

de communication présentent cependant des similitudes avec les problèmes de communication généraux rencontrés en oncologie et en médecine de précision (voir également chapitre 16). Lorsque les professionnels de la santé annoncent de mauvaises nouvelles, ils doivent donner de l'espoir et s'assurer qu'une décision appropriée commune puisse être prise pour ou contre les traitements correspondants.

Pour tirer pleinement profit de la médecine personnalisée, on peut recourir aux «4 P» comme principes directeurs, auxquels s'ajoute un cinquième P spécifique à l'oncologie:

- **Prédictif**: détermination du risque de maladie, du pronostic ou de la réponse à un certain traitement;
- **Personnalisé**: adaptation des thérapies aux risques individuels ou à l'état de santé du patient individuel ou de populations de patients spécifiques;
- **Préventif**: orientation vers la santé/le bien-être; prévention/atténuation des conséquences de la maladie;
- **Participatif**: implication des patients, des fournisseurs de prestations ainsi que des financeurs et des décideurs;
- **Psycho-cognitif**: prise en compte des aspects psychologiques, cognitifs, émotionnels et sociaux chez les patients et leurs proches.

Chapitre 11

Sciences des données dans la médecine personnalisée

La médecine personnalisée cherche à augmenter la probabilité du succès thérapeutique individuel, de réduire le risque d'effets secondaires non souhaitables et de permettre des diagnostics plus précoces et précis ainsi que des mesures de prévention. L'élément clé pour ces améliorations est une stratification haute résolution (à savoir une classification selon des sous-groupes plus homogènes) des individus. Tandis que dans le passé, les indications de thérapie étaient souvent très larges («one drug for all»), la médecine personnalisée tâche d'identifier des groupes de patients plus petits et plus spécifiques qui seront ensuite traités de la même manière, ce qui peut aboutir au groupe d'une seule personne, à savoir à la thérapie individualisée. Cette démarche s'applique autant à la médecine somatique qu'à la médecine psychiatrique.

La stratification repose sur des biomarqueurs, à savoir des grandeurs mesurables pour chaque patiente et suffisamment reproductibles et qui permettent de séparer le groupe de patients initialement hétérogène en sous-groupes plus homogènes. De nombreux biomarqueurs sont issus de mesures moléculaires telles que la génomique, la transcriptomique, la protéomique, la métabolomique, etc. ainsi que de procédés d'imagerie tels que l'IRM, la CT, etc. De plus, les données sur le mode de vie, obtenues par exemple avec des ordinateurs vestimentaires («wearables») prennent de plus en plus d'importance. Dans ce contexte, les antécédents médicaux, idéalement compilés dans le dossier numérique du patient et par conséquent disponibles, revêtent aussi une importance particulière.

Comment fait-on concrètement pour détecter des sous-groupes de patients avec des réponses similaires au traitement? Pour cela, il faut identifier les biomarqueurs parmi l'immense variété des données mentionnées plus haut; il faut donc trouver les mesures qui permettent une classification en sous-groupes plus homogènes. Un tel biomarqueur est par exemple HER2. Les patientes souffrant d'un cancer du sein et dont les cellules tumorales surexpriment ce récepteur peuvent être traitées au trastuzumab (Herceptin®), étant donné que la substance active (un anticorps monoclonal) agit spécifiquement contre ce récepteur. Dans le cas où la tumeur ne surexprime pas HER2, le trastuzumab n'aurait pas d'effet positif, et entraînerait au contraire des dommages. Actuellement, la FDA recense environ 160 de tels biomarqueurs pharmacologiques autorisés pour la stratification en fonction de la réponse au traitement.

11.1. Signatures de biomarqueurs

Un seul biomarqueur n'est souvent pas suffisant pour obtenir une bonne stratification. De nombreuses maladies sont causées par des modifications moléculaires complexes qui ne peuvent être détectées qu'en combinant plusieurs biomarqueurs. Un exemple d'une telle signature de biomarqueur est le test MammaPrint™, autorisé par la FDA en 2007. Ce test pronostique du cancer du sein est composé de 70 gènes qui prédisent ensemble le développement de métastases et le succès d'une chimiothérapie. La stratification selon un groupe à faible risque et un groupe à haut risque a lieu à l'aide d'un score défini comme la somme pondérée des valeurs d'expression des 70 gènes. La description complète de la signature comprend donc, en plus des 70 gènes, également 70 pondérations et une valeur seuil qui sépare le groupe à faible risque du groupe à haut risque.

Seuls des programmes informatiques sont capables d'établir et d'exploiter les signatures de biomarqueurs complexes, car elles contiennent et combinent un grand nombre de paramètres. Les méthodes de la bio-informatique et de l'intelligence artificielle permettent de développer des algorithmes et des logiciels capables de faire des prédictions sur la base de grandes quantités de données, par exemple de profils moléculaires complets. Cette approche est par exemple utilisée dans les tumorboards moléculaires. De nos jours, l'utilisation de profils d'ADN et d'ARN du génome fait partie de la routine pour le diagnostic et le choix de la thérapie en oncologie. Avant de pouvoir exploiter ces données, il est nécessaire de procéder à une analyse statistique approfondie, car les données «omics» comportent souvent d'importantes erreurs suite à l'amplification et au séquençage. Les profils normalisés sont ensuite intégrés et interprétés ensemble quant aux options de thérapie. Le résultat est communiqué par un expert au tumorboard sous la forme d'un rapport bio-informatique clinique.

11.2. Dérivation de modèles prédictifs

Grâce à la bio-informatique et aux sciences des données, il est possible, à partir de volumes importants de données, de dériver des modèles prédictifs qui caractérisent avec une grande précision chaque patient individuel. Les méthodes de l'intelligence artificielle, en particulier celles de l'apprentissage automatique, du «Data Mining» (exploration de données) et celles des statistiques sont utilisées à cet effet. Pour diverses raisons, la découverte de signatures et de modèles de biomarqueurs complexes s'avère être un exercice statistique et arithmétique très difficile. D'une part, les données sous-jacentes sont presque toujours très incomplètes et regorgent d'erreurs à cause des imprécisions de mesure. D'autre part, chaque jeu de données typique contient un nombre de valeurs observées par patient largement supérieur au nombre total de patients. Tandis que le Big Data est généralement simplement associé à un grand volume de données, c'est la grande diversité des données en médecine (par exemple des mesures sur tout le génome, des images composées de nombreux pixels, de longues séries chronologiques fournies par les «wearables») qui complique la détection de schémas. Ainsi, il existe par exemple plus de 10^{200} possibilités pour sélectionner 70 gènes parmi 20'000.

Les procédés d'apprentissage automatique sont des méthodes capables de détecter un ou plusieurs modèles prometteurs parmi ces immenses quantités de données. Dans ce que l'on appelle l'apprentissage automatique contrôlé, ce sont les relations fonctionnelles entre les prédicteurs qui sont apprises. Il s'agit donc des caractéristiques descriptives et d'une variable cible, à savoir de la dimension qui doit faire l'objet de la prédiction. Des variables cibles pertinentes sont des classifications, par exemple en fonction du niveau de risque, des valeurs continues, par exemple en fonction de la gravité de la maladie ou de la durée de survie. Le test MammaPrint™ est donc un modèle linéaire qui établit un classement binaire des tumeurs sur la base de 70 valeurs d'expression génétique. Dans l'apprentissage automatique non contrôlé en revanche, aucune variable cible n'est définie. Le regroupement (*clustering*) consiste par exemple à tenter de regrouper les observations en fonction de leurs caractéristiques afin de former des sous-groupes plus homogènes. Ces deux classes d'apprentissage automatique sont souvent utilisées par la médecine pour identifier des groupes de patients (stratification) et pour construire des modèles prédictifs de la réponse au traitement (prédiction).

11.3. Validation

L'exactitude et la précision des prédictions revêtent une grande importance lorsque des modèles d'apprentissage statistiques et automatiques sont utilisés à des fins de diagnostic et de prédiction du succès thérapeutique. Pour estimer la qualité de la prédiction ou de la stratification, il ne suffit pas de se baser simplement sur le même jeu de données qui a servi à construire le modèle. Il est bien connu que cet évaluateur possède une tendance à la surestimation. En d'autres termes, il va suggérer une plus petite erreur que celle à laquelle il faudrait s'attendre sur la base de nouvelles données non encore considérées. C'est pourquoi une telle validation de modèles cliniques comprend les étapes suivantes: (1) une validation interne du jeu de données initial sur la base d'une validation croisée dans le cadre de laquelle une partie des données sont filtrées successivement pour évaluer la précision de prédiction, puis le modèle est appris sur la base des données restantes; (2) une validation externe sur une cohorte indépendante afin d'éliminer le biais éventuel lié à la composition de la cohorte de découverte initiale; et (3) une validation dans le cadre d'une étude prospective clinique en comparaison avec les normes des soins (*Standard of Care*). Étant donné que l'entier processus est très long et onéreux, il n'y a rien d'étonnant à ce qu'il n'existe à ce jour que peu de modèles validés sur le plan clinique.

11.4. Besoins infrastructurels

La bio-informatique et les sciences des données ne peuvent fournir des modèles utiles à la médecine que lorsque les données disponibles sont conformes aux exigences de quantité et de qualité requises. À cet égard, une harmonisation nationale et internationale des différents types de données et systèmes d'information est indispensable. Par ailleurs, des mesures importantes restent à mettre en œuvre pour l'organisation des données, à savoir la collecte et le traitement de données des patients, de données biologiques de base (y compris de bases de données biologiques) et de données «omics», ainsi que pour les processus structurés. Ceux-ci sont censés garantir une interopérabilité nationale de données cryptées venant des hôpitaux et des structures de recherche, et ce, en tenant compte des dispositions légales, de la sécurité des données et des conditions de protection des données toutes deux optimales.

En Suisse, ces défis ont incité la Confédération en 2016 à mandater l'ASSM avec la construction du «Swiss Personalized Health Network, SPHN». Le SPHN crée l'infrastructure qui permettra aux diverses données de santé d'être exploitées par la recherche et l'innovation.

Pour la période 2017–2020, l'action cible la création d'une infrastructure de données coordonnée sur le plan national. Cela permettra d'harmoniser les systèmes d'information locaux et régionaux et de garantir l'interopérabilité des données. Un autre accent est mis sur les systèmes de gestion de données cliniques qui sont censés assurer la sécurité de l'échange de données de patients (par exemple phénotypes des maladies). À long terme, on peut également envisager l'intégration de bases de données publiques, d'études de cohortes et de données de citoyens sains.

11.5. Limites des sciences des données

Bien qu'il existe déjà de nombreux outils de sciences des données à la base de secteurs entiers de l'économie, il subsiste cependant de nombreuses questions, en particulier concernant les sciences des données biomédicales. Dans les applications médicales par exemple, c'est l'interprétabilité des modèles qui présente un intérêt considérable en plus de la qualité de la prédiction ou de la stratification, car elle contribue à améliorer l'acceptation de prédictions basées sur des modèles. Les méthodes de l'apprentissage automatique sont certes bien appropriées à détecter des motifs complexes dans de grands volumes de données et

à faire des prédictions aussi précises que possible, mais elles ne sont pas en mesure d'élucider des liens mécanistiques ou causaux. La plupart du temps, les procédés reposent sur la découverte de corrélations, mais les corrélations n'impliquent généralement aucune causalité. La recherche de signatures de biomarqueurs aboutit par exemple souvent au constat que de nombreuses signatures différentes atteignent le même niveau optimal de qualité de prédiction. Après la découverte du MammaPrint™, on a par exemple découvert 70 autres signatures génétiques de même qualité. Les modèles qui découvrent des procédés d'apprentissage automatique doivent donc toujours être considérés comme des hypothèses qu'il convient ensuite de vérifier à l'aide d'autres méthodes scientifiques (par exemple Randomized Controlled Trials, «essais comparatifs randomisés», Longitudinal Observational Studies, «études longitudinales»).

La médecine personnalisée ne serait pas concevable sans la bio-informatique et les sciences des données. D'importants jeux de données du génome, d'imagerie et du style de vie ainsi que les progrès de l'intelligence artificielle permettent d'espérer que des progrès médicaux puissent être générés à long terme et durablement et aboutir à de meilleures thérapies et à des coûts réduits. Toutefois, la liste des exemples de réussite dans ce domaine est encore relativement courte. Il est probable que cela est, du moins en partie, la conséquence de la complexité moléculaire de nombreuses maladies et également du fait que même les procédés modernes à haut débit ne permettent que des observations très restreintes de ces systèmes.

Chapitre 12

Médecine personnalisée et santé publique

Les maladies non transmissibles résultent souvent de l'interaction complexe et permanente entre le hasard et des facteurs exogènes et endogènes. Pour la recherche épidémiologique sur leurs causes et les approches qui peuvent en être déduites pour la prévention primaire, la prédiction du risque et le dépistage précoce, deux aspects sont essentiels: la disponibilité de données longitudinales en lien avec des échantillons biologiques prélevés dans un but prospectif ainsi que le grand nombre d'échantillons. Dans le monde entier, on compte ainsi actuellement de plus en plus d'études à long terme portant parfois sur largement plus de 100'000 participants.

Editrice

Académie Suisse des Sciences Médicales (ASSM)
Maison des Académies, Laupenstrasse 7, CH-3001 Bern
mail@samw.ch, www.assm.ch

Conception

Howald Fosco Biberstein, Basel

Traduction

Domínic Nickel, Bern
CVB International, Lausen

Photo de couverture

adobestock – joyt; istock – teekid

Les versions allemandes et françaises (pdf) sont disponibles en ligne sous assm.ch/bases-medicine-personnalisee



Copyright: ©2019 Académie Suisse des Sciences Médicales. Ceci est une publication Open Access, distribuée sous les termes de la licence Creative Commons Attribution (<http://creativecommons.org/licenses/by/4.0/>). Le contenu de cette publication peut donc être utilisé, distribué et reproduit sous toute forme sans restriction, à condition que l'auteur et la source soient cités de manière adéquate.

Recommandation pour citer le texte:

Académie Suisse des Sciences Médicales (2019)
Médecine personnalisée. Bases pour la formation interprofessionnelle prégraduée, postgraduée et continue des professionnels de la santé.
Swiss Academies Communications 14 (6).

ISSN (en ligne): 2297-1823

DOI: <http://doi.org/10.5281/zenodo.3271401>



ODD: Les objectifs internationaux de l'ONU en matière de développement durable

Avec cette publication, l'Académie Suisse des Sciences Médicales apporte une contribution à l'ODD 3: «Permettre à tous de vivre en bonne santé et promouvoir le bien-être de tous à tout âge.»

sustainabledevelopment.un.org
www.eda.admin.ch/agenda2030 → français → agenda 2030
→ 17 objectifs de développement durable