because of their age group and, you know, what their medical condition is.

So the problem for me in the commercial is that you're ((inaudible)) three different clinical entities probably primarily two into the same measure and looking at utilization for very different clinical conditions.

But we do in commercial have the, you know, it's not a bright line between commercial and Medicare, you know, some of our 58 year olds are like that and we've got some fully insured 72 year olds because they are still working.

So just to be able to accurately identify the clinical entity we're evaluating with the measure is probably the most important thing and to get a critical (mass) of data would, as Jim pointed out, probably require the access to the Medicare data.

Sarah Fanta: If there aren't any further comments then I'll go ahead and ask (Robby), if you could please open up the lines for public comments.

Operator: All lines are open.

Sarah Fanta: If anyone listening in right now wants to make a comment, you can go ahead or ask a question. Okay, it appears that we don't have any comments or questions at this time. So I just want to thank everyone for your thoughtful participation and your effort throughout this process.

I know it's really time consuming, but we definitely really appreciate all the work that you've put in and I know that the resources steering committee also really appreciated the detailed review of the measures.

Ashlie Wilbon: Sarah?

Sarah Fanta: Yes?

Ashlie Wilbon: This is Ashlie. Really quickly, just a last piece of guidance on - as you submit your ratings for this measure. While, obviously, I think the opportunity to pass this measure in Medicare data would be great, particularly for this effort, but within the scope of this project and the timing it wouldn't be possible at this point.

So I would like for you guys to, as you're rating it, really think about the measure in the context as it was submitted and be clear in your rationale in the spaces that were provided in the survey for why, you know, you've rated things as you have.

That information will be really helpful for the steering committee as they review the measure and make any final recommendations. So it's evaluating the measure as is, so just wanted to put that out there.

Sarah Fanta: And if you all could complete those evaluations by next Friday, COB, that would be really helpful on our end.

So, thank you to everyone, especially Dr. Weinstein for chairing this task. I think he did a wonderful job and I'm sure everyone would agree with that. So just as a reminder, the Cycle 1 report is going to be going up for comment on August 26, dealing primarily with like the CB and diabetes measures.

So feel free to leave comments on that report and the steering committee will be discussing these measures during their August 30 and 31 in-person meeting. Are there any other questions before we adjourn?

Female: Thank you.

Sarah Fanta: Okay, thanks everyone. Have a great weekend.

Male: Thanks.

Male: Bye.

Male: Thank you.

Male: Bye-bye.

Sally Turbyville: Great job, Sarah and Ashlie and (Lorelei) and (Tarone).

Female: Thanks, Sally.

Female: Thanks, Sally.

Female: Bye.

Female: Good job, Sarah. Thanks a lot.

Sarah Fanta: Oh sure.

Female: Talk to you guys later.

Sarah Fanta: Bye-bye.

Female: Bye.

Sarah Fanta: We're all good, (Robby)?

Operator: Yes.

Sarah Fanta: Okay, thank you very much.

Operator: You're welcome, have a good day.

Sarah Fanta: You too. Bye-bye.

Operator: Bye.

END