

第 4 6 回科学委員会

日時 令和 5 年 7 月 2 6 日 (水)
1 4 : 0 0 ~
場所 PMDA 会議室 1.2 (6 階)
開催形式 ハイブリッド会議

<開会の挨拶>

○事務局（緒方科学委員会・先端科学統括課長）定刻となりましたので、第 46 回科学委員会を開催いたします。本日はお忙しい中、ご出席いただきましてありがとうございます。

はじめに、7月1日付でPMDAにて人事異動がございましたので、紹介させていただきます。RS 統括部長の高橋です。科学委員会・先端科学統括課長の緒方です。それでは、事務局から委員の出席状況の報告と資料の確認をいたします。

委員の出席状況を申し上げます。親委員会は 19 名の委員のうち 19 名にご出席いただいております。全委員の過半数に達しておりますので、設置規程第 7 条の規定に基づき、本委員会の成立をご報告いたします。今回は対面及びウェブのハイブリッド型会議となります。会議室にて 5 名、ウェブにて 14 名の委員にご出席いただいております。

次に配付資料の確認をさせていただきます。不足の資料があれば事務局までお願いします。

次に資料取扱区分表をご覧ください。本日の配布資料である、資料 1, 2 は「取扱注意」、資料 3 は「その他」に該当しておりますので、それぞれ取り扱いにつきましてよろしく願いいたします。

ウェブから参加の先生方におかれましては、通信状況によってはビデオ送信の停止をお願いする可能性がございます。その際にはご協力をお願いいたします。また、ハウリングを防ぐため、マイクは、無効（ミュート）の状態としていただき、発言する際に有効としてください。発言が終わりましたら再度無効（ミュート）に戻していただきますようお願いいたします。

また、今回はウェブ録音から文字を起こして議事録を作成します。速記業者の録音ではないため、議事録作成の際には先生方に御発言の内容を御確認いただくことがあります。その際には御協力のほどよろしくお願いいたします。それでは早速ではありますが、渡邊委員長、議事の進行をよろしくお願いいたします。

<報告書案について>

○渡邊委員長 本日はよろしくお願い致します。進行を務めさせていただきます渡邊です。早速ですが、議事に入っていきたいと思っております。

本日の議事は 2 つあります。専門部会の報告ですが、AI を活用したプログラム医療機器に関する専門部会で報告書案が出来上が

ったということですので、その検討を行い、その後、標的特異性を有する in vivo 遺伝子治療用製品のベクターに関する評価の考え方専門部会について、進捗報告をお願いしたいと思います。

まず、AI を活用したプログラム医療機器に関する報告書案について議論いたします。この専門部会は、2021 年 12 月の第 40 回科学委員会で設置が決定され、2022 年 7 月～今年の 6 月までに合計 6 回の専門部会を開催し、報告書案を取りまとめていただきました。それでは、佐久間部会長より御説明をお願いしたいと思います。よろしく願いいたします。

○佐久間委員 佐久間です。それでは、報告書の概要について資料に基づきお話をさせていただきます。

これは専門部会の委員名簿で、特に読み上げることはいたしません。臨床の先生方、それから医療用のデータベースを扱っていらっしゃる先生方、それから工学系で AI や医療機器などについて専門的にやられている先生方に集まっていたいで議論をさせていただきました。また、一部弁護士の先生が入っていらっしゃいますが、これはやはり個人情報保護のことがありますから、これは科学委員会の範囲をちょっと超えていますので、それについては触れるだけにしております。

目次ですが、科学委員会で、この前の報告書で、こういう深層学習を含めたようなものが出てきたということで、その課題を示しました。これが実際に出てきたことで、どういう動きがあるかということを受けた報告書になっております。国内外の関連動向の分析、機械学習におけるバイアス、それからそこに書いてあるような研究、特に評価データの再利用です。実は、医療データというのは非常に数が限られているので、評価データを再利用することがあり得るだろうと。そのシナリオのときにどういうことが起きるかを考えておかないといけないかを示したものです。5 番目は、文献の報告です。やはり医用画像の開発で 4 番に示したような課題が如実に出てくる例がありますので、その辺りを示しました。一方 6 番目は、先ほど医療データが少ないのでモデルシミュレーションを使って学習データを作る可能性がある。なかなか評価データを作るのは難しいかもしれませんが、例えば機械による影響であるとか模型の問題について考えると。

7 番目は、現在いろいろなデータベースが構築されています。もともとは臨床上の評価をする、医療技術の評価をすることを目的にできたものが多いのですが、実はこれが機械学習に使われると

いう可能性が出てきています。それに関して、どういう留意すべき点があるかということ、課題とともに分析させていただきました。

最後に、こういうことで一言でデータと言っても、学習に使うデータ、開発のときに使う検証データ、それからいわゆる承認であるとか認証のときに使うテストデータの 3 つに分けて、これはどういう関係でなければいけないのかということをもとめさせていただきました。

過去の議論としては、そこに示したとおりで、まずは国際的にもいろいろな動向があるのです。それから、やはり統計的にある偏りをどう持っているかが結構重要な話になるのですが、バイアスというものをもう少し広い概念で捉えるということで議論いたしました。それから、先ほど言いました評価データの再利用をするときに、どういう危険性があるのか、あるいはその解決策としてはどういうことがあるかを考えました。モデルの話をして、それからデータベースの話とするということにしました。対象としては、データベースとして扱ったのは医用画像解析が一番多いわけですね。それから、心電図、脈波などのバイタルサインもこれから出てくるでしょう。いわゆる、スマートフォンやスマートウォッチなどで出てくる可能性がありますので。それから、放射線の計画立案、内視鏡の画像などがあります。対象外にしたのは、ゲノムデータの解析分野については対象外といたしました。これをやっているうちに、今問題になっている大規模言語モデルによる AI というのが出てきて、これを途中から入れるのは難しいということで、今回これはあえて外しました。ただ、一方で、これはまだエラーが多いということも分かっていますので、今後この辺りは出てくるかと思えます。実際に論文では、自然言語で画像を与えて自然言語で答えを出すというものが研究レベルでは出てきていることを把握していますので、これも出てくるだろうと思えます。

国内外の関連動向ということで、まず FDA の動向です。Pre-certification プログラムということで、いわゆる承認の前のプレの段階である程度のリスクを見た上で、それをリアルワールドに入れた上で評価していくということをやってみたわけですね。それで、いろいろなことでのガイダンス案が出てきているということで、アクションプランも出てきていると。ただ、この分野は非常に難しく、実際にこういう考え方でやってみると、まだまだ幾

つか検討すべき課題があるということで、当初 2019 年頃に想定したものではなかなかうまくいかない課題がありそうだ。ただ、やはりこの方向性は今後もあるのだろうけれども、現状においては個別に 1 個 1 個評価していくというようなことを、従来の医療機器的にやっていく必要があるのだろうという動きになっているというように分析しております。

それから、もう 1 つ重要なことは、こういうことに対して Good Machine Learning Practice、GMP に近いようなものになるのですが、そういうことをまとめた文章があります。ここは議論したことと関連するのですが、非常に短い文章なので何を意味しているかということがなかなか分かりにくいことになっております。もう少しそこに対して分かるような形での報告書にするということです。

それから EU は、人工知能に対する整合的規則を作るということで、これは国際規格の動向もあるわけですが、ISO、IEC といったような所でそういう議論が始まっています。ただ、やはり通常の産業用、一般用の AI と違ってリスクがあるので、そこをどう扱うかということについて今、継続的に議論がなされていると思いますが、急速にその辺りも動いてきているかと思えます。

これが今の繰り返しになりますが、ポイントは機械学習、今までの SaMD では市販後に学習するというのがなかなかないのです。しかし、やはりこれが特徴ということで、それを向上させることが可能であるということで、そこに対して市販前の段階でそれを視野に入れたことを考えようとしているわけですが、ここに書いてあるとおり現在の科学的知見では市販後の性能変化を、適用拡大の範囲を想定することが市販前にはなかなか難しいというような分析になっています。判断基準や要件を一様に設定することが現実ではないということで、そういう意味ではリスク・アンド・ベネフィット評価を企業がきちんとやって、今までのような医療機器の評価と同じようなことをしていくということです。

我が国において IDATEN という制度をつくってきています。ここに書いたような、ある程度市販後の変化も容認できるような形にすることを想定しているのですが、これについても記載されていますが、現状こういうことを考えていったときに注意深く見ていく点はあるかと思っています。ただ、やはりこの分野は、どうしてもこういう性格のものを作っていくたいという技術的方向性がありますので、引き続きこの辺りの解釈についての動向を注視し

て、できるだけ効率的な評価をしていくことが求められているのだらうと思っています。

バイアスなのですが、実はバイアスというとデータ自身を持つバイアスというように大体思われるかと思います。これはたとえば言うと、ある病院で取られたデータが他の病院で取られたデータと統計的に同じなのか。これは、実はよく臨床評価のときに患者背景などといいます、そういうことがきちんとできているかということです。もう 1 つは、機器で撮影されたデータ、これも結構機器ごとの特性が微妙に出てくる場合があります。これは特定しないといけないだらうということがあります。それから、年齢、人種などのデータも、それが科学的にどう関わるかということも考察した上でバイアスを考えなければいけないです。それから手法を決めると、どうやってデータを分割するかということがある程度狭められてきます。そうすると、そこで手法特有のバイアスもあるという形になるので、その辺りは見ていく必要があります。

それから、基本的には母集団を標本化するということは、バイアスがあるのだと。それをできるだけ少なくするような標本の取り方が重要だということです。それから、評価の立場でいうと、学習するデータと臨床の場で分類すべきデータの間の偏りをできるだけ少なくすることがポイントだと思うのです。

それから、解析手法を持つバイアスなのですが、開発過程でいろいろなパラメーターのチューニングをすると。そこで、一部はそのときに開発者が持っているバイアスみたいなものが入る可能性もあります。それから、実はそのことを全く気にしていなくても、無意識にデータの選択、手法の選択でバイアスがかかる可能性がある場合がありますので、この辺りが先ほどの Good Machine Learning Practiceなどで参照されている、こういう分野におけるバイアスという考え方を見たものにみられますので、そういう所を参考にしてやっていくことが必要だと思っています。

評価データの再利用ですが、性能の変化は実は学習すれば必ずよくなるか、それは保証されていないのです。学習すると悪くなることもあるということで、それに対する重要なことです。それから、市販前の承認時のテストデータ、又は市販後の新しいテストデータを用いる。これは、実はデータを独立に持っていて、同じデータで評価して与えればいいのかという議論もあるのですが、実は数学的に厳密に見ていくと、そこの持っているテ

ストデータの中の情報が学習過程に入り込むということになります。それが、実はある種のバイアスを生み出すのだということが数学的に、ある程度シミュレーションでも出てきています。ただ、これは余程意図的に、たとえば悪いのですが、試験の対策があつて山を張って、そここのところだけ一生懸命学習するような、わざとそういうことをしたときに出てくるわけで、その点を注意していけば逃れられる可能性があるということを示しています。

そこは 4-3 で、リスクのある再学習法とその対処法です。それから、ではこれはリアルワールドでどういうことが起きているのかと。Kaggle というのは、画像の認識をする 1 つの大規模なコンペティションなのですが、このデータを使って学習してくださいというデータがあつて、最後にそれを評価するデータ。そのときのスコアが、自分たちが学習したときのスコアとテストのスコアとどう違うかという論文があるのです。そういうものを見ていきます。具体的な例を報告書から出させていただきます。

これは、頭部 CT の画像から脳出血の有無を分類する、分類問題です。どちらかというデータ群があつて、データ群はある、なしで分けています。ですから、その区分線を多次元空間で切っていくという感じです。同一テストデータを繰り返し用いて、どういう特徴を使ったらいいかということをしていくと、この左側、右側の図で緑色のものとオレンジ色のものと赤のものがあります。緑色のものは、横軸は Round of adaptivity といって、thresh を組み立てて学習していったと考えていただければ結構です。

そうすると、全体の学習をしていったときに、新しいデータが入ってくると少し色が悪くなったりするのですが、全体としては維持をされていると。それから、黄色のデータは同一のテストを繰り返していくと、同一のテストのデータに対してはよくなったように見えるのです。ところが、実はリアルワールドにやると悪いデータが出てくるのです。学習してテストデータを繰り返し使ったりして、そこだけにアダプティビティを上げていってしまうと、こういう結果が出てくると。一方で、詳しいことは申し上げないのですが、Thresholdout というテストがあります。これは、学習データと学習のときに使う訓練の評価のときに、苦手なデータが出てきたときに、学習データと余りにもそのときの評価が近いようなものについては、これは使わないと。違うときに使うというようなことです。そのときに、そのデータの結果で出てきた評価結果をそのまま使うのではなくて、ちょっとノイズを入れて

使うのです。これはどういうことかということ、データを使って学習するときに、その中にデータの情報を数学的な繰り返しの中に、情報が学習の過程の中に入ってくると。そうすると、ある種の過学習ということ。そういうときに、それを使って行った評価結果に少しノイズを加えるのです。本当は 100 点を取ったのだけれども、95 点であったとか、そういうことをやるわけです。そういうことをしてあげると、その影響は減ってくるということで、完全になくなるわけではないのですが、そういうアルゴリズムを入れると右に書いたとおり、特定のデータを使って訓練をしたときに、リアルワールドの結果に近いような結果が出てくるのです。こういうことをすると、誤解なくできるのではないかということなのですが、これは余りまだ確立された方法ではないのですが、こういうことが議論されているということをご指摘させていただきました。

これは先ほどの Kaggle というもので、横軸はいわゆる Public accuracy ということで公開されたデータに対して一生懸命いいモデルを作るわけです。それに対して、テストということでプライベートな所でやったときにどうなるかということで、あるコンペティションでは非常に相関がある、あるコンペティションでは悪かったということがあるのです。これは悪いといっても、縦軸を見ていただくと、accuracy が 1 ポイントぐらいしか変わらないのです。ですから、そんなに違うわけではないのですが、やはりこのようにリアルワールドと訓練のときのデータでの違いが出るということがあり得るということです。

この論文に出ていたことで、次のデータの話をしてします。横軸にテストセットのサイズを使って、いわゆるパブリックなデータを使ったときの性能とテストをやったときの性能がどう違うかということをやったときに、データが多くなってくるとやはりそれは近くなってくるということで、データが大きいのがいいということは、これは真実だろうと。ただし、医療データの場合にこれをするのはすごく大変なのです。アノテーションスルーなどはすごく負荷の高い作業になります。ですから、データを多く取ることは重要なけれども、やむを得ず限られたデータを使うときには、指摘したようなリスクがあるということをご十分理解した上で、適切な方法を取るべきだろうと思います。

同一のテストデータに評価結果を意図的に利用すると。ですから、ある疾患の成績が悪かったから、このデータをたくさん食べ

させて評価をしたときには、これは気を付けたほうが良いということです。悪くなるということは必ずしも起きるわけではありません。気を付けるということです。それから、そういうことを無意識に利用してしまうようなことを気を付けましょうということです。リスクが低いのは、同一のテストデータへの開発者のアクセス制限をすると。ですから、評価するデータを開発者には見せないで、どんなものをやるかなどということは見せない状態にしておいて、やるということです。やはり、テストデータのサイズを大きくすることは重要だと思っています。

注意点は、このリスクの高低の程度は問題によって異なります。低い行為を同時にした場合に、どちらの影響が優位になるかは、まだよく分からない状況です。この報告書で例示をしていますが、全ては網羅しているわけではありません。ですので、こういう可能性があるということを認識することが、ここではポイントになるかと思っています。リスクが高い行為を 1 回でもすると、実はバイアスが含まれる可能性がある。ただ、そのバイアスが異常に重要な問題を起こすかどうかは、まだよく分からないです。ということで、リスク対策においては、こういう問題を開発側と審査側がしっかり理解して、いいコミュニケーションをしていくことが重要かと思っています。

それから、リスクを低減する工夫をしているのだということを科学的、客観的に説明することが重要だと思っています。それから、リスクにより生じたテストデータの過剰適合の検出や、検出後の対策が、やはり市販後としては重要なことかと思っています。

最近の議論として、論文で出てきたことですが、メタ解析で深層学習を使った医用画像解析の研究はどうやってされているのかということです。臨床試験としてどういうレベルか。結論は、臨床試験としては必ずしも高いものではないというのが論文としては出てきているということかと思っています。例えば、前向きの研究でもケースが少ないとか、独立性が担保されていないとか、いろいろな問題があって、結果として AI を使ったものが優れているということが科学論文では出ているのですが、注意をして見る必要があるだろうということです。ただし、薬に使われているような前向き研究や無作為化研究というのは、本当にこの技術体系の中で最適な方法なのかどうかということについては、まだ議論が必要だろうと思っています。

それから、これはスタンフォードのグループが出したデータな

のですが、画像データを研究ではパブリックに使われているデータを使って、例えば AI のモデルを作ります。実は、パブリックのデータにどういうデータ圧縮が使われているかということについては、余り書かれていないのです。実はデータ圧縮をすると、人間の目には分からないぐらいのところまでデータの欠損があるわけです。それが影響を及ぼして、こういう公共のデータベースを使ったものをリアルワールドのデータで見ると、その違いが如実に出るというようなケースが出てきます。ということで、やはりこれは先ほど言った、バイアスのときに機器や前処理をどうやったということをきちんと記述していくことが重要だということを示しているということで、こういうことが徐々に分かってきたというのが科学的な知見だということになります。

機械学習におけるシミュレーションの利用です。先ほど言ったように、データが少ないということで、数値シミュレーションを使ったモデル、これは *in sillico* 学習と呼ばれるモデルです。これを使って性能を上げるということが出てくるかと思えます。これは、あり得ることだと思います。例えば、実際の歪みを少し入れてみるとか、CT のデータに動きのデータを入れて、モーションブレンダーが入ったらどうなのかとか、そういうことはシミュレートできるわけです。そういうことなどは、物理的にある程度演繹的に分かるものについては、多分できるだろうと。これは、実は科学委員会の数値シミュレーションの報告書とかなり論理的には似ている形になります。したがって、こういうものを使うときには、しっかりバリデートされたものを使うことが重要だということで、6-3 に書いたとおり、実装過程と *verification and validation*、これは数値計算の報告書を見ていただければいいと思いますが、ASME の V&V が有名ですが、そういうものを適用しながらやっていけば、こういうものを使うことはできると。

ただし、これを評価データに使うということについては、まだコントラバーシャルだろうということです。ただ、これも数値シミュレーションの議論であったとおり、しっかりバリデートされていれば、こういうものを使って評価するというのも実はあるかと思うのです。例えば、放射線の計画のシステムなどは、そういうものでやられていることも一部はあるかと思えますので、そういうものの現状の方向性を見ていくことが必要かと思っています。その辺りが数値解析です。

データベースについては、いろいろな手術動画のデータベース、

病理のデータベース、それから心電図のデータベース、といったようなものができつつあります。いろいろやっていくと、きちんとこういうデータベースを作るときに、以下の 10 項目ぐらいのことを想定して作っていく必要があるのだろうということです。データベース構築は何のために作られているのか。それから、収集データの種類・規格、これは何のために使われているかというのは、10. の想定される利用方法と関連するのです。大体のデータベースは、例えば心電図などはレジストリーというものがありますが、レジストリーはこのために作られたものではないのです。しかし、そういうものを使っていくことは今後出てくる可能性がありますので、そのときにこういう点がどう適用可能かということを考える上で、種類、規格や規模、それから紐づけられる医療情報の種類は画像だけではなくて、その背景の電子カルテにあるデータも重要な形になってくるだろうと。

それから、どういう機関で取られたかということです。それから、医療機器、内視鏡、あるいは病理もそうかもしれませんが、手技に関する情報も重要です。デバイスの情報も重要です。それから、アノテーション手法・教師データの作成方法は、どう作られたか。そこを作っていくということをきちんと明らかにしましょうと言っています。それに従って、現状どう作られているかということの説明した形になっていて、データベースにおける課題は後ほど出てくる個人情報などの課題になるのですが、これは実はこの科学委員会の範疇を超えていますので、触れるだけにしていきます。

共通の課題は、やはりデータ容量が大きくて、それをサステナブルに経済的なことも含めて、どうやって管理していくのか。それから、アノテーションデータセット作成の煩雑さも課題になってくるかと思えます。ここが場合によると今後、これを AI でやってしまうということも出てくるかもしれないと思っていますが、そこについては触れておりません。それから、データベースに関連する様々な費用の捻出。

最後に 16 ページになりますが、では、いろいろとデータのことを言ってきましたが、データをどう使うのかということで、開発のときに使うデータ、それから評価のときに使うデータがあります。開発のときには、学習するデータと、それがうまくいったかどうかを見るテストデータがあります。そういうことを考えたときに、それをきちんと意識して使いたいということなのです。そ

れから、では独立性をやりましょうと言ったときに、同一の人は独立性がないからそれを排除するのですかという議論がありましたが、これはやはりないだろうということです。これを厳密に禁止してしまうと、大変なことが起きるということを指摘しています。

これはどういうことかという、例えば 1 人の患者さんが CT をずっと撮っていて再発を見付けると。1 人の患者さんの内視鏡の画像を経過観察をしながら見ていくときに、時間が変わればそれは独立のデータと受け取ってもいいだろうということで、独立性はどこにあるかということをしちんと理解してやっていくことが重要だということです。

学習用データとテストデータは、先ほどのバイアスの問題ということで、分布の同一性が重要だということと、独立性が必要だということです。これは全くコンタミネーションを起こしていない、違うものとしてきちんと管理されているということが重要です。ということで、この同一性があるかということに注意してデータセットを作っていくことが重要だということ、そこでは書かせていただいております。

ただ、最近になって分かってきたことで、深層モデルの仕様不足、under specification という問題が指摘されてきています。これは、まだ解決の方法としてはよく分からないのですが、テストデータでは性能の区別がつかない複数のモデル、これは Amazon や Google など非常に大規模なデータを持っていて、それで幾つかの複数の AI のモデルがあって、その性能は余り区別はつかないのだけれども、それをリアルワールドデータに放り込むと、実際には性能が違って出てくることがあります。これは、先ほどの学習データのバイアスの問題で言ったときの解決の手法で、そこに少しノイズを入れるということが 1 つは解決の方法ではないかということ言われているのですが、ちょっとまだここについてはよく分からないということで、こういう所についてはこういうことが出てきたということを指摘するにとどめさせていただきます。

まとめとしては、そこに書いたとおり、本報告ではデータについて想定すべき注意点を書きました。一方で、審査側と少し議論したところで、実はこのデータについてレギュラトリー的には、こういうデータの信頼性があるということを調査できるように、例えばこれを承認データに使おうとした場合には、そういう資料が提出されることがあるということをしちんと理解しておいてく

ださいということです。それを書きました。

それから、データの匿名化がきちんと適切に行われていること。これは、いろいろな臨床情報をどう使うかということの法的な問題については、日進月歩で変わってきていますので、そういうこともきちんと見ながらやってくださいということを、あえてそこについては注意喚起をすることにしました。

最後の文章ですが、これは非常に重要な技術で出てくるということになりますが、やはり今回、示した留意点を参考にしながら、作る側も冷静に考えていただきたいですし、評価する側もこういう視点で適切にきちんと開発されているものについては、迅速に承認していただけるような考え方の指針になっていけばと思っております。以上です。

○渡邊委員長

どうもありがとうございました。佐久間先生には、また部会の委員の先生方には、AI を活用したプログラム医療機器に関する専門部会報告書詳しく、また分かりやすくおまとめいただいたと思います。「科学委員会は、先端科学技術応用製品へのよりの確な対応を図ることを目的とする」とされていると思いますが、その目的に合致した報告書を作ってください感謝申し上げます。ただいまの御説明、御報告に対して、何かコメント、御意見等がありましたらお願いします。

まず私から質問させてください。データには 3 種類あり、学習データ、検証データ、テストデータということでした。テストデータが加わると、どんどん AI というのは変化していくと言うか、返ってくる答えは、その加わったデータによって変わってくると考えていいのでしょうか。

○佐久間委員

それは学習をするということですか。

○渡邊委員長

そうです。

○佐久間委員

学習して、1 回固定したものにしていれば、その精度は変わってくるわけですがけれども、ここで言っているのは、そういう実際のテストデータと言うか、それを使って学習をしてよりよくするというプロセスを入れたときには、そのテストデータが内々に持っている情報を、そのモデルが取り込んでしまったということで、そこがあるということです。なので、テストデータの結果を、例えばうまくいったか悪かったかということで、細かいところまで示さなければ、それは多分中に入っていないと思います。中岡先生、何かありますか。

○中岡委員

補足的になりますが、結局、素人考え的に我々も考えていたの

ですけれども、同じテストデータを新しく、テストデータを絶対に学習させないで置いておくと。で、その上で新しいデータをかき集めてきて、出来上がったモデルに対して学習を重ねて、その性能をテストデータにおいて、このテストデータはちゃんと性能が上がっているなというように判断すればいいと思っていたのですが、数学的にはそれを繰り返していくと、テストデータを学習することがあり得るそうです。なので、言ってみればカンニングしてしまうみたいな話です。

○佐久間委員　　うまくいくかどうかという情報の中に、その中の情報が入り込むということではないかと。これは非常に複雑な問題なのですけれども、そのときに先ほど言っていたのは、その戻すときに、うまくいったよということ、ちょっとノイズを入れて戻すようなことをすると、実はそのデータのコンタミネーションが抑えられる。

○中岡委員　　数学的に、同じデータを学習しないことになる。数学的に同じデータでテストをしてしまうと、そういうことが起こり得るということで、やはりそうするとテストデータにある程度のバイアスが、テストデータ寄りのバイアスを持った性能になってしまって、そのテストデータに特化した SaMD、AI になりかねない。なので、いろいろそこは議論が分かれているところではあります。そこは業界さんもすごい興味があって、業界さんとしては、当然同じデータを使い回したいわけです。集めていますし、説明が分かりやすいですから。で、そういうようなバイアスが生まれても、そんなに問題にならないだろうということがあるのですけれども、ただ、やはりそこにそういう問題が科学的に潜んでいる以上、使い回しに関しては、何らかの基準と言うか、そういうものは設けないといけないだろうというのが、AI の世界の中でも言われています。

○佐久間委員　　それを完全に禁止するということは、現実には難しいと思います。そういうリスクがあるということ踏まえた上で、先ほどのデータの多様性だとか、分布の同一性だとか、そういうことをやっていて、やむを得ずこういうことをやった結果として、こういうことが起きています、こういう性能が出ていますという説明になるのだと思いますけれども、禁止するということは現実には難しいです。

　　こういうリスクは、それほど大きくはないとは思いますがすけれども、ゼロではないというところがあって、実はそこに示し

たように、これは意図的に、うまくそこに適応するような学習をしてやろうというようなことをやってしまうと、こういうことが起きる可能性があるということで、問題にならないような学習方法を何とか工夫をしてやっていますというようなことになっていただけるといいかなと思います。ここについては、まだまだ研究も進んでくる場所ですので、これが最後の under specification の問題で、使用規則の問題とちょっと絡んでいると思うのです。

いくら大きなデータをやったとしても、非線形の系なので、ちょうど数値計算における曖昧さの定量みたいなところでも問題がありましたよね。そういうものとちょっと近いかなというところでした、実はちょっとデータがずれると、大きく結果が変わるということが出てくる可能性があるのです。そこが数値計算のときは曖昧さという形で議論していたわけですが、このときも最後の議論で、データがいっぱいあって、そのデータの近傍で少し違うものがあつたときに、大きく変化するものが出てくると、ちょっとそこを揺らしてあげて、そこがないようにしてあげるということをするのがいいのではないかという議論もあるということらしいのです。その辺り、まだまだ議論になってくると思います。

○渡邊委員長

ありがとうございました。Web で参加されている先生方も、何か御質問はおありでしょうか。

○中村委員

富山大学の中村です。佐久間先生、まとめてくださって、本当に大変な作業をありがとうございます。非常によくまとめてありまして、私もすごく勉強になりましたし、すばらしいまとめだなと拝聴させていただきました。

私、今年度になってから人間ドックなど、いろいろな健診の仕事もちょっとやるようになりまして、レントゲン写真などをたくさん診る機会がありました。そういうときに、AI のようなものでいろいろと見落としがないように、疑わしいところがパパッと出てきてくれると、すごく有り難いなというように思っていて、こういう技術が臨床に早く進んでほしいと、すごく期待しております。

今回、思ったのですが、例えばデータベース、健診のデータベースでは、正常な患者さんが非常にたくさんありまして、その中から異常を見付けるというのが、健診における AI の活用方法という形になると思います。一方、例えば肺がんを専門に診

療しているような先生方が期待する AI になってきますと、腺がんと小細胞がんの違いとか、そういったようなところが関心になってくると思うのです。

ですから、AI を使うときの目的、ターゲットを考えた上でのデータベースが、また必要になってくると思います。健診の場合だと異常陰影だけでいいのですが、肺がんの種類判別のよう話になってくると、病理までのデータが必要になってきますので、データベースの深さと言うか、目的に応じたデータベースがすごく大事になるのではないのかなということを思いました。その辺りは何か考えていらっしゃるのかなと思ひまして、発言させていただきました。

○佐久間委員

先生がおっしゃったところはすごく重要なポイントで、先ほど言っていた、いわゆる手法のバイアスと言いますか、応用先のバイアスみたいなもの、バイアスというのは悪い意味で使われますが、実は、使う分野が統計的に違うわけですね。そういうことを意識しましょうということなのだろうと思います。

なので、健診の場合は非常に正常な人が多い、大量に出るデータの中のものを見付けるということになると、悪いものがあったときに、その中の更にそれを鑑別していくという問題は全く違う問題ですので、それによってデータベースは変わってくると思います。それがあがるゆえに、ここに書いてあるデータベースの所に、このデータに入っているものはどういうものなのかということの来歴をしっかりとすることが重要で、そうすると、その中からサブサンプリングして作るということも、実はあり得るわけです。そのサブサンプリングが適切なかどうかということは、ちゃんと説明しなければいけないということなので、そういうことになるかと思ひます。

それで、先ほど言った病理との関係とか、血液データとの関係なども見たくなるわけですね。実は、多くのデータベースは、そのところがなかなかうまく紐付けられる構造にはなっていないのが現状ではないかと思ひます。ただ、そういうことがあるということは今後考えていくということが必要で、更に社会としても、せっかくここまで出しているデータですから、そことの関連性を、どうやって法的にも含めて、つないでいくのかというのは、すごく大きな課題かなと思ひますが、まだそこについては十分に体制ができていない状況かと思ひます。ゆえに、ここはこういう書き方をして、この点

に気を付けましょうよという書き方にしているところです。

- 中村委員 健診の作業をしていると、診断を支援してくれる AI のニーズはすごく大きいということを実感しますので、改めてここで申し上げておきたいと思います。ありがとうございます。
- 佐久間委員 それをちゃんとしたものにするために、こういったことを考えましょうということだと思います。
- 渡邊委員長 ありがとうございます。ほかにはいかがでしょうか。
- 堀委員 先ほどのテストデータの考え方ですけれども、データベースによっては、例えば画像関連のデータベースだったら、テストデータもデータベースに全部含めて情報を大きくしていったほうが良いのではないかと思ったのですが、全部普遍的に考えてしまうということでしょうか。
- 佐久間委員 普遍的にというのは。
- 堀委員 テストデータを最後に入れるのは危険があるというような考えだったのですが、例えば画像のデータだったら、テストデータも区別せずにデータベースに加えていくほうが、医学的には妥当なのかなという気がしたのですが。
- 佐久間委員 テストデータというのは、データベースの中にデータがあって、そこをどういう用途で使うか。学習に使うのか、学習に使ったものを、開発の段階でうまく出るかですけれども、そこはある程度閉じた話ですけれども、それに今度は評価をする、承認をするということがあると、ここでやったものではなくて、ほかのものでいいのではないか、それは取らなければいけないということが理想なのですけれども、その理想はできないような状況もあって、そういうところのデータをあらかじめ分けていって、開発者にそのデータを使わせないようにしながらやるといったようなことをするということは、実はあり得るだろうと思います。ただ、そのときに、評価のために申請されたとか、そういうところのデータを使って学習はしていないということを、ある程度準備しておくということは、最低限必要であるということだと思います。それを最低限やったとしても、まだリスクはあるのだけれども、それをやるだけでもかなりよくなるということはあると思います。
- 渡邊委員長 ほかにいかがですか。
- 久米委員 このスライドの一番下に出ているのですが、性能というのは具体的にどういうことなのでしょう。
- 佐久間委員 例えば性能と言うと、ものを名前付けして分類するとか、そ

ういうものがあつたときに、それが正しいかどうかとか、そういうことです。だから、いわゆる機械学習したモデルで、入力が入ってきたときに、正しい答えを出すかどうか、正しくなかつたのか。そういうことです。

○久米委員

医療側から見た場合、先ほどのお話もあつたのですけれども、要するに、たくさんの正常の中のサンプルの中から異常を見付け出すことが、普通私たちがしていることです。そういったことにまで差が出てくるというような事例があるのでしょうか。

○中岡委員

性能の場合は評価型が幾つかあって、ちゃんと正解を、いわゆる悪性腫瘍を悪性と判断できるかどうか、いわゆる正解をちゃんと見いだせるかどうかという観点でのテストと、これは異常ではなくて正常である、いわゆる特異度と言いますが、それがちゃんと検出できるかどうかというところ、その両方のバランスでちゃんと成績を見られているのが現状だと思うのです。その上で、テストデータを使った場合のフォールスポジティブ、フォールスネガティブがどれぐらい出ているかというところまでを含めて、その 4 つの数値を基本的に見て、全体的な性能というものを規定されていると理解しています。

そうすると目的に応じて、例えば先ほど言われたスクリーニング的な、健診のスクリーニング的なもの、あるいはどちらを重視するか、何を重視するかということによって、SaMD の開発方法が決まってきます。ですから、健診でしたら、基本的にはフォールスネガティブという考え方ですよね。そうすると、フォールスネガティブをなるべく最小にするような形で、性能を特化させた製品が生まれてくるということになると思います。そういう形で性能というのを判別するので、ちなみに、先ほど言われた特定の疾患だけを必ず検出できるというわけではないということです。

○久米委員

実はそういったことを思ったのは、先ほど例えば病理とか、ほかの次元のデータで組み合わせたらどうなるのだという話があつたときに、例えば画像を見るためにこういったソフトウェアが、画像の異常な処理があるかないかということだけを見て性能を出しているのか、あるいはほかの、例えば血液のデータとか、そういったことを含めて学習させているのかというのが分からなかつたものですから。

○佐久間委員

多くはそこまでないでしょうね。

- 中岡委員 今は多分、画像の特徴をつかんでやっているという、ディープラーニングの最大のもう 1 つの特徴は、画像における特徴量を抽出して判断できるということなので、そこの中に、血液学的・生理学的なデータとか、そういうものを組み合わせて判断している製品があるかと言うと、まだないと思います。
- 久米委員 分かりました。
- 渡邊委員長 ありがとうございました。Web で御参加されている先生方にも手を挙げていただいております。古矢先生、まず御質問をお願いします。
- 古矢委員 佐久間先生、非常に難しい、でも、とても重要なテーマについて、大変分かりやすく、そして、どういう方向で考えたらいいかということ整理いただきまして、私もよく理解できました。ありがとうございました。
- 今表示されている、3 種類のデータの性質とか使い方が大事だということはよく分かったのですが、一方では、今回の検討チームがお考えになっているのがマシンラーニングが中心ということなのですが、世の中は生成系 AI を使おうという動きがあるかと思います。そうすると、生成系 AI を使って作ったデータ、あるいはデータセットというものが、この 3 種類のデータの中に入ってくるということについては、どのようにお考えになっておられるでしょうか。教えていただければと思います。
- 佐久間委員 そこについては、この検討においては後半に出てきた現象で、十分にまだ考察ができていません。ただ、これも何をもって性能評価するかとか、そのときに使う学習というのがどういことなのか、言語の問題まで入ってきているのですけれども、そこは余りここで議論できるような話ではないと思うのです。その辺りは、まだ今後の課題かなというように思っています。中岡先生、何かございますか。
- 中岡委員 画像系に関しては、比較的以前から、GAN というシステムを使って、そういう模倣的な病理画像を作って学習させたらどうなるかという研究はされています。そういうものを使うことが、果たして適切かどうかということに関しては、今回の報告書の中で、6 章の数値シミュレーションの辺りを、データとして使うことの適正さということはどう考えるかということ、鎮西先生が中心になって検討していただいています。ですから、今でもそういうような生成系を使って学習させようという研究がないわけではないのですが、ただ、現実的に、ものになるかどうか

かということ、言わば科学的な妥当性をもって、それが本当に性能は出ているか、リアルワールドで出るかということに関しては、まだ疑問符が付くのではないかなと思っています。

○古矢委員

非常によく分かりました。私としては、患者のデータを取るといのは手間もお金も掛かりますので、つつい生成系 AI を使ってデータセットを用意するという流れが出てくるのではないかと思ひまして、先生方がおっしゃりたいことはよく分かりました。ありがとうございました。

○渡邊委員長

それでは中江先生、お願いいたします。

○中江委員

大変膨大な報告書を頂きまして、ありがとうございます。この分野は私は専門外ですので、非常にナイーブな質問をして恐縮なのですけれども、私は病理学者ですので、皆さん、画像データ、病理データの話は何人か御質問なさっていましたけれども、病理のほうで、いわゆる機械学習を使った診断システムの構築というのは、かなり前から専門の病院の中でやられていて、恐らくおっしゃっているデータベースの中身は、そういうようなものも加味しておっしゃっているのだと思うのです。

その話をずっと聞いておりますと、診断そのものが難しいということがあるのが 1 つと、もう 1 つはデータベース間と言うか、病院の診断などの統一性というものが、なかなか難しいということがあって、データベースの構築そのものに問題があったり、若しくは同じものを違う表現で別のデータベースでは言っていたり、もっと極端な場合は、1 つのデータベースを構築するとき、データを入れる方によって、同じ病変を違う言葉で言ったり、違う病変を似たような言葉で言ったりして、それがマシンラーニングに非常に支障を来すのだというようなお話を伺ったり聞いたりしたことがあります。そういう統一性と言うか、妥当性と言うのか、そういうことについては、部会としてはどのようにお考えでございましょうか。

○佐久間委員

そこは、今そこに出ている一種のアノテーション手法というところで、そこに関係すると言うように理解しています。正に、先生がおっしゃったような問題があるということで、私もこの議論の中で佐々木先生に教えていただいたのですけれども、ある病理学をある年代でアノテーションして、こうだといったものが、数年経つとその名前が変わってくるということはある。それをちゃんとアップデートしていくようなものを作っていかなければいけないのだけれども、なかなかそこは難しいです。

正に、このような課題があるということを使った上でデータを使うことが必要なのだろうということに対する指摘になっているのだろうと思うのですが、では、それをどう解決するかということ言えば、それはちゃんと整理をしてやるということになるので、これは先ほど言ったデータベースに関連する様々な費用の捻出というところに絡んでくるということで、そこは科学委員会の範囲を越えるところではあるのですが、科学的にはそういうことをやらなければいけないので、どうやってそういうコストを皆さんがシェアするかということ、社会的に考える必要があるのではないかということだと思います。

○中江委員

もう 1 つあります。これは非常にナイーブな質問で恐縮です。最初のほうから出ていることですが、ラーニングを進める上において、あえてノイズを入れることでシェイプアップしていくというお話がありましたが、本当に素人の質問で恐縮なのですが、AI の学習能力を上げるために人がノイズを与えるということは、それ自体がバイアスなのではないかと思うのですけれども、そこはよろしいのでしょうか。つまり、ノイズを入れなければバイアスがかかってくるという御説明をされて、そのデータを見せていただいているのですけれども、それは機械学習の本来の姿なのであって、そこにノイズを入れるということは、それ自体がバイアスを掛けていることにはならないのでしょうか。そこが私は分からないので教えていただければと思います。

○佐久間委員

正確にはお答えできませんけれども、ノイズそのものは、統計的にはかなり相関に近いものを入れているということになるので、そこについては低いだらうと思います。

私の理解は、多次元空間の中にデータがあって、そのデータのデータを例えば区別するということは、それを分ける超平面を作っていく形になると思うのですけれども、超平面の構造が、データに対して非常に敏感に関わりやすいようなことになっていることが、学習の結果として入ってくるとすると、データの僅かな結果が大きな結果に、いわゆるバタフライエフェクトみたいなことが起きる可能性があるとするれば、そのところにちょっとノイズを入れてやることによって、その学習をすると、その全体的な超空間の中のデータの点数の分布に対しては比較的滑らかなものが出てくるということなのだろうと思っていて、その辺りのことだらうというように思っていますので、ここは

完全に私も理解できているわけではないのですけれども、これは過学習と言いますか、あるデータセット群にうまく合うように学習するわけです。その結果として出てきたものが、そういうデータの局所的、あるいは体系的な変化に対して、安定に動くかどうかということに対して議論していくところなのだろうと思います。そのために、ノイズのようなことでということで、要は先ほどの資料不足の問題のところもそうですけれども、いっぱいデータがあるのだから、全部それでいけるのではないかと。ところが、それをうまく分けている超平面みたいな所の構造が、1つのデータが変わると変わってしまったということですよ。そういうものがないようにしていくということで、これを入れているのではないかなというように自分は理解をしています。

○中江委員

ありがとうございました。

○渡邊委員長

私の理解を大分超えてきたのですけれども、テストデータをいつも同じものを使うのではなくて、多様性のあるテストデータを用意しておくということですか。

○佐久間委員

それは毎回新しいデータが取れるのだったら、それでやるのが一番いいのだろうと思うのです。でも、それは現実的に臨床試験をやるということにおいて難しいわけで、そういう相矛盾する中で、どうやって気を付けてやっていくかということが求められていることなのだろうと思います。

○渡邊委員長

中村先生、お願いいたします。

○中村委員

先ほど中江先生から病理のお話くださったデータベースの話についてなのですが、データベースを作る作業という意識をもって、データベース用の言語と言ったらおかしいのですけれども、こういう場合はこういう表現、こういう場合はこういう表現というような形にして、データベースを作るという意識で統一していくということが必要なのではないかなと思いました。

○中江委員

中村先生のお言葉について、1つよろしいでしょうか。

○渡邊委員長

どうぞ。

○中江委員

今の統一言語の話ですけれども、承認申請のことについて FDA が最初に始めた、いわゆるデジタルデータでの提出の話があって、今は PMDA も申請に用いているはずですが、ああいう所で使われているような統一言語は使えないのでしょうか。あるいはそれを使うという取組はあるのでしょうか。

○佐久間委員

私はそこは分からないのですが、PMDA から何かありますか。

○中村委員

今後、データベースを作るという意味で、そういう取り決め

をしていったらいいのではないかと思います。

○佐久間委員

多分そうだと思います。そういう意識をもって作っていくということが重要なのですけれども、現実にはいろいろな流義があるということになっていて、1つあり得ることは、これが正しいかどうかはあれですけれども、実際には研究的には、そういうものを分析して、あるところに標準化に展開すると言うか、転換すると言うか、自動化するという研究もないわけではないです。

そういうことも見ながら、それから私の頭の中にあるのは、これからはそうやっていくのが望ましいのですけれども、過去の膨大な蓄積があるわけです。それをどう使うのかということに関しては、よく考えていかなければいけないところがあって、実は、それはすごく財産だと思うのです。そこら辺をしっかりとやっていくことは必要だなと思います。

○中村委員

先ほど言い忘れたことがありましたので、発言させていただきます。私、大学では生体工学という授業をやっておりまして、その中で調べた情報なのですが、人口当たり普及している CT スキャンとか、MRI とか、恐らくレントゲンの撮影装置などもそうだと思うのですが、日本は世界の中で 1 位か 2 位というぐらいに普及している国なのです。

ということは、アメリカとかヨーロッパと比較しても、日本のデータベースというのは滅茶苦茶大きいと思うのです。これは、ある意味で AI を今後開発していく上においては、日本のすごく大きな強みだと思うので、この辺りを是非活用できるようになったらいいなと感じました。コメントですけれども、佐久間先生に投げてしまって申し訳ないのですが、よろしく願いいたします。これは将来に関わることになるかもしれませんが、よろしく願いします。

○渡邊委員長

ありがとうございました。ほかにいかがですか。青井先生、どうぞよろしく願いします。

○青井委員

神戸大学の青井でございます。佐久間先生、本当にありがとうございました。私にとっては非常に難解な分野ではあるのですけれども、全体としては、いかにしっかりとデータを集める仕組みをするか、それから、間違ってしまうかという文脈であったように理解いたしました。

私が思いましたのは、間違ってしまうと結果的には同じなのかもしれませんが、悪意に対する頑健性ということをお

いました。これは何年か前によく話題になりました画像診断システムを欺くような、患者あるいはその症例を持っている医師が、がんと診断させるには、画像のここをいじれば、がんと診断してくれるというのがこういうのだというのは、米国などでは保険と相俟ってよくやられていると。

一方で、例えば内視鏡画像で、胃がんの患者が減ってきたから増やしたいと思ったら、学習データや、その時点で何らかのことをすれば、目立たない程度に、しかし 1 割ぐらいの売上げが増えるような、そういう悪意が不可能ではないような気もしますし、可能性としてはあって、機械が学習する人工知能の部分というのは、透明化、見える化の人工知能ということも言われていますけれども、なかなか直感的には少なくとも一般の人間には理解し難い部分で、そういうことがありますと、そういういろいろなプロセスの中で、悪意が働かないようにする、それに対する脆弱の反対の頑健であることをどのように担保していくのかとか、その辺りのことは、どのように考えていけばいいのでしょうか。

○佐久間委員

直接の答えになるかどうか分からないのですがけれども、先ほど Good Machine Learning Practice というものを言及させていただきましたけれども、その中の 1 項目にトランスペアレンシーということが書かれています。これは、そういうことをやることに対する透明性を上げておくということです。それが実際に書かれていて、そこが今のことに関連するのだと思います。

それと、最後のまとめの所で書いた、これが評価の段階で信頼性調査を含む審査、プロセスによって、必要になる資料を提出できるように、これはどういうプロセスでデータを作ったとか、どういうプロセスのデータを作ったかということ、そのリスクによると思いますけれども、どこまでやるか。リスクによって、このようなことをやっているわけにはいかないと思いますから、そういうことが出てきて、先ほど言った、いわゆるマシンラーニングするときの透明性の担保ということと、こういうことをもって、今、先生がおっしゃったようなことに対して対抗していくということになるのではないかと思います。

○青井委員

ありがとうございました。

○渡邊委員長

6 枚目のスライドを見せていただけますか。この下から 2 番目の所に、「市販後の性能変化によるリスク・アンド・ベネフィットの評価を企業に委ねることは難しく、その都度、規制当局

が審査することが現時点では現実的」とあります。

これまで PMDA の承認というのは、決まった性能と効果について、承認・評価を与えてきたと思いますが、今後は市販後であっても、その性能はアップグレードを目的とするとは思いますが、今の青井先生の御質問のように、場合によってはダウングレードするかもしれない。そういうようなものに対して、PMDA は対応していかなければいけないということになるのでしょうか。それとも、承認した場合には性能変化というのはない、性能は固定するという事で承認するのでしょうか。

○佐久間委員

いや、そうではないと思います。基本的には IDATEN とか、そういうことから言うと、開発をして、ある段階で想定する性能改善と言うか、そのプログラム、プロセスと言うか、それをちゃんと規定して、それが適正かどうかということを考えます。改善して、それがよくなるということの評価していくことを含めて、それがあれば。今考えているのは、日常的に製品が性能を変えるということを考えているわけです。いろいろなデータが集まってきて、ある時点で、例えば A という疾患でできたことが B まで拡大できますよと。それは自然に現場で働いている部分がずっとなるのではなくて、データを集めてきて、それを製造業者がある程度学習し直して、そういうことで出てきたものに対していくと。

それをある程度妥当な範囲であれば、最初の注入の段階で、そういうプロセスを考えていけばいいということなのですが、なかなかそれが難しいというのが分かってきたところなので、現段階は一定のチェックを、承認が入ったところで新しいものが出てきたときに、軽重はありますけれども、がちんこでやるというわけではないと思うのですが、そこは、その改善プロセスを科学的妥当性と、そこではカバーできないような未知のリスクというものを考えた上で、その評価をしていく。場合によると、それをスキップするということも実はあるのだと思いますけれども。そういうことを決めて、見付けたから OK になってというのではなくて、そこを医療用途にするところから従って、そのことを丁寧に考えていく必要があるのだろうということだと思います。

○中岡委員

補足です。これは、あくまでも Pre-Cert からの学びです。アメリカがやろうとした Pre-Cert のパイロットプログラムというのは、言ってみれば Pre-Cert に合格した企業は、もう審査せず

に出していいと。性能を上げたものとか、そういうものを出していいということを試してみようとしたのです。ところが、実際にやってみると、そのためにはそういうような能力があることを事前に FDA が企業を審査しなければいけないわけです。そういうことをやっていると、実は結構いろいろな問題が出てきて、法律的な問題とかをここでやってしまうとややこしいのですが、現実的な結論としては結局、そういうような審査をするより、そういう企業の内部審査をするよりは、逐一、従来どおり承認審査をしたほうが FDA も楽だ。そちらのほうが現実的であるというような形の結論が書かれていました。ですので、PMDA がどうこうしようと言っているわけではなくて、まず、FDA がそういうことにトライしたのだけれども、FDA もそういうようなことが難しかったので、従前どおりの審査のほうが現状では現実的であるという判断をされたということです。

となると、日本でもなかなか難しい。ただ、今は IDATEN という制度があるので、いわゆる段階的に企業が性能の向上を進めていく上での学びも取り入れた上で、何らかのそういうようなガイダンスを作るという方向に持っていくほうが、日本としては今のところは現実的だろうと。その後、技術が進んでいけば、また更にステップアップすると。先ほど佐久間先生が言われたような形のことも可能になっていくのですが、現状では、そこまで責任を企業に負わすことというのは、それなりに患者にとっても PMDA にとってもリスクがある。規制当局にとってもリスクがあるというところになるかと思えます。

○渡邊委員長 詳しく教えていただいてありがとうございました。それでは岡崎様、お願いいたします。時間の関係で、岡崎様の御質問を最後にしたいと思います。

○岡崎プログラム医療機器審査室長 PMDA のプログラム医療機器審査室の岡崎でございます。質問ではなく、今議論があった点の補足ということになります。これまで、AI を活用して開発された医療機器プログラムについては、既に 20 数品目承認されている現状がございます。やはり画像診断系が多うございまして、今の話題に関連することとしましては、■■■というのが現状です。

ですので、今議論があった IDATEN を活用するというのも、1 つのやり方としては、PMDA 側もあるというように思っていますし、従来どおり、性能が上がったことを確認したことをもって一変申請をしていただいて、承認をしていくという道も■■■

ございますので■■■紹介させていただきました。ありがとうございました。

○渡邊委員長 どうもありがとうございました。それでは、AI を活用したプログラム医療機器に関する専門部会報告書の概要については、以上のとおり非常に丁寧に取りまとめていただきました。科学委員会としては、この形です承ということですのでよろしいでしょうか。Web で御参加の先生は、よろしければ挙手ボタンをクリックしていただければと思います。

(了承)

○渡邊委員長 ありがとうございます。全員の先生にご了承いただきました。科学委員会終了後、報告書の公表までの間に、てにをは等の体裁の修正がある場合は、私が委員長として内容に変更のないことを確認させていただきますので、御一任いただければ有り難いと思います。よろしくをお願いします。

(了承)

○渡邊委員長 どうもありがとうございました。改めて佐久間先生をはじめ、専門部会の先生方に御礼申し上げます。先生方の了承を頂きましたので、この報告書を科学委員会として、PMDA の理事長宛てに提出させていただきたいと思います。PMDA におかれましては、今後の業務に活用していただければ幸いです。どうもありがとうございました。

< 専門部会活動報告について >

○渡邊委員長 それでは、次の議題に進めてまいります。続いては、標的特異性を有する in vivo 遺伝子治療用製品のベクターに関する評価の考え方専門部会についてですが、この部会は今年の 6 月に第 1 回専門部会が開催されました。それでは部会長の久米先生から、活動報告の御説明をお願いしたいと思います。

○久米委員 標的特異性を有する in vivo 遺伝子治療用製品のベクターに関する評価の考え方、この部会長を仰せつかりました久米でございます。ただいまの AI に比べてかなり格調高いものにまだなっていないくて、非常にプリミティブな段階なのですけれども、進捗を御報告いたします。

この部会は、このような先生方を委員として発足いたしました。それで、6 月に第 1 回の部会を行ったところです。

この専門部会設置の背景と目的は、これまでも申し上げているところなので、繰り返しにはなるかと思いますがけれども、も

う一度御説明いたします。

従来、特に遺伝子再生医療等製品の中でも遺伝子治療製品、あるいは細胞加工製品の業界の中で、ex vivo の遺伝子治療製品と言われているものが幾つか上市されてきている、あるいは開発が進んでいるわけなのですけれども、その中で CAR T、キメラ抗原受容体を導入した T 細胞は幾つもの製品が上市されてきて、これまでとは比較にならない大きな成果を上げているところです。基本的には今の製品全て自己由来製品であるということもあって、一旦、患者さんから細胞を頂いて、製造所に持って行って、拡大培養して遺伝子導入してということ、非常に時間とリソースを費やすということになっています。

そのため、非常に限られた施設に運んで行って、あるいはできる患者さんの対象というのは、まだ世界的に見れば限られているということ、何とか打破しなければいけないということで、1つにはドナーを他家というか allo にして、ユニバーサルな CAR T 製品を作ろうという動きもあるのですけれども、これもまだなかなかうまくいっていないということもあって、その中で今まで ex vivo で行われてきた遺伝子導入の操作を、直接 in vivo でやろうというようなことが試みられています。

もともと ex vivo でこういったことをしなければならなかった理由としては、何といても標的特異性、あるいは生体に直接入れてしまった中ではベクターが薄まってしまったりとか、そういったいろいろな障壁があったので ex vivo で行ってきたわけなのですけれども、これを in vivo で行うということになると、量の問題もあれば、やはり何といても、その標的特異性の制御が重要な問題になるということです。

in vivo で行うためには、遺伝子治療のベクター系としてはウイルスベクター、特に今、ex vivo で使われているレンチ、あるいは in vivo で使われているアデノ随伴などと、それから特にワクチンで脚光を浴びています脂質ナノ粒子あるいはポリマー（に包埋された mRNA などの核酸）、こういったものがあるわけで、もう非臨床の段階ではいずれのベクター系を用いても、CAR T が既に、実用化とはいいませんけれども、少なくともフィージビリティ、POC は取れているという段階になっております。

ということで、これからこのような状況の中で、こういったことが実際の臨床応用に向けてどんどん製品開発に進み、あるいは申請の場に出てくるであろうということが考えられますの

で、この場で現状、プラスの側面とマイナスの側面を明らかにした上で、次へのステップを整理することで開発及び審査に資するということが、開発側あるいは審査側にとって非常に有効なことではないのかということが背景と意図になります。

ということで、こういったリスクの考え方を整理することによって、より安全な開発、更に審査員に最新情報を提供するというので、本専門部会が立ち上がりました。

検討のイメージなのですけれども、開発中の標的特異性を有する *in vivo* 遺伝子治療用製品の技術を総体的に俯瞰すると。特に遺伝子導入を行うモダリティとしてのベクターといったもの、特に先ほども言いましたようにレンチ、それからリピッドナノパーティクル、ナノキャリア、そういったものを分けて、その特性を整理すると。

それから、今回の一番の目的は、従来、*ex vivo* で遺伝子治療が行われてきたことを *in vivo* でも行おうということで、特に CAR T が先行しているわけなのですけれども、これからはそれにとどまらず、例えばゲノム編集などの新しい目的でも行おうということ。既に行われているわけですから、そういったものを実際に例を拾って、それぞれについて、ヒトへの投与を許容する品質・非臨床の評価事項を、これまで *in vivo* での遺伝子治療が行われてきた検討に加えて、特に追加する留意点があるかということが一番の大きな主眼になるわけです。そういったことを中心に検討していくと。それから、従来の使用経験も踏まえて、臨床投与での安全性データ等に基づくリスクの考察をするということを中心に、検討していきたいと考えています。

そういったことを踏まえて、報告書の目次の素案を考えてみました。まずは、そのイントロダクションとして、先ほど言いましたような背景ですとか、本課題を今の時点で検討する意義みたいなもの。それから、報告書のスコープなのですけれども、言いましたように、特に *ex vivo* から *in vivo* への転換を図る上で、どのようなことを考えなければいけないのか。

あるいは、一昨日、既にワーキンググループ (WG) の第 1 回を行いました、そこで活発な議論があったのは、実は最初のタイトルは、「*in vivo* 標的指向性」みたいなタイトルだったと思うのですけれども、実際に今回、「特異性」という言葉に変えた 1 つの理由は、実際には特異性を与えるために、少なくとも 3 つの段階があるのではなかろうかと。すなわち、これまでウイ

ルスとか、あるいはナノパーティクルという、与えるものとして、まずは細胞への親和性という意味での、いわゆる指向性ですね。それから細胞に入ってから、その中で、どのような細胞のマシナリーを使うのかということでの特異性、更には、これはもう既にベクター、特にウイルスベクターでは、いろいろ試されてきたことですが、いろいろな細胞に入るのだけでも、その中で組織特異的に発現するような仕組みです。例えば組織特異的なプロモーターですとかエンハンサー、更には、これは DNA の転写のレベルの話なのですが、転写された産物としての RNA では、RNA 干渉を使った特異性の付与みたいな、そういった、少なくとも 3 段階にわたるような特異性の与え方があるのではないかと。これはそれぞれウイルスだったり、それからリピッドナノパーティクル、それぞれによってやり方が違ってくるのではないかとというようなことを踏まえたスコープもちゃんと整理して、読者に与えたほうがいいのではないかと考えております。

それから 2 番、3 番は、そのモダリティーごとに、そこを開発動向と言っていますが、大きな意味での開発動向というよりは、*in vivo* 遺伝子治療用製品としての性格を与えるための、今、言いました標的特異性にどのようなことを要するために、いろいろなことがされているかということの有識者の先生方から挙げていただきたいというようなことになっています。

それで次は、そういったことを受けまして、例えば CAR T、再生医療、更にゲノム編集といったような注目事例それぞれについて、多少一般的なところとの重複は出てくるかとは思いますが、それぞれについて考えなければいけないというようなことを留意点を挙げて、審査に役立てたいということです。

それから、ちょっと順番をどうするかについては、まだ考えているのですが、安全性に対する議論については、一般的に例えば非臨床で考えなければいけないことですか、あるいは臨床試験に入る前に治験の安全性をどうやって保つかというようなことを考えなければいけないと議論しております。

この辺はどこに含めるかということは、構成を考えなければいけないところなのですが、予想されるハザード、そういったもの。それを果たして非臨床でどこまで見られるのか、見られないのかというようなことを考えていきたいと思っております。最後に総括があって参考文献というような構成を、今のところ

は考えているということです。

こういったことを報告にまとめたいというように考えているのですが、どのように議論を進めるかということについては、まずは実際にどんなことになっているのかということを経験、あるいは専門家から御講演いただいてインプットする。実際には、専門部会の下に WG を設置して、WG での議論と、それから専門部会での意見といったものを交互に組み合わせながら、WG でたたき台を作って、それを専門部会に諮って、意見を頂いて、また WG で、というようなことを繰り返して、報告書にまとめていきたいということになります。最終的に、そういったものをまとめて、専門部会で報告書案を確認して、完成させたいということで進める予定です。

このようなスケジュールで大体進めたいと考えているのですけれども、先ほども申しあげましたように、6月に第1回の専門部会で WG の設置を認めていただき、一昨日に第1回目の WG で、専門部会委員の中から指名して執筆担当、あるいは、こういった先生の御講演を頂きたいというようなことを決めまして、本日ここまで検討しましたということで科学委員会、親委員会のほうに御報告を申し上げます。今後、8月、9月に WG と専門委員会を交互に行いつつ進めていきたいと考えているところです。

○渡邊委員長

ありがとうございました。「標的特異性を有する in vivo 遺伝子治療用製品のベクターに関する評価の考え方-in vivo CAR T の開発など」について、専門部会長の久米先生から、この専門部会設置の背景と目的、そして検討の方向性、また、想定される報告書の目次案等や検討課題について御説明いただきました。以上の御報告について、何か御質問があれば、よろしく申し上げます。Web 参加の先生方も何か御意見があれば、挙手ボタンでよろしく申し上げます。

○佐久間委員

私はこの分野は余り専門ではないのですけれども、最近ちょっと出た会で話題のところ、組織のリプログラミングという、何かその組織、いわゆる創傷の治癒のような過程なのですけれども、それをもう少し積極的にして、いわゆる山中先生のものが体外でやっているのに対して、完璧にそこではないのですが、それに近いことを組織の中でやるということを試みる会社、ベンチャーができていくようなことを聞くのですが、そういうものと実は関連してくるのかなと。

- 久米委員 そういった萌芽は確かに出てきていますね。その中で、例えばこの報告書の目次案の中では 4.2 だったり、あるいは 4.3 だったり、そういう所で取り上げたいなと考えております。
- 渡邊委員長 かなり広く網羅していただいていると思っております。ほかにはいかがでしょうか。中村先生、どうぞお願いします。
- 中村委員 実は私、再生医療の研究をずっとやってきておりますが、私はずっと体外で組織や臓器を作ろうと考えております。今ほとんどの再生医療は、体の中に細胞を入れて、あるいはスフェロイドという細胞の塊を入れて、そしてそれが育っていった臓器になるという、そういうストーリーでの研究がほとんどなのですが、例えば組織から臓器を作ろうと思ったときに、いろいろな分化因子、細胞が分化する、あるいは組織を形成するような因子が必ず必要だと私は考えております。しかし、体に埋めてから後はどうなるかというのは、我々は外からは何もできないわけなので。それで、体の外でできるだけ臓器に仕立て上げようというようなことで、体外で組織や臓器をもっともっと育てていくというような研究を今、研究テーマとしてしております。生体内に細胞を入れた後に、必要になってくる細胞増殖因子とか、分化因子とかを体の中に入れるということは、どこでどんな働きをするか全く分からないので、非常に危険です。そんな再生医療はできないぞと思って、できるだけ体の外で組織・臓器を再生・成長させてあげようという研究を今、取り組んでいます。
- そういう観点からすると、*ex vivo* の CAR T 細胞というのは、体外に取り出した T 細胞にいろいろな処置を行って免疫を作らせるという活動になるのですが、それを体の中でやるというのは、これはターゲティングというか、その指向性というのもまだ確立していない状況でやるというのは、非常に危険性が高いのではないかと危惧します。ゲノム編集、及び遺伝子導入などを行うときなども、ドバツとかなり強力な処置を行って、細胞に導入するというような状況なので、それを体の中でやるというのは、ほかの細胞への影響をここで危惧されているのはとても良いポイントだと思うのですが、そういった危険性が滅茶苦茶高まるのではないかなという気がしてならないのです。恐らく、この検討というのは、そういうリスクを踏まえた上で検討されると思っておりますが、とても重要な検討事項だと思いますので、そういった危険性を踏まえた上で、客観的な、そして

正しい判断ができるように、是非進めていっていただきたいなと思います。特に PMDA というのは、患者さんの安全を守らなければいけないという立場だと思うので、是非そういう意識の下でやっていただきたいなと思います。

○久米委員 どうもありがとうございます。重々承知いたしました。

○渡邊委員長 中村先生、どうもありがとうございました。ただいま、中村先生が御指摘なさったような内容や課題についても、この報告書の中で検討していただくということですね。ありがとうございます。ほかにはいかがでしょうか。特にこのほかにはないようですので、この専門部会の御提案については、これで終了させていただきたいと思います。本日は活発な御議論を頂き、どうもありがとうございました。本日予定した議題は以上になりますが、事務局から何かございますか。

<その他>

○事務局（緒方科学委員会・先端科学統括課長） 事務局のほうから1つ、アナウンスをさせていただきたいと思います。まず本日、御了解いただきました、AI を活用したプログラム医療機器に関する報告書につきましては、PMDA の業務の参考にさせていただきほか、PMDA の Web サイトでの公表を予定しております。

また、エクソソームを含む細胞外小胞を利用した治療用製剤に関する専門部会報告書につきましては、既に日本語版が、2023 年の 1 月に PMDA の Web サイトで公表されております。英語版につきましては、論文化を検討いただいているため、2024 年、来年の 1 月までに Web サイトにて公表する予定でございます。以上でございます。

○渡邊委員長 これは論文化というのは、どこかの雑誌に投稿ということになるのですか。

○事務局（緒方科学委員会・先端科学統括課長） はい、その前提で先生方に御協力いただいております。

○渡邊委員長 どうもありがとうございます。次回以降の科学委員会につきまして、もし議題がない場合には、早めに事務局より先生方に御連絡して、その回はスキップするという方針とさせていただきたく存じます。それでは、本日の科学委員会は、これで終了とさせていただきますので、皆さん、どうもありがとうございました。Web 参加の先生方もどうもありがとうございました。