Advanced Inferential MedicineSM

Dennis L. Barbour

Department of Biomedical Engineering
Washington University in St. Louis
St. Louis, Missouri 63130
dbarbour@wustl.edu

ABSTRACT

Traditional medical inference requires explicit determination of a patient's ailment prior to deciding the best treatment option. Evidence-based medicine informs these decisions from the outcomes of previous patients with similar ailments who received similar treatments. Precision medicine seeks to expand this nomothetic framework with many more potential diagnoses, the total possible number of which is inherently limited by the number of similar previous patients available for reference. The concept of making patient-care decisions using idiographic information unique to a particular patient may be a familiar concept to practicing clinicians, but it has no formal role within evidence-based medicine. The collective result is constrained inferential capacity of the dominant medical philosophy, leading to limited effectiveness of individual patient treatment decisions. A means of combining nomothetic and idiographic inference to optimize individual treatment outcomes would be a welcome addition to the modern medical armamentarium. Novel idiographic search algorithms informed by nomothetic prior beliefs can construct predictive models about individual patients that provide rigorous clinical decision support. This advanced medical inference framework exploits modern machine learning to generalize the concept of diagnosis and to make effective treatment decisions with or without definitive etiological understanding of a patient's ailment.

INTRODUCTION

The conventional medical inference chain that begins with patient presentation and ends with patient outcome includes obligatory intermediate steps of observation, diagnosis, prognosis and treatment (**Figure 1**). This sequence of steps has represented the cornerstone of medical practice for centuries (Gordon 1949; Pearn 2011). Each step results in a clinical decision, indicated by arrows. Results of decisions are indicated by abbreviations. Accumulated clinical observations result in a differential diagnosis, which is used to decide an actual diagnosis. The collective natures of the diagnosis and untreated prognosis, coupled with patient values and priorities, lead to the selection of a treatment, which is then delivered and ultimately leads to a measurable outcome. Additional inference and intervention loops are commonly employed (for example, to refine the differential diagnosis using additional observations) but are not considered for the moment.

Figure 1: Medical decisions in the classical patient workup result in observations (Ob) about patient status, a differential diagnosis (dDx) of possible diagnoses, leading to a diagnosis (Dx) with associated untreated prognosis (Px), leading to a treatment (Tx), which results in a future patient outcome (Ou). This pathway is referred to as Traditional Inferential Medicine (TIM).

This classical approach, referred to here as Traditional Inferential Medicine (TIM), is applied universally in different medical traditions. Observations of clinical status are interpreted in light of the underlying medical philosophy (e.g., allopathy, homeopathy, osteopathy), and treatments are selected consistent with that philosophy. Modern medicine is “evidence based,” meaning that treatment decisions are not based on theoretical principles *per se*, but are predominantly guided by outcomes of previous patients with similar diagnoses facing similar treatment options (i.e., the evidence) (Eddy 1986; Sackett, Rosenberg, et al. 1996; Eddy 2005). This philosophy fits squarely within nomothetic epistemology, valuing inference derived from what is shared in common among previous patients and therefore most likely to be generalizable to new patients (Lamiell 1998; Windelband 1998).

Conventional medical inference is highly successful and is being rapidly augmented in the current personalized/precision medicine era. The concept behind precision medicine is to substantially expand the number of potential diagnoses and to match each patient’s diagnosis with the most effective treatment for previous patients with a similar diagnosis (President’s Council of Advisors on Science and Technology 2008; National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease 2011; Collins and Varmus 2015). This concept is not philosophically distinct from the principles of evidence-based medicine.

Diagnostic expansion has great potential to realize an early stated goal of the movement to provide better clinical decisions for patients (Pauker and Kassirer 1987). A key drawback of this framework, however, stems directly from its nomothetic philosophy: *the numerical expansion of unique nomothetic diagnoses is inherently limited by the size of the patient population*. Said another way, in order to make a (nomothetic) evidence-based treatment decision for a particular patient, sufficient similar patients must have been assigned the same diagnosis previously in order to make statistically meaningful comparisons. As the total number of diagnostic categories balloons, fewer similar patients will comprise evidence-based cohorts for any particular patient.

To use one example, whole-genome analysis alone provides more individualized information about a patient than genomes from other patients could ever fully contextualize for treatment decisions, simply due to gene combinatorics. Adding environmental and lifestyle factors pushes this situation even further beyond tractability. By focusing on data emphasizing shared traits, evidence-based medicine induces clinicians either to treat any particular patient as part of a nomothetic cohort (“treating the diagnosis”), or to cross out of the realm of nomothetic evidence toward an individualized, idiographic process whereby information unique to the current patient is used to make treatment decisions. Practicing clinicians grapple with this tension daily, and a framework of formal medical inference that places nomothetic and idiographic

inference on similar foundations for decision making has great potential to overcome the inherent limitation of evidence-based medicine.

A useful process for devising such a new framework involves revisiting classical medical inference. Deploying the same inference chain designed for human brains millennia ago is unlikely to achieve the best possible outcomes in the modern era of Big Data and machine learning. For example, only two irreducible medical decision classes exist: which information to gather and which intervention to enact. All other conventional medical decisions are actually inferential intermediates designed to accommodate limited human reasoning capacity. Revision of this inference chain could be fruitful if inferential processes are aligned to exploit powerful quantitative inference tools and, as a result, eliminate the ubiquitous problem of selecting treatments for patient cohorts instead of individual patients. Current precision medicine efforts are not on this path because of the nomothetic need to group like patients together in order to determine treatment plans.

A proposal for one suitable revision to the traditional medical inference chain is given in **Figure 2**, again omitting additional inference loops. This framework, referred to here as Advanced Inferential MedicineSM (AIM), divides the medical inference chain into an action pathway (top) and an information pathway (bottom). Action decisions answer the question of what to do next (i.e., which tests to order or which treatments to plan) while information decisions provide context and explanation for clinicians and patients. It is possible to use information decisions to determine or constrain action decisions, but a more powerful option exists within AIM.

Figure 2: Medical decisions in an alternate patient workup result in observations (Ob) about patient status followed by an observation model (ObM), which represents both errors in actual observations and predictions of yet-to-be-made observations along with their estimated certainty. A differential diagnosis (dDx) of possible diagnoses followed by a diagnosis (Dx) reside on an information branch and are used to instruct and explain. The observation model itself is used to make the action decision of which treatment (Tx) to deliver, which results in a differential prognosis (dPx) of possible prognoses followed by a treated prognosis (Px) on the information branch. Finally, a patient outcome (Ou) results in the future following treatment. This pathway is referred to as Advanced Inferential MedicineSM (AIM).

Probabilistic inference is a mechanism to encode prior beliefs or pretest probabilities about a patient's state (e.g., suspected diseases, genotype, physiological constraints), as well as observations made regarding the patient's current/past states (e.g., family history, past history, review of systems, physical exam, lab tests) and current/past influencers of patient state (e.g., comorbidities, environment, lifestyle). While traditional medical inference methods segregate classes of observations such as histories, test results and risk factors into different inferential streams, AIM integrates them all within

one framework. It does so by introducing a new concept into the medical inference chain, that of the observation model.

The observation model is a unified probabilistic representation of both observations (e.g., test results) and potential observations (e.g., results of tests not yet conducted), thereby quantifying processes that may influence patient outcome. AIM centralizes the observation model in the same way that TIM centralizes the diagnosis. The observation model contains the information necessary to make useful treatment decisions, and while it can be constructed by conditioning upon diagnostic accuracy, such conditioning is not required.

The introduction of the observation model confers several advantages on AIM relative to TIM. Already alluded to is the patient-customization potential. The observation model for each patient is both unique and probabilistic. This combination of factors allows treatment decisions to be truly individualized to the patient at hand. In other words, outcomes of previous patients diagnosed with the same disease are not strictly necessary in order to make good treatment decisions with AIM, which is a departure from the strategy of evidence-based medicine. AIM allows diagnoses to revert to their original purpose when treatment options were limited—to inform and communicate in shorthand about the current state of the patient. The capability of exploiting etiological and mechanistic insights derived from patient cohorts to develop new or improved treatment strategies remains intact. Meanwhile, individual treatment decisions within AIM are made with all the fine details captured in the observation model.

AIM also provides a formal mechanism to include other relevant information either from fundamental mechanistic principles or from the outcomes of previous patients. These prior beliefs are naturally incorporated into the probabilistic observation model. Furthermore, prior beliefs provide the basis not only for ascribing confidence to observations about the patient, but also for ascribing confidence to potential observations. As a result, the observation model can be used to assess the probability that acquiring particular additional data would improve the likely outcome by revealing a better treatment strategy, thereby establishing best patient evaluation practices and reducing utilization at the same time.

Finally, while the inference path for making treatment decisions in AIM is standardized, the options for how to acquire and value observations have expanded. We could, of course, choose to acquire observations that improve our confidence in treated outcomes, but complex situations may exist where intermediate conclusions represent useful surrogates for outcomes assessments. An example might be when a collection of observations is pathognomonic for a particular diagnosis. In such cases, information decisions could serve as useful inferential tools for improving treatment selection and therefore patient outcomes.

The diagnosis and prognosis are employed in AIM in order to describe the current state of the patient and the most likely future state, respectively. These are both the result of information decisions, where interim conclusions are drawn for additional inference. Other intermediate stages more consistent with the philosophy of AIM are the differential diagnosis and differential prognosis, which are the result of probabilistic inference steps. The differential diagnosis represents the probability of each potential diagnosis and the confidence in that probability. The differential prognosis similarly represents probabilistic patient outcomes and the confidence in those predictions. The advantage of AIM is its formal mechanism to condition action decisions on probabilistic models rather than on intermediate conclusions, thereby superseding the simplifying inferential intermediates inherent to TIM in favor of fully probabilistic models informed by current patient data and previous patient outcomes.

ELEMENTS OF MEDICAL INFERENCE

How individual elements of TIM and AIM are defined is important for understanding how they interact, as well as the differences between the two approaches.

Observations reflect information (or data) gathered about a particular patient. These are all point values defined in their original units. Blood pressure is given in scalar values of mmHG for systolic and diastolic; serum sodium level is given in mg/dL; radiograph readings may be given as the visual observations of a radiologist. The observation component for a patient can be conceptualized as a column vector where each entry is a scalar representing a single observation related to that patient. Observations are generated from tests ordered as a result of action decisions.

The *observation model* is a probabilistic description of observations and potential observations. In the former case, the observation model reflects the uncertainty in actual observations. For example, the error expected from a measurement device introduces uncertainty in the value that was measured compared to the true value. Physiological variability might also lead to the true value at one point in time not reflecting a clear indication of the underlying physiological process. This uncertainty can also be captured in the probabilistic representation of observed evidence.

In the case of potential observations, the observation model reflects the uncertainty of unobserved data based on model assumptions (i.e, prior beliefs) and observed data. For example, the results of a particular blood test may not be available, but a probabilistic prediction of that test result based on the best knowledge available for that patient can still be used for downstream inference. The problem of missing data that plagues large-scale data analytic methods for assessing patients and selecting treatments can therefore be circumvented when downstream inference steps exploit the observation model instead of observations alone. Use of the observation model also overcomes the limitations of relying upon intermediate conclusions such as diagnoses, which themselves are constrained by missing data.

Entries in the observation model consist of probability distributions over the corresponding observations in their original units. Blood pressure is given as a distribution of values in mmHG for systolic and diastolic; serum sodium level is given as a distribution of values in mg/dL; radiograph readings may be given as a distribution of probabilities of radiologist visual observations. The observation model can be conceptualized as a column vector where each entry is a probability distribution reflecting the belief of the actual value of the corresponding measurement for that patient given prior beliefs and observations.

The *diagnosis* represents the single unifying hypothesis regarding anatomical and physiological state that best explains observations made about a patient. For the purposes of this development, potential diagnoses that can be readily differentiated, such as a broken leg and pancreatic cancer, are not considered to be alternatives given the same set of clinical observations. The diagnosis can therefore be conceptualized as a column vector where all entries are 0 except for one that has a scalar value of 1. Diagnosis and phenotype are not the same thing because a genetic disorder can represent a diagnosis before any signs or symptoms of a disease phenotype occur. A diagnosis results from an information decision.

The *differential diagnosis* represents the probability that each of the putative diagnoses is the actual diagnosis. The differential diagnosis can be conceptualized as a column vector where each entry is a probability distribution of probability values (obviously constrained between 0 and 1). A logical decision criterion to infer a diagnosis

from the differential diagnosis is to select the differential diagnosis entry with the highest expected probability. The differential diagnosis is also referred to as the *diagnosis model*.

The *treatment* is the major controllable intervention. The treatment can be conceptualized as a column vector where all entries are 0, reflecting treatment options not selected, except one that has a scalar value of 1, indicating the treatment of choice. Combination therapies can be represented combinatorially such that a single entry represents the combination, or the treatment vector could be formulated in such a way that treatments are not mutually exclusive, allowing multiple entries to take on values of 1. A treatment results from an action decision.

The *prognosis* represents the single most likely outcome at some future time point. Because outcome space is divided into mutually exclusive possibilities, the prognosis therefore can be conceptualized as a column vector where all entries are 0 except one that has a scalar value of 1, indicating the selected prognosis. A prognosis results from an information decision.

The *differential prognosis* represents the probability that each of the potential outcomes actually occurs. Again because outcome space is divided into mutually exclusive possibilities, the differential prognosis can be conceptualized as a column vector where each entry is a probability distribution of probability values reflecting different potential outcomes. The differential prognosis is also known as the *outcome model*.

The *outcome* represents an important patient state at some future time point. Again, because outcome space is divided into mutually exclusive possibilities, the outcome can be conceptualized as a column vector where all entries are 0 except for one that has a scalar value of 1, indicating the actual outcome. Outcomes are expected to be easy to discern by definition because if they were not, they would not be meaningful to the patient. For example, the presence of hypertension in 5 years may not be a meaningful outcome on its own, though number of strokes experienced over the next 5 years would be. For this reason, outcome and phenotype are not the same thing.

ELABORATION OF AIM

The general AIM framework can be cast into several specific formulations intended to accomplish different goals.

AIM1: Optimize Observation Model Accuracy

The simplest use of the AIM framework is to construct as accurate an observation model as possible. A more general way of saying this is to minimize the cost of predicted observations because different errors can have different costs (e.g., misses might be much more costly than false positives). The AIM procedure in this case selects the observations that maximize the inferential capacity of the observation model given the observations. Observation models can be constructed nomothetically from fundamental principles as well as prior patient cohort data, and they become “personalized” with idiographic data collected from a given patient. Inferential capacity for AIM1 is accuracy as quantified by the ability to predict unobserved data. In this case the observation model is trained up using observed data from a patient and evaluated by its accuracy in predicting new observations for that patient.

Figure 3: In AIM1, observations are made to maximize the accuracy of the observation model in making predictions about new observations for the current patient.

AIM2: Optimize Diagnosis Accuracy

An intermediate use of the AIM framework is to determine as accurate a diagnosis as possible. This can be generalized as minimizing the cost of diagnostic decisions. The AIM procedure in this case selects the observations that maximize the inferential capacity of the differential diagnosis given the observations and the observation model. This formulation allows an identical inference process to be applied to patients with different sets of observations because the observation model naturally accounts for missing data. Inferential capacity for AIM2 is accuracy as quantified by the ability to predict the actual diagnosis (quantified probabilistically as the differential diagnosis) for the current patient.

Figure 4: In Aim2, observations are made to maximize the accuracy of the observation model in making predictions about the diagnosis for the current patient.

AIM3.1: Optimize Untreated Prognosis Accuracy

A more advanced use of the AIM framework is to determine as accurate a prognosis (i.e., the least costly prognosis errors) as possible without any treatment. The AIM procedure in this case selects the observations that maximize the inferential capacity of the differential prognosis given the observations and the observation model but no treatment. If prior information exists for a large number of patients from a diverse cohort, the differential prognosis serves as a public health prediction as a function of observation model predictors. Inferential capacity for AIM3.1 is accuracy as quantified by the ability to predict the actual untreated outcome for the current patient.

Figure 5: In Aim3.1, observations are made to maximize the accuracy of the observation model in making predictions about untreated outcomes for the current patient.

AIM3.2: Optimize Prognosis Accuracy

The ultimate use of the AIM framework is to determine as accurate a prognosis (i.e., the least costly prognosis errors) as possible with any treatment. The AIM procedure in this case selects the observations and treatment that maximize the inferential capacity of the differential prognosis given the observations, the observation model and the treatment. Inferential capacity for AIM3.2 is accuracy as quantified by the ability to predict the actual treated outcome for the current patient.

Figure 6: In Aim3.2, observations are made to maximize the accuracy of the observation model in making predictions about treated outcomes for the current patient.

Again, other inference loops not documented here can refine models and decision rules. Outcomes from previous treatments of the current patient or other patients are particularly valuable in this regard. The current development of AIM focuses on a single patient workup procedure.

EARLY APPLICATIONS OF AIM

The AIM framework has been applied to work up patients for perceptual and cognitive status via sequential test batteries. Some examples of such tests include perceptual testing for sensory processing disorders, neuropsychiatric testing for mental status, and cognitive testing for dementia. These tests fall into the broad class of psychometrics, which is a well-established field employing substantial mathematical

rigor developed over many decades and combined with empirical findings to yield validated results (Fechner 1860; Gescheider 2015; Prins 2016).

The main problem with psychometric inference methods, however, is that they are generally underpowered relative to the amount of data that can be acquired and used for estimation. The key limitation in this case is the bottleneck of serial sampling because people can only be asked one question at a time. Neural processes are extremely high-dimensional spaces, and performing inference about neural state with relatively small amounts of data requires both efficient and effective methods. A similar statement can be made about medical inference generally. Consequently, the first application of AIM1 has been to dramatically improve the estimation process behind a particular perceptual estimation problem, that of hearing thresholds.

The pure-tone threshold audiogram represents the gold standard evaluation of hearing function. Conventional methodology is to deliver a series of tones to a patient, some of which can be heard and some not heard. The exact collections of heard or unheard tones are used to estimate hearing thresholds accurately. The standard test of this type in use today and codified by international practice standards was developed in the 1940's. These standards stipulate which tones to deliver under which conditions and in which order, exactly analogous to practice standards guiding diagnostic test selection in other clinical scenarios.

Figure 7 depicts a novel active machine learning procedure that delivers probe tones selected for estimating accurate hearing thresholds, reflecting AIM1. This method is at least as accurate and considerably more efficient than existing methods (Song, Wallace, et al. 2015; Song, Garnett, et al. 2017; Song, Sukesan, et al. 2017). AIM1 is completely compatible with TIM because it essentially represents an efficient way to obtain an accurate test result. Any decision about whether a patient has a particular pathology must be made following all data collection and audiogram estimation in this case.

Figure 7: Hearing estimation using the AIM1 active sampling framework uses tone delivery that maximizes information about the underlying function itself. Normal hearing is represented on the left; noise-induced hearing loss on the right. Blue plusses represent heard tones; red diamonds, unheard tones. Grayscale represents the true hearing function.

Figure 8 depicts the same two patient conditions and the same test configuration. The difference this time is that the tones are selected (and observation of patient responses are recorded) in such a way as to rule in or out the narrow notch that represents noise-induced hearing loss (Gardner, Malkomes, et al. 2015). Neither of these two test results provides an accurate estimate of hearing threshold itself, yet the tests terminated at 99% confidence that either no notch existed (left) or that a notch truly did exist (right). The key realization in this example is that the best tests to order (and corresponding observations to make) to optimize a particular diagnostic decision may not result in the most accurate observation model, but can achieve arbitrary confidence

about which diagnosis is appropriate to ascribe. This graphical example of AIM2 illustrates how the observations that best serve one inference goal may not be the observations that best serve another goal. It also clarifies how test results that are not particularly accurate or informative in raw terms can still be extremely useful when taking into account additional nomothetic information, such as disease variations and prevalence.

Figure 8: Hearing estimation using the AIM2 active sampling framework uses tone delivery that maximizes the accuracy of diagnosis. Normal hearing is represented on the left; noise-induced hearing loss on the right. Blue pluses represent heard tones; red diamonds, unheard tones. Grayscale represents the true hearing function.

Workup for hearing loss under TIM performs all downstream inference using observations about hearing thresholds themselves. This is the reason that observation accuracy and reliability are so critical with current methods. By accounting for measurement and estimation error in the observation model, quantifiable objectives can be achieved in AIM with lower-quality observations. The ultimate goal of AIM3 is to make good treatment decisions, which in this framework can be done even with observations that by themselves would incompletely reduce uncertainty about a patient's physiology or disease state under TIM.

PHILOSOPHICAL SHIFT OF AIM

While many elements of AIM are novel relative to TIM, the relocation of the diagnosis out of the direct inference chain between observations and outcomes most clearly reflects AIM's philosophical shift. Diagnoses represent a taxonomy of disease that has propelled the steady advancement of medical practice since antiquity and has led to the heights of accomplishment permeating medical practice today (National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease 2011; Pearn 2011). The main point of AIM is not to discard those advances, but to generalize them. Treatment decisions are made in every medical practice without definitive diagnoses. *AIM intends to empower good treatment decisions in the face of this uncertainty while preserving the ability to make great decisions when definitive diagnoses are possible.*

The departure of AIM from traditional patient workup procedures can be understood by considering its proposed alteration in information organization. Historically, taxonomies have proven extremely useful if not outright necessary for organizing information too extensive to be processed all at once by one brain. Scholars have relied upon vast amounts of information to be organized into books and classified by subject on library shelves. Professionals have relied upon client files to be ordered by topic into folders, drawers and cabinets. Early computer file systems and email software mirrored

this type of organization, with electronic files and mail messages only able to reside within one electronic folder at a time. Organization of information in this case can readily be accomplished by a straightforward hierarchical abstraction. All life, for example, can be organized into a hierarchical taxonomy (Linnaeus 1759; International Commission on Zoological Nomenclature 1999). Diseases can be organized similarly (Osler 1892).

Modern computer file systems and email software typically still allow this type of mutually exclusive information classification, but now also enable non-mutually exclusive tags, labels or modifiers to be attached to individual files or messages. In this way, an organizational hierarchy need not be conceived prior to data collection. The best organization for particular retrieval purposes can be constructed algorithmically as data are accrued, with desired information then retrieved by a textual search. Modern algorithms determine the relevance of search results under these conditions and are so good at their task that the concept of typing a natural language query to find one entry out of millions of records not taxonomically organized is now second nature to modern computer and smartphone users. A hierarchical taxonomy is one way to organize information; allowing other organizational structures to develop naturally in ways not readily comprehensible to the human brain has the potential to improve the quality of both inference and decisions.

These analogies are relevant in this case because working up a patient is essentially a search task. Most of the necessary data have not yet been observed at the outset of a workup, however, so search algorithms cannot simply organize existing data. Instead, they must anticipate what would be the most useful information to observe for any particular goal. The observation model implements this capability with a flexibility that a diagnostic taxonomy cannot. *The value of such an alternative becomes apparent upon realizing that a workup is not fundamentally a search for what is wrong, but a search for the best course of action.*

Attempts to provide more flexibility in the patient workup by assessing uncertainty typically assign diagnoses probabilistically. The disease taxonomy in such cases is represented via the differential diagnosis (Pauker and Kassirer 1987; Pearn 2011). AIM2 exploits this framework, only not to make treatment decisions directly from diagnoses. Instead, diagnostic inference in AIM2 is used to train decision rules that partition observation model space into subspaces reflecting optimal treatment decisions. Clinical decisions can then be based more flexibly on combinations of observations rather than on an intermediate diagnostic decision, probabilistic or not. *The key argument in support of AIM is that the meaningful classes in observation space can be predicted outcomes from treatments instead of diagnostic decisions.* A particular treatment might be appropriate for a wide variety of observations, just like an email message might have many relevant labels instead of simply residing within a single mutually exclusive mail folder.

The diagnosis has been an obligate intermediate in medical treatment for so long that relocating it may seem counterintuitive or even counterproductive. That is because it is an extremely valuable tool when the only inference engine available is the human brain. Humans cannot accurately compute serial conditional probabilities, and the diagnosis is an intermediate step that allows clinicians to compartmentalize these calculations. This inferential intermediate works well when diagnoses can be definitive or nearly definitive (i.e., high confidence), but suffers obvious limitations when a clear diagnosis is not readily forthcoming.

The AIM formulation also automatically accommodates comorbidities, which would lead to different informative observations compared to cases where other or no comorbidities exist. Finally, by using the observation model rather than the diagnosis for decision making, both nomothetic prior information and idiographic active observations

about a particular patient can be used to plan treatment strategies. Recognition that combining nomothetic and idiographic information is needed to achieve the best patient outcomes is acknowledged or at least debated in some fields, such as psychiatry (Schafer 1999; Insel, Cuthbert, et al. 2010; Pilecki and McKay 2016; Shean 2016), but formal methods to reconcile them have generally been lacking.

SUMMARY

A summary of AIM using updated nomenclature is depicted in **Figure 9**. Notably absent are information decisions, which in TIM serve as intermediates to accommodate limited human inferential capacity. In AIM, the only decisions are action decisions and inference remains probabilistic in the models until a decision is to be made.

The observation, diagnosis and outcome models themselves are initially informed by previous patients, represented as prior beliefs. Additional observations of the current patient are used to update the models specifically for that patient. In this way nomothetic data and idiographic data are both formally incorporated into the inference chain. All 3 models are inherently probabilistic, with the degree of certainty about any particular model quantity reflected in the associated probabilities. The outcome model essentially feeds a clinical decision support system predicting outcomes given treatment choices and the observation model. Treatments are selected in conjunction with this information and patient values/priorities. Which observations to make about the patient in order to deliver better decisions are determined in conjunction with predictions from the observation model, the diagnosis model, the outcome model, or some combination of the 3. Actual outcomes in the current patient can be used to refine the outcome model for future patients.

Figure 9: Patient workup using the full AIM framework including the observation model, the diagnosis model and the outcome model. Clinical decisions are made according to the indicated AIM strategies with the primary goal of selecting the best treatment for a particular patient at a particular point in time.

VISIONS OF CLINICAL FUTURES

If AIM is successful at empowering future clinicians and patients to improve outcomes, some of today's recalcitrant clinical questions may be at least partially resolved. Should Patient X receive a prostate-specific antigen (PSA) test? Should Patient Y receive a screening mammogram? Should Patient Z's papillary thyroid carcinoma be resected or monitored? Answers to these questions very likely differ by patient. For example, the relationship among serum PSA level, treatment selection, and survival appears to have a high variance (Greene, Albertsen, et al. 2009). AIM provides

a framework to quantify both the relationship and uncertainty among these variables, proceeding individually for each patient and taking into account other relevant patient-specific variables.

The examples above from the cancer domain are particularly relevant today because alternative clinical workup strategies exist with the express goal of individualizing diagnoses and treatments. Molecular medicine promises to identify genetic factors contributing to cancers and improve the selection of therapeutic agents specifically targeting exactly the etiology faced by individual patients. This tremendous potential is leading the charge for personalized/precision medicine generally and represents a promising extension of evidence-based medicine.

As noted earlier, however, purely nomothetic regimes of medical inference like these are inherently limited by the quantity and similarity of previous patients. AIM's goal is to provide the ability to couple rigorous nomothetic inference with rigorous idiographic inference in order to improve outcomes in individual patients. In the PSA example, the high variance may indicate that serum PSA is a predictive test for some patients and not for others or is influenced by latent predictors not fully accounted for. Beginning with nomothetic prior beliefs about other patients, idiographic search procedures during a workup can be used to determine which potential predictors in Patient X are actual predictors. This individualized information is captured in the probabilistic models of AIM and used along with personal priorities to decide which prostate treatment, if any, is appropriate for Patient X.

CONCLUSIONS

AIM can be thought of as a generalization of TIM appropriate for modern machine-assisted clinical decision support. To achieve recognizable TIM for a patient using the AIM procedure, clinical decisions would be made using the AIM1 framework until some observation criterion is achieved, then using the AIM2 framework until some diagnostic criterion is achieved, then using the AIM3 framework until some outcome criterion is achieved. Ideally in the full AIM framework, however, clinical decisions will be made using the highest AIM subframework that is feasible. For example, it is better in terms of utilization and time efficiency to order tests that are most helpful at differentiating which treatment would be most effective given existing knowledge about a patient (AIM3). If inadequate outcome models exist, however, AIM2 could be useful for ordering tests until a confident diagnosis model can be constructed. If diagnostic clarity is unlikely to be achieved, tests could be ordered that are intended to be informative about patient status with the hope that the accumulation of a sufficient number of these observations might lead to a better treatment decision. The entire AIM framework is intended to map the relationships among treatments (inputs), observations (states) and outcomes (outputs).

AIM represents the germ of a medical philosophy that leaves human experts in charge of the most critical clinical decisions and all significant clinical information transfer among clinic, patients and other interested parties. The inference steps relegated to machines in AIM are those that humans are inherently challenged to implement at the proficiency required for extending clinical outcomes to the next level of performance. The result is a conceptual platform upon which to organize ongoing large-scale efforts to standardize electronification of medical records, collect Big Data about patients and populations, and deploy advanced inference techniques to assist clinicians and patients in jointly making good decisions that improve healthcare outcomes.

ACKNOWLEDGEMENTS

I would like to thank Mitchell Botney, Roman Garnett, Soo Kim, Jay Piccirillo, Harold Lehmann and Chad Sylvester for helpful comments on an early draft of this manuscript.

GLOSSARY

All definitions are given in the context of medical inference.

Action decision: a decision that results in the selection of a diagnostic test or a therapeutic treatment for a patient

Assumption: belief about the state of a patient or an outcome before evidence is collected about that patient; same as prior belief

Data decision/diagnostic test decision: an action decision that results in a particular diagnostic data-collection procedure being performed for a patient

Decision: an evaluation of information and distillation to a single best variant

Diagnosis (Dx): the best description of a patient's current medical state

Diagnosis model (DxM): a probabilistic description of possible medical states that may describe the patient currently; same as differential diagnosis

Differential diagnosis (dDx): a probabilistic description of possible medical states that may describe the patient currently; same as diagnosis model

Differential prognosis (dPx): a probabilistic description of possible personal and/or medical states that may describe the patient at a future time; same as outcome model

Idiographic: referring to trends unique to a particular individual

Inference: making a determination by mathematically combining assumptions and observations

Inferential capacity: the ability to perform accurate inference or make good decisions, typically quantified with a loss function or utility function

Information decision: an inferential decision that results in an intermediate conclusion intended to clarify a condition or support an action decision

Medical state: a condition of a patient that falls within the standard scope of pathology

Nomothetic: referring to trends people share in common

Observation (Ob): measurement or other data collection about a patient that may be used to improve treatment decisions; examples include patient history (pHx), family history (fHx), review of systems (Rs), physical exam (Pe), test results (Te), genetic factors (Ge), environmental factors (En), and lifestyle factors (Li)

Observation model (ObM): probabilistic representation of observed and unobserved data about a patient along with uncertainty

Outcome (Ou): a patient's personal or medical state at a future time

Outcome Model (OuM): a probabilistic description of possible personal and/or medical states that may describe the patient at a future time; same as differential prognosis

Personal state: a condition of a patient that is non-medical and therefore unlikely to be represented in a diagnosis, such as poverty or happiness

Prior belief: belief about the state of a patient or an outcome before evidence is collected about that patient; same as assumption

Probabilistic inference: combining prior beliefs and observed data into a representation of a quantity of interest embodied by a random variable

Prognosis (Px): the best description of a patient's future personal and/or medical state

Prognosis Model (PxM): a probabilistic description of possible personal and/or medical states that may describe the patient at a future time; same as differential prognosis

Treatment (Tx): an intervention designed to improve health and/or reduce disease burden

Treatment decision/therapeutic decision: an action decision that results in a particular treatment option being applied

Unobserved data: measurement or other data collection about a patient that has not yet occurred and can only be represented probabilistically by combining prior beliefs and observed data

BIBLIOGRAPHY

- Collins, F. S. and Varmus, H. (2015). "A new initiative on precision medicine." *N Engl J Med* 372(9): 793-795.
- Eddy, D. M. (1986). "Successes and challenges of medical decision making." *Health Aff (Millwood)* 5(2): 108-115.
- Eddy, D. M. (2005). "Evidence-based medicine: a unified approach." *Health Aff (Millwood)* 24(1): 9-17.
- Fechner, G. T. (1860). *Elements of Psychophysics*. New York, Holt, Rhinehart & Winston.
- Gardner, J., Malkomes, G., Garnett, R., Weinberger, K. Q., Barbour, D. and Cunningham, J. P. (2015). "Bayesian active model selection with an application to automated audiometry". *Adv Neural Inf Process Syst*, Morgan Kaufmann.
- Gescheider, G. A. (2015). *Psychophysics: the fundamentals*. Hove, United Kingdom, Psychology Press.
- Gordon, B. L. (1949). *Medicine Throughout Antiquity*, F.A. Davis Co.
- Greene, K. L., Albertsen, P. C., Babaian, R. J., Carter, H. B., Gann, P. H., Han, M., Kuban, D. A., Sartor, A. O., Stanford, J. L., Zietman, A. and Carroll, P. (2009). "Prostate specific antigen best practice statement: 2009 update." *J Urol* 182(5): 2232-2241.
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C. and Wang, P. (2010). "Research domain criteria (RDoC): toward a new classification framework for research on mental disorders." *Am J Psychiatry* 167(7): 748-751.
- International Commission on Zoological Nomenclature (1999). *International Code of Zoological Nomenclature*. London, The International Trust for Zoological Nomenclature, 1999, c/o The Natural History Museum.
- Lamiell, J. T. (1998). "'Nomothetic' and 'idiographic' - Contrasting Windelband's understanding with contemporary usage." *Theory & Psychology* 8(1): 23-38.
- Linnaeus, C. (1759). *Systema Naturae*. Stockholm, Laurentius Salvius.
- National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease (2011). *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, DC, The National Academies Press.
- Osler, W. (1892). *The Principles and Practice of Medicine*. New York, D. Appleton and Company.
- Pauker, S. G. and Kassirer, J. P. (1987). "Decision analysis." *N Engl J Med* 316(5): 250-258.
- Pearn, J. (2011). "Differentiating diseases: the centrum of differential diagnosis in the evolution of Oslerian medicine." *Fetal Pediatr Pathol* 30(1): 1-15.
- Pilecki, B. and McKay, D. (2016). "Evidence-Based Therapies, Evidence-Based Practice, and the Intersection of Nomothetic and Idiographic Foundations of Psychotherapy Research and Application: A Reply to Shean." *Psychodyn Psychiatry* 44(1): 25-38.
- President's Council of Advisors on Science and Technology (2008). *Priorities for Personalized Medicine*. Washington, DC, Executive Office of the President Office of Science and Technology Policy.
- Prins, N. (2016). *Psychophysics: A Practical Introduction*, Academic Press.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B. and Richardson, W. S. (1996). "Evidence based medicine: what it is and what it isn't." *Bmj* 312(7023): 71-72.
- Schafer, M. (1999). "Nomothetic and idiographic methodology in psychiatry--a historical-philosophical analysis." *Med Health Care Philos* 2(3): 265-274.
- Shean, G. (2016). "Psychotherapy Outcome Research: Issues and Questions." *Psychodyn Psychiatry* 44(1): 1-24.

- Song, X. D., Garnett, R. and Barbour, D. L. (2017). "Psychometric function estimation by probabilistic classification." *J Acoust Soc Am* 141(4): 2513-2525.
- Song, X. D., Sukesan, K. A. and Barbour, D. L. (2017). "Bayesian active probabilistic classification for psychometric field estimation." *Atten Percept Psychophys*.
- Song, X. D., Wallace, B. M., Gardner, J. R., Ledbetter, N. M., Weinberger, K. Q. and Barbour, D. L. (2015). "Fast, continuous audiogram estimation using machine learning." *Ear Hear* 36 (6): e326-e335.
- Windelband, W. (1998). "History and natural science." *Theory & Psychology* 8(1): 5-22.