Diagnoses Without Names

Challenges for Medical Care, Research, and Policy

Michael D. Lockshin
Mary K. Crow
Medha Barbhaiya
Editors

Springer

Diagnoses Without Names


HSS
HOSPITAL FOR SPECIAL SURGERY

Michael D. Lockshin • Mary K. Crow
Medha Barbhaiya
Editors

Diagnoses Without Names

Challenges for Medical Care, Research, and Policy

Springer

Editors
Michael D. Lockshin
Weill Cornell Medicine
Hospital for Special Surgery
New York, NY, USA

Mary K. Crow
Weill Cornell Medicine
Hospital for Special Surgery
New York, NY, USA

Medha Barbhaiya
Weill Cornell Medicine
Hospital for Special Surgery
New York, NY, USA

ISBN 978-3-031-04934-7 ISBN 978-3-031-04935-4 (eBook)
https://doi.org/10.1007/978-3-031-04935-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To all the patients whose experiences taught and will continue to teach all of us.

Foreword

The author has no conflict of interest to report.

For many centuries, persons dedicated to helping others who are sick have sought ways to classify and group illnesses in an attempt to better diagnose and treat them. Eventually, medical science focused on the classification of diseases came to be called nosology, derived from the Greek words *nosos* (disease) and *-logia* (-study of). The related field nosography sought to enable diagnostic labels for conditions and syndromes. A diagnosis is a consensus label that synthesizes a patient's subjective and objective findings. The label, with no formal boundaries, can describe symptoms (chronic pain syndrome), physical findings (asthma), laboratory tests (infections), or molecular biology (the newly described lupus-like VEXAS syndrome).

Diagnostic labels are time-limited and evolve as cumulative medical and scientific knowledge leads to greater insights. For example, in the nineteenth century, the science of microbiology transformed the concept of infectious diseases and the electrocardiogram changed the concept of heart attack. In the twentieth century, medical imaging such as roentgenographs (now called radiographic images), followed by computerized tomography, magnetic resonance imaging, and positron emission tomography, revolutionized concepts of and priorities for assigning diagnoses. In the twenty-first century, improved ways to accumulate large datasets related to cell biology, such as cytometry by time of flight (CyTOF); gene expression, such as single cell RNA sequencing; large-scale genotyping; and other approaches have led to more sophisticated ways to subset, cluster, redefine, and identify new diagnoses. New illnesses such as HIV disease, antiphospholipid syndrome, alpha-gal syndrome (tick bite-induced red meat allergy), and immunoglobulin 4-related diseases (IgG4-RD) have reset doctors' vocabularies.

Importantly, not all patients' illnesses fall within definitional rules. Physicians once acknowledged uncertainty in medical charts. Rule out diagnoses (R/O MI, rule out myocardial infarction) and vague, symptom-based diagnoses (FUO, fever of unknown origin) were commonly accepted diagnoses in mid-twentieth century medical charts. Assuming they were honestly applied, tests and treatments were reimbursed for all patients, whether or not the diagnosis was certain, until the 1980s. Then, new administrative rigor required doctors to chart diagnoses and based reimbursement policies on diagnosis-related group (DRG) codes. Also, in the 1980s, increasing rigor of patient selection for clinical research and clinical trials led to development of specific, but often conflicting, diagnostic criteria that exclude patients

whose diagnoses are uncertain. More recently, DRG codes have given way to the International Classification of Diseases (ICD) codes, currently 19,000 items. ICD codes can reflect uncertainty or ambiguity, but physicians who code for uncertainty find their requests for tests and treatments denied, doctors now rarely document uncertainty in medical records. Doctors, patients, insurers, clinical and basic scientists, public health administrators, lawyers, published press, media, social media, and public conversation use different diagnosis definitions for different purposes. When they ignore diagnostic uncertainty, they speak of different things.

This book began with an idea expressed by Dr. Michael Lockshin in the 1980s (in press interviews, in talks to medical audiences, and later in books for lay audiences) that contemporary medical discourse poorly serves and systematically ignores patients whose diagnoses are uncertain. The persistence and seeming increasing acceptance of the idea led to a recent workshop, *When a diagnosis has no name: Uncertainty and opportunity*, sponsored by The Barbara Volcker Center at the Hospital for Special Surgery. At the workshop, stakeholders from diverse fields discussed their views and priorities regarding diagnostic uncertainty. A summary paper briefly and without specific participant attribution outlines the workshop's points, conclusions, and recommendations [1].

This book is a product of the workshop. Workshop participants provide here the details of their data, opinions, and priorities, with three goals: to bring forward an open, all stakeholder-based conversation on diagnostic uncertainty; to develop a vocabulary for diagnoses acceptable to all; and thereby to improve science, communication, and patient outcomes. Better understanding of disease pathogenesis is leading to improved molecular classification of human disease, such has been reported in immune-mediated inflammatory diseases [2]. Hopefully, these scientific and clinical advances, paired with better collaboration among physicians, scientists, patients, healthcare regulators, third party payers, and legislators, will build on the framework outlined in this book, to achieve the common goal of improving patients' lives worldwide.

References

1. Lockshin MD, Crow MK, Barbhaiya M. When a diagnosis has no name: uncertainty and opportunity. ACR Open Rheumatol. 2021. https://doi.org/10.1002/acr2.11368.
2. Schett G, McInnes IB, Neurath MF. Reframing immune-mediated inflammatory diseases through signature cytokine hubs. N Engl J Med. 2021;385(7):628–39. https://doi.org/10.1056/NEJMra1909094. PMID: 34379924.

S. Louis Bridges
Physician-in-Chief and Chair of the Department of Medicine and Division of Rheumatology, Hospital for Special Surgery (HSS), New York, NY, USA

Chief of the Division of Rheumatology at New York-Presbyterian/Weill Cornell Medical Center (WCM), New York, NY, USA

Franchellie M. Cadwell Professor of Medicine at HSS, Joseph P. Routh Professor of Rheumatic Diseases in Medicine at WCM, New York, NY, USA

Preface: Introduction—When A Diagnosis Has No Name

Diagnosis, the founding principle of medicine, underlies all aspects of patient care: evaluation, intervention, prognosis, communication, research, and public policy.

In common conversation, a diagnosis name is binary (present or not) and unambiguous; it describes a known pattern of symptoms, laboratory tests, and biological phenomena. When patients' findings do not fit definitions, their diagnoses are uncertain. As a result, they lose access to laboratory tests and treatments; they are excluded from administrative and public health documents and from research studies; they lose dignity in their interactions with physicians, friends, and families.

To bring attention to diagnostic uncertainty, in April 2021, the Barbara Volcker Center at the Hospital for Special Surgery convened a 2-day virtual workshop, "When A Diagnosis Has No Name," that asked experts representing the fields of patient care, basic and clinical medical research, industry, federal regulatory agencies, insurers, medical philosophy, public media, law, hospital managers, and patients to analyze diagnostic uncertainty. The goal was to improve medical care, research, and administration. The workshop deconstructed the concept of diagnostic uncertainty, debated the definitions and purposes of diagnosis, and made recommendations regarding uncertainty.

The workshop asked stakeholders to answer four questions:

- What is your definition of "diagnosis"?
- For what purposes do you use diagnosis names?
- Why are diagnoses uncertain?
- Can we quantify and use uncertainty in patient care, science, and administration?

The workshop participants' responses are the chapters in this book.

The Definition of Diagnosis

Diagnosis names rest on quantitative and qualitative interpretations of published patterns of symptoms, physical findings, and laboratory tests. In different medical fields, there are no specific rules regarding the evidence required to assign a diagnosis name; different stakeholders use different criteria to assign the same diagnosis

name. Existing criteria seldom weigh endogenous or exogenous variables that influence disease expression. In both the patient's course and the science that underlies use of a diagnosis name, definitions of diagnoses change over time. Diagnoses are not binary; they are time-restricted points on an analogue scale.

Broadly defined diagnoses, which include long-term evaluations, new technological methods, and large databases of patients, are important for individual patient care. Narrow, exclusionary, and time-limited diagnoses are needed to perform studies and generalize diagnosis-based information for patient groups.

The Purposes of Diagnosis

Stakeholders use different definitions for different purposes (Fig. 1).

Lawyers, practicing physicians, and patients use *inclusive* biologically based definitions to describe disease mechanisms and to personalize treatments for individuals. For these stakeholders, molecular mechanisms more than clinical phenotypes justify use of a diagnosis name (Table 1).

To minimize variability, researchers use sociologically-based *exclusionary* definitions to study typical but not atypical patients (those who fail to meet diagnostic or classification criteria). Administrators use exclusionary definitions to create policies for populations of patients, classify illnesses, and count patients for public

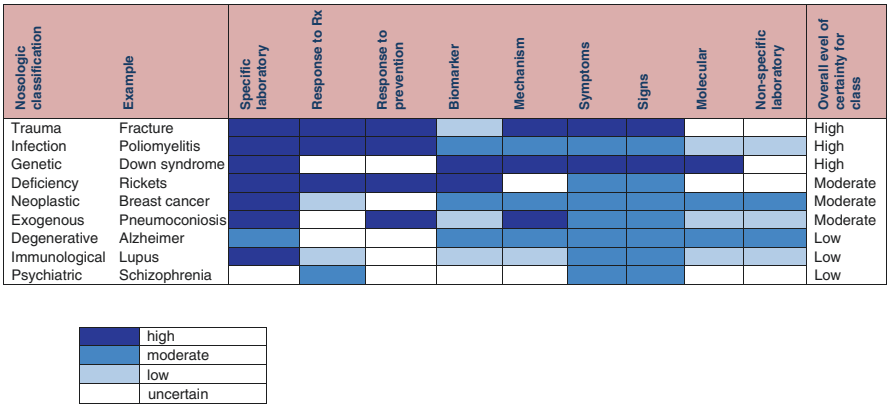

Fig. 1 Estimated levels of certainty for data elements that, summed, result in relative certainty for each nosologic class. Estimates are those of the author; they are not based on systematic study of this topic. They suggest that diagnoses of traumatic, infectious, and genetic nosologies have the highest likelihood of certainty; of deficient, neoplastic, and exogenous nosologies have intermediate likelihood of certainty; and of degenerative, immunological, and psychiatric nosologies have least likelihood of certainty

Table 1 Assumed purposes for which different stakeholders use diagnosis names, listing which types of patients are included or excluded in their counting of patients and the types of questions they ask of the data

		Definition		
Stakeholder	Purpose	Includes	Excludes	Questions
Basic scientist	Standard definition upon which to base studies and rules	Typical	Atypical	Nosology Etiology Mechanism
Clinical scientist	Standard definition upon which to base studies and rules	Typical	Atypical	Categorization for identification, prognostication Intervention to understand mechanisms Treatment trials
Epidemiologist	Standard definition upon which to base studies and rules	Typical and atypical	Overlap, change	Risk identifiers Data for public health analysis Public health policy making
Educator, media, editor	Standard definition upon which to enhance public understanding	Typical and atypical	Overlap, change	Clarity of communication
Administrator	Standard definition upon which to base studies and rules	Typical and atypical	Overlap, change	Funders of science Funders of care Public health planning Hospital administration Social services Legal services
Lawyer	Assign responsibility	None	None	Negotiate definition
Clinician	Create hypothesis upon which to base interventions and billing	Typical and atypical	None	Efficient improvement of patient health
Patient and family	All of the above	Typical and atypical	None	Efficient improvement of health

reimbursement, intervention, research, and public health policies. Most public discourses prioritize sociologic over biologic purposes for using a diagnosis name[1].

To assign diagnoses, physicians evaluate data in nine domains: symptoms, signs, non-specific laboratory tests, specific laboratory tests, response to therapy, prevention, biomarkers, biologic mechanisms, and molecular signatures (Fig. 1). Data in each domain vary in their objectivity or subjectivity. Objective data points like positive blood cultures or biopsies yield diagnoses with high certainty, and those

[1] The experience of the SARS-CoV2/COVID-19 epidemic, in which, even though the etiology was the same, patients varied widely in phenotype and biotype, stimulated many participants to reevaluate the rules by which they make diagnoses and assign them levels of certainty.

with subjective data points like pain yield diagnoses with low certainty. By weighing relative objectivity and subjectivity of the data points, it is possible to rank nine nosologic classes according to likelihood of certainty. Thus trauma, infection, and genetic classes have the highest likelihood of certainty; deficiencies, neoplastic, and exogenous classes have moderate likelihood of certainty; and degenerative, immunologic, and psychiatric classes have least likelihood of certainty.

Workshop participants identified four causes of diagnostic uncertainty:

- *Diagnostic and classification criteria* set *qualitative* rules about who has a diagnosis. (In an administrative sense, patients with otherwise identifiable disease who do not have specified abnormalities do not have that diagnosis.) Criteria set *quantitative* thresholds (a patient whose blood test is abnormal but below a threshold does not have the diagnosis). Because criteria do not offer alternative diagnoses, patients' diagnoses have no names.
- An illness' *concatenation* may not provide criteria-defining information because symptoms did not appear within a criteria-defined time frame or because a future discovery will explain what is not yet understood.
- *Heterogeneity*, in which symptoms occur in atypical order or with atypical manifestations, leads to exclusion of patients from diagnosis definitions without providing alternative diagnosis names.
- Stakeholders select different *times of onset*—susceptibility, trigger factor, first symptoms, first medical consultation, or first fulfillment of criteria—when they use a diagnosis name. In medical offices, physicians use unofficial names such as "pre-diagnosis" or "incomplete diagnosis" to converse among themselves, but patients who do not meet consensus criteria are ineligible for study or benefits available to those who do.

Can We Quantify and Use Uncertainty?

New biological science can provide molecular explanations for disease mechanisms; new computer and statistical methods can reassign diagnosis names for populations; AI mining of large data sets that can compare atypical patients with typical; and natural language processing can scan EMRs. Uncertainty can be quantified, and diagnoses can be assigned probabilities of sensitivity and specificity. Individual and groups of patients can be stratified by probability of diagnoses.

These are the challenges: stakeholders must speak a common language; they must include the role of uncertainty when they make decisions; and they must consider diagnoses to not be binary truths but somewhat ambiguous, time-dependent points on an analogue scale.

Lessons from the Workshop

Diagnostic uncertainty is anomaly that causes both problems and opportunities in patient care, science, and administration.

The workshop identified these problems:

- Uncertainty results when diagnoses do not have standard definitions, when illnesses vary over time, when exogenous factors modulate phenotype, and when narrowly defined biologic processes explain features of illnesses of different nosologic classes.
- Uncertainty impedes prognostication, documentation, interventions, communication, public policy, and social identities for patients.

The Workshop identified these opportunities:

- Communicate more effectively by distinguishing sociologic and biologic definitions of diagnoses (or selecting a common vocabulary)
- Create diagnosis policies, in democratic groups, that include all relevant medical specialties, patients, and public stakeholders
- Quantify diagnostic uncertainty and stratify patient groups by degrees of certainty
- Accept that not all diagnoses are binary, but are points in time in an illness journey
- Simultaneously mine large data sets that contain undiagnosed and excluded patients to validate in the real world hypotheses generated by studies on narrowly defined patients
- Provide to patients whose diagnoses are uncertain equal access to benefits available to others

New York, NY, USA Michael D. Lockshin

Acknowledgments

The Barbara Volcker Center for Women and Rheumatic Diseases sponsored the conference that was the source of this book. Wanda Napolitano, Senior Events & Marketing Coordinator, Professional Education, Hospital for Special Surgery, ably handled the logistics of the conference and the book.

Contents

Contributors

Karen H. Costenbader Harvard Medical School, Division of Rheumatology, Inflammation and Immunity, Brigham and Women's Hospital, Boston, MA, USA

Peter Croft Primary Care Centre Versus Arthritis and Centre for Prognosis Research, School of Medicine, Keele University, Keele, Staffordshire, UK

Pat Croskerry Dalhousie University Medical School, Halifax, NS, Canada

David C. Fajgenbaum Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Joseph J. Fins Division of Medical Ethics, Weill Medical College of Cornell University, Weill Cornell Medical College, New York, NY, USA

Adjunct Faculty, The Rockefeller University, New York, NY, USA

Visiting Professor of Law and Solomon Center Distinguished Scholar in Medicine, Bioethics and the Law, Yale Law School, New Haven, CT, USA

Richard Furie Zucker School of Medicine at Hofstra/Northwell, Division of Rheumatology, Northwell Health, Great Neck, NY, USA

Allan Gibofsky Weill Cornell Medicine, Hospital for Special Surgery, New York, NY, USA

Jerome Groopman Medicine Harvard Medical School, Experimental Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

Paul K. J. Han Division of Cancer Control and Population Sciences, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

Judith A. James Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA

Departments of Medicine and Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA

S. Sam Lim Department of Medicine, Division of Rheumatology, Emory University School of Medicine, Atlanta, GA, USA

Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA

Peter E. Lipsky Ampel BioSolutions and RILITE Research Institute, Charlottesville, VA, USA

Michael D. Lockshin Weill Cornell Medicine, Hospital for Special Surgery, New York, NY, USA

D. Douglas Miller Medical College of Georgia, Augusta, GA, USA

Paola Mina-Osorio Aurinia Pharmaceuticals Inc., Rockville, MD, USA

Lars Noah Levin College of Law, University of Florida, Gainesville, FL, USA

David S. Pisetsky Duke University Medical Center, VA Medical Center, Durham, NC, USA

Jillian Rose Hospital for Special Surgery, New York, NY, USA

Jane E. Salmon Department of Medicine, Hospital for Special Surgery, Weill Cornell Medicine, New York, NY, USA

Andrew Schafer Department of Medicine, Weill Cornell Medical College, New York, NY, USA

New York-Presbyterian Hospital/Weill Cornell, New York, NY, USA

Miriam Solomon Philosophy Department, Center for Bioethics, Urban Health and Policy, Temple University School of Medicine, Philadelphia, PA, USA

Catriona A. Wagner Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA

Vera Wilde Berlin, Germany

Sule Yavuz Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden

Part I

What Is a Diagnosis?

Chasing My Cure: Lessons Learned from My Rare Illness

1

David C. Fajgenbaum

Introduction

As a physician-scientist, rare disease advocate, son of a brain cancer patient, and rare disease (Castleman disease) patient myself, I appreciate the primacy of establishing a diagnosis and the challenges that arise when one cannot be reached. In fact, Castleman disease provides an excellent example of the evolution of a constellation of symptoms towards becoming a named illness without an established definition and diagnostic criteria to one with evidence-based diagnostic criteria and treatment approaches. I present a framework for this evolution for Castleman disease in Table 1.1. Defining a disease includes assigning a name to a constellation of symptoms, appropriately sub-classifying a disease into clinically relevant subgroups, establishing a disease-specific ICD code, and developing diagnostic criteria. Treating a disease includes elucidating underlying mechanisms, performing trials of novel approaches, advancing treatments towards FDA approval, and establishing treatment guidelines.

Before the mid-1950s, patients with Castleman disease would have fallen within our concept of an illness without a name. In 1954, Benjamin Castleman first described these patients with "angiofollicular lymph node hyperplasia" and inflammatory symptoms based on characteristic lymph node histopathological changes that he observed [1]. Dr. Castleman struggled to convince others that what he described in these patients differed from atypical Sjögren syndrome, Mikulicz disease, or lymphoma.[1] By the 1970s, the medical community began to accept the existence of this disease, which enabled identification of additional subtypes and

[1] Lockshin MD, personal communication. Dr. Lockshin was a student under Dr. Castleman during some of these years.

D. C. Fajgenbaum (✉)
Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA
e-mail: davidfa@pennmedicine.upenn.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_1

Table 1.1 Framework for evolution of Castleman disease based on analysis of published papers during the cited eras

	Pre-1950	1950–1969	1970–1989	1990–2009	2010–2022
Defining the disease					
Named disease	No	Yes	Yes	Yes	Yes
Disease sub-classifications	No	No	Poor	Poor	Uniform
Specific ICD code	No	No	No	No	Yes
Diagnostic criteria	No	No	No	No	Yes
Treating the disease					
Mechanistic understanding	Poor	Poor	Poor	Partial	Partial
Trial of novel approach	No	No	No	Yes	Yes
FDA-approved treatment	No	No	No	No	Yes
Treatment guidelines	No	No	No	No	Yes

phenotypes. Flendrig et al. described the plasma cell, the hyaline vascular, and the "intermediate" (or mixed) histopathological variants [2, 3]. Further descriptions over the years provided insight into the broad range of etiologies, presentations, treatments, and outcomes across Castleman disease patients [3, 4]. By the mid-1980s, Castleman disease was divided into unicentric Castleman disease (UCD), which involved a single enlarged lymph node or region of lymph nodes, and multicentric Castleman disease (MCD), which involved multiple enlarged lymph node stations and more severe inflammatory symptoms [5, 6]. Despite progress towards a named illness and clinically relevant subtypes, no diagnostic criteria or established treatments existed. Thus, the disease had been named but the ability to identify patients associated with that disease was still very challenging. Co-occurrence with and overlap between the plasma cell neoplasm polyneuropathy, organomegaly, endocrinopathy, monoclonal plasma cell disorder, skin change (POEMS) syndrome (also known as Takatsuki or Crow-Fukase), and MCD was also noted in the 1980s. Then, in the 1990s, progress was made to identify the etiology underlying MCD in a fraction of patients [7, 8]. The monoclonal plasma cells causing POEMS were proposed to be causing the MCD in these rare cases. Human herpes virus-8 (HHV8) was identified as the etiological driver of HIV-positive and some HIV-negative MCD cases. This discovery led to a highly sensitive diagnostic test for the small fraction of HHV8-associated MCD patients, but the vast majority of UCD and MCD patients continued to have poorly understood mechanisms, no diagnostic tests, no diagnostic criteria, and no treatment guidelines. In the 2010s, Takai et al. recognized a severe form of the most common subtype of MCD called HHV8-negative or idiopathic MCD (iMCD) in which patients had a homogeneous constellation of abnormal laboratory tests and clinical features that he called thrombocytopenia, ascites, reticulin fibrosis, renal dysfunction, and organomegaly (TAFRO) syndrome [9, 10]. It was not until the 2010s when I co-founded the Castleman Disease Collaborative Network (CDCN) that a unified classification system, diagnostic criteria, and treatment guidelines were established for the remaining majority of Castleman disease patients with UCD and iMCD [11]. These steps were also closely linked to the development and approval of the first FDA-approved therapy for any form of Castleman disease and further advances in research and treatment detailed below. Five-year overall

survival of iMCD has improved from approximately 65% in 2012 to approximately 75% in 2021 [12, 13]. Herein I present more details on my diagnostic journey and the critical steps that led to this progress.

My Diagnostic Journey

When I first set out to be a doctor, I had already borne witness to incurable disease and inconsolable sadness—my mother died of brain cancer when I was in college—but I was still optimistic about the power of science and medicine to find answers and cures for all medical challenges. I basically believed that for every problem in the world, there are surely people working diligently—in workshops near and far—to solve it. Or perhaps they've already solved it. That faith has perverse effects, especially in medicine. Believing that nearly all medical questions are already answered means that all you need to do is find a doctor who knows the answers. They can make the diagnosis and identify the treatment. You just need to find them. And as long as you believe these doctors are working diligently on those diseases for which there are not yet answers, there is no incentive for us to try to push forward progress for these diseases when they affect us or our loved ones.

I know better now and I suspect anyone reading this chapter knows better as well. One thing I've learned about doctors is that every one of us who puts on a white coat has a fraught relationship with the concept of authority. Of course, we all train for years and years to have it. We all want it. And we all seek to be the trusted voice in the room when someone else is full of urgent questions such as an unclear diagnosis or an untreatable condition. And the public expects near omniscience from physicians. But at the same time, all of our training and experience instills in us a kind of realism about what is and what is not ultimately possible. Not one of us knows all there is to know. Not even nearly. We may perform masterfully from time to time—and a select few may really be masterful at particular specialties—but by and large we accept our limits. It's not easy. Because beyond those limits are mirages of omnipotence that torture us: a life we could have saved, a cure we could have found. A drug. A diagnosis. A firm answer.

The truth is that no one knows everything, but that's not really the problem. The problem is that, for some things, no one knows anything, nothing is being done to change that, and sometimes medicine can be frankly wrong. The 2021 workshop, "When a diagnosis has no name," focused on these largely neglected diseases and why and how patients with ambiguous diagnoses differ from those with similar, criteria-fulfilling illnesses. During my opening session talk, I shared my personal journey and the lessons I learned along the way.

During my third year of medical school, I presented with a 2-week-long history of constitutional symptoms, multicentric lymphadenopathy, and abdominal pain progressing rapidly to anasarca, organomegaly, thrombocytopenia, eruptive cherry hemangiomatosis, and multi-organ failure. My diagnostic team could not figure out what my diagnosis was, and I continued to deteriorate. I was terrified and shocked that no one could figure out what was ailing me.

Unbeknownst to me, a family friend gave my dad the cellphone number of someone at the National Institutes of Health. My dad didn't know who this doctor was or what he did, but that didn't deter him and he wasn't interested in asking. This wasn't a social call. He had heard this doctor would be helpful, and he wanted answers about his son. My dad called at least once a day, often keeping this busy doctor on the phone for 30 to 45 min at a time, and he would shout into the phone about the latest developments: "Hey, Foochi, I've got more results I want your thoughts on." Then he would rattle off results and questions. I later asked my dad who this "Dr. Foo-chi" was that he had called so much, and he didn't know, so I availed myself of Google and was mortified. It was Dr. Anthony (Tony) Fauci—*the* Dr. Tony Fauci—director of the National Institute of Allergy and Infectious Diseases and one of the most revered physician-scientists in the world. Fauci was a presidential adviser, he'd helped develop George W. Bush's President's Emergency Plan for AIDS Relief—he'd won a *Presidential Medal of Freedom*. My dad had never been one for credentials and certainly didn't care about them now. He would have done anything to help get a diagnosis for his son, even hound a director at the NIH. A lymph node was eventually resected and sent to the Mayo Clinic because the attending pathologist did not recognize the abnormal appearance. The Mayo clinical pathologist recognized the histopathological features and indicated that HHV-8-negative/idiopathic multicentric Castleman disease was likely, but that clinical evaluation and exclusion of overlapping conditions should be performed. Unfortunately, no diagnostic criteria or guidance existed to help the clinical team to confirm the diagnosis. Given the absence of evidence in support of an alternative diagnosis, I was diagnosed with idiopathic multicentric Castleman disease (iMCD).

The only thing known about the mechanisms underlying iMCD was that an inflammatory cytokine called interleukin-6 (IL-6) seemed to be very important. Thankfully, an experimental drug that neutralizes IL-6 was undergoing a clinical trial, and I was granted emergency compassionate use. My diagnosis was a fundamental step for me to receive this potentially lifesaving treatment. Unfortunately, it did not show an immediate effect. Given my critical condition, I was administered my *last rites* and given a combination of chemotherapies in a last-ditch effort to save my life. It worked, for now.

When I learned that there were no treatments in development for iMCD other than the IL-6 inhibitor and that it only works in about one-third of patients, I asked my doctor questions about what caused the disease and how it could be stopped. His answers **"**no one knows**"** made me realize that for some diseases finding the expert doesn't mean that they'll have all the answers. The world's expert only knows as much as the world knows. I dedicated my life to advancing research and treatment for Castleman disease. I would need to begin conducting laboratory research into Castleman disease and create the CDCN to help accelerate progress on a global scale.

When I got out of the hospital, I began looking into the state of research for iMCD. I learned that only about $10 k was invested into iMCD translational research each year. The NIH had never granted funding to iMCD. There were no registries, biobanks, cell lines, or animal models. Two advocacy organizations existed, but there was no infrastructure or plan in place to advance research. The disease was

very poorly understood. No diagnostic criteria of specific ICD-10 code existed. Some treatments were being used off-label such as the chemotherapies that saved my life, but there were no treatment guidelines and no FDA-approved drugs. Though iMCD had a name, it needed a clear definition and diagnostic criteria to serve as a foundation for further research.

Defining Castleman Disease and Advancing the Field

I co-founded the CDCN in 2012 and established a Scientific Advisory Board to guide our scientific direction. In 2013, the CDCN's Scientific Advisory Board prioritized the establishment of evidence-based, patient-guided, expert consensus diagnostic criteria. An international working group comprising 34 pediatric and adult hematopathology, hematology/oncology, rheumatology, immunology, and infectious diseases experts in iMCD and related disorders representing 8 countries on 5 continents, including 2 physicians that are also iMCD patients, was assembled. The CDCN assembled clinical data for 244 iMCD patients as well as 88 lymph node tissue biopsies for histopathologic review. One hundred twenty-eight cases came from a systematic literature review of pathology-based iMCD, where HHV-8 was excluded and individual clinical data were available, 37 cases were submitted by working group members, and 79 were from a randomized controlled study of siltuximab in subjects with symptomatic iMCD (NCT01024036).

An international symposium sponsored by the CDCN and University of Pennsylvania Orphan Disease Center was held on November 20–21, 2015, in Philadelphia, Pennsylvania, with 21 expert participants, and a follow-up meeting was held on December 6, 2015, in Orlando, Florida, with 19 participants. All votes were anonymous and 75% agreement was needed to pass an individual decision. The final criteria vote required 100% consensus. Literature reviews and expert interviews were performed to select a hybrid Delphi method and nominal group technique (NGT) approach to guide criteria development [14]. Clinical and laboratory parameters were chosen for consideration from literature review and expert nomination via the Delphi method in advance of the meetings. NGT was used during the meetings to select parameters through group discussion and secret ballots to achieve consensus. A team of expert hematopathologists examined hematoxylin and eosin-stained lymph node slides from 88 cases with a presumptive diagnosis of iMCD and graded the 5 histopathologic features using a scale of 0–3. The team expanded during the working group meeting to include additional hematopathologists. The group reviewed each case simultaneously at a multihead microscope until a majority of reviewers voted on a grade for each feature. The average grade for each histopathologic feature assigned during review was calculated and compared between subtypes by two-way analysis of variance using a generalized linear model. Three of the 88 submitted pathology cases had insufficient tissue to be fully assessed. At the conclusion of the meetings, the newly established diagnostic criteria were applied separately to cases that met both major criteria from the literature review, submitted cases, and NCT01024036, to evaluate the number of reported minor criteria required

for the case definition. We also calculated response to siltuximab in NCT01024036 based on the number of minor criteria. These diagnostic criteria were published in *Blood* in 2017.

More than 60 years had passed since Benjamin Castleman's first published paper about the disease. Now physicians finally had a checklist to use when considering a diagnosis of iMCD.

This was a critical step in the trajectory of iMCD. You can't treat or save a single patient's life if you can't properly diagnose the disease. Another problem with not having diagnostic criteria is that incorrectly diagnosing people when they don't have Castleman disease sets back research and drug development, because these patients skew the results of studies. As expected, the new criteria have greatly sped up the time to diagnosis for patients and systematized their identification for research.

Beyond establishing diagnostic criteria, the CDCN advanced significant progress for Castleman disease from 2012 to 2021. Specifically, over $1.4 M was invested into iMCD research which led to an additional $10.1 M in funding from external organizations for iMCD research. A registry has over 1000 patients, including over 400 with in-depth clinical histories. A unified physician and patient community, specific ICD-10 code, biobank, and unified research agenda have helped to advance research and understanding of Castleman disease. There are now significant disease awareness, diagnostic criteria, treatment guidelines, one FDA-approved treatment, and two more drugs in development.

Building upon the CDCN's foundation, I personally used proinflammatory cytokine panels, quantitative serum proteomics, flow cytometry, pathway analyses, and immunohistochemistry to identify a pharmacologically targetable disease pathway and began testing a novel treatment approach on myself, which is saving my life [15]. I had the chance to share this journey through a book I wrote called *Chasing My Cure: A Doctor's Race to Turn Hope into Action* [16].

Conclusion

The experience contributing to Castleman disease's evolution from a named but undefined disease towards a named and defined disease has uncovered several important lessons. First, there is incredible power when patients drive research. No one is more motivated or committed to making an impact than patients. Second, all stakeholders must join together. Physicians, researchers, patients, loved ones, advocates, industry officials, and regulators must work together to identify and define diseases and to advance care for all. Third, sometimes solutions can be hiding in plain sight. The drug that is keeping me alive was sitting at my nearby pharmacy for the 3 years that I was in and out of the hospital, but no one thought to use it. We had to define the disease and unlock pathogenesis to identify this new treatment approach. How many more drugs are waiting to be linked to named and unnamed diseases that could be lifesaving? Each of us has the opportunity to contribute to

naming, defining, and advancing treatments. I hope you'll work to identify other key stakeholders within your individual patient networks, within your given disease areas, and across diseases to make an impact. Patients, like me, are waiting.

References

1. Castleman B, et al. Localized mediastinal lymphnode hyperplasia resembling thymoma. Cancer. 1956;9(4):822–30.
2. Flendrig JA, Schillings PHM. Benign giant lymphoma: the clinical signs and symptoms and the morphological aspects. Folia Med Neerl. 1969;12:119–20.
3. Keller AR, et al. Hyaline-vascular and plasma-cell types of giant lymph node hyperplasia of the mediastinum and other locations. Cancer. 1972;29(3):670–83.
4. Gaba AR, et al. Multicentric giant lymph node hyperplasia. Am J Clin Pathol. 1978;69(1):86–90.
5. Frizzera G, et al. A systemic lymphoproliferative disorder with morphologic features of Castleman's disease. Pathological findings in 15 patients. Am J Surg Pathol. 1983;7(3):211–31.
6. Weisenburger DD, et al. Multicentric angiofollicular lymph node hyperplasia: a clinicopathologic study of 16 cases. Hum Pathol. 1985;16(2):162–72.
7. Lachant NA, et al. Multicentric angiofollicular lymph node hyperplasia (Castleman's disease) followed by Kaposi's sarcoma in two homosexual males with the acquired immunodeficiency syndrome (AIDS). Am J Clin Pathol. 1985;83(1):27–33.
8. Soulier J, et al. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. Blood. 1995;86(4):1276–80.
9. Takai K, et al. Thrombocytopenia with mild bone marrow fibrosis accompanied by fever, pleural effusion, ascites and hepatosplenomegaly [in Japanese]. Rinsho Ketsueki. 2010;51(5):320–5.
10. Iwaki N, et al. Clinicopathologic analysis of TAFRO syndrome demonstrates a distinct subtype of HHV-8-negative multicentric Castleman disease. Am J Hematol. 2016;91(2):220–6.
11. Fajgenbaum DC, et al. International, evidence-based consensus diagnostic criteria for HHV-8-negative/ idiopathic multicentric Castleman disease. Blood. 2017;129(12):1646–57.
12. Dispenzieri A, et al. The clinical spectrum of Castleman's disease. Am J Hematol. 2012;87(11):997–1002.
13. Cohen AB, et al. Clinical characteristics, treatment patterns, and overall survival of real-world patients with idiopathic multicentric Castleman disease. J Clin Oncol. 2021;39(suppl 15):7048.
14. Nair R, et al. Methods of formal consensus in classification/diagnostic criteria and guideline development. Semin Arthritis Rheum. 2011;41(2):95–105.
15. Fajgenbaum D, et al. Identifying and targeting pathogenic PI3K/AKT/ mTOR signaling in IL-6 blockade–refractory idiopathic multicentric Castleman disease. JCI. 2019.
16. Fajgenbaum D. Chasing my cure: a Doctor's race to turn Hope into action (Ballantine). 2019.

A Pragmatic Approach to Diagnostic Categorization

2

Miriam Solomon

Introduction

The ancient philosopher Plato expected classificatory systems to "carve nature at its joints," revealing the correct categories into which all objects, organisms, processes, and properties fall. This theoretical ideal of discerning an absolute system of "natural kinds" often guides scientists in many disciplines, including medicine. The International Classification of Diseases (ICD), currently approaching its 11th Edition, can be thought of as aiming toward such a goal. But a more nuanced look at practices of classification reveals different purposes guiding different classificatory systems.

Medical diagnosis serves goals that are more specific and variable than the metaphor of "carving nature at its joints" indicates. Possible goals include explanation of symptoms, determination of appropriate treatment, giving a prognosis, communication of medical information to other healthcare providers, establishing eligibility for clinical trials, application for reimbursement of healthcare expenses, and/or maintenance of public health records. Sometimes the same classification can serve more than one goal, and sometimes the same goal can be achieved with more than one classification. It is simplest to have just one system of classification but finding such a single system might involve compromise among different goals and settling for a pragmatic solution. At times we make those compromises; at other times we use more than one system of classification.

Moreover, our knowledge of human diseases is growing and changing and thus we should expect classification systems to change and improve over time. Any system of classification needs to include plans for incorporating updates. Because there

M. Solomon (✉)
Philosophy Department, Center for Bioethics, Urban Health and Policy, Temple University School of Medicine, Philadelphia, PA, USA
e-mail: msolomon@temple.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_2

are administrative, educational, and research costs to making changes, the scientific bar for proposed changes should be set quite high.

What Are the Goals of Diagnosis?

Diagnosis has a pivotal role in healthcare. Before diagnosis, a patient can only be described and treated symptomatically. After diagnosis, much more is possible: symptoms may be explainable, targeted treatments can be recommended, prognosis can often be given, eligibility for clinical trials can be determined, information about the patient can be transmitted efficiently in referral a specialist, healthcare reimbursement can be provided, and sick leave can be granted.

Because of this pivotal role, diagnosis is critical. Errors in diagnosis can have far-reaching negative consequences. Although a physician is considered to have the knowledge and authority to diagnose, errors in diagnosis are common: the recent National Academy of Medicine Report on Improving Diagnosis in Health Care stated that "diagnostic errors account for 6 to 17 percent of hospital adverse events" [1]. Errors can involve wrong diagnoses, delayed diagnoses, and missed diagnoses.

"Having a diagnosis" (whether correct or not) has become a necessarily prerequisite to being recognized as ill and even a necessary prerequisite to having one's symptoms taken seriously. In this way, diagnosis confers credibility on the patient's complaints and can justify their eligibility for healthcare reimbursement, sick leave, disability status, etc.

Consequences of Diagnosis

Diagnosis *medicalizes* conditions. What this means is that it becomes appropriate to address the patient's condition through the lens of medical observation and/or intervention, rather than as an example of normal human variation or change over time. Medicalization suggests that there is something amiss with the patient's condition, something that we should aim to correct. So, for example, instead of viewing higher blood pressure in older adults as a normal consequence of aging, it is medicalized as "hypertension" and treated in order to reduce the risk of stroke and other complications. Pharmaceutical companies have benefited from the expansion of diagnostic categories and the invention of new diagnostic categories because this produces a greater market for their drugs. (Some examples in which financial incentives were particularly strong are premenstrual dysphoric disorder, social phobia, baldness, and osteopenia.) It follows that it is important to be aware that there may be interested parties playing a role in discussions about diagnostic change.

A diagnosis can sometimes provide people diagnosed with the condition a *social identity.* This can be important when a disease is life-changing because patients and their families can connect with others sharing a particular illness experience and exchange ideas, provide support, and/or engage in activism. Disease categories such as fibromyalgia and chronic fatigue syndrome are good examples of this. Social

identities are usually neutral or positive but can be negative when a condition is stigmatized. The emotional valence of the social identity can affect a patient's willingness to accept a diagnosis.

What Are the Kinds of Diagnostic Categories?

Traditional diagnostic categories identify observed syndromes (defined by symptoms and signs) whose cause(s) is unknown. Examples are chronic fatigue syndrome and schizophrenia. These days, due to greater scientific understanding of some syndromes, many diagnostic categories group diseases by their causes or underlying mechanisms. For example, the category HIV-AIDS identifies illnesses caused by the HIV virus; the category of scurvy identifies the illness of Vitamin C deficiency; the category of acromegaly identifies the results of excessively high levels of growth hormone; and the category of PKU identifies the genetic cause of phenylketonuria. Causes can be microbial, biochemical, physiological, or genetic, just to name some common kinds of causes. Some diagnostic categories describe pathological/cellular findings. For example, diagnosis of non-small cell lung cancer is made from biopsy findings (causes of such pathology may be multiple or unknown), and diagnosis of lymphangiomyomatosis is from high-resolution chest computerized tomography (CT) scans. Occasionally, diagnostic categories are (also) influenced by administrative goals. For example, the diagnostic category "end-stage renal disease," while defined physiologically as a glomerular filtration rate of less than 15, includes diseases with different causes and pathologies, which have kidney failure and the need for dialysis in common. It is an important category administratively because it signals Medicare coverage for dialysis in the USA [2].

Over the history of medicine, traditional syndromic diagnoses have often been replaced and refined with causal, physiological, or pathological diagnoses, as more is learned about the diseases. For example, the diagnosis of dropsy (edema) has been replaced by two different physiological diagnoses, kidney failure and heart failure [2]. Likewise, some pathological diagnoses are being replaced, or supplemented, with biochemical and/or genetic characterizations, such as triple-negative breast cancer and HER-2-positive breast cancer.

How Does Diagnosis Happen?

Diagnosis requires three things: a physician-patient interaction, an appropriate process of reasoning, and a suitable diagnostic category. The physician-patient interaction involves communication of symptoms from the patient to the physician, detection of signs in clinical examination, and sometimes test results. The process of reasoning is that of differential diagnosis, in which the physician considers possible diagnoses and uses symptoms, signs, and test results to narrow down the diagnosis. A suitable diagnostic category is one that satisfies the goals of diagnosis, as explored in this paper.

What Are the Consequences of Not Having a Suitable Diagnostic Category?

There are multiple consequences for patients, physicians, medical research, and public health when an illness does not fall under a suitable diagnostic category. Without a diagnosis, the illness does not fall under an established category of disease for which there are treatment guidelines. Treatment can only be symptomatic and/or experimental. Similarly, without a diagnosis, the illness does not fall under an established category of disease for which we have prognostic information. Both treatment and prognosis are accompanied by uncertainty.

In addition, the absence of a diagnosis usually means that there is no explanation of the patient's symptoms and signs, which can be frustrating for both patients and physicians. Especially when symptoms and signs are mostly subjective (e.g., pain, exhaustion), the patient's complaints may not be taken seriously by the healthcare profession or by employers or families. It may even be difficult for a patient to obtain healthcare coverage without a diagnosis ("no name, no claim"). The patient may be unnecessarily isolated and unable to find others with similar conditions for social and practical support.

Patients without diagnoses are rarely included in clinical research trials or public health records. It can be difficult for healthcare practitioners to communicate to one another about a patient without a diagnosis, since all the details need to be communicated for a full picture; it is much simpler to communicate a diagnostic category.

Factors Contributing to Diagnostic Difficulties

Diagnostic difficulties can arise in a number of circumstances. Different diseases with common symptoms may be difficult to distinguish from one another. Diseases that change in character over time may not permit diagnostic stability. The lack of a "gold standard" test for the presence of a disease also leads to diagnostic uncertainty. In general, diseases with *complex and heterogeneous causes and manifestations* (such as cancers, autoimmune diseases, psychiatric diseases, and amyloid diseases), especially those that *develop and change over time*, can be difficult to understand well enough to conceptualize and distinguish from one another and from other diseases. In such cases, using broader and less specific categories may be a good strategy, although when a category gets too nonspecific it becomes less useful for research purposes. On the other hand, when a category is too specific—say, it singles out one of several possible processes leading to symptoms—it may make recruitment for clinical trials difficult as well as limiting the patient pool to which the research is relevant. Moreover, making a category too specific risks getting it wrong through focus on a specific hypothesized mechanism that may not turn out to play a significant role. It can make sense to work with more than one diagnostic system, as we do with, e.g., discussion of cancers in which we make use of both the traditional tumor-node-metastasis staging classification and classifications in terms of particular biochemical markers.

Often, diagnostic criteria for inclusion in research studies are stricter than diagnostic criteria for clinical use. This means that two different sets of diagnostic criteria are in use. This can lead to translational difficulties. It is better to have the same criteria—or at least easily translated criteria—for research and clinical contexts.

Diagnostic categories ending in "not otherwise specified" ("NOS") are ubiquitous for many kinds of disease, such as "peripheral T-cell lymphoma, not otherwise specified." They can be used when there is lack of information for a more precise classification, or they can be used when the information about the patient's disease does not fit with any existing diagnostic category. The "NOS" classification is helpful administratively, ensuring healthcare reimbursement and inclusion in public health statistics. However, if too many cases fall under "NOS," this is an indication that current diagnostic categories are inadequate.

A different kind of diagnostic difficulty occurs when a patient rejects a physician's diagnosis. Then the diagnosis becomes a point of tension between physician and patient. Patients are in a difficult epistemic position when they disagree with their physician about a diagnosis. Their best option is to ask for a second opinion. If this second opinion differs from the first, then there is expert disagreement about a diagnosis, which of course leads to uncertainty about treatment and prognosis.

General Recommendations for Diagnostic Change

Successfully proposing new diagnostic categories, or successfully making changes to existing diagnostic categories, needs both authoritative suggestions and uptake of these suggestions by stakeholders. These days, this is generally achieved through a high-profile evidence-informed consensus process, in which recognized experts and stakeholders are brought together to deliberate and seek consensus on a reasonable recommendation. Often, relevant evidence is incorporated through a systematic evidence review that is conducted in advance and distributed to all participants. Participants in the consensus process are chosen for the expertise that they bring, and needed expertise includes specialist research expertise, specialist clinical expertise, and expertise in patient experience (patients and their representatives). For the broadest uptake, there should be the broadest representation of perspectives among the participants, consistent with keeping the group a manageable size.

Consensus processes, like all group decision processes, should be managed democratically and allow dissenting views to be expressed. They should take the time to consider all points of view together with the available evidence. Traditionally, consensus conferences take 2–3 days and are composed of 8–20 experts who meet in person and aim to come to agreement on several specific questions. (For more details about medical consensus conferences, see [3].) These pandemic days, virtual meetings have taken the place of in-person meetings; time will tell whether they can work as well.

In general, diagnostic categories should be broad enough that they address a significant population yet narrow enough to generate projectible findings. If a diagnostic category is too narrow, it will be difficult to gather enough data, and findings

will not apply widely enough to be worthwhile. If a diagnostic category is too broad, it will be easy to gather enough data, but there will be little in the way of generalizations about prognosis and treatment.

For coordination across communities and countries, it is best if the diagnostic system is shared as widely as possible. Thus, a diagnostic system is not really the place for a new and controversial theoretical framework because it will not receive general uptake and may not last long enough to be worth the effort of trying to persuade the clinical and research communities to use it. Diagnostic systems tend to be a little more conservative than the leading edge of research, and this is appropriate.

Given the imperfect state of the science and the multiple goals of diagnosis, it may be difficult to settle on one best diagnostic system. It is reasonable to *satisfice* instead and to choose one of several possible comparably good diagnostic systems. In some cases, it may be reasonable to use more than one diagnostic system, when no one system captures all that is diagnostically relevant. For example, it is reasonable to use both the traditional tumor-node-metastasis framework and biochemical marker classifications for some cancers.

Examples

A couple of historical examples illustrate some of the points in this essay.

Alzheimer's disease was at first (when described by Emil Kraepelin in 1910) a category reserved for early onset dementia. The discovery of similar pathologies in early onset and later onset dementia led to the extension of the category to all dementias with similar pathology (amyloid plaques, tangled tau proteins, and shrinking of brain tissues) by the end of the twentieth century. Stages of Alzheimer's disease were, and still are, demarcated in terms of symptoms as pre-clinical, mild cognitive impairment, early, middle, and late stages. The discovery that other kinds of pathology can lead to similar symptoms has led to efforts to distinguish Alzheimer's disease from other neurodegenerative diseases. Most recently, there have been attempts to integrate biomarkers (such as findings in cerebrospinal fluid) and brain imaging studies (PET scans) to improve the accuracy of diagnosis of Alzheimer's disease and identify the disease before it manifests clinically [4].

Disruptive mood dysregulation disorder (DMDD) is a new category in DSM 5.0 (2013), introduced primarily to reduce iatrogenic harm. It takes the place of pediatric bipolar disorder, a diagnosis which was found to be both scientifically deficient and harmful to patients. The harm was an epidemic of over-prescription of atypical antipsychotics in juveniles, producing severe metabolic side effects. DMDD is a provisional category, introduced without much understanding or characterization of the phenomenon, designed in order to signal that these juvenile behavioral phenomena are not related to bipolar disorder or even psychosis. The normal standards for introducing a new diagnostic category in DSM were relaxed in order to prevent more harm to patients [5].

Conclusions

Diagnosis is pivotal in healthcare because so much follows from it. It needs to satisfy clinical, research, administrative, communicative, educational, explanatory, and psychological goals. I have suggested a pragmatic approach to selecting a classificatory system that keeps all the goals in mind.

References

1. National academy of medicine report on improving diagnosis in health care; 2015. see https://www.ncbi.nlm.nih.gov/books/NBK338596/.
2. Peitzman S. From dropsy to Bright's disease to end-stage renal disease. Milbank Quart. 1989;67(Suppl 1):16–32. PMID: 2682170.
3. Solomon M. Making medical knowledge. Oxford University Press; 2015.
4. Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, DeKosky ST, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol. 2014;13(6):614–29.
5. Solomon M, Kendler K. The problem of aggregating validators for psychiatric disorders. J Nerv Ment Dis. 2021;209(1):9–12.

How Diagnoses Are Assigned

3

Pat Croskerry

A critical feature of cognitive function is our ability to appreciate features and similarities of elements of the external world around us (objects, life-forms, events, concepts, ideas) and categorise them into types, groups, or classes on the basis of shared features and characteristics. It is a process that brings order and meaning to our perception and is essential for survival. Notably, the mid-eighteenth-century Swedish naturalist, Carl Linnaeus, established an orderly way of understanding the biology of our environment by creating a systematic approach towards classification of its essential elements: animal, vegetable and mineral kingdoms. This systemising capacity enables us to analyse and understand patterns in our respective environments. Without it our world would be chaos.

Classification in Medicine

Aspects of health and disease were originally attributed to the gods with individual gods associated with different features: Asclepius was the god of medicine and healing; Iaso the goddess of cures, remedies and practices of healing; Panacea – the goddess of universal remedy; and so on. In ancient Greece, Hippocratic physicians posited that the body was a shell containing four humours: blood, phlegm, black bile and yellow bile. This notion proved to be surprisingly long-lived, appearing as late as the sixteenth century in the works of Shakespeare, and some residue of this classification persists in the current descriptors of demeanour – phlegmatic and sanguine.

As more interest and understanding of anatomy developed, disease began to be connected to specific physical aspects of the human body, and by the fifteenth century lists of discrete disease entities appeared, notably with the *London Bills of*

P. Croskerry (✉)
Dalhousie University Medical School, Halifax, NS, Canada

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_3

Mortality, John Graunt's pioneer study of medical statistics and demography in 1662, attributing causes of death of 121 males and 111 females to such things as consumption, drowned, griping the guts, suddenly, and worms; further refinements followed. Boissier de Sauvages's first comprehensive and systematic classification of diseases, *Nosologia Methodica* (1763), listed 10 major classes of diseases, broken down into 44 orders, 315 genera and 2400 species. In 1785, William Cullen of Edinburgh produced a classification *Synopsis nosologiae methodicae* which was adopted into general use. At an International Statistics Congress in London in 1860, Florence Nightingale called for a further classification of diseases based on hospital morbidity statistics, and in 1893 the statistician Bertillon published the first international classification of diseases *Classification of Causes of Death* containing 44 entries. The beginnings of an ICD were born and came with a recommendation that it be revised every 10 years. The French government was responsible for revisions 1 to 5 with the fourth revision appearing in 1928 and the fifth in 1938. The sixth revision in 1948 was titled the *International Statistical Classification of Diseases, Injuries and Causes of Death* and the World Health Organization henceforth assumed responsibility for preparing and publishing the revisions from this edition onwards. The current ICD 11 was published in 2018. It contained about 19,000 diseases. A detailed history of the development of disease classification and of the ICD is available [1, 2].

Disease, Illness and Sickness

Most patients have the expectation that their doctor will be able to diagnose their complaint from their symptoms and signs. The doctor's frame of reference is based on an accepted classification of the various known diagnoses. Efforts at classification of diagnoses have been directed towards identifying disease, but not necessarily illness or sickness. Cassell made the distinction that illness is something the patient has, whereas disease is something that organs have [3]. More often than not illness will be associated with disease, but not necessarily. Heart attacks and deadly cancers may be 'silent' with no apparent manifestation of either disease or illness. As well, the quantitative definition of disease may change over time, e.g. national guidelines on hypertension. Conversely, illness symptoms may be highly apparent but with no measurable organ changes. The third term 'sickness' describes more a social or societal identity which is adopted by the individual or may be defined by others in relation to the behaviour of the individual, for example, where the patient has been diagnosed with a disease but does not feel ill, or changes in behaviour arise from a sick role within the family context or with regard to health insurance, compensation [4] or other circumstances. Of the three concepts, disease seems to be the most 'anatomicopathological' lending itself to objective assessment and scientific methods that are expected to reveal an ultimate cause and would be most favoured by medical minds.

Pattern Recognition

Pattern recognition is a critical feature of perception upon which all animal species depend for survival. Dogs have an exceptionally advanced sense of smell with a 40-fold greater number of olfactory receptors than humans, while bird's vision is about five-fold better than that of humans. Catfish are said to have one of the most advanced skills in taste and touch, and moths the best sense of hearing. Humans enjoy an overall high level of superior pattern processing (SPP) capability that is fundamental to many of the unique features of human cognitive performance, such as reasoning, language, intelligence, invention, and imagination [5], and may involve all five senses. One of Hippocrates' criteria for diagnosing diabetes was to taste the patient's urine. 'No other organ system or organ of the human body provides so much information by its excretion as does the urinary system', he wrote. While taste is used sparingly in modern medicine, clinicians often refer to their 'gut' feeling, an amorphous sixth sense of intuition. Intuition is a widely used term with varying meanings [6] but in this context suggests some ill-defined, subcWonscious awareness, suspicion or hunch that may contribute to the diagnostic process.

Systematising of patterns is seen as the most basic and essential of human skills. It is a drive towards analysing and understanding systems, governed by rules and which operate according to logic and scientific laws. Essentially, systematising allows us to predict how systems will behave [7]. Thus, pattern seeking is a starting point for systematising and understanding ill-health. The assignment of any disease in medicine begins with a pattern recognition process that leads to a systematising process based on symptoms the patient may be relating, signs they may be showing or information from other sources.

Development of Clinical Expertise

Most animals depend exclusively on their genes for pattern recognition. Weaver birds select specific materials to build complex nests in particular locations. All of the patterns essential to the process, nest material, species of tree and shape and design of nest are accomplished with no prior experience, so it must be encoded in their DNA. In contrast, while some human behaviours have similar genetic determinants, most pattern recognition is acquired through learning. Children by age 2–4, for example, can recite the alphabet – a sequence of 26 unrelated letters. Like most learning it is accomplished through a process of association where particular sensory patterns are associated with and come to trigger specific responses. In the course of medical education, medical students begin with a scaffold of foundational knowledge (anatomy, physiology, pathophysiology) upon which causal networks are developed. These networks explain the causes and consequences of disease and become encapsulated into diagnostic knowledge such that disease prototypes may be developed in which the signs and symptoms of a particular disease may be recognised. Pathognomonic findings such as the rash of herpes zoster are highly representative patterns.

Other labels for encapsulated knowledge are syndromes, stigmata and other characteristics, all of which are reinforced through repetition in training and further experience. The next stage is a reorganisation of encapsulated knowledge to form illness scripts. These contain important, clinically relevant information about the enabling conditions for disease. Learners match the scripts to the patient presentation, and instantiation occurs in the course of verification of the script, i.e. a patient's illness may be accepted or rejected as a valid instance of the illness script that characterises a particular disease [8, 9]. One of the problems inherent in traditional medical training is the emphasis placed on prototypical presentations with insufficient attention paid to atypicality and the associated uncertainty that goes with it.

The Diagnostic Process

Making a diagnosis often appears deceptively simple and straightforward but actually is quite complex [10]. Much of the complexity occurs in the early phase at the point when the patient engages with the healthcare system and the process of information gathering, integration and interpretation occurs [11]. The diagnostic process begins with a person experiencing a health problem that may be already known or new. Through self-referral or other means, they may engage with the healthcare system and become a patient. Members of the healthcare team gather, integrate and interpret information through a clinical history and interview and a physical exam. A working diagnosis is usually formulated and diagnostic testing is done as needed – bloodwork, urinalysis, imaging, EKG and other tests. Referral and consultation with other specialists may occur to refine and develop the diagnosis which is then communicated to the patient. A planned path of care is established between the providers and the patient. Outcomes of the treatment are monitored and treatment revised as necessary. The process has been examined in detail [12].

Over 50 factors have been identified at this stage that cluster into 6 major groups that affect the diagnostic process [10]. Significant overlap and interactions occur between factors from different clusters.

A. *Characteristics of the individual clinician:* Intellect, knowledge, gender, ethnicity, experience, age, culture and religion
B. *Properties of their personality, intelligence and cognition:* Active, open-minded thinking, personality, logicality, critical thinking, rationality, metacognition, experientiality, perseverance, mindfulness, reflection, reflective coping, lateral thinking, adaptiveness and need for cognition
C. *Ambient work conditions:* Hunger, sleep deprivation, thirst, stress, cognitive load, fatigue, affective state, sleep debt
D. *Workplace features:* System design, communication, ergonomic factors, resource allocation, information technology, scheduling and team factors
E. *Features of the disease process:* Symptoms, signs, context, familiarity, onset, progression, prototypicality, pathognomonicity, manifestness, comorbidities and mimics
F. *Features of the patient:* Patient, family, caregivers, culture, context, friends, other patients

It is estimated that the information gathered from the history and physical examination alone will yield the correct diagnosis – about 88% of the time [13]. Pattern recognition, mostly through auditory and visual cues, is key at this stage. One obvious heuristic appears to prevail at this point: 'common things are common' and 'when you hear hoofbeats, think of horses, not zebras'. As the experience of the clinician increases, speed and accuracy will generally increase too.

Much will depend on how familiar clinicians are with the disease, its presentation and how it evolves, i.e. how manifest, representative, prototypical and pathognomonic it is (Cluster E). *Manifestness* is an important property of the disease at the outset. It refers to what is obvious or clearly apparent to an observer, or something revealed in an obvious manner. Thus, an injury that results in a dinner-fork deformity of the wrist manifests a diagnosis of Colles fracture. *Prototypical* refers to symptoms and/or signs that are expected for a particular disease; thus, chest pain or discomfort; light-headedness; nausea or vomiting; jaw, neck, back, arm or shoulder pain or discomfort; and shortness of breath are collectively prototypical of an acute coronary syndrome, possibly presaging myocardial infarction. *Pathognomonic* refers to specific or characteristic distinctive signs where a particular disease is present beyond any reasonable doubt. Pathognomonic signs are the acme of SPP. *Progression* refers to how rapidly different features of the disease concatenate, evolve and link together as the diagnostic process unravels. Fast and coherent linkages make for strong convictions at the outset but the process may be less compelling (and more frustrating for the patient) when it is drawn out over time.

The Manifest Continuum

A major issue with current estimates of around 19,000 diseases is that there simply aren't enough symptoms to go around, at least in any one-to-one way, so many diseases will share the same symptoms. Chest pain generates about 25 possibilities and headache about 300. Approximately 150–200 common signs and symptoms have been described [14]. In his handbook of common complaints, Ely suggests that 99% of diagnostically challenging symptoms can be covered by a short list of 63 common complaints [15]. Further, in some instances just a few symptoms may be suggestive of a disease, while in others disease may occur with none. Several studies suggest that 20–40% of all heart attacks are silent and unrecognised at the time they occur, probably more in elderly women [16]. In other instances the disease may present atypically. Chest pain, considered by doctors and the lay public alike as the cardinal symptom of a cardiovascular problem, may be absent in over a third of cases [17, 18]. Another issue is that signs and symptoms are not static and may evolve over time. A 'tincture of time' strategy may well be a felicitous strategy in the earlier stages of the process.

In other clusters of symptoms, no specific disease has been identified, suggesting several possibilities: Firstly, the disease may simply not yet have been discovered. New genomic techniques such as exome sequencing have been extremely helpful in identifying new diseases, about 200 of which are added each

year. Secondly, it may be a rare disease (in the USA defined as affecting less than 200,000 people) and goes unrecognised, except perhaps in academic or specialty centres. Or, the disease may be sufficiently common but cannot be categorised within existing classification schemes, i.e. with the usual explanations being excluded, the illness has no name. For a variety of reasons, it is important to attach a name or label to this group. It increases the visibility of the illness, patients feel better with the knowledge that their condition is known and recognised, the discomfort that some physicians may have in dealing with uncertainty may be diminished, and the likelihood of an effective treatment may be increased. A consensus is emerging to label the patients having the illness as those with persistent physical signs (PPS).

The key issue here is what the disease or illness is doing. Absent, incomplete or overlooked symptoms and signs will lead to the diagnostic process being compromised to some extent and will increase the likelihood of misdiagnosis. We can visualise a spectrum of illness manifestness [19], ranging from high signal/noise ratios, e.g. pathognomonic presentations, to low signal/noise ratios, e.g. persisting symptoms without physical findings, where at one end reside the exclusive pathognomonic few, while at the other are a variety of less well-defined conditions (Table 3.1). Whereas recognisable patterns of pathognomonic disease have a high signal-to-noise ratio, with the signal being easily distinguished from competing noise and readily detected, illnesses with no name have a low signal-to-noise ratio and are difficult to identify [20]. Not surprisingly, the manifestness of disease bears a close relationship to the difficulty with which it is diagnosed. Knowing that laboratory investigations lead to the diagnosis of about 10% of the time [13], it is also not surprising that the less manifest the disease, the more testing is likely as clinicians cast an ever-widening net in attempts to associate the symptoms with an already known disease. This is complicated further by an increasing likelihood that, as more testing is done, more results will be outside the normal range, as well as increasing numbers of false positives and false negatives.

Table 3.1 General terms used and examples of illnesses without a name

General terms	Examples
Somatic symptom disorder	Chronic fatigue syndrome
Medically unexplained symptoms	Fibromyalgia
Bodily distress disorder	Irritable bowel syndrome
Complex physical symptoms	Tension headache
Non-specific somatoform functional syndrome	Chronic widespread pain
Functional somatic disorder/syndrome	Temporomandibular disorder
Functional symptoms	Tension headache
Persistent physical symptoms	Pelvic pain
Cryptogenic syndrome	Multiple chemical sensitivity
Supratentorial	Non-cardiac chest pain
Idiopathic	Atypical facial pain
Non-organic	Chronic insomnia
Culture specific syndrome	Chronic dyspepsia
Complex physical symptoms	

When the Illness Has No Name

Many people experience unexplained symptoms of illness. Often, these will resolve with no consequence but others will persist to challenge both the patient and the diagnostician. If no medical explanation is found by the treating physician or other healthcare providers, or if the cause remains contested, they may be classified as PPS or by a variety of other overlapping terms (Table 3.1). Examples of conditions that fall into these categories are also given in Table 3.1. Estimates of prevalence vary but may account for 15–30% of all primary care consultations [21] and are assigned more frequently to females [22]. This is not an insignificant problem. In 2008–2009, the National Health Service (NHS) in England estimated it spent about 10% of its total budget (£3 billion) for the working age population dealing with the diagnosis and treatment of PPS [23].

The diagnostic process may be confounded in other ways when the disease is simulated by the patient, such as in factitious disorders or in malingering. Further complications occur with disease-mongering – when diseases are invented or promoted in order to profit from their treatment [24]. The obvious beneficiaries here being pharmaceutical companies, alternate practitioners and others, including physicians and to some extent the patients themselves. The sceptical view is that, in effect, this strategy *provides a name where there is no illness.* An essential feature of this practice involves shifting the boundaries between normality and abnormality. Thus, aspects of behaviour towards the extremes of a normal distribution may be seen as abnormal, and a degree of discomfort or suffering may be imputed that is exaggerated, i.e. advertising a common symptom as a serious disease. Often the number of people alleged to have the disease is inflated. Some examples suggested to be products of disease mongering are restless leg syndrome, irritable bowel syndrome, female sexual dysfunction, testosterone deficiency, erectile dysfunction, hypoactive sexual desire disorder, attention-deficit/hyperactivity disorder, premenstrual dysphoric disorder, social anxiety disorder and others. The issue is not whether these entities exist or not within the normally distributed spectrum of human behaviour, but that they are abnormal and require pharmacological treatment.

The aetiology of PPS and other similar collections of symptoms is complex. The underlying disease, where it exists, typically defies classification according to traditional diagnostic methods. A default bias is often evident in that poorly understood illnesses have a tendency to be seen as having a psychiatric origin pending the development of a better understanding of their pathophysiology, or another explanation emerges [25]. A common theme appears to be an association with psychological issues, perhaps the later repercussions of childhood trauma, or somehow connected to psychological traits of anxiety and/or depression, especially in patients inclined to somatise. There appears to be a major overlap with psychological stress or disturbance. Those suffering from anxiety or depression and/or who focus excessively on their body might be particularly prone to these symptoms [26]. This appears to be the case with non-cardiac chest pain, where there may be no apparent organic pathology to explain cardiac-like symptoms [27]. Thus, psychosomatic (somatic illness caused or exacerbated by mental stress and distress) as well as

somatopsychic aetiologies (mental disorders caused or exacerbated by somatic disorders) may be involved, with possible interplay between the two. It is of interest that while the list of purely psychosomatic illnesses appears to be shrinking, that of somatopsychic disorders is expanding [25].

Occasionally, powerful undertones of delusion are apparent or have been inferred. For example, Koro is an acute state of anxiety lasting up to days or longer, associated with the mistaken belief that the genital organs are shrinking and/or retracting, in the absence of any measurable changes [28]. It is classified as a culture-bound delusional syndrome. In contrast, another, Morgellons disease (MD), is the belief that strange multi-coloured fibres are emerging from skin lesions along with a variety of other symptoms and signs. A major study of 115 cases concluded the diagnosis was likely a 'delusional infestation' [29], and a French paper referred to it as a 'disease transmitted by the media'. [30] However, more recent studies have discovered that the condition is actually due to a Borrelia spirochaetal infection, similar to the bacteria associated with Lyme disease and syphilis. The authors cautioned against involving mental health status in the diagnosis [31]. It has since been staged in detail [32].

The Problem with Assigning a Label

To extend an earlier point, there appear to be some advantages in assigning an illness with no name to a particular category such as PPS. It allows the illness to be treated as a specific entity and be explicitly included on a differential diagnosis list, with no specific attribution, and indicates a status of 'not-yet-diagnosed'. It may also remove stigmatisation by distinguishing it from psychiatric disorders. As has been noted [24], a bias has prevailed historically to classify poorly understood illness as having a psychiatric basis and accompanying psychiatric diagnostic label. Creed et al. [33] have proposed various criteria by which a suitable term might be evaluated. They suggest that the term:

A. Is acceptable to patients
B. Is acceptable and usable by doctors and other healthcare professionals, making it likely that they will use it in daily practice
C. Does not reinforce unhelpful dualistic thinking
D. Can be used readily in patients who also have pathologically established disease
E. Can be adequate as a stand-alone diagnosis
F. Has a clear core theoretical concept
G. Will facilitate the possibility of multidisciplinary (medical and psychological) treatment
H. Has similar meaning in different cultures
I. Is neutral with regard to aetiology and pathology
J. Has a satisfactory acronym

In a later study of a lay sample of 844 healthy adults in the UK about their preferences for terminology, the most popular term was PPS [34].

However, the downside is that any labelling of patients can give a false sense of comfort and understanding, as well as constraining further thinking. The behaviour of certain patients sometimes attracts undesirable epithets such as 'drug-seeking' and 'frequent flyer' which may lead to a limited assessment. Labels such as PPS are less pejorative but are still labels which may have the effect of limiting the diagnostic process, at times dangerously (Box 3.1).

Box 3.1 Assigning Labels to Patients (Adapted from Croskerry [35])

A 28-year-old presents to the emergency department (ED) with lower abdominal pain. She has had numerous visits to the ED over the last few years with similar complaints and is known to have visited all four EDs in the city in the same evening. She has been described as a 'frequent flyer' but not believed to be drug-seeking. She has been referred for assessment to gynaecology, urology, gastroenterology and psychiatry. Despite numerous investigations including ultrasound, abdominal CT and hysterosalpingogram, no organic cause has been found for her discomfort. In a note circulated to staff, the head of the ED has described her as having a somatoform disorder. Nevertheless, the staff continue to refer to her as a 'frequent flyer'.

The ED is extremely busy. The emergency physician (EP) who assesses her has seen her several times in the past. Her abdominal discomfort is mostly on the left side. Her vital signs are stable. She has a soft abdomen with good bowel sounds and vague tenderness in the left lower quadrant. Her urinalysis is normal.

The EP reassures her that she doesn't have a urinary tract infection and does not appear to have any serious condition. He comments on the extensive work-ups she has had in the past for similar symptoms. She tells him that she feels different this time, but he reminds her that she has said that numerous times before. She responds well to further reassurance and he discharges her from the ED.

The following afternoon, she is brought back to the same ED having collapsed at the local mall. She is pale and hypotensive. Bloodwork shows a hemoglobin of 6 and a positive pregnancy test. Pelvic ultrasound revealed a complex adnexal mass on the left side with a large amount of free fluid. She is diagnosed with a ruptured ectopic pregnancy and taken immediately to the operating room. She made an uneventful recovery.

Recommendations for Management

The setting for optimal management of patients with persistent physical symptoms should be one in which the clinician is not operating under significant time pressure, fatigue, stress or cognitive loading, all of which may diminish frontal lobe function

and lead to the increased use of heuristics. When possible, extra time should be set aside for a planned visit of a patient known to have this condition, to allow for a more analytical approach. As the maxim holds 'short-cuts are often the long way home'.

A. Reasonable efforts should be made to exclude other competing, plausible diagnoses or explanations. Importantly, diagnoses should not be made until sufficient information has been unpacked. Psychiatric possibilities should be considered but care should be taken they do not lead to premature diagnostic closure. This cannot be stressed enough given that it took over 400 years for Morgellons disease to be correctly diagnosed.
B. Given that common diseases may present atypically, or even with none of their usual symptoms and signs, the differential diagnosis should be broader than usual, and active consideration given to more remote possibilities. Thus, the differential diagnosis should be extended to include more rare possibilities as well as atypical presentations. Unless the pre-test possibility is very low, this needs to be more than simply passive consideration.
C. Clinicians need to be aware of and recognise the entity of illnesses which have no name. Many practising clinicians may not be aware of it nor its prevalence. There should be an option to explicitly include it on the differential diagnosis.
D. Consideration should be given towards enhancing the usual strategies for clinical decision making. While 'routine expertise' will normally allow approximately 85% of usual diagnoses to be made correctly, it may be necessary to augment the routine process using an 'adaptive expertise' approach that employs additional strategies [10]. One in particular, lateral thinking, achieves greater flexibility, innovation and creativity. It involves a conscious effort to detach oneself from orthodox, vertical stepwise reasoning and conceive new approaches to the problem-solving task at hand [36]. This is important because medical training often focuses on prototypical, textbook presentations of disease, whereas in clinical practice non-specific, atypical presentations may occur and are more likely to be missed, especially in the elderly [37]. Complex clinical problems where the usual clues to what is going on are obfuscated may require a different approach.

   Another lateral thinking strategy involves 'provocation and movement'. Provocation involves taking a provocative statement such as 'thinking the opposite'. Playing the devil's advocate establishes a similar but more general contrary position that promotes further debate about an issue and may mitigate premature closure. Such provocations are said to have 'movement value' in that they encourage movement away from a known or conventional idea toward novel ideas that might otherwise not receive consideration and may be useful, for example, with atypical presentations. Lateral thinking mitigates the cognitive bias *vertical line failure* characterised as predictable, orthodox styles that emphasise economy, efficacy and utility but which carry the inherent penalty of inflexibility [38]. Vertical thinking works well in simple situations; however, even when things look simple and straightforward, it is important to remain

open-minded and simply ask the question 'What else might this be?' It provides an opportunity to break out of vertical thinking silos and move laterally towards other potential solutions. De Bono described lateral thinking as being particularly appropriate for asymmetric patterns – those that are not predictable, often non-specific and atypical. Again, such presentations are common in medicine and are especially problematic in the elderly. Lateral thinking is an antidote to inflexibility and can be mobilised against the cognitive miser function, the natural tendency of the brain to minimise cognitive effort and engage the more error-prone intuitive mode of decision making. Kahneman's WYSIATI (what you see is all there is) captures this idea well [39].

E. Psychological sources of somatic complaints should be considered, which may require some additional psychological training in general practitioners [22]. Does the patient focus excessively on bodily complaints? Is there extreme anxiety over symptoms, frequent assessment of the body for signs of abnormality or preoccupation with thoughts that even mild symptoms might be suggesting serious disease, or is the patient repeatedly visiting healthcare professionals for evaluation and failing to be reassured despite often exhaustive testing, or does the patient believe that health professionals are not taking them seriously? These characteristics may be associated with certain personality types. Consideration might be given to administering a brief assessment such as the eight-item Somatic Symptom Scale (SSS-8) [40]. However, practitioners should always keep in mind that some serious medical conditions (e.g. pulmonary embolus) may be associated with hyperadrenergic stimulation.

F. Finally, while all specialties generally attempt to understand disease and illness in terms of an alteration of the structure-function relationships within the organ of focus, such that hepatologists correlate clinical observations of hepatic failure with changes at a cellular level in the liver, and nephrologists relate changes in urinary symptoms with nephron function, there may be no such recourse for illnesses that have no name. The brain is undoubtedly involved but it is a highly complex organ with great structural and functional diversity. Its degree of complexity far exceeds that of any other organ in the human body and is said to be the most complex substance on the planet. Methods for observing more subtle changes in brain structure are, as yet, poorly developed although the future may hold some promise.

Whereas traditionally, investigators have looked at anatomically defined brain lesions to map neurologic symptoms to specific regions, it appears that many neurologic, psychiatric and other symptoms show greater correspondence with networks of connected regions. Recent work on the human connectome project, which aims to build a network map (connectome) of anatomical structure and function connectivity, is in the process of developing a high-resolution wiring diagram for the human brain, using functional neuroimaging [41]. New MRI scanners with tenfold the resolution of past scanners have revealed the basic pattern of connectivity to be a 3-D grid-like structure, with no diagonals. Ultimately, it is expected this will lead to a greater understanding of brain disorders and disease.

Conclusions

While most patients will receive a correct diagnosis from their clinician most of the time, a significant minority will be undiagnosable, in part due to constraints of current methodology as well as to the complexity of the diagnostic process itself. Some improvements in clinical decision making may be achieved by adopting the augmenting strategies of adaptive expertise that have been proposed.

These current limitations need to be recognised and patients with an undiagnosable illness need a designation, even if it is only to say they are undiagnosable. Of the various labels that have been used thus far, persistent physical symptoms (PPS) appear suitable in that it is non-attributional and avoids the potential stigma of labelling as a psychiatric illness.

References

1. https://www.who.int/classifications/icd/en/HistoryOfICD.pdf.
2. Côté RA, Robboy S. Progress in medical information management. Systematized nomenclature of medicine (SNOMED). JAMA. 1980;243:756–62.
3. Cassell EJ. Illness vs. Disease. Presented at: 1980 medical students symposium on the medical mosaic health care in a changing world, sponsored by McMaster University Medical Student Society; January 18–19, 1980; Hamilton, Ontario, Canada.
4. Seidlein AH, Salloch S. Illness and disease: an empirical-ethical viewpoint. BMC Med Ethics. 2019;20:5.
5. Mattson MP. Superior pattern processing is the essence of the evolved human brain. Front Neurosci. 2014;8:1–17.
6. Croskerry P. The rational diagnostician. In: Croskerry P, Cosby K, Graber M, Singh H, editors. Diagnosis: interpreting the shadows. CRC Taylor and Francis Group: Boca Raton, Fla; 2017.
7. Baron-Cohen S. Autism: the empathizing-systemizing (E-S) theory. Ann N Y Acad Sci. 2009;1156:68–80.
8. Feltovich PJ, Barrows HS. Issues of generality in medical problem solving. In: Schmidt HG, de Volder ML, editors. Tutorials in problem-based learning: new directions in training for the health professions. Assen, Netherlands: Van Gorcum; 1984. p. 128–42.
9. Ericsson KA, Krampe RT, Tesch-Romer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev. 1993;100(3):363–406.
10. Croskerry P. Adaptive expertise in medical decision making. Med Teach. 2018;23:1–6.
11. National Academies of Sciences, Engineering and Medicine. Improving diagnosis in health care. Washington, DC: The National Academies Press; 2015.
12. Croskerry P, Cosby K, Graber M, Singh H. Diagnosis: interpreting the shadows. CRC Taylor and Francis Group: Boca Raton, Fla; 2017.
13. Peterson MC, Holbrook JH, Hales D, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. West J Med. 1992;156:163–5.
14. https://en.wikipedia.org/wiki/list_of_medical_symptoms. Accessed 17 Mar 2021.
15. Ely JW, Graber M, Croskerry P. Checklists to reduce diagnostic errors. Acad Med. 2011;86:307–13. See also Ely JW. 63 common complaints in the outpatients setting (unpublished handbook, 2013).
16. Leening MJG, Elias-Smale SE, Felix JF, Kors JA, Deckers JW, Hofman A, Stricker BHC, Witteman JCM. Unrecognised myocardial infarction and long-term risk of heart failure in the elderly: the Rotterdam Study. Heart. 2010;96:1458–62.

17. Brieger D, Eagle KA, Goodman SG, Steg PG, Budaj A, White K, Montalescot G. Acute coronary syndromes without chest pain, an underdiagnosed and undertreated high-risk group: insights from the global registry of acute coronary events. Chest. 2004;126(2):461–9.
18. Arenja N, Mueller C, Ehl NF, Brinkert M, Roost K, Reichlin T, Sou SM, Hochgruber T, Osswald S, Zellweger MJ. Prevalence, extent, and independent predictors of silent myocardial infarction. AMJ. 2013;126(6):515–22.
19. Croskerry P. The English patient in Croskerry P. The cognitive autopsy: a root cause analysis of medical decision making. New York, NY: Oxford University Press; 2020. p. 55–63.
20. Croskerry P, Petrie DA, Reilly JB, Tait G. Deciding about fast and slow decisions. Acad Med. 2014;89(2):197–200.
21. Kirmayer LJ, Groleau D, Looper KJ, Dao MD. Explaining medically unexplained symptoms. Can J Psychiatr. 2004;49(10):663–72.
22. JCPFMH. Guidance for commissioners of Services for People with medically unexplained symptoms. Joint Commissioning Panel for Mental Health. Royal College of General Practitioners and the Royal College of Psychiatrists; 2017.
23. Patel M, Kirsty J, Moss-Morris R, Husain M, Ashworth M, Frank P, Ferreira N, Mosweu I, McCrone P, Hotopf M, David A, Landau S, Chalder T. Persistent physical symptoms reduction intervention: a system change and evaluation (PRINCE)—integrated GP care for persistent physical symptoms: protocol for a feasibility and cluster randomised waiting list, controlled trial. BMJ Open. 2019;9. https://doi.org/10.1136/bmjopen-2018-025513.
24. Payer L. Disease-mongers: how doctors, drug companies, and insurers are making you feel sick. New York: J. Wiley; 1992. 978-0471543855
25. Bransfield RC, Friedman KJ. Differentiating psychosomatic, somatopsychic, multisystem illnesses and medical uncertainty. Healthcare. 2019;7:114. https://doi.org/10.3390/healthcare.
26. Kozlowska K. Stress, distress, and bodytalk: co-constructing formulations with patients who present with somatic symptoms. Harv Rev Psychiatry. 2013;21(6):314–33.
27. Budtz-Lilly A, Vestergaard M, Fink P, Carlsen AH, Rosendal M. Patient characteristics and frequency of bodily distress syndrome in primary care: a cross-sectional study. Br J Gen Pract. 2015;65(638):617–23.
28. Chowdhury AN. The definition and classification of Koro. Cult Med Psychiatry. 1996;20(1):41–65.
29. Pearson ML, Selby JV, Katz KA, Cantrell V, Braden CR, et al. Clinical, epidemiologic, histopathologic and molecular features of an unexplained dermopathy. PLoS One. 2012;7(1):e29908.
30. Misery L. Morgellons syndrome: a disease transmitted by the media. Ann Dermatol Venereol. 2013;140:59–62.
31. Middelveen MJ, Fesler MC, Stricker RB. History of Morgellons disease: from delusion to definition. Clin Cosmet Investig Dermatol. 2018;11:71–90.
32. Middelveen MJ, Martinez RM, Fesler MC, Sapi E, Burke J, Shah JS, et al. Classification and staging of Morgellons disease: lessons from syphilis clinical. Cosmet Investigat Dermatol. 2020;13:145–64.
33. Creed F, Guthrie E, Fink P, Henningsen P, Rief W, Sharpe M, White P. Is there a better term than 'Medically unexplained symptoms'? J Psychosom Research. 2010;68:5–8.
34. Marks EM, Hunter MS. Medically unexplained symptoms: an acceptable term? Brit J Pain. 2015;9:109–14.
35. Croskerry P. Clinical decision making in emergency medicine. In: Higgs J, Jensen GM, Loftus S, Christensen N, editors. Clinical reasoning in the health professions, Ch 26. 4th ed. New York: Elsevier; 2019. p. 285–94.
36. De Bono E. Lateral thinking: a textbook of creativity. London: Penguin Life; 1991.
37. Emmett KR. Nonspecific and atypical presentation of disease in the older patient. Geriatrics. 1998;53:50–60.
38. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med. 2003;78:775–80.
39. Kahneman D. Thinking fast and slow. New York: Farrar. Straus and Giroux; 2011.

40. Gierk B, Kohlmann S, Kroenke K, Spangenberg L. The somatic symptom Scale-8 (SSS-8) a brief measure of somatic symptom burden. JAMA Int Med. 2014;174(3):399–40.
41. Fox MD. Mapping symptoms to brain networks with the human connectome. NEJM. 2018;379(23):2237–45.

Toward Molecular Diagnoses for Autoimmune Rheumatic Diseases

4

Judith A. James and Catriona A. Wagner

Introduction

Autoantibodies occur in at least 14–18% of the adult population [1–3], and about 5–8% of the US population suffers from an autoimmune disease [4]. Many patients struggle for years, knowing they have symptoms that limit their lives but unable to find a diagnosis. Although select autoantibody specificities are enriched in certain systemic rheumatic diseases, these often lack sensitivity and/or specificity, resulting in no definitive diagnostic tests for these disorders and the use of consensus-driven classification criteria to guide clinical trial inclusion and research studies and even to aid in diagnoses. This lack of diagnostic precision leads to uncertainty for patients, their families, and their providers, as well as diagnosis and treatment delays.

As one example, SLE is a complex, multi-organ autoimmune disease with unpredictable periods of flare and remission [5]. Although no diagnostic criteria currently exist, SLE is typically characterized by clinical and/or serologic parameters defined by classification criteria [6–9]. However, SLE is highly heterogeneous in its presentation and progression, and early symptoms are often nonspecific and may overlap with those of other rheumatic diseases, such as Sjögren's syndrome (SS) and rheumatoid arthritis, making diagnosis challenging. Furthermore, in some cases, patients present with only a few SLE features and never meet classification criteria, with

J. A. James (✉)
Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA

Departments of Medicine and Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
e-mail: jamesj@omrf.org

C. A. Wagner
Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA
e-mail: Catriona-Wagner@omrf.org

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_4

most not going on to major organ involvement [10, 11]. As early diagnosis and treatment are associated with reduced flares and improved outcomes [12–14], establishing tools that facilitate early and accurate diagnosis, prognosis, and treatment selection is essential to improving the clinical care of these patients. Other disease entities, such as cancer and clinical genetics, have addressed the issues currently faced in rheumatic diseases by implementing molecular diagnostics to aid precision treatment and improved outcomes [15, 16].

Molecular diagnoses identify a disease by detecting individual biologic molecules, such as genes and proteins, in a tissue or fluid. Molecular tools are frequently used to diagnose genetic conditions, infectious diseases, and hereditary cancers and to select effective cancer treatments based on molecular markers on tumors, such as estrogen receptors and HER2 in breast cancer [15]. In autoimmune systemic rheumatic diseases, moving from a symptom-based to molecular-based diagnosis (or a hybrid of the two) may be more effective at identifying individuals with shared molecular pathways that may be amenable to treatment with a specific class of drugs. However, the extensive heterogeneity of these disorders suggests multiple molecular pathways contribute to these diseases, and it is unlikely that an individual molecular biomarker will apply to all patients. Therefore, high-throughput, multidimensional measurements and machine learning are necessary to identify molecular pathways with the most utility to diagnose and stratify patients into more homogenous molecular groups.

Molecular Clustering in Autoimmune Rheumatic Disease

Efforts have been made to stratify autoimmune rheumatic diseases into homogenous molecular groups of patients. Using SLE as an example, previous studies have stratified SLE patients based on autoantibody [17–19], clinical [20–23], gene expression [24–30], soluble mediators, or immune cell [31] profiles individually. In a large cohort of pediatric SLE patients, longitudinal analyses of blood transcriptome profiles defined seven patient clusters based on five distinct immune signatures correlating with disease activity, including plasmablasts, type I IFN, neutrophils/myeloid cells, and lymphocytes [26]. In addition, combined genetic and transcriptomic profiling of whole blood from 142 SLE patients identified a susceptibility and activity signature [28]. The susceptibility signature persisted during remission, suggesting constant immune activation, and the activity signature differentiated SLE patients with inactive and active disease and was enriched in genes related to oxidative phosphorylation and cell metabolism [28]. Notably, neutrophil signatures in both studies were associated with active lupus nephritis [26, 28], which is associated with increased morbidity and mortality. A recent study characterized the immune cell profiles of juvenile-onset SLE patients using machine learning and identified an immune cell signature specific to juvenile-onset SLE, which was further stratified into four different patient clusters associated with disease activity [31]. Together, these studies demonstrate the importance of stratifying SLE patients based on molecular phenotypes. However, although some transcriptome

signatures can distinguish between most patients with SLE and healthy controls [28, 32], the signatures are not inclusive of all SLE patients and the diagnostic potential of these signatures in patients who do not align with SLE classification criteria is still unknown.

We hypothesized that integrating gene expression, soluble mediator, autoantibody, and clinical data with machine learning modeling would more effectively define homogenous patient subsets based on molecular phenotypes. Using this approach [33], we defined seven unique molecular profiles associated with different immune pathways based on both soluble mediator profiles and previously defined gene expression modules [26, 34–36]. In one cluster, termed Cluster 4, patients exhibited significantly higher IFN, inflammation, and neutrophil module scores; the highest levels of the soluble mediators IP-10, MIG, APRIL, TNFRI/II, and IL-10; increased DNA binding; and slightly higher disease activity. In contrast, Cluster 4 had reduced B and T cell signatures. Surprisingly, clinical features commonly used to subset SLE patients, such as the affected organs, cumulative ACR criteria, and autoantibodies, were similar across molecularly distinct clusters with different patterns of immune activation, suggesting that different pathogenic mechanisms may lead to similar clinical outcomes [33].

We used a similar approach to stratify primary Sjögren's syndrome (pSS) patients and identified three clusters with similar clinical features but significantly different IFN transcriptional module signatures [37]. One cluster, termed Cluster 2, had strong IFN and inflammation signatures, with increased IP-10, MIG, BLyS, and LIGHT levels, similar to Cluster 4 in SLE. As SLE and pSS share similar clinical and serological features, it is often difficult to distinguish between the two diseases. Therefore, these similar molecular profiles may indicate that some of these patients with divergent diagnoses are more molecularly similar than different, and analyzing molecular profiles may help more precisely diagnose these patients. Alternatively, these findings may suggest that Cluster 4 is a distinct subset of autoimmunity with different clinical presentations, including features of SLE or pSS. Consistent with this hypothesis, a recent study found that patients with systemic rheumatic diseases were stratified into three clusters based on transcriptome and methylome data, independent of clinical diagnosis, demonstrating shared molecular phenotypes [38]. Therefore, we hypothesize that Cluster 4 may be present in other manifestations, including patients who do not meet current classification or diagnostic criteria.

Patients with Symptoms Who Do Not Meet Current Classification Guidelines: Lessons from Incomplete Lupus Erythematosus (ILE) and Undifferentiated Connective Tissue Disease (UCTD)

As is centric to a conference with the title "When the Illness Has No Name," many patients exhibit clinical symptoms or serological evidence of systemic rheumatic diseases but do not fulfill classification criteria. Patients whose symptoms do not fall into a specific autoimmune disease are termed undifferentiated connective tissue

disease (UCTD), while those whose symptoms meet some SLE criteria are termed incomplete lupus erythematosus (ILE). Similar to SLE and other rheumatic diseases, patients with UCTD and ILE experience heterogeneous clinical and serologic manifestations, resulting in unclear diagnoses and treatment plans. There is also heterogeneity in the progression of UCTD and ILE. While most patients maintain a relatively mild disease course with limited involvement of major organs, 10–50% transition to SLE or another rheumatic disease within 5 years of onset [39–45], suggesting that ILE is an early stage of SLE in some patients. In addition, a subset of patients without formal SLE classification develops serious clinical manifestations that result in hospitalization in some cases [10, 44, 46]. Stratifying patients with ILE and UCTD by the risk of transition or major clinical disease would allow early interventions and intensive follow-up for high-risk patients to delay or prevent transition while minimizing unnecessary treatments and follow-ups for those likely to maintain a mild disease. However, there are currently no standard predictors of disease transition or severe clinical disease for ILE and UCTD patients, resulting in disease uncertainty, patient anxiety, and improper care [47].

Previous studies of ILE and UCTD patients have identified demographic, clinical, and serological factors associated with disease transition, such as younger age, malar rash, renal involvement, altered T cell frequencies, and anti-double-stranded DNA, anti-Sm, and anti-cardiolipin autoantibodies [40, 42, 44, 45, 48–50]. In addition, half of ILE patients express elevated levels of IFN-associated genes and an increased IFN score compared to healthy controls [51, 52], identifying a potential diagnostic marker for a subset of these patients. In a recent study, patients at risk for connective tissue disease who transitioned to SLE or pSS had higher IFN scores at baseline compared with those who did not transition [45], suggesting IFN signatures may also be a predictive biomarker in ILE and UCTD. However, due to the heterogeneity of ILE and UCTD, it is unlikely that IFN signature alone will be sufficient to diagnose and subset these patients.

Our lab previously measured soluble mediators and autoantibodies in samples from preclinical SLE patients to determine predictive profiles of SLE transition [53]. We found that a gradual increase in innate and T helper-associated immune pathways and a decrease in regulatory soluble mediators precede SLE classification [53]. In addition, models incorporating IL-5, IL-6, MIG/CXCL9, and antinuclear autoantibodies accurately distinguish preclinical SLE patients from healthy controls [53]. Similar alterations in soluble mediators are also observed in first-degree relatives of SLE patients who transition to SLE; however, in these patients, SCF, TGFβ, and ACR scores are independent predictors of SLE transition [54]. SLE patients also exhibit dysregulation of inflammatory and regulatory mediators up to 12 weeks before disease flare, and a combined soluble mediator score reliably identifies impending disease flare in both European American and African American SLE patients [55, 56]. Therefore, similar immune perturbations may also help identify ILE and UCTD patients at risk of SLE transition. However, extensive, multidimensional molecular evaluation with genetic, genomic, and immunomic information, partnered with clinical and serologic data, is needed to define molecular signatures to help accurately diagnose and classify patients with ILE and UCTD.

Uncertain Disease Classification in Understudied Populations

Clinical manifestations of rheumatic diseases can differ by sex and race/ethnicity, potentially delaying diagnosis in underrepresented patients. As an example, the incidence and prevalence of SLE are higher in American Indian populations compared to other racial and ethnic groups [57, 58]. American Indian patients also have higher disease activity at diagnosis, damage accrual, and increased morbidity rates, with more frequent vasculitis and nephritis [57, 59]. Early clinical SLE diagnosis is especially challenging in American Indian patients as they exhibit atypical clinical presentations and serological profiles and are more likely to have concurrent rheumatic diseases or symptoms [60–62]. As a result, approximately 28–47% of American Indian rheumatic disease patients fail to meet ACR classification criteria and are left without an official diagnosis [63, 64]. Therefore, it is essential to identify molecular markers of systemic autoimmune rheumatic disease in American Indian patients to optimize early diagnosis and help guide treatment.

Due to the differences in American Indian clinical and serological presentations, unique immune pathways may contribute to disease pathogenesis in these patients. In a recent study, Catalina et al. found that SLE gene expression signatures differed significantly by race, indicating the involvement of different molecular pathways in different races/ethnicities [65]. Specifically, American Indian ancestry was associated with increased inflammasome, erythrocyte, and unfolded protein response signatures and decreased IFN, T cell, and MHC class II signatures compared to European and African ancestries [65]. We recently stratified American Indian rheumatic disease patients based on serum autoantibody, cytokine, and chemokine expression using machine learning and identified five patient clusters with distinct immune signatures, which were associated with different clinical characteristics [66]. Interestingly, the clusters did not stratify based on clinical diagnoses, providing further evidence that similar immune pathways may result in different clinical manifestations and highlighting the importance of determining these immune signatures to optimize diagnosis and treatment of patients with rheumatic diseases.

Conclusion

Defining molecular phenotypes of patients with SLE and other rheumatic diseases may allow for a more precise diagnosis of disease and disease state, helping patients address current issues of uncertainty. Molecular diagnoses are essential in patients of different races/ethnicities and sex who may have unique clinical and serological features. In addition, patients currently diagnosed with different rheumatic diseases may belong to larger homogeneous groups based on immune pathways, aiding clinical trials and treatment selection and discovery.

References

1. Satoh M, Chan EKL, Ho LA, Rose KM, Parks CG, Cohn RD, et al. Prevalence and sociodemographic correlates of antinuclear antibodies in the United States. Arthritis Rheum. 2012;64:2319–27.
2. Dinse GE, Parks CG, Weinberg CR, Co CA, Wilkerson J, Zeldin DC, et al. Increasing prevalence of antinuclear antibodies in the United States. Arthritis Rheumatol. 2020;72:1026–35.
3. Dillon CF, Weisman MH, Miller FW. Population-based estimates of humoral autoimmunity from the U.S. National Health and Nutrition Examination Surveys, 1960–2014. PLoS One. 2020;15:e0226516.
4. NIH Autoimmune Diseases Coordinating Committee. Progress in autoimmune diseases research; 2005. https://www.niaid.nih.gov/sites/default/files/adccfinal.pdf. Accessed 2 Jun 2021.
5. Tsokos GC. Systemic lupus erythematosus. New Engl J Med. 2011;365:2110–21.
6. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1982;25:1271–7.
7. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1997;40:1725.
8. Petri M, Orbai A, Alarcón GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the systemic lupus international collaborating clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum. 2012;64:2677–86.
9. Aringer M, Costenbader K, Daikh D, Brinks R, Mosca M, Ramsey-Goldman R, et al. 2019 European league against rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. Ann Rheum Dis. 2019;78:1151–9.
10. Aberle T, Bourn RL, Munroe ME, Chen H, Roberts VC, Guthridge JM, et al. Clinical and serologic features in patients with incomplete lupus classification versus systemic lupus erythematosus patients and controls. Arthritis Care Res (Hoboken). 2017;69:1780–8.
11. Aberle T, Bourn RL, Chen H, Roberts VC, Guthridge JM, Bean K, et al. Use of SLICC criteria in a large, diverse lupus registry enables SLE classification of a subset of ACR-designated subjects with incomplete lupus. Lupus Sci Med. 2017;4:e000176.
12. Doria A, Zen M, Canova M, Bettio S, Bassi N, Nalotto L, et al. SLE diagnosis and treatment: when early is early. Autoimmun Rev. 2010;10:55–60.
13. Oglesby A, Korves C, Laliberté F, Dennis G, Rao S, Suthoff ED, et al. Impact of early versus late systemic lupus erythematosus diagnosis on clinical and economic outcomes. Appl Health Econ Health Policy. 2014;12:179–90.
14. Faurschou M, Dreyer L, Kamper A, Starklint H, Jacobsen S. Long-term mortality and renal outcome in a cohort of 100 patients with lupus nephritis. Arthritis Care Res (Hoboken). 2010;62:873–80.
15. Sokolenko AP, Imyanitov EN. Molecular diagnostics in clinical oncology. Front Mol Biosci. 2018;5:76.
16. Bienvenu T, Lopez M, Girodon E. Molecular diagnosis and genetic counseling of cystic fibrosis and related disorders: new challenges. Genes (Basel). 2020;11:619.
17. Kessel A, Rosner I, Halasz K, Grushko G, Shoenfeld Y, Paran D, et al. Antibody clustering helps refine lupus prognosis. Semin Arthritis Rheu. 2009;39:66–70.
18. Artim-Esen B, Çene E, Şahinkaya Y, Ertan S, Pehlivan Ö, Kamali S, et al. Cluster analysis of autoantibodies in 852 patients with systemic lupus erythematosus from a single center. J Rheumatol. 2014;41:1304–10.
19. Pacheco Y, Barahona-Correa J, Monsalve DM, Acosta-Ampudia Y, Rojas M, Rodríguez Y, et al. Cytokine and autoantibody clusters interaction in systemic lupus erythematosus. J Transl Med. 2017;15:239.
20. Jacobsen S, Petersen J, Ullman S, Junker P, Voss A, Rasmussen JM, et al. A multicentre study of 513 Danish patients with systemic lupus erythematosus. I. Disease manifestations and analyses of clinical subsets. Clin Rheumatol. 1998;17:468–77.

21. Font J, Cervera R, Ramos-Casals M, Garcia-Carrasco M, Sentis J, Herrero C, et al. Clusters of clinical and immunologic features in systemic lupus erythematosus: analysis of 600 patients from a single center. Semin Arthritis Rheum. 2004;33:217–30.
22. To C, Mok C, Tang S, Ying S, Wong R, Lau C. Prognostically distinct clinical patterns of systemic lupus erythematosus identified by cluster analysis. Lupus. 2009;18:1267–75.
23. Peschken CA, Katz SJ, Silverman E, Pope JE, Fortin PR, Pineau C, et al. The 1000 Canadian faces of lupus: determinants of disease outcome in a large multiethnic cohort. J Rheumatol. 2009;36:1200–8.
24. Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann WA, Espe KJ, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. Proc Natl Acad Sci U S A. 2003;100:2610–5.
25. Lyons PA, McKinney EF, Rayner TF, Hatton A, Woffendin HB, Koukoulaki M, et al. Novel expression signatures identified by transcriptional analysis of separated leucocyte subsets in systemic lupus erythematosus and vasculitis. Ann Rheum Dis. 2010;69:1208–13.
26. Banchereau R, Hong S, Cantarel B, Baldwin N, Baisch J, Edens M, et al. Personalized immunomonitoring uncovers molecular networks that stratify lupus patients. Cell. 2016;165:551–65.
27. Toro-Domínguez D, Martorell-Marugán J, Goldman D, Petri M, Carmona-Sáez P, Alarcón-Riquelme ME. Stratification of systemic lupus erythematosus patients into three groups of disease activity progression according to longitudinal gene expression. Arthritis Rheumatol. 2018;70:2025–35.
28. Panousis NI, Bertsias GK, Ongen H, Gergianaki I, Tektonidou MG, Trachana M, et al. Combined genetic and transcriptome analysis of patients with SLE: distinct, targetable signatures for susceptibility and severity. Ann Rheum Dis. 2019;78:1079–89.
29. Figgett WA, Monaghan K, Ng M, Alhamdoosh M, Maraskovsky E, Wilson NJ, et al. Machine learning applied to whole-blood RNA-sequencing data uncovers distinct subsets of patients with systemic lupus erythematosus. Clin Transl Immunol. 2019;8:e01093.
30. Sandling JK, Pucholt P, Rosenberg LH, Farias FHG, Kozyrev SV, Eloranta M-L, et al. Molecular pathways in patients with systemic lupus erythematosus revealed by gene-centred DNA sequencing. Ann Rheum Dis. 2021;80:109–17.
31. Robinson GA, Peng J, Dönnes P, Coelewij L, Naja M, Radziszewska A, et al. Disease-associated and patient-specific immune cell signatures in juvenile-onset systemic lupus erythematosus: patient stratification using a machine-learning approach. Lancet Rheumatol. 2020;2:e485–96.
32. Hamilton JA, Wu Q, Yang P, Luo B, Liu S, Li J, et al. Cutting edge: intracellular IFN-β and distinct type I IFN expression patterns in circulating systemic lupus erythematosus B cells. J Immunol. 2018;201:2203–8.
33. Guthridge JM, Lu R, Tran LT, Arriens C, Aberle T, Kamp S, et al. Adults with systemic lupus exhibit distinct molecular phenotypes in a cross-sectional study. EClinicalMedicine. 2020;20:100291.
34. Chiche L, Jourde-Chiche N, Whalen E, Presnell S, Gersuk V, Dang K, et al. Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type I and type II interferon signatures. Arthritis Rheumatol. 2014;66:1583–95.
35. Banchereau R, Cepika A-M, Banchereau J, Pascual V. Understanding human autoimmunity and autoinflammation through transcriptomics. Annu Rev Immunol. 2016;35:337–70.
36. Jourde-Chiche N, Whalen E, Gondouin B, Speake C, Gersuk V, Dussol B, et al. Modular transcriptional repertoire analyses identify a blood neutrophil signature as a candidate biomarker for lupus nephritis. Rheumatology (Oxford). 2017;56:477–87.
37. James JA, Guthridge JM, Chen H, Lu R, Bourn RL, Bean K, et al. Unique Sjögren's syndrome patient subsets defined by molecular features. Rheumatology (Oxford). 2019;59:860–8.
38. Barturen G, Babaei S, Català-Moll F, Martínez-Bueno M, Makowska Z, Martorell-Marugán J, et al. Integrative analysis reveals a molecular stratification of systemic autoimmune diseases. Arthritis Rheumatol. 2021;73:1073–85.
39. Greer JM, Panush RS. Incomplete lupus erythematosus. Arch Intern Med. 1989;149:2473–6.

40. Vila L, Mayor A, Valent A, Garc M, Vila S. Clinical outcome and predictors of disease evolution in patients with incomplete lupus erythematosus. Lupus. 2000;9:110–5.
41. Swaak AJ, Van de Brink H, Smeenk RJ, Manger K, Kalden JR, Tosi S, et al. Incomplete lupus erythematosus: results of a multicentre study under the supervision of the EULAR Standing Committee on International Clinical Studies Including Therapeutic Trials (ESCISIT). Rheumatology (Oxford). 2001;40:89–94.
42. Hallengren CS, Nived O, Sturfelt G. Outcome of incomplete systemic lupus erythematosus after 10 years. Lupus. 2004;13:85–8.
43. Laustrup H, Voss A, Green A, Junker P. SLE disease patterns in a Danish population-based lupus cohort: an 8-year prospective study. Lupus. 2010;19:239–46.
44. Daabil MA, Massarotti EM, Fine A, Tsao H, Ho P, Schur PH, et al. Development of SLE among "potential SLE" patients seen in consultation: long-term follow-up. Int J Clin Pract. 2014;68:1508–13.
45. Yusof MYM, Psarras A, El-Sherbiny YM, Hensor EMA, Dutton K, Ul-Hassan S, et al. Prediction of autoimmune connective tissue disease in an at-risk cohort: prognostic value of a novel two-score system for interferon status. Ann Rheum Dis. 2018;77:1432–9.
46. Rúa-Figueroa Í, Richi P, López-Longo FJ, Galindo M, Calvo-Alén J, Olivé-Marqués A, et al. Comprehensive description of clinical characteristics of a large systemic lupus erythematosus cohort from the spanish rheumatology society lupus registry (RELESSER) with emphasis on complete versus incomplete lupus differences. Medicine (Baltimore). 2015;94:e267.
47. Ugarte-Gil MF, Alarcón GS. Incomplete systemic lupus erythematosus: early diagnosis or overdiagnosis? Arthritis Care Res (Hoboken). 2016;68:285–7.
48. Olsen NJ, Li Q-Z, Quan J, Wang L, Mutwally A, Karp DR. Autoantibody profiling to follow evolution of lupus syndromes. Arthritis Res Ther. 2012;14:R174.
49. Szodoray P, Nakken B, Barath S, Csipo I, Nagy G, El-Hage F, et al. Altered Th17 cells and Th17/regulatory T-cell ratios indicate the subsequent conversion from undifferentiated connective tissue disease to definitive systemic autoimmune disorders. Hum Immunol. 2013;74:1510–8.
50. Bourn R, James JA. Preclinical lupus. Curr Opin Rheumatol. 2015;27:433–9.
51. Lambers WM, de Leeuw K, der Meer BD, Diercks GFH, Bootsma H. Interferon score is increased in incomplete systemic lupus erythematosus and correlates with myxovirus-resistance protein A in blood and skin. Westra J Arthritis Res Ther. 2019;21:260.
52. Li QZ, Zhou J, Lian Y, Zhang B, Branch VK, Carr-Johnson F, et al. Interferon signature gene expression is correlated with autoantibody profiles in patients with incomplete lupus syndromes. Clin Exp Immunol. 2010;159:281–91.
53. Lu R, Munroe ME, Guthridge JM, Bean KM, Fife DA, Chen H, et al. Dysregulation of innate and adaptive serum mediators precedes systemic lupus erythematosus classification and improves prognostic accuracy of autoantibodies. J Autoimmun. 2016;74:182–93.
54. Munroe ME, Young KA, Kamen DL, Guthridge JM, Niewold TB, Costenbader KH, et al. Discerning risk of disease transition in relatives of systemic lupus erythematosus patients utilizing soluble mediators and clinical features. Arthritis Rheumatol. 2017;69:630–42.
55. Munroe ME, Vista ES, Guthridge JM, Thompson LF, Merrill JT, James JA. Proinflammatory adaptive cytokine and shed tumor necrosis factor receptor levels are elevated preceding systemic lupus erythematosus disease flare. Arthritis Rheumatol. 2014;66:1888–99.
56. Munroe ME, Vista ES, Merrill JT, Guthridge JM, Roberts VC, James JA. Pathways of impending disease flare in African-American systemic lupus erythematosus patients. J Autoimmun. 2017;78:70–8.
57. Peschken CA, Esdaile JM. Systemic lupus erythematosus in north American Indians: a population based study. J Rheumatol. 2000;27:1884–91.
58. Ferucci ED, Johnston JM, Gaddy JR, Sumner L, Posever JO, Choromanski TL, et al. Prevalence and incidence of systemic lupus erythematosus in a population-based registry of American Indian and Alaska native people, 2007–2009. Arthritis Rheumatol. 2014;66:2494–502.
59. Hurd K, Barnabe C. Mortality causes and outcomes in indigenous populations of Canada, the United States, and Australia with rheumatic disease: a systematic review. Semin Arthritis Rheum. 2018;47:586–92.

60. Scofield RH, Fogle M, Rhoades ER, Harley JB. Rheumatoid arthritis in a United States Public Health Service Hospital in Oklahoma: serologic manifestations in rheumatoid arthritis vary among tribal groups. Arthritis Rheum. 1996;39:283–6.
61. Kheir JM, Guthridge CJ, Johnston JR, Adams LJ, Rasmussen A, Gross TF, et al. Unique clinical characteristics, autoantibodies and medication use in native American patients with systemic lupus erythematosus. Lupus Sci Med. 2018;5:e000247.
62. Guthridge CJ, Gross T, Quintero M, Kheir JM, Levin J, Bourn RL, et al. Expanded autoantibody profiles for subsetting of native American, African American, and European American patients with systemic lupus erythematosus. ACR Open Rheumatol. 2020;2:415–23.
63. Atkins C, Reuffel L, Roddy J, Platts M, Robinson H, Ward R. Rheumatic disease in the Nuu-Chah-Nulth native Indians of the Pacific northwest. J Rheumatol. 1988;15:684–90.
64. Gaddy JR, Vista ES, Robertson JM, Dedeke AB, Roberts VC, Klein WS, et al. Rheumatic disease among Oklahoma tribal populations: a cross-sectional study. J Rheumatol. 2012;39:1934–41.
65. Catalina MD, Bachali P, Yeo AE, Geraci NS, Petri MA, Grammer AC, et al. Patient ancestry significantly contributes to molecular heterogeneity of systemic lupus erythematosus. JCI Insight. 2020;5:e140380.
66. Slight-Webb S, Guthridge CJ, Kheir J, Chen H, Tran L, Gross T, et al. Unique serum immune phenotypes stratify Oklahoma Native American rheumatic disease patients. Arthritis Care Res (Hoboken). 2021;10.1002/acr.24795.

Part II

Purposes of Diagnosis

Diagnostic Uncertainty in Drug Development

5

Paola Mina-Osorio

Introduction

Over decades, the pharmaceutical industry developed therapies that changed the management of primary care diseases, causing therapeutic areas such as cardiovascular medicine to mature. This drove a shift in the industry's R&D focus toward more complex areas, including immunology. Most immune-mediated inflammatory diseases (IMIDs) are extremely complex biologically and clinically, and therefore, they are associated with high levels of scientific and clinical uncertainty [1].

The drug development process has not evolved to match that complexity. In contrast to areas like oncology [2], where more than 40 precision-oncology drugs are now on the market, there has been less innovation in clinical trial design in immunology. Methodologies such as molecular profiling to enhance patient selection and stratification and predict treatment response have proven extremely difficult to advance in this space [3]. In addition, the information used to identify and validate novel targets in the early stages of the drug development process often comes from academic research and animal models that cannot recapitulate all aspects of human disease [4]. It is not until a new compound reaches clinical stages that we attempt to incorporate the true complexity of the disease under study. There is much to learn regarding endotypes in IMIDs (Fig. 5.1) to understand their influence on treatment outcomes.

Therapeutic strategies that target immune pathways that are only active in a subset of patients or individual pathways in diseases with multifactorial etiology will continue to have limited success. The breakthroughs made in the late 1990s and early 2000s with the introduction of biologics for diseases such as rheumatoid arthritis have not been followed by incremental successes, particularly in the most

P. Mina-Osorio (✉)
Aurinia Pharmaceuticals Inc., Rockville, MD, USA
e-mail: pminaosorio@auriniapharma.com

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_5

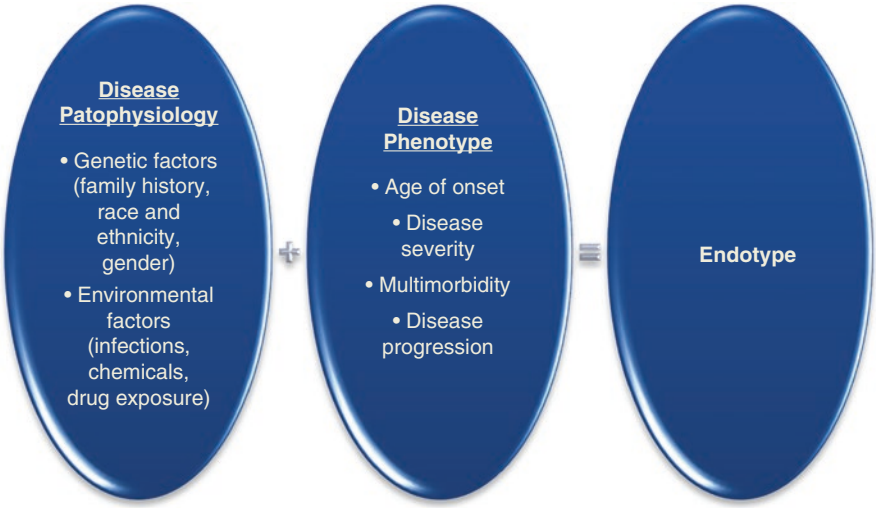

Fig. 5.1 IMIDs have multifactorial pathophysiology involving genetic and environmental factors. This complexity translates into heterogeneous clinical presentations (phenotypes). The term "endotype" refers to the molecular mechanisms underlying observable disease characteristics (phenotypes). Big data that integrates all aspects of the endotype using a systems biology approach must drive the development of drugs that address immune dysregulation occurring in multiple pathways and could be used either as monotherapy or in combination. (Modified from [5])

complex diseases where only marginal benefits or a complete lack of efficacy has been observed time and again, with few approved drugs in decades [6, 7].

The lack of representation of real-world populations in randomized-controlled clinical trials (RCTs) is a significant issue, particularly relevant in diseases in which such subpopulations are more frequently or more severely affected by the disease (Fig. 5.2a) [8].

The role of newer generations of drug developers will be to address issues such as the lack of representativeness in clinical trials of complex diseases and the urgent need to advance modern methodologies to appropriately incorporate diagnostic uncertainty into drug discovery and development, in a cost-effective manner.

Defining Diagnostic Uncertainty in Drug Development

If we define diagnostic uncertainty as a "subjective perception of an inability to provide an accurate explanation of a patient's health problem" [9], then we can say with great certainty that this definition has no place in the current model of drug development. A clinician is not always certain about the diagnosis before initiating treatment [10]. In contrast, the perception of diagnostic certainty is a prerequisite to initiating treatment in clinical trials, which have the overarching goal of demonstrating that the investigational compound is efficacious and safe in a relatively homogeneous patient population. This is one of the premises for the development of disease

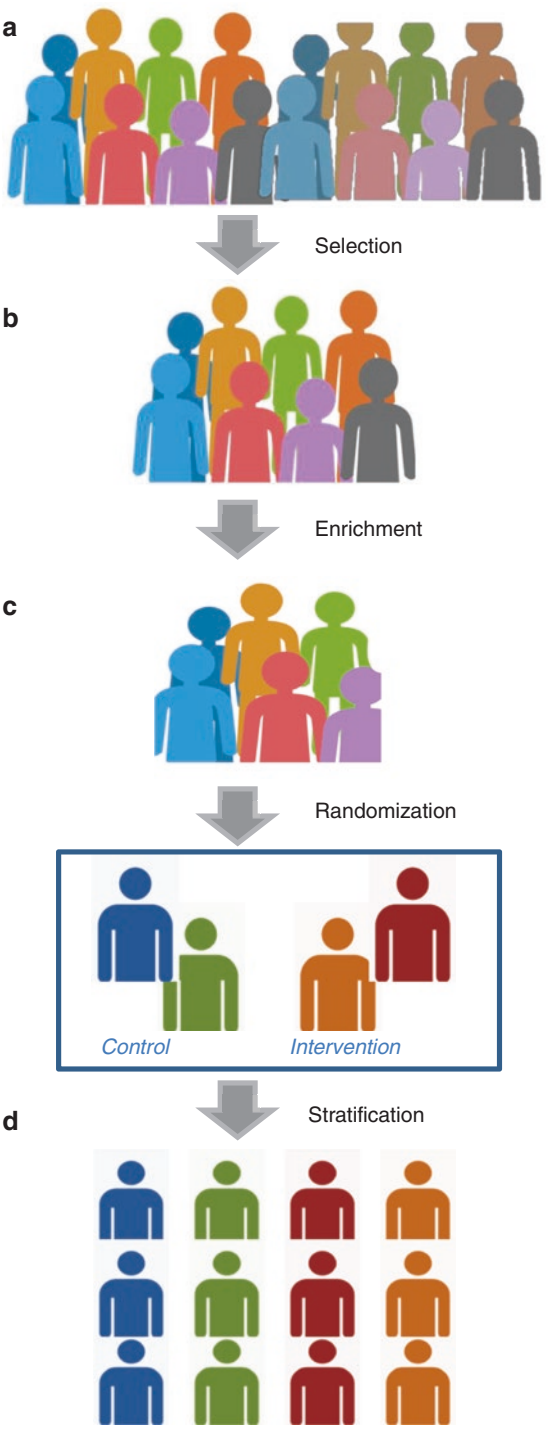

Fig. 5.2 Lack of representativeness in clinical trials of complex diseases. (**a**) Real-world patient populations are heterogeneous. Patients selected to participate in clinical trials tend to have specific, nonrepresentative characteristics such as better access to healthcare, higher health literacy, and low racial and ethnic diversity. (**b**) Patient enrichment practices further exclude patients who may be representative of the real-world population, e.g., minority patients who are not compliant or patients with difficult-to-treat disease who do not fulfill the classification criteria or whose disease is not currently active. (**c**) Randomization attempts to equally distribute patients into study arms and increase internal validity. However, study groups are never completely homogeneous because not all aspects of the endotype are taken into account. (**d**) Patient stratification methods are limited to a few variables to avoid increasing the sample sizes. Analysis of patient subgroups is usually conducted at the end of the trial once all data has been collected. Insufficient samples prevent us from making unequivocal conclusions for the greater population creating uncertainty around the efficacy and safety of novel treatments among certain subsets of patients

classification criteria, which are considered different from diagnostic criteria in that they are not designed to be used for clinical diagnosis in individual patients, but to capture pathognomonic disease characteristics to group patients into homogeneous populations for clinical research [10–12].

The use of homogeneous populations is useful because the statistical power (i.e., probability of rejecting the null hypothesis) of the study increases by decreasing the variability. However, relevant for this book, homogeneity is difficult to achieve in clinical trials of complex diseases. The application of stringent criteria results in the biased selection of patients with "typical" clinical presentation and the exclusion of many others (Fig. 5.2a, b) [13]. Consequently, the evaluation of new therapies in patients who do not meet study entry criteria (Fig. 5.2a) is usually conducted empirically in the clinic [14], using treatments off-label and with significant access barriers.

Randomization after study entry is one of the methods used to ensure homogeneous populations across the treatment arms and to increase internal validity in clinical trials. However, we do not conduct clinical trials of randomly selected patients at the outset because of the various types of patient enrichment and selection derived from the inclusion and exclusion criteria (Fig. 5.2c) and because of the large number of variables (endotypes) that may not end up equally distributed across treatment arms. This is aggravated by small sample sizes and difficulties recruiting patients into trials of rare diseases. For this reason, stratification (e.g., stratified randomization or stratified proportional sampling) is a critical aspect of clinical research of complex diseases that can be implemented during the design instead of the analysis phases of clinical trials (Fig. 5.2d).

Additional characteristics of classification and response criteria that result in poor clinical trial representativeness and generalizability are listed in Table 5.1.

Addressing the Lack of Representativeness in Clinical Trials

The question of how to increase the external validity of clinical research of complex diseases without further decreasing productivity and increasing costs is as complex as the diseases that we are trying to address. Some options focusing in data collection and clinical trial design innovation are listed below and in Fig. 5.3**.**

Real-World Evidence

Big data analytics are an absolute requirement for artificial intelligence (AI) applications to continue to evolve to be able to design trials with higher external validity and to develop better approaches to patient selection, recruitment, and stratification.

Real-world data (RWD) are defined as any data related to the patient's health status that comes from sources other than traditional clinical trials. Common sources include electronic medical records (EMR), insurance claims, patient-generated

Table 5.1 Reasons for the lack of external validity in RCT – challenges and opportunities

	Challenges	Opportunities
Patient enrichment practices that decrease the representativeness of RCTs	Patient enrichment practices decrease the variability of the study patients but also the generalizability of the data. Some examples that may be part of inclusion/exclusion criteria or may unconsciously be implemented by investigators include: Practical: e.g., selecting patients who are more compliant or have more access to healthcare; selecting only patients who are naïve to therapy or have failed certain therapies Prognostic: e.g., selecting patients at high risk of disease progression or poor long-term outcomes Predictive: e.g., selecting patients who are more likely to respond to a therapy	Data-driven approaches can improve patient selection and recruitment Adaptive designs in which the sample sizes, design features, and even patient populations are modified in a prospectively planned manner according to information obtained during the course of the trial Enriching populations with well-defined endotypes to develop drugs that address discrete aspects of a systemic disease [15]
Diagnostic/ classification criteria are time limited and must be fulfilled at the time of enrollment	In clinical research, the diagnosis begins when patients are eligible to receive investigational drug. IMIDs are frequently relapsing-remitting in presentation, yet patients are only eligible to enter a study when they have "active" disease. Most RCTs do not provide information about the ability of therapeutic interventions to prevent disease onset or subsequent periods of disease exacerbation. This is often due to their short, cost-efficient duration. The concept of preventive trials in immunology has gained strength in recent years, but patient selection and recruitment have been difficult [16]	RWD collected before patients become eligible and after the trial concludes will generate RWE to increase our understanding of disease trajectories and other outcomes Trial designs with observational phases before and after trial initiation and use of RWD to complement RCT
Endpoints that are often binary, categorical, and quantitative	Clinical trial endpoints must be measurable within a reasonable period for all or a high proportion of patients, and prespecified thresholds of response must be met. Study participants either fulfill all endpoint criteria or are considered nonresponder. The therapeutic effects of novel classes of therapies may be different when administered at specific time points in the trajectory of disease progression, or they may take longer to appear in some patients	Continuous measures for evaluation of treatment outcomes, for example, time-to-event endpoints, area under the curve, or other overarching outcomes more frequently Incorporating observational periods at the end of a trial to follow patients who did not meet thresholds for response at prespecified endpoints

(continued)

Table 5.1 (continued)

	Challenges	Opportunities
Diagnostic/ classification tools are rarely used in real-world clinical practice	Investigators are expected to be proficient at applying complex and often unfamiliar classification and response tools that are not routinely used in clinical practice outside of highly specialized academic centers Most registration trials are conducted globally, and there are often language barriers and international variations in the use of diagnostic tools and access to care. Assuming that single-site/ investigator training sessions are sufficient to enable the trial can result in lower quality of data collection	Novel types of simple and efficient methods to evaluate disease activity and response to treatment are being developed in several IMIDs [17] Cloud-based technologies can improve the quality of investigator's trainings by simulating in-person meetings Wearables and other electronic data collection devices to limit the dependence on in-person visits and intermittent assessments
Rarely incorporate information on endotypes	IMIDs have multifactorial etiology, and patient populations are heterogeneous. Many aspects of the endotype (Fig. 5.1) influence both the trajectory of disease progression and the response to treatment with an investigational compound [18]. Yet, they are not always taken into account in early drug discovery or when defining clinical trials due to sample sizes, cost, lack of reliable biomarkers, and enrolment difficulties. We often rely on post hoc analyses of small subsets of patients to try to provide answers to questions that are relevant in the clinic where these patient endotypes can be common	Big data will eventually inform AI applications that can help take into account multiple variables in the analysis of response to treatment in discrete patient subgroups according to their underlying pathophysiology Supplementing RCT with RWD including molecular profiles using, for example, "master observational trials" [19]

data, and data from disease registries that can be enhanced with molecular annotation. These data can generate insights into pathophysiology mechanisms, unmet needs, disease phenotypes, and the clinical and economic impact of specific diseases on patients and healthcare systems. In addition, RWD of increasing quality will provide information on the real-world management of patient populations before and after exposure to the therapeutic intervention.

Data collection of this complexity is currently associated with high cost and analysis capability needs. However, the use of AI applications will eventually result in lower failure rates and better patient outcomes. Costs will continue to decrease over time as their use spreads and technology and data sharing practices improve [20]. Several companies have begun projects that allow data sharing from clinical trials, including those that have failed. Sharing molecular and phenotypic data from the placebo groups of all IMID trials should be encouraged.

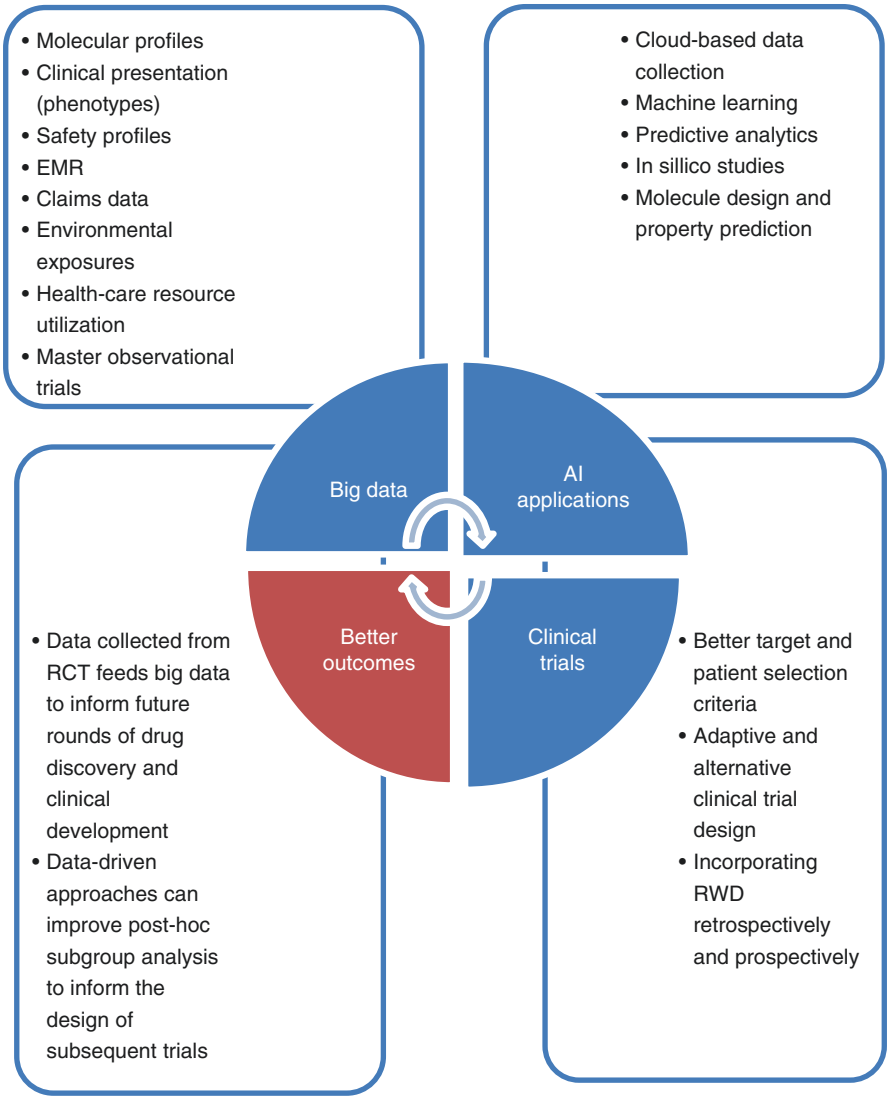

Fig. 5.3 The modern cycle of innovation in drug development**. Big data** from multiple sources, including RWD and existing RCT data, can fill knowledge gaps related to patient endotypes. This information can then be used to evolve **AI applications** such as machine learning, which can then be implemented to improve patient selection and stratification criteria in clinical trials and to discern the influence of these variables on responses to treatment with novel agents. **Clinical trial innovation** results in more high-quality data to fuel the cycle. This iterative cycle results in higher efficiency in drug development and **better outcomes**

There are limitations associated with certain types of RWD such as claims and EMR, including coding errors. Still, significant improvements in data entry and analysis have been made in recent years [21, 22]. It is also now possible to match

these datasets to multi-omics data, an approach that is already being implemented retrospectively and prospectively in large patient cohorts.

One of the challenges associated with the collection of RWD from observational studies is that some of these studies are sometimes designed using the same classification criteria for patient enrolment as a clinical trial. Reproducing clinical trial inclusion will not improve our understanding of natural history and disease endotypes. Except for drug registries which must follow on-label practices, we must design disease registries that can enrich the types of data collected, adding missing information such as molecular signatures [23].

It is also critical to harmonize the data collection [24] and be transparent in data disclosure to make it possible to compare or even integrate datasets from different studies. Data integration and collaborative efforts by which data from most patients with a disease are captured into a single database are ideal and achievable in many IMIDs [25].

Clinical Trial Optimization

The history of RCTs is relatively short. In fact, the essential requirements of a meaningful clinical trial that we use in modern drug development were not agreed upon until the 1980s [26]. It is time to re-evaluate how we measure the efficacy and safety of novel therapies in complex diseases and design cost-efficient trials with higher external validity.

The trial designs described below are not without challenges, but modern data-driven patient selection and recruitment methods will continue to increase their feasibility and likelihood of success. Importantly, they require flexibility in licensing practices and collaborative work between investigators, institutions, and pharma.

Trials Within Trials "Nested" and "clustered" clinical trials can maintain the traditional structure of an RCT while incorporating cohorts of patients with characteristics of special interest or by segregating patient groups into clusters based on their phenotype/endotype. By studying subgroups as part of a PoC[1] or exploratory cohort within the main trial, investigators can directly compare outcomes in subgroups of patients without the tremendous cost and complexity of a separate trial in such populations. The sample size calculations must be approached carefully to generate sufficient evidence to increase the likelihood of success of subsequent studies [27]. We have already seen a similar approach applied in SLE, perhaps inadvertently, after a post hoc analysis of a subset of patients with specific organ involvement in a pivotal trial resulted in a subsequent trial and approval in that subset of patients [28].

Adaptive Clinical Trials The FDA defines them as a clinical trial design that allows for prospectively planned modifications to one or more aspects of the design based on accumulated data from subjects in the trial. For example, this type of

[1] Proof-of-concept study: the minimum number of experiments/studies that provide critical data

design allows for the use of a statistical model to randomize patients to treatment arms according to the probability of higher efficacy by using ongoing response and biomarker data. In its guidance for industry on enhancing the diversity of clinical trial populations [12], the FDA states: "When developing clinical trial protocols, work to ensure that eligibility criteria serve the goal of having a representative sample of the population for whom the drug has been developed and examine each exclusion criterion to determine if it is needed to help assure the safety of trial participants or to achieve the study objectives. This includes the possibility of modifying the study population using adaptive clinical trial designs."

In Silico Trials Four of the FDA's top eight strategic priorities to improve R&D output are related to modeling and simulation. This is since the 2018 Appropriations Bill instructed the agency to expand the use of in silico clinical trials. Using computer simulations to predict outcomes on a virtual population is a method that can be applied at all stages of drug development [29], including clinical trial optimization. In silico approaches offer the possibility of analyzing billions of data points from the literature and patient data, including demographics, treatment history, multi-omics, and many more. Although this approach is in its infancy in IMIDs because it requires a good understanding of the disease state under study, in silico trials permit the assembly of disease models to explore early assumptions and optimize the design of RCTs. Several companies are already implementing this approach. In fact, the first FDA approval of a new drug indication under the model-informed drug development (MIDD) recently took place [30].

Observational to Interventional Trials Transitional trials that allow the investigators to follow the disease progression in a group of patients, enrolling them in the interventional portion of the trial when they meet specific criteria. This model provides pre-flare information and data on differences in response to treatment according to the natural history of the disease in individual patients or groups of patients. *Master observational trials* are a clinical trial modality that incorporates all these aspects [19].

RWD Within RCT As mentioned above, supplementing RCT data with real-world data will be an essential step toward a better understanding of patient populations, endotypes and long-term efficacy, safety, and healthcare resource utilization in patients treated with novel therapies. Observing patients before and after they participate in an interventional study would capture the full disease course and provide with valuable information about the impact of multiple aspects of the patient's history on responses to treatment.

Keeping the End in Mind: The Patient at the Center of the Drug Development Process

It is also critical to invest in patient-reported outcome research [31]. As patient organizations become more involved in drug development, they sometimes disapprove that physicians and scientists create definitions of disease and response to treatment without involving patients until drugs are ready to go into the market.

In a drug development environment that is more patient-centric than ever, this is evidently not due to a lack of interest by industry sponsors but instead, traditional drug discovery processes around the science, with research teams with no clinical expertise and minimal participation of physicians in early discovery efforts.

Modern drug development strategies take into account the patient's most bothersome symptoms in the process of identification of new drug targets and develop patient-reported outcomes (PROs) to evaluate trial feasibility and efficacy of new drugs [23, 32]. Novel methods to measure disease activity also take PROs into account [17, 31, 33].

Finally, our clinical development teams must remember that payers often use clinical trial criteria to restrict access to therapies. This affects patients' access to care, especially among patients with more complex disease presentations and higher diagnostic uncertainty who do not meet such criteria.

Conclusions

In conclusion, we need to incorporate diagnostic uncertainty into a drug development process that has been built around certainty, causing too many clinical trials in complex diseases to fail due to variability resulting from the heterogeneity of the patient population. We must learn from that heterogeneity.

Throughout this chapter, I have emphasized the need to innovate drug development by using big data to evolve an extensive range of artificial intelligence solutions that can increase volume and accuracy in data collection, improving our approaches to target and patient selection. In clinical stages, it is time to revisit clinical trial design to incorporate alternative and adaptive methodologies and evaluate patients in real-world settings. The goal is to complement data from RCT with big data that can help increase our understanding of disease pathophysiology, disease progression, and, eventually, disease prevention.

I acknowledge that many of these AI applications are not ready for prime time, especially those that could be used during clinical stages of drug development and in complex diseases, but the progress made in the last few years is remarkable. In fact, when one looks at the trends that will significantly impact pharmaceutical drug development in the future, the list includes novel biomarkers, PROs, real-world data, regulatory shifts, and AI applications such as predictive analytics and machine learning (Fig. 5.4).

New generations of academic scientists, patient organizations, and drug developers must raise awareness and collaborate in the collection and integration of big data

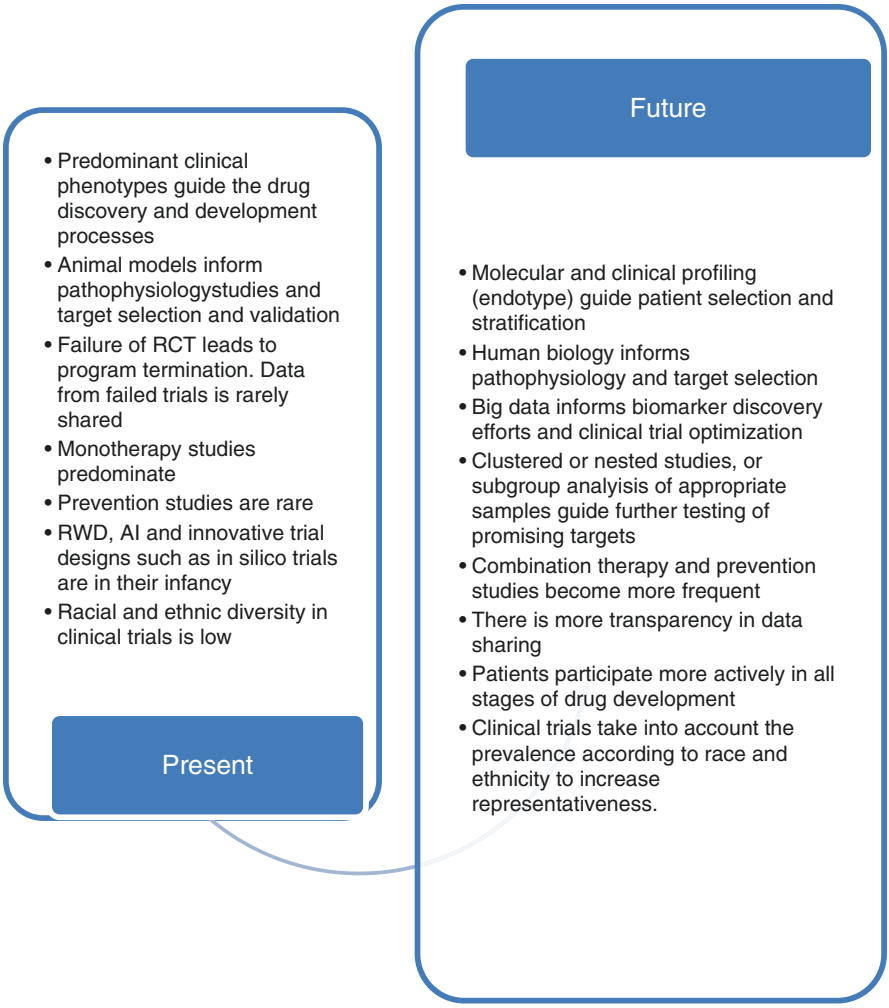

Fig. 5.4 Present and future of the drug development process in IMIDs. Clinical trial optimization and innovation and the growth of artificial intelligence applications and novel data analytic tools, including machine learning fueled by the influx of big data, will undoubtedly change how we identify and evaluate new drugs

that can fuel novel computational methods that can increase clinical trial success, bringing more novel treatments to patients in need [34]. We can only realize the impact of these technologies with strong collaboration between academia, pharma, regul*a*tors, payers, and, importantly, patients.

Disclaimer The opinions expressed in this chapter are those of the author and do not represent her employer's.

References

1. Mina-Osorio P. Next-generation therapies and technologies for immune-mediated inflammatory diseases, in progress in inflammation research. Cham: Springer; 2017. p. 1. online resource (XI, 202 pages 8 illustrations, 6 illustrations in color
2. Tsimberidou AM, et al. Review of precision cancer medicine: evolution of the treatment paradigm. Cancer Treat Rev. 2020;86:102019.
3. Catalina MD, et al. The pathogenesis of systemic lupus erythematosus: harnessing big data to understand the molecular basis of lupus. J Autoimmun. 2020;110:102359.
4. Mina-Osorio P. Review: basics of drug development in rheumatology. Arthritis Rheumatol. 2015;67(10):2581–90.
5. Irvine AD, Mina-Osorio P. Disease trajectories in childhood atopic dermatitis: an update and practitioner‘s guide. Br J Dermatol. 2019;181(5):895–906.
6. Dall'Era M, et al. Current challenges in the development of new treatments for lupus. Ann Rheum Dis. 2019;78(6):729–35.
7. Pena Y, et al. Establishing consensus understanding of the barriers to drug development in lupus. Ther Innov Regul Sci. 2020;54(5):1159–65.
8. Falasinnu T, et al. The representation of gender and race/ethnic groups in randomized clinical trials of individuals with systemic lupus erythematosus. Curr Rheumatol Rep. 2018;20(4):20.
9. Bhise V, et al. Defining and measuring diagnostic uncertainty in medicine: a systematic review. J Gen Intern Med. 2018;33(1):103–15.
10. Yazici H, Hatemi G, Yazici Y. Diagnostic criteria as separate from classification criteria: a perpetual motion machine? Comment on the article by Aggarwal et al. Arthritis Care Res (Hoboken). 2016;68(7):1049.
11. Aggarwal R, et al. Distinctions between diagnostic and classification criteria? Arthritis Care Res (Hoboken). 2015;67(7):891–7.
12. June RR, Aggarwal R. The use and abuse of diagnostic/classification criteria. Best Pract Res Clin Rheumatol. 2014;28(6):921–34.
13. Averitt AJ, et al. Translating evidence into practice: eligibility criteria fail to eliminate clinically significant differences between real-world and study populations. NPJ Digit Med. 2020;3:67.
14. Enrichment strategies for clinical trials to support determination of effectiveness of human drugs and biological products. Guidance for industry; 2019, US Department of Health and Human Services Food and Drug Administration.
15. Rovin BH, et al. Efficacy and safety of voclosporin versus placebo for lupus nephritis (AURORA 1): a double-blind, randomised, multicentre, placebo-controlled, phase 3 trial. Lancet. 2021.
16. van Boheemen L, et al. How to enhance recruitment of individuals at risk of rheumatoid arthritis into trials aimed at prevention: understanding the barriers and facilitators. RMD Open. 2021:**7**(1).
17. Askanase AD, et al. Development and content validity of the Lupus Foundation of America rapid evaluation of activity in lupus (LFA-REAL): a patient-reported outcome measure for lupus disease activity. Health Qual Life Outcomes. 2019;17(1):99.
18. Hingorani AD, et al. Improving the odds of drug development success through human genomics: modelling study. Sci Rep. 2019;9(1):18911.
19. Dickson D, et al. The master observational trial: a new class of master protocol to advance precision medicine. Cell. 2020;180(1):9–14.
20. The DiscovEHR partnership. 30 May, 2021; Available from: https://www.geisinger.org/precision-health/mycode/discovehr-project.
21. Mc Cord KA, Hemkens LG. Using electronic health records for clinical trials: where do we stand and where can we go? CMAJ. 2019;191(5):E128–33.
22. Stamford white paper: the future of electronic health records. Stamford Medicine White Papers; 2018.

23. Paller AS, et al. Protocol for a prospective, observational, longitudinal study in paediatric patients with moderate-to-severe atopic dermatitis (PEDISTAD): study objectives, design and methodology. BMJ Open. 2020;10(3):e033507.
24. Chalmers JR, et al. Report from the fifth international consensus meeting to harmonize core outcome measures for atopic eczema/dermatitis clinical trials (HOME initiative). Br J Dermatol. 2018;178(5):e332–41.
25. Foundation, T.M.M.R. The MMRF CoMMpass℠ Study. Available from: https://themmrf.org/finding-a-cure/our-work/the-mmrf-commpass-study/.
26. Junod SW. FDA and clinical drug trials: a short history. In: Davies M, Kerimani F, editors. A quick guide to clinical trials. Washington: Bioplan, Inc; 2008. p. 25–55.
27. Placzek M, Friede T. Clinical trials with nested subgroups: analysis, sample size determination and internal pilot studies. Stat Methods Med Res. 2018;27(11):3286–303.
28. Dooley MA, et al. Effect of belimumab treatment on renal outcomes: results from the phase 3 belimumab clinical trials in patients with SLE. Lupus. 2013;22(1):63–72.
29. [video] 2017 cited YouTube. Available from: https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/division-applied-regulatory-science.
30. Somberg JC, et al. Model-informed development of Sotalol loading and dose escalation employing an intravenous infusion. Cardiol Res. 2020;11(5):294–304.
31. NIH. PROMIS® (Patient-Reported Outcomes Measurement Information System). Available from: https://commonfund.nih.gov/promis/index.
32. Mosor E, et al. I would never take preventive medication! Perspectives and information needs of people who underwent predictive tests for rheumatoid arthritis. Arthritis Care Res (Hoboken). 2020;72(3):360–8.
33. Yosipovitch G, et al. Peak pruritus numerical rating scale: psychometric validation and responder definition for assessing itch in moderate-to-severe atopic dermatitis. Br J Dermatol. 2019;181(4):761–9.
34. Gajra A, Zettler ME, Feinberg BA. Randomization versus real-world evidence. N Engl J Med. 2020;383(4):e21.

Confronting the Inevitability of Diagnostic Uncertainty Across Multiple Legal Domains

6

Lars Noah

If the names are unknown,
knowledge of the things also perishes.
—Linnaeus.[1]

Introduction

Do legal institutions share any of the blame for an intolerance toward diagnostic uncertainty? For instance, courts may resolve claims of negligent misdiagnosis in malpractice litigation, and they often must decide whether physicians may testify about diagnostic judgments as expert witnesses in all manner of personal injury lawsuits. Regulatory agencies must make risk-benefit judgments about therapeutic products seeking licensure in ways that implicate the manner of labeling diseases as do efforts at public health surveillance. Programs for dispensing disability benefits and statutory protections against discrimination also regularly look to the judgments of physicians. Ultimately, however, most of these contexts tolerate a good deal of diagnostic imprecision. Resistance to uncertainty springs primarily from health insurers, and government-run insurance programs undoubtedly enjoy tremendous clout; moreover, even when largely left to private ordering, courts may become involved when disputes arise over the language in contracts between insurers and patients or providers.

[1] Carolus Linnaeus, Philosophia Botanica, aphorism 210 (1751). Although remembered primarily as a botanist, the father of modem taxonomy in the biological sciences also had trained as a physician and included diseases among his subjects for classification: Carolus Linnaeus, G*enera* M*orborum* (1763).

L. Noah (✉)
Levin College of Law, University of Florida, Gainesville, FL, USA
e-mail: noah@law.ufl.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_6

I first tackled these issues more than 20 years ago, identifying the numerous ways that legal institutions influenced the diagnostic process [1]. As explained in that lengthy article, agencies and courts may create various incentives for assigning disease labels to conditions and claimants, which had encouraged both nosologic creativity and diagnostic dishonesty—in effect, when a non-illness gets a (disease-like) name, which represents the flipside of the issue that we now have before us. Indeed, legal institutions seemed far too accepting of novel syndromes and dubious diagnoses, so I urged them to become more critical consumers of what biomedical researchers and healthcare professionals had to offer in the course of certifying the nature and severity of parties' complaints.

In the last couple of decades, hints of skepticism have started to appear in court decisions, but judges and administrative officials hardly became terribly restrictive in demanding that claimants demonstrate unimpeachable diagnoses of widely accepted diseases before offering recourse for their alleged illnesses and injuries. The general public also seems to have become more cognizant of diagnostic uncertainty, whether from first-hand experiences as frustrated patients or vicariously when major newspapers and popular television programs draw attention to cases of mysterious illnesses, and such awareness cannot but help affect the resolution of disputes that land in the courts.

Preliminary Thoughts Prompted by the Pandemic

Extensive media coverage of the novel coronavirus has offered us a stark illustration of the uncertainties and difficulties that society must confront when new contagious threats emerge. We have learned, for example, that 5 years earlier the World Health Organization (WHO) adopted nomenclature guidelines designed to guard against stigmatizing particular animal reservoirs (e.g., bird or swine flu), people (e.g., Legionnaires' disease), or points of origin (e.g., Middle Eastern Respiratory Syndrome (MERS), West Nile virus, Lyme disease, Ebola, Zika, or even the Spanish flu pandemic of over a century ago). This gave us the somewhat less memorable "COVID-19" moniker for the latest scourge, caused by the novel infectious agent now designated as SARS-CoV-2 so as to distinguish it from the coronavirus originally associated with severe acute respiratory syndrome (SARS) in 2003. After the variants that initially emerged became known by their countries of origin, the WHO called for the use of non-stigmatizing letters of the Greek alphabet (e.g., Delta and Omicron) while subvariants follow the Nextstrain clade designations for mutations (e.g., BA.2.12.1).

In a sense, assigning a label to persons infected with the novel virus represented the easy part. Understanding the disease has proven to be far more difficult. An ability to diagnose COVID-19 (i.e., identify positive cases) hardly answers more fundamental questions about its etiology (e.g., routes of transmission and patient risk factors), which then drive public health interventions, or more meaningful questions about its prognosis (e.g., likely course of the illness and potential treatments), which will impact choices about clinical care. Already, one perplexing constellation of

symptoms found in rare pediatric cases, which resembled Kawasaki disease, required a new label: "multisystem inflammatory syndrome in children" (MIS-C). Its connection to COVID-19 remains, however, a mystery. In addition, some patients who recover from the infection have experienced the so-called long haul symptoms, which now goes under the more formal and cumbersome banner "post-acute sequelae of SARS-CoV-2" (PASC).

Contested Vital Signs, Pre-diseases, and Beyond

For a far more mundane version of this issue, consider the continued uncertainty about appropriate thresholds for basic vital signs. For instance, disagreement persists about what qualifies as normal body temperature. More than a century ago, a convention developed of treating 37°C (or 98.6°F) as normal and 38°C (or 100.4°F) as fever, leaving ambiguity about how to characterize readings that fall between those somewhat arbitrary points. Indeed, newer research suggests setting the thresholds a bit lower. Even trivial variations (e.g., 99° vs. 100°F) can have profound impacts on public health efforts—such as screening of travelers and others during a pandemic—or insofar as fever and other vital signs appear among the criteria embedded in schema used to classify complex diseases.

Similarly, blood pressure has become a moving target. In 2017, the American College of Cardiology and the American Heart Association revised their guidelines: in the past, readings below 140/90 mm Hg counted as normal (or at least not worrisome); now, these organizations treat only those readings below 120/80 as normal (unless they go too low), while higher readings that remain below 130/80 would qualify as "elevated" blood pressure, below 140/90 as "Stage I" hypertension, and readings that exceed either of the old systolic or diastolic thresholds would count as "Stage II" hypertension (and, if those readings exceeded either 180 or 120 respectively, it represented a hypertensive crisis). More recently, evidence of gender differences suggests further lowering these thresholds for women. These sharp demarcations assume, of course, accuracy in readings, but research has found a tendency to overstate blood pressure when this vital sign gets measured during routine patient checkups.

More importantly, hypertension has transitioned from an asymptomatic risk factor in heart disease and stroke to become a condition worthy of treatment in its own right, and elevated blood pressure readings that do not quite make the cut have joined a growing list of so-called pre-diseases. Even if the identification of, let us say, "pre-hypertension" or "pre-diabetes" does nothing more than encourage patients to make healthy lifestyle changes, such labels also may cause undue anxiety in those supposedly afflicted. No doubt rebranding carcinoma as "Stage 0" cancer may get people to take more seriously a possible precursor to a genuine malignancy, but the hype can also harm patients if it prompts unnecessary treatment of benign conditions.

Although not formally regarded as a vital sign, health professionals routinely measure the height and weight of their patients, and these together allow for

calculation of a body mass index (BMI), which had originated as a research tool rather than a diagnostic measure designed for clinical use. The latest numbers from the US Centers for Disease Control and Prevention (CDC) indicate that more than 40% of adults exceed the BMI thresholds for obesity, and almost 10% qualify as morbidly obese. Rather than amounting to a symptom of certain metabolic disorders (or, more typically, inactivity coupled with excessive caloric intake), or qualifying as a risk factor (in the development of a variety of other diseases such as Type II diabetes), obesity now evidently counts as a disease or condition in its own right. A number of entities have jumped on this bandwagon, including the American Medical Association (AMA), various federal public health agencies, and even the Internal Revenue Service (IRS). Moreover, thanks to this re-characterization, the US Food and Drug Administration (FDA) apparently has become more willing to approve prescription of weight-loss products, though it recently had to withdraw another drug in this category because of serious new safety concerns.

These nosologic inventions aside, placing patients within the relevant boxes generally does not present any serious difficulties. Blood pressure readings, BMI tables, and laboratory tests for blood sugars and so forth can situate an individual within the arbitrary thresholds that separate normal, pre-disease, and disease states. In addition, even as genomics has upended conventional understandings of cancer, it also has come with the tools necessary to differentiate between various malignancies that previously had instead focused solely on tissue site.

Emphasizing Etiology and Prognosis

Legal institutions generally care less about a patient's present diagnosis than closely affiliated questions regarding etiology (past) and prognosis (future). Thus, an illness without a name poses far less of a difficulty than an illness without a known cause—instead of "atypical" forms of a recognized disease, it becomes a problem of designating as "idiopathic" a disease of uncertain origin. In contrast, labeling an illness as "iatrogenic" (or, better yet, drug-induced) helpfully builds an assessment of causality right into the diagnostic label.

Similarly, forecasting the uncertain trajectory of a known illness in terms of its likely duration and severity becomes essential when courts award damages to cover anticipated future harm or make disability determinations. A firm disease label will, no doubt, assist in making such predictions because the available research base has sorted the afflicted according to such arbitrary boxes, but judges grant significant latitude to medical experts when they testify about both etiology and prognosis. Even deciding whether a patient qualifies as "terminally ill" (conventionally understood as having less than 6 months left to live) may have importance in making judgments about end-of-life options pursuant to state laws governing the right to try still investigational drugs or authorizing physician aid-in-dying, but clinicians can only offer best guesses in making such prognoses. Of course, etiology and prognosis represent essential aspects of the broader diagnostic enterprise—diagnosis attempts to discern the underlying cause of a patient's

complaint, while etiology seeks to identify the (external) cause of that (internal) cause. Conflict may arise, however, when medical experts attempt to paper over difficulties associated with these discrete inquiries by resorting to nosologic games or offering clever diagnoses.

Demanding Diagnostic Prowess

In one limited context, legal institutions expect an arguably unrealistic degree of diagnostic precision. Physicians may face liability for negligently misdiagnosing their patients. Typically, this category of malpractice claims focuses on allegedly unreasonable failures to diagnose a condition later detected in a patient. Less frequently, false positives (such as incorrectly advising a patient that she/he tested HIV+) may prompt litigation, though primarily seeking emotional distress damages. Because hindsight cannot but help to heavily influence the resolution of such cases, courts may end up demanding an unattainable level of diagnostic acumen from physicians. If, however, healthcare professionals and researchers take care to eschew undue precision when formulating diagnostic criteria in favor of retaining flexibility, then this wiggle room may better guard against the prospect of getting unfairly second-guessed should litigation subsequently arise.

If a malpractice plaintiff managed to produce sufficient expert testimony demonstrating that a reasonable health professional would have made the correct diagnosis, then the defendant would owe compensation for any injuries caused to the patient by the delay in learning the true nature of their affliction. Historically, this latter requirement made recovery unlikely for those patients who faced poor odds when originally seeking out medical care, which meant that they would find it impossible to prove that the delayed diagnosis probably caused their ultimate adverse outcome even if it further reduced their already low likelihood of avoiding it. In the past few decades, however, a majority of jurisdictions have changed their rules to allow for partial recoveries in such "loss-of-a-chance" cases, making negligent misdiagnosis claims more enticing for victims or their estates to bring when their initial medical exams missed something important. The central question, however, does not depend on whether the patient had any particular type of cancer, or indeed what may have originally caused it; instead, the resolution of these cases turns on offering reliable prognoses at different stages of the disease [2].

In other respects, tort doctrines seem more willing to embrace diagnostic uncertainty. In malpractice litigation, for example, custom generally defines the standard of care, which tolerates departures from evidence-based medicine (EBM), including deviations from FDA-approved labeling for therapeutic products [3]. In lawsuits against manufacturers of prescription drugs and devices, special rules exist to protect their design choices in recognition of the inevitable variability among patients for whom healthcare professionals might select such treatments [4]. In short, most courts understand that, when asked to evaluate therapeutic encounters after the fact, one size doesn't fit all.

Expertise and Etiology

When physicians seek to testify as experts in litigation, however, a peculiar old rule continues to surface. They often must couch their opinion about a contested issue in terms of "a reasonable degree of medical certainty" or words to that effect. Although this inexplicable requirement has attracted its share of criticism, it persists in many jurisdictions. Nonetheless, assuming that medical experts parrot such words when testifying, courts show them tremendous deference; indeed, physicians have largely managed to escape the increasingly stringent rules of admissibility applied to other types of experts [5].

Even so, when it comes to requirements for establishing causation in tort litigation or other legal contexts, the diagnostic enterprise has encountered some judicial resistance. For instance, the federal courts recently rejected lawsuits alleging that the Mirena® (levonorgestrel) IUD caused pseudotumor cerebri, also known as idiopathic intracranial hypertension (IIH), a rare and serious condition marked by increased cerebrospinal fluid pressure in the skull. Two decades earlier, after epidemiologists debunked the alleged link between silicone-gel breast implants and various autoimmune diseases, personal injury lawyers recast their clients' injuries as "atypical" forms of these same illnesses. Courts rebuffed this maneuver, though usually not because of any real questions about the correctness of these revised diagnoses; instead, they held that the plaintiffs could not demonstrate that the use of the implants had in fact caused atypical variants of traditional diseases. More recently, some have relabeled the condition as "breast implant illness" (BII), which the FDA has described as involving a variety of vague symptoms such as fatigue, memory loss, rash, "brain fog," and joint pain. Unlike atypical lupus or scleroderma, BII builds an assumption of causation right into the label, but at least for now it seems even further removed from any recognizable disease.

Similarly, labels such as "asbestosis" and "silicosis" denominate otherwise indistinguishable forms of pulmonary fibrosis based largely on exposure to a particular agent in the workplace, thereby ascribing causality in the very name of these conditions. Indeed, many occupational diseases share this tendency, but it can appear in other settings as well. For instance, the CDC began keeping tabs on what it has designated as "e-cigarette, or vaping, product use associated lung injury" (EVALI), which no doubt will soon find its way into tort litigation against sellers of these devices. Still more recently, researchers have christened rare cases of blood clotting associated with certain vaccines against COVID-19 as "vaccine-induced immune thrombotic thrombocytopenia" (VITT).

Perils in Prognostication

Courts have expressed skepticism about other efforts to use clever diagnoses as an end run around doctrines governing compensation, though these do not involve problems of establishing causation so much as identifying a present—rather than a possible future—injury. For example, in the course of granting summary judgment

to the manufacturer of the injectable contraceptive drug Depo-Provera® (medroxyprogesterone acetate) on failure-to-warn claims, a pair of courts concluded that "osteopenia" (a.k.a. pre-osteoporosis) does not qualify as an injury or disease. Similarly, at least three federal appellate courts have confronted claims that workers had developed "beryllium sensitization" (BeS), a precursor to chronic beryllium disease (CBD), but two of them rejected the lawsuits as lacking proof of any compensable damages. In effect, these judges declined to allow diagnoses of a mere predisposition as a means for eliding uncertainties in making a prognosis about the likelihood of developing the actual disease long after exposure.

Nonetheless, such pushback remains the exception rather than the rule, as evidenced by the growing willingness of courts to recognize the still vaguer and more all-encompassing "multiple chemical sensitivity" (MCS) label, though typically in contexts other than tort litigation. In a related vein, rather than getting tangled up in heated nosologic debates about "Gulf War Syndrome," Congress decided to simply amend the VA statute to authorize disability payments under certain circumstances to Persian Gulf veterans "suffering from a chronic disability resulting from an undiagnosed illness." Post-traumatic stress disorder (PTSD), though among civilians rather than service members, offers the clearest illustration from tort litigation: plaintiffs asserting claims for the negligent infliction of emotional distress—or pain and suffering damages resulting from a physical injury—have found courts more receptive when presented with a PTSD label [6]. In these cases, the diagnostic label does double duty: in addition to certifying the genuineness of a present emotional injury that otherwise defies corroboration, PTSD supplies the etiological basis for establishing causation. Along similar lines, "post-traumatic fibromyalgia syndrome" has begun to turn up as a forensic diagnosis in tort litigation.

Inadvertent Consequences of Undue Inventiveness

Judicial willingness to tolerate nosologic creativity might, however, come back to haunt the personal injury bar. If courts began treating pre-diseases as present injuries, then the statute of limitations may begin running once someone should have discovered that they suffered from such a supposed affliction; if, instead, they waited until manifesting a genuine disease many years later, then the resulting delay in filing a lawsuit may lead to the dismissal of their claims.

A similar dynamic might confound judicial resolution of insurance coverage disputes. Policies frequently depend on the diagnosis of some disease. More than a quarter of a century ago, the Nebraska Supreme Court held that "breast-ovarian carcinoma syndrome" represented an illness under the terms of a health insurance policy and, therefore, that radical prophylactic surgery should have been reimbursed as medically necessary. The court rejected the health insurer's argument that this syndrome amounted to nothing more than a genetic predisposition—diagnosed from nothing more than the patient's family history of breast and ovarian cancer—for the development of a disease at some indefinite point in the future. Although the syndrome was not based on any detected chromosomal abnormality, the court noted

that women with the plaintiff's family history had at least an even chance of eventually developing one of these cancers.

As genetic screening becomes more sophisticated, issues of this sort will become even more complicated. Indeed, insurers might attempt to turn the tables on policyholders with a previously identified predisposition to a certain disease by denying coverage for the treatment of the actual manifestation of that disease as a preexisting condition. The Affordable Care Act prevents health insurers from behaving in this fashion, but its fate remains in doubt, and the law does not bar such a practice with respect to other types of insurance.

Looking Beyond the Courtroom

Insurers and others may look to disease classifications and diagnostic criteria designed for public health surveillance purposes. When the WHO announced the labeling of COVID-19, or the CDC inaugurated EVALI, they did so in order to assist with the tracking of emerging health threats; similarly, public health agencies may denominate strains of infectious diseases based on their lack of responsiveness to available treatment options, such as methicillin-resistant Staphylococcus aureus (MRSA), vancomycin-resistant enterococci (VRE), and multi-drug-resistant (MDR) tuberculosis, or worse still when it becomes extensively drug-resistant (XDR). In effect, treatability becomes an aspect of the prognosis in the course of defining such illnesses. Over time, the diagnostic criteria used for surveillance purposes may become more refined—as happened, for instance, with the CDC's tracking of acquired immune deficiency syndrome (AIDS). Conversely, the WHO periodically produces the *International Classification of Diseases* (ICD), now in its 11th edition, and the *ICD-11* continues the proliferation of codes seen with its immediate predecessor. Public and private health insurers have raised the stakes associated with these undertakings by deciding to tie coverage decisions to such manuals.

The federal government has expressed enthusiasm about "precision medicine," which promises (among other things) to see that the right drug gets to the right patient. The labeling for pharmaceutical agents has begun to reflect genetic markers for predicting efficacy. Although pharmacogenomics may facilitate targeted drug development, we cannot forget about the "imprecise" multitudes that also need medical care [7]. Even before the push for personalized medicine, restrictive enrollment criteria made access to clinical trials dependent on first getting a fairly precise diagnosis. Moreover, women of reproductive age traditionally could not enroll in clinical trials, which aligns with the view in some preventive health circles that this population should be regarded as "pre-pregnant."

As desperate patients have clamored for access to promising investigational products, legislators and regulators have created various avenues—including "compassionate use" and treatment INDs—for those deemed ineligible to enroll in ongoing clinical trials, but we need to do a better job of collecting and considering this messy data. In 2016, Congress directed the FDA to consider using "real world"

evidence to support post-approval revisions to product labeling. For an agency accustomed to demanding randomized controlled trials, such an initiative represented a potentially dramatic departure from its ingrained habits. Then again, in its post-market surveillance activities, the FDA has long confronted precisely such indeterminacy in watching for red flags about previously licensed therapeutic agents; moreover, when they submit MedWatch reports of suspected adverse events, physicians tend to more candidly communicate ambiguity than when they confidently record their diagnoses in patient charts.

Insofar as legal institutions continue to permit the off-label use of therapeutic products, they tolerate diagnostic uncertainty. Although health insurers have resisted this practice to some extent, a far greater threat to physician autonomy comes from the prospect of direct restrictions on off-label use, which would, for instance, prevent the selection of a pharmaceutical agent for a patient in the absence of a qualifying diagnosis. The opioid crisis recently has prompted some states to move in this potentially worrisome direction [8].

Final Musings About Trans-scientific Enterprises

Taxonomists devote their energies to sorting and labeling natural phenomena. Even as genomic sequencing has fundamentally changed the manner of identifying different species of animals and plants, the goal remains the same. Putting aside the occasional hybrids, chimeric forms, and mutations, natural scientists persist in their classificatory exercise. Moreover, among members of the same species, we still routinely differentiate between male and female (or juvenile and adult) in spite of growing recognition that these binary categories do not capture the full range of sex identification (or developmental maturation).

It should come as no great surprise then that taxonomic efforts in medicine can only go so far. After all, illnesses are hardly static; they may have life cycles just like individuals do, which means that a diagnosis offers only a snapshot of what a patient may experience over the course of their disease. Legal institutions frequently need help answering medical questions that look further backward and forward in time, and it is here that firm diagnostic labels (when available) have a good deal less to offer.

Lastly, the contested labeling of diseases hardly exhausts the range of terminological disagreements that may affect modern medical practice. For instance, and somewhat remarkably, embryologists and others working in the field of reproductive medicine evidently cannot settle on precise labels for different stages of early human development. For lawyers, these choices may impact the resolution of a dizzying array of issues: constitutional personhood, statutory restrictions related to reproductive medical practice, the applicability of regulatory protections for research subjects, licensing of products used before and during pregnancy, the availability of public funding for research, insurance coverage, patentability, the capacity to assert personal injury claims, and the disposition of cryopreserved reproductive tissues upon death or divorce. Advocates and decision-makers will latch on to

whatever labels best suit their particular purposes; if, however, the varied options supplied by the research community do not come in handy for some nonscientific purpose, then these legal and political actors will simply make up something else. Obviously, a feedback mechanism means that these institutions will influence what would otherwise represent an entirely dry exercise in physiology (akin to taxonomy), but rising above the fray and achieving a scientific consensus on such terminology will hardly resolve the many contentious questions confronted in other domains [9].

References

1. Noah L. Pigeonholing illness: medical diagnosis as a legal construct. Hastings Law J. 1999;50(2):241–307.
2. Noah L. An inventory of mathematical blunders in applying the loss-of-a-chance doctrine. Rev Litigation. 2005;24(2):369–408.
3. Noah L. Medicine's epistemology: mapping the haphazard diffusion of knowledge in the biomedical community. Arizona Law Rev. 2002;44(2):373–466.
4. Noah L. This is your products liability restatement on drugs. Brooklyn Law Rev. 2009;74(3):839–926.
5. Noah L. Letter. Standards for medical expert testimony. JAMA. 2002;288(23):2971.
6. Noah L. Comfortably numb: medicalizing (and mitigating) pain-and-suffering damages. Univ Michigan J Law Reform. 2009;42(2):431–80.
7. Noah L. The coming pharmacogenomics revolution: tailoring drugs to fit patients' genetic profiles. Jurimetrics J. 2002;43(1):1–28.
8. Noah L. State regulatory responses to the prescription opioid crisis: too much to bear? Dickinson Law Rev. 2020;124(3):633–67.
9. Noah L. A postmodernist take on the human embryo research debate. Connecticut Law Rev. 2004;36(4):1133–61.

The FDA and the Drug Development Process

7

Allan Gibofsky

Introduction

The FDA can trace its origins back to the creation of the Agricultural Division in the Patent Office in 1848 [1]. Its origins as a federal consumer protection agency began with the 1906 Pure Food and Drug Act [2], which was passed by Congress in response to public outcry about the unhygienic conditions in the Chicago stockyards described in Upton Sinclair's book *The Jungle* [3]. As a result, the Agency (then called the US Department of Agriculture Bureau of Chemistry) initially focused on regulating the interstate transport of food which had been "adulterated"; however, the Agency also had the authority to regulate the interstate marketing of "adulterated" drugs, in which the "standard of strength, quality, or purity" of the active ingredient was not either stated clearly on the label or listed in the official United States Pharmacopeia. The act also banned "misbranding" of food and drugs. In 1927, the Bureau of Chemistry's regulatory powers were reorganized under a new USDA body, the Food, Drug, and Insecticide organization, shortened to the Food and Drug Administration (FDA) 3 years later.

In 1937, over 100 people died after using a drug formulated with a toxic, untested solvent [4]. In response, Congress passed the Food, Drug, and Cosmetic Act (FD&C Act) into law on June 24, 1938 [5]. This increased federal regulatory authority over drugs by mandating a pre-market review of the safety of all new drugs, as well as banning false therapeutic claims in drug labeling. This law has been extensively amended in subsequent years; however, it remains the current central foundation of FDA regulatory authority.

A. Gibofsky (✉)
Weill Cornell Medicine, Hospital for Special Surgery,
New York, NY, USA
e-mail: gibofskya@hss.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_7

The specific regulation of drugs is the division of the FDA known as the Center for Drug Evaluation and Research (CDER). The specific role of CDER is to ensure that drugs marketed in this country are safe and effective. Contrary to the belief of some, CDER does not develop or test drugs (although the Center's Office of Testing and Research does conduct limited research in the areas of drug quality, safety, and effectiveness). As will be discussed below, it is the responsibility of the company seeking to market a drug to conduct preclinical and clinical trials to demonstrate that the drug is both safe and effective.

Stages of Drug Development and Review

The steps of drug development and review occur in nine sequential steps: new idea/basic research, non-clinical data, investigational new drug application (IND), phase 1 (toxicity testing), phase 2 (efficacy testing), phase 3 (at least two controlled studies demonstrating low risk/benefit ratios), approval of new drug application/biologic (NDA) or biologics license application (BLA), detailed risk management and mitigation strategy (REMS), and phase 4 (post-release follow-up for complications). Details of these steps follow.

Investigational New Drug Application (IND)

A company seeking to develop a drug for human use must first do preliminary preclinical studies of toxicity and pharmacokinetics in an appropriate animal model. At this point, the company will submit an investigational new drug application (IND), outlining the proposed plan for clinical testing in humans. The FDA will then decide if it is reasonably safe to proceed.

Clinical studies in humans cannot begin until the IND is approved by the FDA. In addition, the FDA will require review by an Institutional Review Board (IRB), an independent panel usually composed of both scientists and non-scientists and charged with oversight of clinical research. The IRB determines if the proposed study meets both scientific and ethical standards, including that participants have given appropriate consent, are fully informed of the risks of participation, and that the researchers conducting the study have procedures in place to protect harm to patients, to the extent possible.

Phase 1

These studies are usually conducted in healthy volunteers, the objective being to determine side effects and human pharmacokinetics. These studies are small, usually ranging from 20 to 80 subjects. The emphasis of phase 1 studies is safety.

Phase 2

Assuming that phase 1 studies do not reveal unacceptable toxicity, the company may move to phase 2 studies, where the emphasis is on effectiveness. In this phase, the company obtains preliminary data on whether the drug works in people who have a certain disease or condition. Phase 2 studies are almost always controlled trials, i.e., patients receiving the drug are compared with similar patients receiving a different treatment, either an inactive substance (placebo) or a different drug. Phase 2 studies continue to evaluate safety, as well as short-term side effects are studied. Typically, the number of subjects in phase 2 studies ranges from several dozen to a few hundred. At the end of phase 2, the company and the FDA will meet to come to an agreement on the protocol for larger studies in the next phase.

Phase 3

If phase 2 studies demonstrate efficacy and no unacceptable safety signals are seen, phase 3 studies may begin. Phase 3 studies continue to obtain data on more both safety and effectiveness. In this phase, different populations and different dosages are often tested, as well as using the drug in combination with other drugs. The number of patients in a phase 3 usually ranges from several hundred to several thousand.

New Drug Application (NDA)

Assuming that the efficacy to safety ratio is acceptable in phase 2, the company will then file a new drug application (NDA) (if the drug is a biologic agent, the application is called a BLA or biologics license application). This is a formal request that the FDA review all of the preclinical and clinical data collected in phases 1, 2, and 3 and approve the drug for marketing in the United States. This application will also contain the specifics of manufacturing and packaging.

Once submitted, the FDA has 60 days to decide whether to accept it for review. The FDA can refuse to accept an NDA that it deems incomplete or inadequate. When accepted, the NDA is reviewed by the FDA in detail, in accordance with the provisions of the Prescription Drug User Fee Act (PDUFA). The FDA expects to review and act on at least 90 percent of NDAs for standard drugs no later than 10 months after the accepted application.

Phase 4: Postmarketing Requirement and Commitment Studies

A company may be required to do one or more studies by the FDA as conditions of approval and marketing. Postmarketing safety studies may also be done as part of a voluntary overall scientific plan by the company or by independent

investigators. Both required and independent investigator postmarketing studies continue to gather additional information about a product's safety and efficacy.

What Does FDA Approval Mean?

FDA approval means that the preclinical and clinical trial data submitted to the FDA has been reviewed by internal experts (and often by external experts as well) and that the clinical benefits of use outweigh the possible risks. As a general rule, the FDA requires the company that filed the NDA to submit results from (at least) two well-designed clinical trials in the population for which the drug is intended for use. This requirement is to eliminate the possibility that the results of the first trial are not the result of chance. (For rare diseases, the Agency may accept data from only one clinical trial, if the evidence of safety and efficacy is convincing.)

Once approved, the drug must then get a "label." The wording of the drug label is the result of extensive discussions between the FDA and the company. At a minimum, the approved label must include the approved indications for which the Agency has approved use ("the on-label indications") and the risks and benefits of the drug. In some instances, a more detailed risk management and mitigation strategy (REMS) may be required. Among the additional elements that the Agency may require are limited distribution by designated suppliers, prescriber certification, patient education programs, and a mandatory registry of all patients to whom the drug has been prescribed.

In many instances, drugs are prescribed for conditions OTHER than those listed in the approved label. This is known as "off-label" use, and while this practice is not illegal for the prescriber, the company may not market, encourage, or otherwise support this practice.

Accelerated Approval

In some cases, the approval of a new drug may be expedited if the company can show that the drug will treat a serious or life-threatening condition and that it provides lower risk and/or higher therapeutic benefit than currently used therapies. This "accelerated approval" [6] pathway may be requested by the company if the drug is meant to treat a chronic disease and thus an extended period of time is needed to fully determine its safety and efficacy. If approved via this pathway, the FDA will generally require postmarketing clinical trials as well.

The accelerated approval pathway was established in 1992. Since then, many drugs have successfully been brought to market this way to treat life-threatening diseases and conditions. These include many of the antiretroviral drugs used to treat HIV/AIDS as well as oncologic agents.

Drug Development Designations

There are several approaches used by the Agency to encourage the development of certain drugs, especially those that may represent a novel available treatment for a disease or condition, or those thought to have a significant benefit over existing drugs. These approaches, or designations, are meant to address specific clinical needs, and a new drug application may receive more than one designation, if appropriate. They consist of:

1. Fast track [7], a process designed to facilitate the review of NDAs. The company can request the fast track process if it can show that the data submitted will fill a significant unmet medical need tracking can get important new drugs to the patient earlier.
2. Breakthrough therapy [8] designation can be requested by the company for the review of drugs that are intended to treat a serious condition, and preliminary clinical evidence indicates that the drug may demonstrate substantial improvement over available therapy. A drug with breakthrough therapy designation is also eligible for the fast track process.
3. Priority review [9] means that the FDA has committed to review the IND and render a decision within 6 months (as compared to an average of 10 months for a standard review).

Summary

The FDA is charged with reviewing the data submitted from pharmaceutical manufacturers seeking approval to develop and market drugs to patients in the United States. Consistent with its charge from Congress, the Agency has developed a series of well-publicized review processes, thus ensuring that appropriate standards of safety and efficacy are met for any drug submitted for approval. In this manner, the FDA fulfills its historical mission.

Conflict of Interest Consultant- AbbVie, Acquist Pharma, Lilly, Novartis, Pfizer, Samumed
Speaker-AbbVie, Amgen, Lilly, Pfizer
Stock Ownership (joint with spouse)- AbbVie, Johnson& Johnson, Pfizer, Regeneron
Royalties-UptoDate

References

1. https://www.fda.gov/about-fda/fda-history. Accessed 30 Mar 2021.
2. Pure Food and Drug Act. United States statutes at large (59th Cong., Sess. I, Chp 3915); 1906. p. 768–72.
3. Sinclair, Upton. The Jungle. Project Gutenberg. Retrieved 8 Feb 2021.
4. https://www.fda.gov/files/about%20fda/published/The-Sulfanilamide-Disaster.pdf. Accessed 10 Feb 2021.

5. Federal Food, Drug, and Cosmetic Act, 21 U.S.C. §§ 301–392 (Suppl. 5 1934).
6. 21 CFR 314.510.
7. https://www.fda.gov/patients/fast-track-breakthrough-therapy-accelerated-approval-priority-review/fast-track. Accessed 2 Apr 2021.
8. https://www.fda.gov/patients/fast-track-breakthrough-therapy-accelerated-approval-priority-review/breakthrough-therapy. Accessed 2 Apr 2021.
9. https://www.fda.gov/patients/fast-track-breakthrough-therapy-accelerated-approval-priority-review/priority-review. Accessed 2 Apr 2021.

Part III

Assigning

8 Diagnosis of Systemic Lupus Erythematosus in the Age of Precision Medicine

Sule Yavuz and Peter E. Lipsky

Introduction

The concept of diagnosis, or the art or act of identifying a condition from its signs and symptoms, has been discussed for nearly 5000 years. In fact, the concept of diagnosis antedated that of medicine. It was well developed before there was any relationship to treatment and was, in fact, originally more related to prognosis. Since the development of scientific-based medicine following the Flexner Report of 1910, diagnosis has been the centerpiece of medicine, and we recognize many great physicians because of their diagnostic skills [1, 2].

Of all medical disciplines, rheumatology especially prided itself as the bastion of great diagnosticians. We were often the last resort of patients with an array of hard-to-explain signs and symptoms and can often affix a label to a patient, even if the label is only a restatement of the array of signs and symptoms without much in the way of physiologic or therapeutic implication. In addition, sometimes when a diagnosis could not be made, frustration may have resulted in discounting or even stigmatizing the patient. Over the past few decades, diagnosis has become more complex with an array of molecular and imaging technologies that have often changed medical nosology and affected the approach to diagnosis. In addition, the advent of newer and more effective targeted therapies has increased the importance of accurate diagnosis. As a result of all of these changes, diagnosis has evolved from a concept in medicine to one that has numerous other implications, including psychological, sociological, and also political and economic. Herein, we will discuss the concept of diagnosis and future directions in systemic lupus erythematosus (SLE), a most challenging disease in rheumatology.

S. Yavuz (✉)
Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden

P. E. Lipsky
Ampel BioSolutions and RILITE Research Institute, Charlottesville, VA, USA

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_8

SLE

SLE is a complex, prototypic autoimmune disease characterized by loss of tolerance and sustained autoantibody production. Strong genetic influences have been demonstrated in family and twin studies [3–5]. Moreover, data shows that ancestry affects not only incidence and prevalence as well as renal involvement, but also molecular pathways and autoantibody profiles, that results in differential response to treatments in SLE [6–9].

The vastly diverse nature of the disease presents immense challenges to physicians for diagnosis and treatment. Despite improved prognosis over the last 50 years, the chance of being dead at the age of 35 for a patient diagnosed with SLE at the age of 20 is still one in seven. In a recent study, SLE ranked tenth in the leading cause of death in women between 15 and 24 years of age and is the only chronic inflammatory disorder, ranking higher than diabetes mellitus or HIV [1, 2]. Therefore, early diagnosis and introducing the best suitable treatment for the patients are of utmost importance to improve prognosis further.

Diagnosis in SLE relies heavily on the physician's clinical judgment based on a combination of clinical signs/symptoms and available clinical tests (most frequently ANAs). Physicians' experiences are important in solving the problem of lupus diagnosis. However, diagnostic delays and misdiagnosis are inevitable in a disease with such heterogeneity, as several conditions mimic SLE. A survey of more than 2500 UK lupus patients in 2014 showed that the mean time between patients' first awareness of SLE symptoms and actual diagnosis was 6.4 years and half of the patients reported that they had been misdiagnosed initially [10].

Given the considerable heterogeneity in SLE, most efforts have been directed toward developing more precise classification criteria that aim to assemble cohorts that are representative of the majority with disease for clinical research. Scientifically, classification criteria target high specificity to classify accurately even if it means a trade-off of lower sensitivity, which is more important for diagnosis. The 1997 revised version of the 1982 ACR criteria is highly specific (96%) for classification when at least 4 of 11 criteria are positive [11]. However, sensitivity was remarkably lower compared to the criteria of the Systemic Lupus International Collaborating Clinics (SLICC) (83% vs. 97%), but the specificity dropped off to 84% in the latter [12]. The development process of the recent ACR/EULAR SLE classification aimed to improve sensitivity compared to ACR 1997 criteria as well as applicability to early or new onset lupus without compromising specificity and focusing on true autoimmune disease. The 2019 European League against Rheumatism (EULAR)/American College of Rheumatology (ACR) SLE classification criteria reached the combination of high sensitivity and specificity of 96.1% and 93.4%, respectively, with comparable sensitivity to SLICC criteria in capturing early disease (Table 8.1). However, it is important to emphasize that these are classification criteria used for research purposes and not diagnostic criteria employed in clinical practice. If employed as diagnostic criteria, they will identify only the most stereotypical patients, leaving many with fewer features of lupus undiagnosed and often frustrated.

Table 8.1 Comparative sensitivity and specificity of systemic lupus erythematosus (SLE) classification criteria factored by disease duration

	Sensitivity			Specificity		
Disease duration	ACR1982/1997 criteria	95%CI SLICC 2012 criteria	EULAR/ ACR 2019 criteria	ACR1982/1997 criteria	95%CI SLICC 2012 criteria	EULAR/ ACR 2019 criteria
<1 year	0.56 0.21–0.86	0.89 0.52–0.99	0.89 0.52–1.00	0.92 0.74–0.99	0.92 0.74–0.99	0.92 0.74–0.99
1 to <3 years	0.81 0.72–0.88	0.98 0.93–1.00	0.97 0.92–0.99	0.95 0.92–0.99	0.88 0.80–0.94	0.96 0.90–0.99
3 to <5 years	0.81 0.70–0.90	0.91 0.82–0.97	0.96 0.88–0.99	0.94 0.87–0.98	0.89 0.80–0.94	0.99 0.94–1.00
≥5 years	0.84 0.80–0.87	0.97 0.96–0.99	0.96 0.94–0.98	0.93 0.90–0.95	0.81 0.76–0.85	0.93 0.89–0.95

Adapted from Ref. [38]
ACR American College of Rheumatology, *EULAR* European League Against Rheumatism, *SLICC* Systemic Lupus International Collaborating Clinics

The repositioning of anti-nuclear antibody (ANA) as an obligatory entry criterion for the 2019 SLE classification criteria has spurred vigorous debates [6–8]. Targeting individuals with true autoimmunity was the impetus behind the proposal and acceptance of ANA as a key entry criterion. An ANA titer of 1:80 has been shown to have a sensitivity of 98%, and although rare at onset, ANA-negative SLE cases may exist, which may cause a small subset of SLE patients to be unclassified [13, 14]. If the classification criteria were used for diagnosis, this would be a problem, but it is less of an issue for classification for research purposes [15, 16]. Of note, physicians should understand the assay used to detect ANA since these techniques have intrinsic differences and may provide disparate results. For example, the results of an indirect immunofluorescence (IIF) assay using human epithelial type 2 cells (HEp-2- IIFA) may vary in different laboratories because of its dependence on visual reading of antibody patterns [17]. In addition, an IIF assay may be disadvantageous because of its low specificity at low antibody titers [18]. The disease duration and treatment may also affect ANA seroconversion [19]. Interestingly, the recent phase II belimumab trial showed that 29.5% of established SLE patients were found to be negative for ANA at enrollment, raising the possibility that ANA positivity may reflect an immunologically active state and patients with positive ANA may respond to some therapies differently [20, 21].

Although data have shown that autoantibodies, such as ANA, appear in the blood as early as 9.4 years (mean 3.3 years) before the clinical onset of SLE, given its low specificity, and that it is also positive in up to 20% of healthy individuals [22], screening ANA-positive patients with nonspecific symptoms is usually ineffective. Some autoantibodies are highly specific for SLE diagnosis, such as

anti-double-stranded DNA (anti-ds DNA), anti-Sm, and anti-ribosomal P, but they are less sensitive. For example, anti-dsDNA is an important biomarker for SLE diagnosis and disease activity, and its prevalence ranges between 50% and 75% depending on the assay, disease state, and ethnicity [6, 23, 24]. Indirect assays to measure complement proteins or their cell-bound activation products have proven to be informative and reliable [25]. The assay detecting cell-bound complement activation products *(Exagen, Vista, CA, USA),* which outperforms anti-dsDNA by up to 48% in terms of sensitivity, may be helpful in supporting the diagnosis of SLE [25].

Can Genomics and Transcriptomics Be Used to Diagnose SLE?

The advent of high-throughput genotyping, coupled with contemporary bioinformatics approaches and modeling, has significantly improved the understanding of the pathophysiology of several multigenic complex diseases, including SLE [26–28]. So far, approximately 100 genetic susceptibility loci at genome-wide significance have been identified in SLE, some of which are shared with other autoimmune disorders. The individual gene effects, however, are somewhat small (relative risk <2), and unlikely to assist in diagnosis or predict outcome when utilized individually [29, 30].

Genetic risk scores (GRS) are numeric scores that combine a large number of disease-associated genetic variants that are weighted by SLE risk odds ratios and reflect the disease-associated genetic load in an individual patient [31]. Therefore, the idea of utilizing GRS as a tool for predicting disease susceptibility and outcome has become a tantalizing approach that has been explored by several groups in SLE. However, some reports had limitations because of sample size or were restricted to limited sets of single nucleotide polymorphisms (SNP) based on immunochip analysis [32–35].

Not testing results across ancestries is another caveat of most studies. Chen L et al. performed a GRS analysis for SLE across Chinese and European populations. Utilizing three European and two Chinese GWAS datasets and training on a dataset for one population, they tried to predict SLE in the other dataset [32]. Perhaps not surprisingly, they found the most SLE predicted SNPs were enriched in patients with kidney involvement, indicating that most SLE-associated variants also confer risk for lupus nephritis. Another takeaway from this study was the correlation between GRS and age of onset in lupus, which corroborated in both European and Chinese populations, albeit it was independent of renal involvement. Notably, another large independent European GWAS also showed that higher GRS is associated with renal disease and SLE onset at a younger age [35]. This group also demonstrated that a high GRS had the potential of predicting patient outcomes.

In summary, these studies may render an example of incorporating GRS information into the clinical diagnosis of an SLE patient and may assist in diagnosis of lupus nephritis early. However, the GRS studies are still in their early stages and their role in diagnosis of generalized lupus remains to be determined.

Finally, the question remains whether a genetic diagnosis will equate with a clinical diagnosis and how both false positives and false negatives will affect the perception of the utility of the GRS approach. Since there remains debate about whether the genetic tendency will establish the diagnosis of lupus, the utility of the GRS will require considerable debate and eventually a consensus to be accepted.

As autoimmunity precedes overt clinical disease [22], signs and symptoms can be relatively nonspecific in the early stages of SLE. Thereby, early diagnosis can be challenging. A longitudinal study of more than 9000 SLE patients showed that early diagnosis (<6 months) and early adequate treatment result in fewer numbers of flares, low hospitalization rate, and low lupus-related medical costs compared to the matched SLE patients who were diagnosed later than 6 months [36]. This study underlines again the importance of early diagnosis and intervention to prevent damage and eventually mortality. Genetic risk factors are widely known for many diseases; however, their translation into clinical practice is still in its infancy. Knevel et al. have developed a GRS (G-PROB) using genome-wide significant variants ($p \leq 5\times10^{8}$) from previously published genome-wide association studies (GWAS) and tested its potential as a diagnostic tool in a set of patients with inflammatory arthritis (rheumatoid arthritis, systemic lupus erythematosus, spondyloarthropathy, psoriatic arthritis, and gout) [37]. Coupled with good discriminatory capacity (area under the curve (AUC), 0.69–0.84), it could single out a likely diagnosis for 45% of patients with a positive predictive value of 0.64 that could be further improved with the addition of serologic data. Despite only being tested in Caucasian cohorts, the results of this study demonstrate the potential clinical utility of the GRS for diagnosis, especially when incorporating serologic findings, such as ANA, and, possibly, clinical manifestations.

Transcriptomic analysis might also contribute to diagnosis. Based on a meta-analysis of 40 independent publicly available gene expression studies containing 7471 transcriptomic profiles, Haynes et al. identified a core gene set (93-gene signature, SLE MetaSignature) that is dysregulated in patients with SLE and distinguishes SLE from other relevant rheumatic disorders and infections [38]. They further validated the SLE MetaSignature in a prospective study comprising patients with juvenile-onset SLE, juvenile idiopathic arthritis, and healthy subjects. This study demonstrates the potential value of the integration of gene expression studies into the clinic as a means to improve the diagnosis of SLE.

Despite these promising advances in genetic and molecular pathogenesis, lupus remains a clinical diagnosis. Regardless of the elegant genetic and molecular advances, it remains uncertain how this information will be integrated into the process of assigning a diagnosis to an individual patient.

Gene Expression Studies and Organ Involvement

In addition to molecular diagnosis of SLE, several groups have attempted to utilize transcriptomic data from blood, purified T and B cells, myeloid cells, and, although less common, cells from tissue in order to stratify patients based on their molecular

signatures or predict disease activity. Grouping patients based on molecular signatures could also be a successful strategy for clinical trials, which may pave the way for personalized precision medicine.

One of the key features in SLE is the prominent expression of interferon (IFN)-inducible genes, an interferon gene signature (IGS) regardless of disease activity [39, 40]. An IGS may be induced by both type I and type II IFNs, in which type I IFNs with 13 IFN-alpha genes (A1, A2, A4, A5, A6, A7, A8, A10, A13, A14, A16, A17, and A21) and IFNB1, IFNW1, and IFNE are likely the major contributors. Type II IFN, IFNG, also induces an IGS through its distinct receptor, but its role in the SLE pathogenesis has been largely deduced from in vitro studies [40–42]. Using the weighted gene expression network analysis (WGCNA), a bioinformatics approach to derive gene modules in the dataset based on co-expression [43], IGS has been investigated in detail by looking at differences in various blood cells from patients with SLE and compared with the results of patients with other autoimmune rheumatic diseases and healthy volunteers [44]. The results demonstrate that a T-cell-specific module is exclusively expressed in SLE, whereas monocyte and neutrophils may be present similarly in other diseases and controls [44, 45]. Together with the hypomethylation of type I IFNs in naïve CD4 + T cells, these results suggest that type I IFN T-cell signaling may contribute to SLE disease pathogenesis [46].

In order to understand how disease activity affects gene expression and if it helps to group patients or disease manifestations, longitudinal gene expression studies are needed. In this context, Banchereau et al. performed a blood transcriptome profile of a longitudinal cohort of pediatric SLE patients [47]. They were able to group SLE patients into seven subsets, where each group was associated with a specific gene module. Of those, the identified neutrophil transcripts were enriched in patients with active lupus nephritis. They also found a robust plasmablast signature that was associated with disease activity, and the signal was stronger in African ancestry patients [47]. Although easier to obtain, blood transcriptome analyses provide a more general picture, especially when cell-specific transcriptome differences are targeted. Moreover, access to matched transcriptome data in the whole blood and tissue would provide a better understanding of how to interpret the differences or changes in cell populations as well as signaling pathways detected in many SLE studies. Labonte et al. developed an in-house tool based on differentially expressed T-cell receptor genes (TCR), immunoglobulin genes, and HLA genes in most SLE studies then created a Biologically Informed Gene Clustering (BIG-C) platform utilizing more than 40 SLE and control microarray datasets [48]. Gene set variation analysis employing the IFNA2, IFNB1, IFW1, IFNG, TNF, IL12, and the IFN core signature genes demonstrated prominent expression of IFNB1 and IFNW1 signatures differentially associated with organ involvement. The researchers found strong IFNB1 enrichment in skin and synovium in comparison to those in the kidneys in SLE patients [40]. This result is particularly interesting as several case reports show that drug-induced SLE arises with positive dsDNA after treatment with IFNB1 in multiple sclerosis [49, 50]. Besides proposing IFNB1 as an intriguing treatment target of SLE, the results of this study also demonstrated that IGS is less likely to

correlate with the disease activity because of prolonged expression of IGS in monocytes.

Lupus nephritis is a leading cause of morbidity and mortality of SLE. Although advances have been made through immunologic discoveries and genetic association studies in SLE, the outlook for patients with LN has not improved dramatically over the years, as ~10% still progress to end-stage renal disease (ESRD) [51, 52]. Renal biopsy is the gold standard for treatment decisions; however, the International Society of Nephrology/Renal Pathology Society (ISN/RPS) classification is limited by reliance on only histologic findings by mostly light microscopy without integrating recent molecular insights. Conventionally used biomarkers such as proteinuria or serologic markers have a limited ability to predict renal prognosis adequately, and persistent proteinuria can be secondary to residual activity, chronic damage, or comorbid conditions. Nevertheless, no robust markers have been identified yet to replace kidney biopsy, despite the extensive search for blood or urinary biomarkers.

Single-cell RNA sequencing (sc-RNA-seq) is a powerful unbiased approach to overcome the aforementioned limitations of the bulk analysis, such as defining cell types that link to observed gene expressions as well as provide new insights into the diverse mechanisms involved in the pathogenesis of tissue injury in kidneys. Recent studies show the potential of transcriptomic profiling of skin biopsies as a biomarker of lupus nephritis by performing sc-RNA-seq in lupus kidney and skin tissues [53, 54]. These studies also demonstrate that residential cells, such as kidney epithelial cells, in addition to infiltrating cells contribute to the LN disease progression.

Given that several inflammatory autoimmune diseases share common nonspecific symptoms early in the disease stage, comparing gene expression profiles of these conditions might be informative in clinical practice to segregate and treat early [55–58]. In a recent study, the gene expression profiles of SLE synovium were interrogated by using knee synovia samples from SLE, rheumatoid arthritis (RA), and osteoarthritis (OA) patients. Bioinformatic analyses revealed a myeloid-cell-mediated inflammation that governs the immunopathogenesis of lupus arthritis [59]. Upregulated differentially expressed genes in RA, on the other hand, indicated T cells, B cells, NK, NKT cells, and the other lymphocytes. In another study, 91% of IFN-inducible genes were differentially expressed in systemic sclerosis (SSc) as in SLE within the same platform compared to healthy individuals [60]. A subset of SSc patients who were also grouped as “lupus-like” phenotype showed type I IFN and plasma cell signatures. They also found a correlation between the type I IFN signature and the presence of lymphopenia, anti-topoisomerase, and anti-U1RNP antibodies.

In summary, the recent state-of-the-art technologies have advanced the understanding of the underlying molecular heterogeneity in SLE. Simultaneously, analyzing data from multiple sources including various cells and tissues at the cellular, molecular, and protein level will be important in the future to stratify diseases into clinically relevant groups. Leveraging these advances provides the chance to devise molecular tools that improve SLE diagnosis and help to predict early organ involvement.

Summary and Conclusions

Currently, the diagnosis of SLE involves the use of clinical tools that have not changed in many years. Because of their imprecision and a lack of consensus on what constitutes lupus in the clinic, many patients remain undiagnosed for protracted periods of time. Recently, genetic and molecular tools have been developed that afford the possibility of improving the precision of lupus diagnosis. Whether these will evolve to the point of clinical utility and whether they will be embraced by the clinical community remain major challenges for the field and patients living with recognized or undiagnosed lupus.

References

1. Ludmerer KM. Commentary: understanding the Flexner report. Acad Med. 2010;85(2):193–6.
2. Duffy TP. The Flexner report–100 years later. Yale J Biol Med. 2011;84(3):269–76.
3. Shai R, Quismorio FP, Li L, Kwon OJ, Morrison J, Wallace DJ, et al. Genome-wide screen for systemic lupus erythematosus susceptibility genes in multiplex families. Hum Mol Genet. 1999;8(4):639–44.
4. Kuo CF, Grainge MJ, Valdes AM, See LC, Luo SF, Yu KH, et al. Familial aggregation of systemic lupus erythematosus and Coaggregation of autoimmune diseases in affected families. JAMA Intern Med. 2015;175(9):1518–26.
5. Deapen D, Escalante A, Weinrib L, Horwitz D, Bachman B, Roy-Burman P, et al. A revised estimate of twin concordance in systemic lupus erythematosus. Arthritis Rheum. 1992;35(3):311–8.
6. Catalina MD, Bachali P, Yeo AE, Geraci NS, Petri MA, Grammer AC, et al. Patient ancestry significantly contributes to molecular heterogeneity of systemic lupus erythematosus. JCI Insight. 2020;5(15).
7. Feldman CH, Hiraki LT, Liu J, Fischer MA, Solomon DH, Alarcón GS, et al. Epidemiology and sociodemographics of systemic lupus erythematosus and lupus nephritis among US adults with Medicaid coverage, 2000–2004. Arthritis Rheum. 2013;65(3):753–63.
8. Peschken CA, Katz SJ, Silverman E, Pope JE, Fortin PR, Pineau C, et al. The 1000 Canadian faces of lupus: determinants of disease outcome in a large multiethnic cohort. J Rheumatol. 2009;36(6):1200–8.
9. Pons-Estel BA, Catoggio LJ, Cardiel MH, Soriano ER, Gentiletti S, Villa AR, et al. The GLADEL multinational Latin American prospective inception cohort of 1,214 patients with systemic lupus erythematosus: ethnic and disease heterogeneity among "Hispanics". Medicine (Baltimore). 2004;83(1):1–17.
10. Morgan C, Bland AR, Maker C, Dunnage J, Bruce IN. Individuals living with lupus: findings from the LUPUS UK Members Survey 2014. Lupus. 2018;27(4):681–7.
11. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1997;40(9):1725.
12. Petri M, Orbai AM, Alarcón GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the systemic lupus international collaborating clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum. 2012;64(8):2677–86.
13. Leuchten N, Hoyer A, Brinks R, Schoels M, Schneider M, Smolen J, et al. Performance of antinuclear antibodies for classifying systemic lupus erythematosus: a systematic literature review and meta-regression of diagnostic data. Arthritis Care Res (Hoboken). 2018;70(3):428–38.
14. Choi MY, Clarke AE, St Pierre Y, Hanly JG, Urowitz MB, Romero-Diaz J, et al. Antinuclear antibody-negative systemic lupus erythematosus in an international inception cohort. Arthritis Care Res (Hoboken). 2019;71(7):893–902.

15. Aggarwal R, Ringold S, Khanna D, Neogi T, Johnson SR, Miller A, et al. Distinctions between diagnostic and classification criteria? Arthritis Care Res (Hoboken). 2015;67(7):891–7.
16. Aringer M, Johnson SR. Classifying and diagnosing systemic lupus erythematosus in the 21st century. Rheumatology (Oxford). 2020;59(Suppl 5):v4–v11.
17. Pisetsky DS, Lipsky PE. New insights into the role of antinuclear antibodies in systemic lupus erythematosus. Nat Rev Rheumatol. 2020;16(10):565–79.
18. Mahler M, Meroni PL, Bossuyt X, Fritzler MJ. Current concepts and future directions for the assessment of autoantibodies to cellular antigens referred to as anti-nuclear antibodies. J Immunol Res. 2014;2014:315179.
19. Acosta-Mérida A, Isenberg DA. Antinuclear antibodies seroconversion in 100 patients with lupus. Clin Exp Rheumatol. 2013;31(4):656.
20. Wallace DJ, Stohl W, Furie RA, Lisse JR, McKay JD, Merrill JT, et al. A phase II, randomized, double-blind, placebo-controlled, dose-ranging study of belimumab in patients with active systemic lupus erythematosus. Arthritis Rheum. 2009;61(9):1168–78.
21. Putterman C, Pisetsky DS, Petri M, Caricchio R, Wu AHB, Sanz I, et al. The SLE-key test serological signature: new insights into the course of lupus. Rheumatology (Oxford). 2018;57(9):1632–40.
22. Arbuckle MR, McClain MT, Rubertone MV, Scofield RH, Dennis GJ, James JA, et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. N Engl J Med. 2003;349(16):1526–33.
23. Pisetsky DS, Spencer DM, Rovin B, Lipsky PE. Role of ANA testing in the classification of patients with systemic lupus erythematosus. Ann Rheum Dis. 2019; https://doi.org/10.1136/annrheumdis-2019-216259.
24. Aggarwal A. Role of autoantibody testing. Best Pract Res Clin Rheumatol. 2014;28(6):907–20.
25. Putterman C, Furie R, Ramsey-Goldman R, Askanase A, Buyon J, Kalunian K, et al. Cell-bound complement activation products in systemic lupus erythematosus: comparison with anti-double-stranded DNA and standard complement measurements. Lupus Sci Med. 2014;1(1):e000056.
26. Vasquez-Canizares N, Wahezi D, Putterman C. Diagnostic and prognostic tests in systemic lupus erythematosus. Best Pract Res Clin Rheumatol. 2017;31(3):351–63.
27. Giannopoulou EG, Elemento O, Ivashkiv LB. Use of RNA sequencing to evaluate rheumatic disease patients. Arthritis Res Ther. 2015;17:167.
28. Taylor KE, Chung SA, Graham RR, Ortmann WA, Lee AT, Langefeld CD, et al. Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. PLoS Genet. 2011;7(2):e1001311.
29. Langefeld CD, Ainsworth HC, Cunninghame Graham DS, Kelly JA, Comeau ME, Marion MC, et al. Transancestral mapping and genetic load in systemic lupus erythematosus. Nat Commun. 2017;8:16021.
30. Tsokos GC, Lo MS, Costa Reis P, Sullivan KE. New insights into the immunopathogenesis of systemic lupus erythematosus. Nat Rev Rheumatol. 2016;12(12):716–30.
31. Brown MA, Aletaha D. Genetic risk scores in inflammatory arthritis: a new era? Nat Rev Rheumatol. 2020;16(10):545–6.
32. Chen L, Wang YF, Liu L, Bielowka A, Ahmed R, Zhang H, et al. Genome-wide assessment of genetic risk for systemic lupus erythematosus and disease severity. Hum Mol Genet. 2020;29(10):1745–56.
33. Webb R, Kelly JA, Somers EC, Hughes T, Kaufman KM, Sanchez E, et al. Early disease onset is predicted by a higher genetic risk for lupus and is associated with a more severe phenotype in lupus patients. Ann Rheum Dis. 2011;70(1):151–6.
34. Gianfrancesco MA, Balzer L, Taylor KE, Trupin L, Nititham J, Seldin MF, et al. Genetic risk and longitudinal disease activity in systemic lupus erythematosus using targeted maximum likelihood estimation. Genes Immun. 2016;17(6):358–62.
35. Reid S, Alexsson A, Frodlund M, Morris D, Sandling JK, Bolin K, et al. High genetic risk score is associated with early disease onset, damage accrual and decreased survival in systemic lupus erythematosus. Ann Rheum Dis. 2020;79(3):363–9.

36. Oglesby A, Korves C, Laliberté F, Dennis G, Rao S, Suthoff ED, et al. Impact of early versus late systemic lupus erythematosus diagnosis on clinical and economic outcomes. Appl Health Econ Health Policy. 2014;12(2):179–90.
37. Knevel R, le Cessie S, Terao CC, Slowikowski K, Cui J, Huizinga TWJ, et al. Using genetics to prioritize diagnoses for rheumatology outpatients with inflammatory arthritis. Sci Transl Med. 2020;12(545).
38. Haynes WA, Haddon DJ, Diep VK, Khatri A, Bongen E, Yiu G, et al. Integrated, multicohort analysis reveals unified signature of systemic lupus erythematosus. JCI Insight. 2020;5(4).
39. Rönnblom L, Alm GV, Eloranta ML. The type I interferon system in the development of lupus. Semin Immunol. 2011;23(2):113–21.
40. Catalina MD, Bachali P, Geraci NS, Grammer AC, Lipsky PE. Gene expression analysis delineates the potential roles of multiple interferons in systemic lupus erythematosus. Commun Biol. 2019;2:140.
41. Munroe ME, Lu R, Zhao YD, Fife DA, Robertson JM, Guthridge JM, et al. Altered type II interferon precedes autoantibody accrual and elevated type I interferon activity prior to systemic lupus erythematosus classification. Ann Rheum Dis. 2016;75(11):2014–21.
42. Jackson SW, Jacobs HM, Arkatkar T, Dam EM, Scharping NE, Kolhatkar NS, et al. B cell IFN-γ receptor signaling promotes autoimmune germinal centers via cell-intrinsic induction of BCL-6. J Exp Med. 2016;213(5):733–50.
43. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity. 2008;29(1):150–64.
44. Flint SM, Jovanovic V, Teo BW, Mak A, Thumboo J, McKinney EF, et al. Leucocyte subset-specific type 1 interferon signatures in SLE and other immune-mediated diseases. RMD Open. 2016;2(1):e000183.
45. Alarcón-Riquelme ME. New attempts to define and clarify lupus. Curr Rheumatol Rep. 2019;21(4):11.
46. Coit P, Jeffries M, Altorok N, Dozmorov MG, Koelsch KA, Wren JD, et al. Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poising of interferon-regulated genes in naïve CD4+ T cells from lupus patients. J Autoimmun. 2013;43:78–84.
47. Banchereau R, Hong S, Cantarel B, Baldwin N, Baisch J, Edens M, et al. Personalized Immunomonitoring uncovers molecular networks that stratify lupus patients. Cell. 2016;165(6):1548–50.
48. Labonte AC, Kegerreis B, Geraci NS, Bachali P, Madamanchi S, Robl R, et al. Identification of alterations in macrophage activation associated with disease activity in systemic lupus erythematosus. PLoS One. 2018;13(12):e0208132.
49. Bonaci-Nikolic B, Jeremic I, Andrejevic S, Sefik-Bukilica M, Stojsavljevic N, Drulovic J. Anti-double stranded DNA and lupus syndrome induced by interferon-beta therapy in a patient with multiple sclerosis. Lupus. 2009;18(1):78–80.
50. Sladkova V, Mares J, Lubenova B, Hlustik P, Kanovsky P. Drug-induced systemic lupus erythematosus in interferon beta-1b therapy. Neuro Endocrinol Lett. 2011;32(1):4–6.
51. Davidson A, Aranow C, Mackay M. Lupus nephritis: challenges and progress. Curr Opin Rheumatol. 2019;
52. Yu F, Haas M, Glassock R, Zhao MH. Redefining lupus nephritis: clinical implications of pathophysiologic subtypes. Nat Rev Nephrol. 2017;13(8):483–95.
53. Der E, Suryawanshi H, Morozov P, Kustagi M, Goilav B, Ranabothu S, et al. Tubular cell and keratinocyte single-cell transcriptomics applied to lupus nephritis reveal type I IFN and fibrosis relevant pathways. Nat Immunol. 2019;20(7):915–27.
54. Der E, Ranabothu S, Suryawanshi H, Akat KM, Clancy R, Morozov P, et al. Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis. JCI Insight. 2017;2(9).
55. Boers M. Understanding the window of opportunity concept in early rheumatoid arthritis. Arthritis Rheum. 2003;48(7):1771–4.

56. Finckh A, Liang MH, van Herckenrode CM, de Pablo P. Long-term impact of early treatment on radiographic progression in rheumatoid arthritis: a meta-analysis. Arthritis Rheum. 2006;55(6):864–72.
57. Maas K, Chan S, Parker J, Slater A, Moore J, Olsen N, et al. Cutting edge: molecular portrait of human autoimmune disease. J Immunol. 2002;169(1):5–9.
58. Olsen N, Sokka T, Seehorn CL, Kraft B, Maas K, Moore J, et al. A gene expression signature for recent onset rheumatoid arthritis in peripheral blood mononuclear cells. Ann Rheum Dis. 2004;63(11):1387–92.
59. Hubbard EL, Catalina MD, Heuer S, Bachali P, Robl R, Geraci NS, et al. Analysis of gene expression from systemic lupus erythematosus synovium reveals myeloid cell-driven pathogenesis of lupus arthritis. Sci Rep. 2020;10(1):17361.
60. Assassi S, Mayes MD, Arnett FC, Gourh P, Agarwal SK, McNearney TA, et al. Systemic sclerosis and lupus: points in an interferon-mediated continuum. Arthritis Rheum. 2010;62(2):589–98.

9 The Impact of Antinuclear Antibody Testing on the Naming and Misnaming of Disease

David S. Pisetsky

The naming of an illness occurs in two distinct dimensions: the global and the personal. In the global dimension, naming provides an essential categorization to advance the study of disease and its management. In this dimension, the naming of an illness involves a construct developed by experts to establish validated criteria for diagnosis and classification. The naming of an illness (i.e., creation of a diagnostic category) is the foundation of scientific medicine, with research over time revising and refining any proposed criteria.

In contrast to the global dimension, the personal dimension involves the individual patient, with the application of criteria determined by the individual provider. The provider who does the naming may or may not be an expert in the particular clinical situation in question, especially for conditions that are rare or uncommonly encountered. Furthermore, the data that can inform appropriate naming may not be available in the timeframe needed. Thus, the naming of an illness in the real word is often tentative and imprecise.

The number of names and diagnostic categories for illnesses has proliferated dramatically in recent years with the advent of molecular techniques to subset illnesses into ever more narrow categories [1]. Indeed, precision or personalized medicine approaches signify the inadequacy of existing names to guide effective treatment. In a world of genetic and genomic testing, the molecular mechanisms of disease (e.g., patterns of aberrant gene expression) may be more relevant than the traditional disease name in developing and prescribing new treatments for diseases that may affect different tissues and organ systems. This approach can also lead to the development of tissue-agnostic agents and the conduct of basket trials involving several different conditions [2, 3].

Despite the burgeoning number of names and diagnostic categories, many patients simply do not fit well into existing categories, leading to uncertainty. This uncertainty

D. S. Pisetsky (✉)
Duke University Medical Center, VA Medical Center, Durham, NC, USA
e-mail: david.pisetsky@duke.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_9

can impair the relationship between patients and providers, provoke extensive and unrevealing diagnostic workups, and send patients in sometimes frantic searches to find a provider willing to provide a name for their signs and symptoms. In this situation, an inability to name the illness can be both obstructive and destructive.

Among medical subspecialties, rheumatology, in particular, cares for many patients for whom the naming of illness is problematic. The most established diagnostic categories encompass a wide range of signs and symptoms (e.g., pain, fever, depression) that are common in the general population [4]. Every illness has a threshold for findings to allow diagnosis. For rheumatologic illnesses, this threshold is often vague and long periods of time can pass before the evidence for disease is decisive. As a result, an illness can be named either too early or too late. It can also be given the wrong name.

A disease of protean manifestations, systemic lupus erythematosus (SLE or lupus) is often the subject of incorrect naming. SLE primarily affects women and can range from relatively mild joint pains to devastating neurologic disease; glomerulonephritis is a common source of morbidity and mortality [5]. The pattern of disease can vary markedly among different racial and ethnic groups. As any rheumatologist can attest, many patients who carry the diagnosis of lupus probably don't have this condition. These considerations do not diminish the severe symptoms of patients thought to have lupus; they only suggest that lupus is the wrong name.

For lupus, the designation of the wrong name often results from reliance upon laboratory tests whose characteristics are not widely appreciated. The prime example of a test that is either "misunderstood or misbegotten" is the antinuclear antibody test or ANA [6]. Antinuclear antibodies are directed toward diverse macromolecules in the cell nucleus [7, 8]. ANA positivity is now required for patient classification since studies suggest that 95–99% of patients with SLE express an ANA at some point in their illness [9–11]. Even though the ANA test has been used for over 60 years, its performance is subject to variability and inconsistency and its result subject to misinterpretation. The so-called lupus test is not a test for lupus. Indeed, the ANA test may not be a test for any disease.

In view of the importance of ANA testing to both the naming (and misnaming) of illness, I would like to provide a perspective on current serological testing and suggest ways it can be used to develop new nomenclature.

The Problems of ANA Determination

Box: Issues with ANA Testing in the Naming of Illness

- Lack of standardization.
- High frequency of false-positive results.
- Uncertain frequency of false-negative results.
- Lack of quantitation.
- Uncertain interpretation of cytoplasmic staining.

ANA testing involves a variety of assay formats that each has advantages and disadvantages. The most venerable approach is the indirect immunofluorescence assay, denoted as the IIF or IFA. In this assay, serum is incubated with a microscope slide to which is fixed an organ slice or cell line as a source of nuclei. After incubation with an immunofluorescent anti-immunoglobulin reagent, the next step is visual inspection under the microscope to assess the binding in terms of a titer as well as the pattern (e.g., homogeneous, speckled). The pattern reflects the distribution of the target antigen within the cell nucleus, with the pattern providing information concerning the specificity of the IgG ANA present [12, 13].

Current versions of the IFA utilize a long-term cell called HEp2 since most of the relevant target antigens of disease-related ANAs are abundant in this cell. The main advantage of this assay relates to the high frequency of positivity in patients with SLE and related connective tissue diseases (CTDs), also called autoantibody (or ANA)-associated rheumatic diseases (AARDs). Beyond availability of a fluorescence microscope, the assay does not require any specialized equipment and is well within the capabilities of hospital and clinical testing laboratories. Furthermore, even though the IFA is designed to detect antibodies to the cell nucleus, the assay allows identification of antibodies to cytoplasmic antigens; antibodies to cytoplasmic antigens can also be biomarkers for lupus. Of note, when a sample with cytoplasmic binding is called ANA negative, an opportunity for naming can be missed.

The IFA has two main disadvantages that impact on the naming process. The first is the very high frequency of assay positivity in the otherwise healthy population [14]. Depending on the kit used, as many as 15–20% of the healthy population can be ANA positive. The frequency of positivity is twice as high in women as men and has a peak age of around 30–40 years. Since SLE primarily affects women in this age group, confusion can result if the test is used to evaluate women with vague or non-specific symptomatology. Interestingly, the frequency of ANA positivity appears to be increasing in the population [15].

The basis of the high frequency of ANA expression in the population is unknown. To the extent that ANA positivity signifies immune disturbance, the human immune system may have an unfortunate propensity to develop autoreactivity. A less dire or worrisome explanation for the high frequency of IFA reactivity is technical. Perhaps the fixation conditions for slide preparation denature or otherwise modify proteins so that they resemble foreign proteins in immunological reactivity.

The other technologies for ANA detection utilize recombinant or purified proteins as a source of nuclear antigens. Because of advances in molecular biology, the molecular identity of most of the target antigens relevant in rheumatology is now known and specific immunoassays are available. Of these approaches, multiplex assays allow the simultaneous measurement of antibodies to a series of cloned or purified proteins by a LINE assay or an addressable laser bead immunoassay (ALBIA) [16, 17]. ALBIAs allow detection of antibodies to antigens for SLE, Sjogren's syndrome, myositis, and progressive systemic sclerosis. Usually, results are provided as either positive or negative except for anti-DNA for which anti-DNA levels are valuable for assessing disease activity.

While, in general, the specificity of antibodies producing ANA positivity by otherwise healthy people is unknown, one exception is an antigen system called DFS70. DFS stands for dense fine speckled which is the characteristic pattern of staining

associated with antibodies to a protein called DFS70 or lens-epidermal derived growth factor. The presence of these antibodies can be recognized by IFA or by a specific immunoassay. Studies have indicated that, while anti-DFS70 antibodies can appear in a variety of conditions, they are not increased in patients with CTDs, including SLE. Thus, the finding of either DFS staining or antibodies to DFS70 could suggest that the patient does not have a CTD [18, 19].

The main disadvantage of using a multiplex assay like an ALBIA is that the assay is not really an ANA assay [6, 8, 16, 17]. Since only a small number of antibodies can be measured, many ANA specificities relevant to diagnosis are missed. Clinical testing laboratories, however, like the ALBIA because these assays are high throughput and do not require a dedicated technician skilled in reading IFA patterns. For many in the field, however, the IFA remains the gold standard since it can detect a broad range of specificities.

The positioning of the IFA as a gold standard is not as solid as often considered since variation between kits is substantial and many patients with SLE can be negative in one assay and positive in another [20–22]. These inconsistencies can be reduced by testing the same sample by more than one assay type (e.g., an IFA and ALBIA) but this approach is often not possible because of issues of costs or assay availability. Given the serious impact of an incorrect diagnosis of SLE (either way, missing the diagnosis of SLE or making the diagnosis in someone without the disease), the cost of seemingly redundant testing seems well justified.

Symptomatology

Autoimmune diseases often start gradually and serological findings can predate clinical findings. By using samples from a biobank repository assembled by the US military, Arbuckle et al. showed that individuals with SLE begin to express characteristic antibodies years before the diagnosis, with the number of specificities increasing over time [23]. The period of time can be termed pre-autoimmunity in distinction to autoimmunity when signs and symptoms accompany serological abnormalities and diagnostic or classification criteria are met [24–26]. While pre-autoimmunity is a fascinating subject, in the real world, it can lead to ambiguity and uncertainty about naming.

Consider a hypothetical case of a Ms. Jones, a 41-year-old woman with symptoms of fatigue and arthralgia. She notices headaches and does not feel like herself. She is worried about her condition since her mother had rheumatoid arthritis which started in a similar way.

Ms. Jones sees her general internist who orders a battery of tests including an ANA by immunofluorescence; the IFA is negative as is the rheumatoid factor and the anti-CCP. The provider reassures Ms. Jones and prescribes ibuprofen.

The symptoms persist, and Ms. Jones, dissatisfied with the first provider, goes to another. This provider repeats the ANA which is now positive. The provider says that he is concerned that Ms. Jones has lupus. After reading about lupus on the Internet, Ms. Jones becomes frightened.

Ms. Jones is referred to a rheumatologist who orders an ANA. This time, a multiplex assay is used. The ALBIA is positive for anti-Ro but is negative by the IFA used as part of "reflex testing" to confirm the multiplex assay. The rheumatologist says she is uncertain about the diagnosis but, because of the positive anti-Ro, suggests the diagnosis of undifferentiated connective tissue disease. Ms. Jones is now confused as well as frightened since she was first told she may have lupus and now receives another diagnosis. She is also angry and discouraged that providers cannot figure out what is wrong.

This case is hypothetical but illustrates the difficulties when the naming of illness depends upon a test that is not well standardized and is subject to variability. The case also illustrates the problems that can arise when the performance characteristics of tests are not well understood. Which is the most informative: the anti-Ro by multiplex, the one positive IFA, or the two negative IFA tests? Anti-Ro can be detected in low amounts by an ALBIA but may be missed by an IFA depending on the kit used, accounting for the negative IFA reflex assay. In reality, the actual serological profile of Ms. Jones is not clear although such information would be valuable in determining whether she has early stages of a CTD including SLE and is, thus, in a state of pre-autoimmunity.

The Issue of Nomenclature

Whether justified or not, ANA testing is very commonly performed in the evaluation of patients with a host of signs and symptoms ranging from rash to low back pain to depression. It can also be a part of the general screen for musculoskeletal disease even when the pretest probability for SLE is low. For many of these patients, the test will be false-positive, often leading to referral to a rheumatologist who may perform additional, sometimes costly, tests to understand the significance of the serology. The situation with false positivity is so extreme that some healthcare systems have considered prohibiting generalists from even ordering the ANA. On the other hand, for a very few individuals, the positive ANA is an early sign of disease, a harbinger of more serious events in the future. For these individuals, the ANA has functioned successfully as an antecedent biomarker since early treatment can perhaps attenuate disease and reduce damage.

Another approach to nomenclature (i.e., naming) could improve the use of ANA testing. For SLE as well as other CTDs, serology can be interpreted in a probabilistic way, inferring a likelihood of disease and not its presence. The likelihood increases depending on the number and kind of other serological disturbances present as well as the nature of signs and symptoms. In the future, genomic analysis as well as flow cytometric analysis of cell immune populations may provide adjunctive biomarkers but these technologies are not yet ready for widespread use [27].

The existing nomenclature involves terms like undifferentiated connective tissue disease (UCTD) to encompass serological disturbances and certain signs and symptoms; while indicative of some type of disease, the findings in someone considered to have a UCTD are not decisive or specific enough to allow a diagnosis. Despite the

frequent use of the term UCTD, its meaning seems nebulous. It is not clear whether UCTD denotes a final state (i.e., the differentiation has already occurred) or whether further differentiating is in the offing. I doubt that the term differentiating connective tissue disease would catch on but, perhaps, it would be more accurate.

In the past, the diagnosis of rheumatoid arthritis included stages of possible (or equivocal), probable, definite, and classical [28]. The diagnostic criteria, however, gave way to a simpler classification system in view of better serological markers (e.g., anti-citrullinated protein antibodies or ACPA, also known as anti-cyclic citrullinated peptide antibodies or anti-CCP) [29]. The importance of early aggressive therapy provided an impetus to create the new criteria to allow the use of disease modifying anti-rheumatic drugs in the earliest phases of disease. With therapy guided by treat-to-target principles, classical disease could actually disappear.

For SLE, some current disease names (e.g., preclinical lupus, incomplete lupus, non-classical lupus) indicate that diagnosis and classification can be uncertain and tentative, with the presence of ANA positivity a major determinant of these names. Given the likelihood that ANA testing will continue unabated in the future, I would argue that a categorization of serological findings based on stages of disease (possible, probable, definite, and classical) would advance scientific inquiry. Such a categorization could also facilitate communication between the patient and provider as well as underpin more effective programs of prevention and treatment.

With well-standardized assays, serology would be a valuable adjunct to help name an illness at its earliest stages in the presence of certain signs and symptoms. In terms of serology, a positive ANA is possibly lupus; a positive ANA and positive anti-DNA are probably lupus; a positive ANA with anti-DNA and anti-Sm is definitely lupus. Low C3 and C4 along with an array of ANA specificities (e.g., anti-DNA, anti-Sm, anti-RNP, anti-Ro) and complement split products would signify classical disease [30]. Rather than positing an ANA as a requirement for the diagnosis or classification of SLE, ANA testing could be used to define a risk or likelihood of disease depending on the signs and symptoms, even if non-specific or vague.

Whether insurers or professional organizations would accept such a nomenclature system is speculative. Its acceptance by patients and providers is also unknown. Nevertheless, in settings where illness has no name, immunological testing has the potential to provide unique prognostic and diagnostic information for the individual patient. Hopefully, when used rationally and wisely, ANA testing can help name illness and, thereby, relieve the distress that uncertainty can cause for so many patients.

References

1. Ginsburg GS, Phillips KA. Precision medicine: from science to value. Health Aff (Millwood). 2018;37(5):694–701.
2. Tsimberidou AM, Fountzilas E, Nikanjam M, Kurzrock R. Review of precision cancer medicine: evolution of the treatment paradigm. Cancer Treat Rev. 2020;86:102019.
3. Seligson ND, Knepper TC, Ragg S, Walko CM. Developing drugs for tissue-agnostic indications: a paradigm shift in leveraging cancer biology for precision medicine. Clin Pharmacol Ther. 2021;109(2):334–42.

4. Pisetsky DS, Clowse MEB, Criscione-Schreiber LG, Rogers JL. A novel system to categorize the symptoms of systemic lupus erythematosus. Arthritis Care Res. 2019;71:735–41.
5. Kaul A, Gordon C, Crow MK, Touma Z, Urowitz MB, van Vollenhoven R, et al. Systemic lupus erythematosus. Nat Rev Dis Primers. 2016;2:16039.
6. Pisetsky DS. Antinuclear antibody testing: misunderstood or misbegotten? Nat Rev Rheumatol. 2017;13:495–502.
7. Meroni PL, Schur PH. ANA screening: an old test with new recommendations. Ann Rheum Dis. 2010;69:1420–2.
8. Pisetsky DS, Lipsky PE. New insights into the role of antinuclear antibodies in systemic lupus erythematosus. Nat Rev Rheumatol. 2020;16(10):565–79.
9. Leuchten N, Hoyer A, Brinks R, Schoels M, Schneider M, Smolen J, et al. Performance of antinuclear antibodies for classifying systemic lupus erythematosus: a systematic literature review and meta-regression of diagnostic data. Arthritis Care Res. 2018;70:428–38.
10. Aringer M, Costenbader K, Daikh D, Brinks R, Mosca M, Ramsey-Goldman R, et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. Arthritis Rheumatol. 2019;71:1400–12.
11. Choi MY, Clarke AE, St Pierre Y, Hanly JG, Urowitz MB, Romero-Diaz J, et al. Antinuclear antibody-negative systemic lupus erythematosus in an international inception cohort. Arthritis Care Res. 2019;71:893–902.
12. Wiik AS, Hoier-Madsen M, Forslid J, Charles P, Meyrowitsch J. Antinuclear antibodies: a contemporary nomenclature using HEp-2 cells. J Autoimmun. 2010;35(3):276–90.
13. Damoiseaux J, Andrade LEC, Carballo OG, Conrad K, Francescantonio PLC, Fritzler MJ, et al. Clinical relevance of HEp-2 indirect immunofluorescent patterns: the International Consensus on ANA patterns (ICAP) perspective. Ann Rheum Dis. 2019;78(7):879–89.
14. Tan EM, Feltkamp TE, Smolen JS, Butcher B, Dawkins R, Fritzler MJ, et al. Range of antinuclear antibodies in "healthy" individuals. Arthritis Rheum. 1997;40(9):1601–11.
15. Dinse GE, Parks CG, Weinberg CR, Co CA, Wilkerson J, Zeldin DC, et al. Increasing prevalence of antinuclear antibodies in the United States. Arthritis Rheumatol. 2020;72(6):1026–35.
16. Hanly JG, Su L, Farewell V, Fritzler MJ. Comparison between multiplex assays for autoantibody detection in systemic lupus erythematosus. J Immunol Methods. 2010;358(1–2):75–80.
17. Bruner BF, Guthridge JM, Lu R, Vidal G, Kelly JA, Robertson JM, et al. Comparison of autoantibody specificities between traditional and bead-based assays in a large, diverse collection of patients with systemic lupus erythematosus and family members. Arthritis Rheum. 2012;64(11):3677–86.
18. Carbone T, Pafundi V, Tramontano G, Gilio M, Padula MC, Padula AA, et al. Prevalence and serological profile of anti-DFS70 positive subjects from a routine ANA cohort. Sci Rep. 2019;9(1):2177.
19. Yumuk Z, Demir M. Clinical value of anti-DFS70 antibodies in a cohort of patients undergoing routine antinuclear antibodies testing. J Immunol Methods. 2020;480:112754.
20. Pisetsky DS, Rovin BH, Lipsky PE. New perspectives in rheumatology: biomarkers as entry criteria for clinical trials of new therapies for systemic lupus erythematosus: the example of antinuclear antibodies and anti-DNA. Arthritis Rheumatol. 2017;69:487–93.
21. Pisetsky DS, Spencer DM, Lipsky PE, Rovin BH. Assay variation in the detection of antinuclear antibodies in the sera of patients with established SLE. Ann Rheum Dis. 2018;77(6):911–3.
22. Pisetsky DS, Thompson DK, Wajdula J, Diehl A, Sridharan S. Variability in antinuclear antibody testing to assess patient eligibility for clinical trials of novel treatments for systemic lupus erythematosus. Arthritis Rheumatol. 2019;71(9):1534–8.
23. Arbuckle MR, McClain MT, Rubertone MV, Scofield RH, Dennis GJ, James JA, et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. N Engl J Med. 2003;349(16):1526–33.
24. Olsen NJ, Karp DR. Autoantibodies and SLE: the threshold for disease. Nat Rev Rheumatol. 2014;10(3):181–6.
25. Bourn R, James JA. Preclinical lupus. Curr Opin Rheumatol. 2015;27(5):433–9.

26. Aberle T, Bourn RL, Munroe ME, Chen H, Roberts VC, Guthridge JM, et al. Clinical and serologic features in patients with incomplete lupus classification versus systemic lupus erythematosus patients and controls. Arthritis Care Res (Hoboken). 2017;69(12):1780–8.
27. Slight-Webb S, Smith M, Bylinska A, Macwana S, Guthridge C, Lu R, et al. Autoantibody-positive healthy individuals with lower lupus risk display a unique immune endotype. J Allergy Clin Immunol. 2020;146(6):1419–33.
28. Ropes MW, Bennett GA, Cobb S, Jacox R, Jessar RA. 1958 revision of diagnostic criteria for rheumatoid arthritis. Bull Rheum Dis. 1958;9(4):175–6.
29. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum. 2010;62(9):2569–81.
30. Olsen NJ, Karp DR. Finding lupus in the ANA haystack. Lupus Sci Med. 2020;7(1):e000384.

In the Box or Out of the Box

10

Jane E. Salmon

Inclusion and exclusion criteria may be the most important aspect of a clinical study or trial. Lumping or splitting determines which patients can ultimately use a risk stratification algorithm or a therapy. Systemic lupus erythematosus (SLE) is a prototype illness that displays pros and cons of lumping and splitting. Among all SLE studies, those concerning pregnancy are most illustrative.

In 1872, the Viennese dermatologist Moritz Kaposi, MD, described SLE in a paper, "New Contributions to Knowledge of Lupus Erythematosus" [1] in these words: "Edematosis and thickening, glandular, painful swelling of the skin and tissues around the joints" occurs; he mentions "ripping, tearing, deep bone pains, especially in the main bones, the tibia, forearm and carpal bones." The patients had the symptoms of an "intense, generalized, feverish disease. They laid on the back, had a hot, dry, cracked tongue, general prostration, disturbed consciousness. Over the course of two to three weeks, coma, stupor, and death occurred under increasing brain disorder." Among 15 female cases of lupus erythematosus, he saw this picture five times, and three times death occurred with the symptoms described. In this work, Kaposi provided the first description of the systemic nature of lupus and its severe constitutional and visceral manifestations. He distinguished "lupus erythematosus discoides," a chronic cutaneous disease, from SLE, which he described as "the prognostically more serious one to watch." He was inclusive in his characterization of the signs and symptoms of disease and specified many of the current criteria for SLE [2]. To group together patients with such varied manifestation was bold and prescient. As a consequence of his grouping, we continue to struggle with how to stratify, prognosticate, and personalize therapies for lupus patients. I study SLE pregnancy, a circumstance that illustrates the problem.

J. E. Salmon (✉)
Department of Medicine, Hospital for Special Surgery, Weill Cornell Medicine, New York, NY, USA
e-mail: salmonj@hss.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_10

In addition to the variables associated with SLE itself, pregnant patients have high rates of preeclampsia, fetal and neonatal death, and fetal growth restriction. Identifying women destined for complications remains challenging and limits our ability to counsel and care. Do we use broad or narrow criteria to address these issues? The criteria we use to enroll patients in observational studies, the goals of which are to develop a risk stratification tool, will eventually be applied to a subset of patients for interventional studies to prevent pregnancy complications, not necessarily to all affected patients. In the paragraphs that follow, I describe how we selected and changed criteria throughout the study and the lessons we learned.

In designing the PROMISSE (**P**redictors of P**r**egnancy **O**utcome: Bio**m**arkers **i**n Antiphospholipid **S**yndrome and **S**ystemic Lupus **E**rythematosus) study, I was challenged with regard to defining inclusion and exclusion criteria. The PROMISSE study is a prospective multicenter, multiracial, multiethnic observational study to identify markers that predict poor pregnancy outcome in patients with antiphospholipid antibodies (aPL) and/or SLE. We enrolled nearly 500 patients with SLE and/or aPL antibodies from 2001 to 2013. SLE patients were required to meet American College of Rheumatology (ACR) criteria [3, 4]. The definition of aPL positivity and its association with adverse pregnancy outcomes (APO) evolved through the study. Initially, aPL positivity was defined as values of [1] anti-cardiolipin antibodies (aCL) and/or [2] anti-B2 glycoprotein I (B2GPI) antibodies greater than lab normal range and/or [3] lupus anticoagulant (LAC) positive. Fifty-three percent of those referred to PROMISSE with a history of positive aPL in local labs were negative for aPL in the study core laboratory [5]. These patients were excluded. In addition, only 12% of women who did not have SLE but with aCL <40 u/mL had APO, whereas 29% of women with IgG aCL antibody ≥40 u/mL had APOs. It became clear that if we wanted to identify biomarkers of APO in PROMISSE, it would be necessary to enrich for patients more likely to have APOs. With approval of our study monitoring board, we changed inclusion criteria to classify as aPL-positive-only patients with aCL or anti-B2GPI IgG or IgM ≥40 u/mL and/or LAC. This decision led to more exclusions.

We also learned that LAC positivity was the most powerful predictor of APOs [5]. All APO cases with IgG aCL <40 u/mL were LAC positive; among those with IgG aCL ≥40 u/mL, the APO rate was 8% in the LAC-negative group compared to 43% in the LAC-positive group. And when we considered only LAC, 3% of LAC-negative women had an APO, whereas 39% of LAC-positive patients had an APO.

But what about the women with histories of pregnancy complications or thromboses, both of which increased risk for APO) and aPL antibodies who did not have a positive LAC? If they are excluded from studies, we will not be able to evaluate biomarkers of APO in this group.

Another group of patients excluded from PROMISSE were women with active SLE. We did not want to conflate biomarkers of SLE activity with those associated with placental dysfunction. We published a series of papers describing clinical features and biomarkers measured in early pregnancy that were highly predicative of APO in SLE patients [4, 6, 7]. But many patients for whom we might apply our findings were not included in PROMISSE.

To accomplish our goal to improve *real-world clinical decision making*, enable physicians to reassure patients at low risk, and identify patients at elevated risk who can be enrolled in future adverse pregnancy prevention trials, we adopted a more inclusive approach. Our new studies externally validate the prediction models for APO developed from PROMISSE in independent prospective SLE cohorts that collect longitudinal clinical, laboratory, and pregnancy outcome data as in PROMISSE but without exclusions [3]. PROMISSE gave us the tools to identify high-risk pregnancies. We used these data to design an interventional trial.

The IMPACT trial is an open-label Phase II trial using the TNF inhibitor certolizumab. The rationale for the trial is based on evidence that TNF-α signaling inhibition in animal models of obstetric antiphospholipid syndrome (APS) rescues pregnancies [8]. The trial asks whether TNF-α blockade added to a regimen of heparin and low-dose aspirin during pregnancy reduces the rate of fetal death and/or preterm delivery due to preeclampsia and placental insufficiency in women meeting clinical and laboratory criteria for APS and with a positive LAC.

Identifying treatments to prevent poor pregnancy outcomes requires rigorous studies in well-defined populations. Our inclusion criteria make enrollment difficult; these patients are rare. As we recruit patients, I wonder about those we have excluded. In a pregnant woman with LAC who has not yet developed clinical evidence of APS, will this pregnancy be the event that makes her meet criteria for APS? Could our intervention have prevented it? What can we offer LAC-negative patients with APS and two late pregnancy losses?

If IMPACT meets its endpoint, and we learn that TNF-α blockade will prevent placental dysfunction in women with APS and LAC, we will have discovered a treatment for a specific group of patients. Can other types of patients also benefit? There are no effective therapies to prevent preeclampsia and placental insufficiency. The challenge that follows successful studies is how to advise and treat individuals who do not meet the study criteria. We know that some patients excluded from trials might have benefited from the intervention, but we do not know which ones. We have been rigorous in the design of the IMPACT trial, and we hope to be able to be generous in application of results.

Inclusion and exclusion criteria may be the most important aspect of a clinical study or trial, as lumping or splitting determines which patients can use a risk stratification algorithm or a therapy.

It is valid to accept studies with rigorous exclusion criteria because identification of treatments to prevent poor pregnancy outcomes requires rigorous studies in well-defined populations. But to improve real-world clinical decision making regarding all pregnant patients with SLE or APS, studies must be inclusive in order that the lessons learned can be applied to the broadest range of patients. Both inclusive and exclusive studies must be more transparent about how they have lumped or split and must emphasize the limitations of the use of a diagnosis name.

References

1. Kaposi KM. Neue Beiträge zur Kenntnis des Lupus erythematosus. Arch Dermatol Syphilis. 1872;4:36–78.
2. Holubar K, Fatović-Ferencić S. Cazenave, Kaposi and lupus erythematosus. A centennial and a sesquicentennial. Dermatology. 2001;203:118–20.
3. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [Letter]. Arthritis Rheum. 1997;40:1725.
4. Buyon JP, Kim MY, Guerra MM, et al. Predictors of pregnancy outcomes in patients with lupus: a cohort study. Ann Intern Med. 2015;163(3):153–63. https://doi.org/10.7326/M14-2235.
5. Lockshin MD, Kim M, Laskin CA, et al. Prediction of adverse pregnancy outcome by the presence of lupus anticoagulant, but not anticardiolipin antibody, in patients with antiphospholipid antibodies. Arthritis Rheum. 2012;64(7):2311–8. https://doi.org/10.1002/art.34402.
6. Kim MY, Buyon JP, Guerra MM, et al. Angiogenic factor imbalance early in pregnancy predicts adverse outcomes in patients with lupus and antiphospholipid antibodies: results of the PROMISSE study. Am J Obstet Gynecol. 2016;214(1):108.e1–e14. https://doi.org/10.1016/j.ajog.2015.09.066.
7. Kim MY, Guerra MM, Kaplowitz E, et al. Complement activation predicts adverse pregnancy outcome in patients with systemic lupus erythematosus and/or antiphospholipid antibodies. Ann Rheum Dis. 2018;77(4):549–55. https://doi.org/10.1136/annrheumdis-2017-212224.
8. Berman J, Girardi G, Salmon JE. TNF-alpha is a critical effector and a target for therapy in antiphospholipid antibody-induced pregnancy loss. J Immunol. 2005;174(1):485–90. https://doi.org/10.4049/jimmunol.174.1.485. PMID: 15611274.

11 Ever-Evolving Disease Classification Criteria for Clinical Trials and Studies: The Case of Systemic Lupus Erythematosus

Karen H. Costenbader

Introduction

A multisystem autoimmune disease that can cause inflammatory arthritis, disfiguring rashes, fevers, cytopenias, and more, systemic lupus erythematosus (SLE) is perhaps the most heterogeneous of the autoimmune rheumatic diseases. It is called the great mimicker and it is said that no two patients have the same disease course. Certainly, recognizing and diagnosing SLE takes experience and treating it is still a medical artform. Moreover, even lupus-focused rheumatologists, let alone other rheumatologists and general medical specialists, do not always agree on the diagnosis. The disease develops rapidly with an explosive onset of involvement in multiple organ systems in some individuals, and in others the symptoms and lab abnormalities develop slowly and insidiously over years. Most but not all patients with SLE have antinuclear antibodies (ANAs), at least at one point in time and often in high titer in speckled or homogeneous patterns. However, most people who have positive ANAs do not have SLE, as they are common in many other conditions and in healthy individuals. To further complicate matters, many of SLE's signs and symptoms and lab abnormalities are also present in related connective tissue diseases such as Sjögren's syndrome, rheumatoid arthritis, scleroderma, and mixed connective tissue disease (MCTD), in which there are features of more than one autoimmune connective tissue disease, as well as undifferentiated connective tissue disease (UCTD), in which the physician can only say that one of these conditions may be developing. Not surprisingly, there are no current diagnostic criteria for SLE. Patients with early or non-classical symptoms are often labeled as having "incomplete lupus," "suspected lupus," potential lupus, and myriad-related and imprecise terms (Table 11.1) [1].

K. H. Costenbader (✉)
Harvard Medical School, Division of Rheumatology, Inflammation and Immunity, Brigham and Women's Hospital, Boston, MA, USA
e-mail: KCostenbader@bwh.harvard.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_11

Table 11.1 Terms used to describe early and non-classical SLE

Lupus-like syndrome ("overlap," subset, and/or variant syndrome
Undifferentiated connective tissue disease (UCTD)
Antiphospholipid antibody syndrome with SLE features
Mixed connective tissue disease (MCTD) with SLE features
Occult lupus
Pseudo-lupus
Borderline lupus
Latent lupus
Incipient lupus
Incomplete lupus
Possible lupus
Probable lupus
Potential lupus
Suspected lupus

Adapted from reference [1]

Table 11.2 Disease diagnosis versus classification criteria

Diagnosis criteria	Classification criteria
Aim: Individual prognosis and therapy	Aim: Homogenous group (research)
Many different pieces of information	Feasible set of objective criteria
Sensitivity issues critical (therapeutic decisions)	Sensitivity often not as important
Diagnosis can be questioned again	Specificity at one time important

What Is the Difference Between Disease Diagnosis and Classification? And Why Do We Need Classification Criteria for Clinical Research and Trials?

While for many years, and still to this day in other diseases, the concepts of diagnosis and classification were synonymous, in SLE they are and should be kept distinct as their conflation has several ramifications. The most important distinction is that diagnosis refers to the unique patient and their individual physician or care team at a particular point in time and diagnosis has important implications, including the therapies offered and prognosis discussed. The ACR Committee on Classification Criteria has addressed this confusion, stating: "Diagnosis is the 'determination of the cause or nature of an illness by evaluation of the signs, symptoms and supportive tests in an *individual* patient'." [2] Often the diagnosis given by one physician does not agree with that given by another and even the diagnosis provided by a single physician may change over time. Classification criteria on the other hand are to be used to describe populations, not individuals. "Classification criteria are standardized definitions that are primarily intended to create well-defined, relatively homogenous cohorts for clinical research; they are not intended to capture the entire universe of possible patients, but rather to capture the majority of patients with key shared features of the condition." [2] Thus, the goal of classification differs from that of diagnosis (Table 11.2).

Classification criteria are intended for the purpose of clinical trials and research to ensure we have a homogeneous population of subjects that is widely understood. While they are not intended to be used for diagnosis, they should have high

agreement with diagnosis. And, while the treating clinician can use the universe of available data in attempting to arrive at the correct diagnosis, classification criteria should include a smaller set of widely available criteria, such that collection of the data informing them is feasible worldwide. Arguably the most important aspects of classification criteria are that they can be applied by anyone to any subject at one point in time and they are highly specific for the condition, creating a population above the threshold that is well-described and highly reproducible for studies. However, it is particularly challenging to create highly accurate classification criteria for heterogeneous and overlapping conditions such as autoimmune rheumatic diseases, the epitome of which is SLE. As no classification criteria are ever 100% sensitive or specific, there are individual patients who may have diagnoses of SLE without meeting classification and those who might be classified without having a formal diagnosis. Lastly, as medical knowledge evolves and diagnostic modalities evolve in tandem, but usually lagging a bit behind, classification criteria evolve and need to be updated (Table 11.3). Thus, classification and diagnosis are distinct concepts that must remain clearly separated at least for complex rheumatic diseases, but information derived from the process of developing classification criteria can shine light on the diagnosis and pathogenesis of disease and vice versa [3].

The Evolution of Classification Criteria for Systemic Lupus Erythematosus

The history of the development of criteria for classification of SLE has been summarized recently [4]. The 1971 American Rheumatism Association preliminary criteria were the first to be developed [5]. They included 14 criteria manifestations and were based on expert opinion. The 1982 American College of Rheumatology (ACR) criteria for the classification of SLE were authored by Tan et al. and revised the 1971 preliminary criteria "to incorporate new immunologic knowledge and improve disease classification." [6] They included malar and discoid rashes, photosensitivity, oral ulcers, arthritis, serositis, renal and neurological involvement, as well as positive fluorescence ANA and antibodies to native DNA and Sm antigen [6]. They also included an LE cell preparation and a false-positive test for syphilis. Raynaud's phenomenon and alopecia were removed in 1982, due to low sensitivity and specificity.

These 1982 criteria were then updated in 1997 by a letter to the editor first-authored by Hochberg, stating that the diagnostic and therapeutic criteria committee of the ACR had reviewed the 1982 criteria and recommended the deletion of LE cell prep (because of its rarity and infrequent use in clinical practice) and the addition of a positive finding of anti-phospholipid antibodies (based on abnormal IgG or IgM anticardiolipin antibody or a positive test for a lupus anticoagulant, in addition to the false-positive syphilis test) [7]. Four or more of the 11 criteria were considered positive.

The concept of weighting the criteria for SLE, with more weight assigned to some classification criteria than others, was first introduced in 1984 [8] and

Table 11.3 Comparison of evolving classification criteria for systemic lupus erythematosus over the years

Criterion	ACR 1971 [6]	ACR 1982 [4]	ACR 1997 [5]	SLICC 2012 [7]	EULAR/ACR criteria 2019 [8]
	≥4 criteria (each category counts as one independent criterion)	≥4 criteria (each category counts as one independent criterion)	≥4 criteria (each category counts as one independent criterion)	≥4 criteria (each category counts as one independent criterion) ≥1 clinical criterion and ≥1 immunologic criterion OR biopsy-proven lupus nephritis and + ANA or + anti-dsDNA	Entry criterion: + ANA of ≥1:80 on Hep2IF (or equivalent) Additive and weighted criteria: >1 clinical criterion and ≥ 10 points
Malar rash	Malar rash	Malar rash	Malar rash	Acute cutaneous rash	Acute cutaneous lupus **(weight = 6)**
Discoid rash	Discoid rash	Discoid rash	Discoid rash	Chronic cutaneous rash	Subacute cutaneous OR discoid lupus **(weight = 4)**
Raynaud phenomenon	Raynaud phenomenon				
Alopecia	Alopecia			Nonscarring alopecia	Non-scarring alopecia **(weight = 2)**
Photosensitivity	Photosensitivity	Photosensitivity	Photosensitivity		
Oral/nasal ulcers	Oral or nasal ulcers	Oral or nasal ulcers	Oral or nasal ulcers	Oral or nasal ulcers	Oral ulcers **(weight = 2)**
Arthritis	Arthritis without deformity	Nonerosive arthritis in at least two peripheral joints	Nonerosive arthritis in at least two peripheral joints	Synovitis involving at least two joints *or* tenderness in at least two joints with at least 30 min of morning stiffness	Joint involvement **(weight = 6)**
Serositis	A. Pleurisy	A. Pleurisy	A. Pleurisy	A. Pleurisy	Pleural or pericardial effusion **(weight = 5)**
	B. Pericarditis	B. Pericarditis	B. Pericarditis	B. Pericarditis	Acute pericarditis **(weight = 6)**

Renal disorder	A. Profuse proteinuria	A. Profuse proteinuria	A. Profuse proteinuria	A. Profuse proteinuria	Proteinuria >0.5 g/24 h **(weight = 4)**
	B. Cellular casts	B. Cellular casts	B. Cellular casts	B. Red blood cell casts	Renal biopsy class II or V lupus nephritis **(weight = 8)**
					Renal biopsy class III or IV lupus nephritis **(weight = 10)**
Neurologic	A. Psychosis	A. Psychosis	A. Psychosis	A. Psychosis	Psychosis **(weight = 3)**
	B. Convulsions	B. Seizures	B. Seizures	B. Seizures	Seizure **(weight = 5)**
				C. Mononeuritis multiplex	Delirium **Weight = 2)**
				D. Myelitis	
				E. Peripheral or cranial neuropathy	
				F. Acute confusional state	
Hematologic	A. Hemolytic anemia	A. Hemolytic anemia	A. Hemolytic anemia	A. Hemolytic anemia	Autoimmune hemolysis **(weight = 4)**
	B. Leukopenia	B. Leukopenia	B. Leukopenia	B. Leukopenia	Leukopenia **(weight = 3)**
	C. Thrombocytopenia	C. Thrombocytopenia	C. Thrombocytopenia	C. Lymphopenia	Thrombocytopenia **(weight = 4)**
		D. Lymphopenia	D. Lymphopenia	D. Thrombocytopenia	

(continued)

Table 11.3 (continued)

Criterion	ACR 1971 [6]	ACR 1982 [4]	ACR 1997 [5]	SLICC 2012 [7]	EULAR/ACR criteria 2019 [8]
Immunologic	A. LE cells	A. LE cells	A. False-positive STS	A. Anti-DNA	ANA at a titer of ≥1:80 on Hep-2 cells or an equivalent positive test (ever) required for entry criteria
	B. False-positive STS	B. False-positive STS	B. Anti-DNA	B. Anti-Sm	Anti-Sm **(weight = 6)**
		C. Anti-DNA	C. Anti-Sm	C. APS	Anti-cardiolipin antibodies OR anti-B2GP1 antibodies OR lupus anticoagulant **(weight = 2)**
		D. Anti-Sm	D. APS	C1. Lupus anticoagulant	
			E. Lupus anticoagulant	C2. False-positive rapid plasma reagin	
				C3. At least medium anticardiolipin ab	
				C4. Anti-β_2-glycoprotein I	
				D. Low complement (C3, C4, CH_{50})	Low C3 OR low C4 **(weight = 3)**
					Low C3 AND low C4 **(weight = 4)**
				E. Direct coombs test in the absence of hemolytic anemia	
Antinuclear antibody	Positive ANA	Positive ANA	Positive ANA	Positive ANA	ANA at a titer of ≥1:80 on Hep-2 cells or an equivalent positive test (ever) required for entry criteria
Constitutional					Fever **(weight = 2)**

revisited in the Boston Weighted Criteria, in which a renal biopsy with lupus nephritis (World Health Organization histologic classes 3–6) was given a heavy weight allowing classification, and points in the renal domain were not additive [9]. In 2012, the SLE International Collaborating Clinics (SLICC) group developed and published a further revised vision of the SLE classification criteria based on expert opinion with validation [10]. This version expanded the dermatologic manifestations to include acute cutaneous and subacute cutaneous lupus, as well as chronic cutaneous lupus, photosensitivity, oral and nasal ulcers, and nonscarring alopecia. It also expanded the immunologic criteria to include low complement and a direct test for Coomb's. They also specified that lupus glomerulonephritis alone sufficed for classification as SLE [10].

Since 2012, there was increasing recognition that patients with early SLE should be identified and classified for inclusion and clinical studies and trials [11]. The sensitivity of the 1997 updated ACR criteria was shown to be low in early SLE: 76% among individuals with up to 5 years of SLE duration and rose steadily over time as individuals accumulated organ involvement and thus criteria manifestations to greater than 94% in those who had had SLE for over 20 years [12]. Just as there is concern about delays in diagnosis leading to poor outcomes, so too does waiting for a long duration of SLE to accumulate criteria and allow classification limit inclusion in clinical trials and studies. Low sensitivity and early disease would exclude early SLE cases from trials and research studies and prevent testing of new compounds in those with early onset disease that could and thus limit our ability to understand early SLE pathogenesis and the ability to conduct trials and test treatments in early disease.

The European League Against Rheumatism (EULAR) received a proposal from Dr. Martin Aringer and then partnered with the ACR to provide support to the development of new international criteria to replace and update the updated ACR 1997 and the SLICC 2012 criteria. This was an international consensus to set a threshold for classifying definite SLE for clinical trials and research. The aims were also to validate these criteria as potentially more sensitive and specific than the 1997 updated ACR criteria and to evaluate if they were useful in classifying earlier disease. The 2019 EULAR/ACR classification criteria for SLE were developed by a process combining expert consensus and data-driven methods. This methodology has been used in development of other classification criteria including those for rheumatoid arthritis, scleroderma, gout, and IgG4 disease more recently and is currently thought to be the state of the art for criteria development, although this too will evolve with time. An international expert panel was formed and the development process proceeded in four phases, including (1) criteria generation [13]; (2) criteria reduction including a Delphi and nominal group technique exercise [14]; (3) item reduction, weighting, and threshold score determination [15, 16]; and (4) refinement and validation [17]. The process was iterative and took approximately 3 years.

A first multicenter Delphi exercise generated 145 new items from a large group of SLE international experts and then went through a process of reducing these items, eliminating duplicate and rare manifestations [14]. The key features for

classification of both early and established SLE included characteristic autoantibodies, specific renal features, and skin manifestations. Of note, 85% of the expert group stated they would positively classify a patient as having SLE with only a renal biopsy showing lupus nephritis, although there was debate within the group about the specificity of the histology of lupus nephritis in the absence of known SLE. A multicenter study of the presentations of patients newly diagnosed with SLE compared to SLE "mimickers" was helpful in informing the inclusion of fever, but not Raynaud's or sicca symptoms [18].

Developing and refining the new candidate criteria for classification was perhaps the longest of the phases with meetings to determine precise definitions and consider the relative sensitivity and specificity of non-overlapping items and organizing them into new domains. A systematic meta-regression revealed that the specificity and sensitivity for an ANA titer cutoff of 1:80 were optimal for inclusion of SLE patients, reducing the non-specificity of low titer ANA tests [19]. Thus, it was decided that an ANA of 1:80 titer or greater by immunofluorescence on Hep2 cells (later altered to "or the equivalent" as not all ANA assays are performed on Hep2 cells these days) would be required as an entry criterion. This was to be at least once, not necessarily at the most recent date. The group met for a 2-day multicriteria decision analysis process using 1000 Minds™ [16]. During this meeting, experts were presented with multiple forced-choice decisions in which the patient presented was conceived as identical in all other aspects except for two items. The relative weighting of all criteria items was then found through multiple paired decisions.

The 2-day multicriteria decision analysis meeting led to a set of preliminary SLE classification criteria with relative weights. The classification schema required items in both the clinical domains and immunologic domains, as well as having the inclusion criterion of at least 1:80 ANA positivity. There were no exclusion criteria, although an "attribution rule" was adopted. The attribution rule stated that for each of the items it would only be counted and given points if it was thought at the time of application of the criteria that the manifestation was at least it's likely to be lupus or more likely to be lupus than another cause. For example, an inflammatory arthritis would have had to be attributed to as SLE rather than to gout or rheumatoid arthritis. A continuous scale of increasing probability of SLE was produced with a threshold set at great and equal to 83 points above which all experts agreed that SLE was present. After further adjustment by their relative weights, these domains and criteria were adjusted such that the threshold for classification became greater than equal or to 10 points. In an extremely large international validation effort, new data from independent newly collected patients with possible SLE were collected for a two-step process of further refining and validation [17]. Data from 1001 patients were included in the derivation cohort, followed by data from in 1270 patients in validation cohort. The final gold standard for both groups was an international independent set of four SLE experts who were simply asked "Would you classify this patient as having SLE – yes or no?". The final new 2019 EULAR/ACR criteria had a sensitivity of 96.1% vs. 82.8% for the 1997 ACR criteria and a specificity of 93.4% vs. 93.4% for 1997 ACR criteria [17].

Cases Below the Threshold

Establishing a preliminary threshold for classification of SLE, above which 21 experts in the 2-day multicriteria decision analysis meeting were in consensus on classification as SLE, the international SLE panel then turned its attention to the possibility of "classifying" cases below the threshold into a more homogeneous group of likely SLE patients. Conceiving of classification on a scale of increasing likelihood of SLE and affording the opportunity to study the spectrum of SLE likelihood, including those cases that fall just below the classification threshold, the question they posed was, "Could we also find a lower threshold below which SLE was highly unlikely?", thus creating a group of intermediate probability of SLE, who might be candidates for observational studies or low-risk interventional trial (e.g., a trial of lifestyle change or non-toxic medication to prevent progression to SLE) [16].

The experts attempted to find a consensus for lower threshold of very low probability for classification. They discussed that individuals with scores falling between these definite and the lower thresholds might be candidates for inclusion in observational studies or SLE prevention trials. These patients could be considered as "potential" SLE and the group was asked "Would you feel comfortable enrolling this patient in an SLE prevention trial using a non-toxic medication (e.g., hydroxychloroquine)?". Experts individually rated the 52 cases below the upper threshold score (then 83) as "probable SLE," "possible SLE," or "unlikely SLE" (Fig. 11.1). The score of the case for which ≥70% indicated "unlikely SLE" = 27. 7 cases (14%) included in this exercise would be classified as "unlikely SLE" based on this lower threshold [16]. The remaining 86% would potentially be candidates for inclusion into observational or preventive studies. However, a very key piece of information on these cases was lacking, precluding pursuing this level as a second threshold: symptom duration. Experts soon acknowledged that having a few minor signs and symptoms of SLE for 20 years was a very different scenario than having them newly appearing in the past couple of years.

Patients below the definite SLE threshold are an extremely heterogeneous group including those with early SLE and a short duration of symptoms, as well as mild or *forme fruste* SLE, subjects with single organ involvement, SLE-like and related conditions, as well as the bona fide SLE cases that are still below the threshold. With our knowledge of the molecular underpinnings of SLE and the advancement of diagnostic techniques, including the cell-bound complement activation products, cytokines, chemokines, and novel autoantibodies, there are many exciting opportunities for research [20]. We have seen that the 2019 criteria perform well in patients with early disease and in many different populations [21]. It will be interesting to examine the duration and evolution of SLE signs, symptoms, and criteria over time. Studies are ongoing to follow subjects with molecular biomarkers and signatures to understand the heterogeneity of pre-SLE and that of lupus-related diseases.

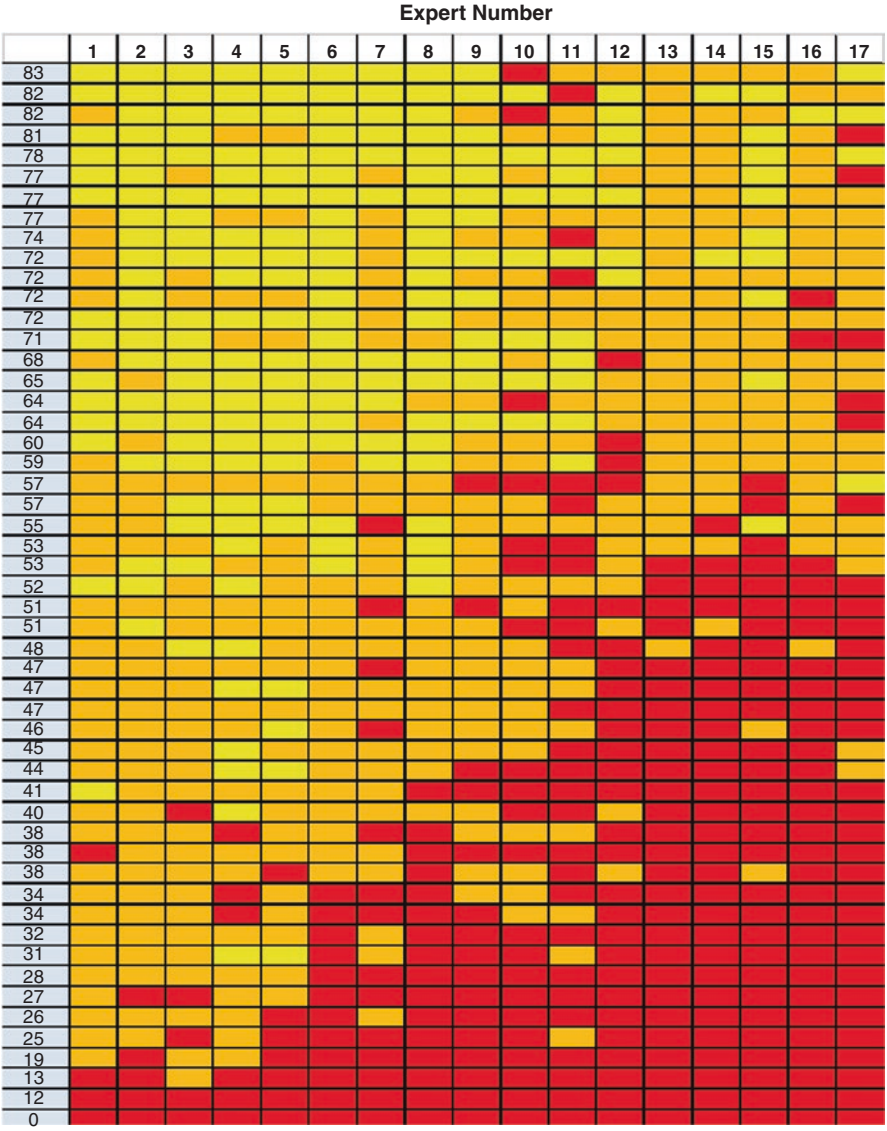

Fig. 11.1 Exercise to consider a lower threshold for classifying as probable, possible, and unlikely SLE. SLE expert panel members anonymously labeled SLE cases. Among 82 cases below the threshold, 52 cases had unique combinations of criteria (illustrated). Rows represent individual cases. Each column represents one SLE expert panel member. Yellow: probable SLE; orange, possible SLE; red, unlikely SLE. The gray column on the left is the average number of points awarded by the expert. These patient scenarios illustrate how diversely individual experts diagnose patients with SLE. (Adapted from reference [16])

Summary

Classification criteria fulfill a different purpose than does diagnosis, although the two have similar challenges: subjects below the threshold for either and the evolution of our diagnostic capabilities over time. Classification criteria for the purposes of clinical research are often used for teaching about SLE and its multisystem involvement, so it should have good face validity. However, it should always be recalled that they are not intended for diagnosis and the matter of diagnosis rests in the qualified treating physician's hands. Classification criteria necessarily evolve over time, as does our understanding of disease pathogenesis, subtypes, disease evolution, biomarker, and lab test availability, e.g., replacement of LE cell prep and the biologic false-positive test for SLE. It would be ideal to be able to use classification criteria as widely internationally as possible and for as long a time period as possible to be able to compare clinical trials and studies.

Acknowledgments The author acknowledges and thanks Ray Naden for his patience and brilliant insights, Sara Tedeschi and Corinne Sinnette for their unrivaled competence and data management skills, Martin Aringer and Sindhu Johnson for their leadership, and EULAR and ACR for their support of the 2019 SLE Classification Criteria effort.

References

1. Panush RS, Greer JM, Morshedian KK. What is lupus? What is not lupus? Rheum Dis Clin N Am. 1993;19:223–34.
2. Aggarwal R, Ringold S, Khanna D, Neogi T, Johnson SR, Miller A, Brunner HI, Ogawa R, Felson D, Ogdie A, Aletaha D, Feldman BM. Distinctions between diagnostic and classification criteria? Arthritis Care Res (Hoboken). 2015;67:891–7.
3. Aringer M, Johnson SR. Classifying and diagnosing systemic lupus erythematosus in the 21st century. Rheumatology (Oxford). 2020;59:v4–v11.
4. Aringer M. EULAR/ACR classification criteria for SLE. Semin Arthritis Rheum. 2019;49:S14–S7.
5. Cohen AS, Reynolds WE, Franklin EC, Kulka JP, Ropes MW, Shulman LE, Wallace SL. Preliminary criteria for the classification of systemic lupus erythematosus. Bull Rheum Dis. 1971;21:643–8.
6. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, Schaller JG, Talal N, Winchester RJ. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1982;25:1271–7.
7. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1997;40:1725.
8. Clough JD, Elrazak M, Calabrese LH, Valenzuela R, Braun WB, Williams GW. Weighted criteria for the diagnosis of systemic lupus erythematosus. Arch Intern Med. 1984;144:281–5.
9. Costenbader KH, Karlson EW, Mandl LA. Defining lupus cases for clinical studies: the Boston weighted criteria for the classification of systemic lupus erythematosus. J Rheumatol. 2002;29:2545–50.
10. Petri M, Orbai AM, Alarcon GS, Gordon C, Merrill JT, Fortin PR, Bruce IN, Isenberg D, Wallace DJ, Nived O, Sturfelt G, Ramsey-Goldman R, Bae SC, Hanly JG, Sanchez-Guerrero J, Clarke A, Aranow C, Manzi S, Urowitz M, Gladman D, Kalunian K, Costner M, Werth VP, Zoma A, Bernatsky S, Ruiz-Irastorza G, Khamashta MA, Jacobsen S, Buyon JP, Maddison P, Dooley MA, van Vollenhoven RF, Ginzler E, Stoll T, Peschken C, Jorizzo JL, Callen JP,

Lim SS, Fessler BJ, Inanc M, Kamen DL, Rahman A, Steinsson K, Franks AG Jr, Sigler L, Hameed S, Fang H, Pham N, Brey R, Weisman MH, McGwin G Jr, Magder LS. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum. 2012;64:2677–86.

11. Aringer M, Dorner T, Leuchten N, Johnson SR. Toward new criteria for systemic lupus erythematosus-a standpoint. Lupus. 2016;25:805–11.
12. Ines L, Silva C, Galindo M, Lopez-Longo FJ, Terroso G, Romao VC, Rua-Figueroa I, Santos MJ, Pego-Reigosa JM, Nero P, Cerqueira M, Duarte C, Miranda LC, Bernardes M, Goncalves MJ, Mourino-Rodriguez C, Araujo F, Raposo A, Barcelos A, Couto M, Abreu P, Oton-Sanchez T, Macieira C, Ramos F, Branco JC, Silva JA, Canhao H, Calvo-Alen J, Rheumatic Diseases Registry of the Portuguese Society of R, Registry of Systemic Lupus Erythematosus Patients of the Spanish Society of R. Classification of systemic lupus erythematosus: Systemic Lupus International Collaborating Clinics Versus American College of Rheumatology Criteria. A Comparative Study of 2,055 patients from a real-life, International Systemic Lupus Erythematosus Cohort. Arthritis Care Res (Hoboken). 2015;67:1180–5.
13. Johnson SR, Khanna D, Daikh D, Cervera R, Costedoat-Chalumeau N, Gladman DD, Hahn BH, Hiepe F, Sanchez-Guerrero J, Massarotti E, Boumpas DT, Costenbader KH, Jayne D, Dorner T, Kamen DL, Mosca M, Ramsey-Goldman R, Smolen JS, Wofsy D, Aringer M. Use of consensus methodology to determine candidate items for systemic lupus erythematosus classification criteria. J Rheumatol. 2019;46:721–6.
14. Schmajuk G, Hoyer BF, Aringer M, Johnson SR, Daikh DI, Dorner T, Committee SLEccs, the International SLEepoti. Multicenter Delphi exercise to identify important key items for classifying Systemic Lupus Erythematosus. Arthritis Care Res (Hoboken). 2018;70:1488–94.
15. Tedeschi SK, Johnson SR, Boumpas D, Daikh D, Dorner T, Jayne D, Kamen D, Lerstrom K, Mosca M, Ramsey-Goldman R, Sinnette C, Wofsy D, Smolen JS, Naden RP, Aringer M, Costenbader KH. Developing and refining new candidate Criteria for Systemic Lupus Erythematosus Classification: an international collaboration. Arthritis Care Res (Hoboken). 2018;70:571–81.
16. Tedeschi SK, Johnson SR, Boumpas DT, Daikh D, Dorner T, Diamond B, Jacobsen S, Jayne D, Kamen DL, McCune WJ, Mosca M, Ramsey-Goldman R, Ruiz-Irastorza G, Schneider M, Urowitz M, Wofsy D, Smolen JS, Naden RP, Aringer M, Costenbader KH. Multicriteria decision analysis process to develop new classification criteria for systemic lupus erythematosus. Ann Rheum Dis. 2019;78:634–40.
17. Aringer M, Costenbader K, Daikh D, Brinks R, Mosca M, Ramsey-Goldman R, Smolen JS, Wofsy D, Boumpas DT, Kamen DL, Jayne D, Cervera R, Costedoat-Chalumeau N, Diamond B, Gladman DD, Hahn B, Hiepe F, Jacobsen S, Khanna D, Lerstrom K, Massarotti E, McCune J, Ruiz-Irastorza G, Sanchez-Guerrero J, Schneider M, Urowitz M, Bertsias G, Hoyer BF, Leuchten N, Tani C, Tedeschi SK, Touma Z, Schmajuk G, Anic B, Assan F, Chan TM, Clarke AE, Crow MK, Czirjak L, Doria A, Graninger W, Halda-Kiss B, Hasni S, Izmirly PM, Jung M, Kumanovics G, Mariette X, Padjen I, Pego-Reigosa JM, Romero-Diaz J, Rua-Figueroa Fernandez I, Seror R, Stummvoll GH, Tanaka Y, Tektonidou MG, Vasconcelos C, Vital EM, Wallace DJ, Yavuz S, Meroni PL, Fritzler MJ, Naden R, Dorner T, Johnson SR. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. Arthritis Rheumatol. 2019;71:1400–12.
18. Mosca M, Costenbader KH, Johnson SR, Lorenzoni V, Sebastiani GD, Hoyer BF, Navarra S, Bonfa E, Ramsey-Goldman R, Medina-Rosas J, Piga M, Tani C, Tedeschi SK, Dorner T, Aringer M, Touma Z. Brief report: how do patients with newly diagnosed Systemic Lupus Erythematosus present? A multicenter cohort of early Systemic Lupus Erythematosus to inform the development of new classification criteria. Arthritis Rheumatol. 2019;71:91–8.
19. Leuchten N, Hoyer A, Brinks R, Schoels M, Schneider M, Smolen J, Johnson SR, Daikh D, Dorner T, Aringer M, Bertsias G, Systemic Lupus Erythematosus Classification Criteria Steering C. Performance of antinuclear antibodies for classifying Systemic Lupus Erythematosus: a systematic literature review and meta-regression of diagnostic data. Arthritis Care Res (Hoboken). 2018;70:428–38.

20. Ramsey-Goldman R, Alexander RV, Massarotti EM, Wallace DJ, Narain S, Arriens C, Collins CE, Saxena A, Putterman C, Kalunian KC, O'Malley T, Dervieux T, Weinstein A. Complement activation in patients with probable Systemic Lupus Erythematosus and ability to predict progression to American College of Rheumatology-Classified Systemic Lupus Erythematosus. Arthritis Rheumatol. 2020;72:78–88.
21. Johnson SR, Brinks R, Costenbader KH, Daikh D, Mosca M, Ramsey-Goldman R, Smolen JS, Wofsy D, Boumpas DT, Kamen DL, Jayne D, Cervera R, Costedoat-Chalumeau N, Diamond B, Gladman DD, Hahn B, Hiepe F, Jacobsen S, Khanna D, Lerstrom K, Massarotti E, McCune J, Ruiz-Irastorza G, Sanchez-Guerrero J, Schneider M, Urowitz M, Bertsias G, Hoyer BF, Leuchten N, Tani C, Tedeschi SK, Touma Z, Schmajuk G, Anic B, Assan F, Chan TM, Clarke AE, Crow MK, Czirjak L, Doria A, Graninger WB, Halda-Kiss B, Hasni S, Izmirly PM, Jung M, Kumanovics G, Mariette X, Padjen I, Pego-Reigosa JM, Romero-Diaz J, Rua-Figueroa I, Seror R, Stummvoll GH, Tanaka Y, Tektonidou MG, Vasconcelos C, Vital EM, Wallace DJ, Yavuz S, Meroni PL, Fritzler MJ, Naden R, Dorner T, Aringer M. Performance of the 2019 EULAR/ACR classification criteria for systemic lupus erythematosus in early disease, across sexes and ethnicities. Ann Rheum Dis. 2020;79:1333–9.

Prognosis: A Framework for Clinical Practice When Patients Have 'Symptoms with No Diagnosis'

12

Peter Croft

Diagnosis in Clinical Practice

The traditional cornerstone of clinical practice is the triad of diagnosis, treatment and prognosis [1].

> Fred, a previously fit 55-year-old schoolteacher, is learning what people mean by 'unbearable pain'. Overnight, his right big toe, after niggling uncomfortably for a day or two, has become dramatically red, swollen and very painful. His wife drives him to see Kate, their family doctor. Fred's history and Kate's examination indicate a diagnosis of gout and a course of an anti-inflammatory drug to relieve the acute attack. Kate discusses what is likely to happen next (short-term prognosis): 'This should settle quickly, Fred'. Once the pain and inflammation have settled, Kate arranges a blood test to measure Fred's urate level. Gout attacks occur when serum levels of this metabolic waste product are high and urate crystals form in joints. Fred's level is above normal, and a repeat test is arranged for some months' time. During that period, Fred experiences another severe attack, his urate remains high, and Kate investigates and rules out underlying diseases that can cause gout. At this stage she discusses what is likely to happen in the future (long-term prognosis): 'If we don't do something to bring down your urate levels, then you are likely to get more of these attacks and, what's more, these raised levels can cause other problems – in your kidneys, for example'. So Fred agrees to start daily allopurinol, a drug known to lower urate and prevent acute gout attacks and (importantly for Fred) with a well-established safety record when used long-term for prevention. The dose of Fred's allopurinol is adjusted against his regular urate measurements until those are normal. He remains free of further acute attacks.

This is clinical expertise and scientific medicine in action: the process of identifying and labelling a pathological abnormality underlying a patient's illness (disease diagnosis) and selecting a treatment targeting that disease. Fred's prognosis is determined by both the diagnosis (short-term effects of acute inflammation and long-term

P. Croft (✉)
Primary Care Centre Versus Arthritis and Centre for Prognosis Research, School of Medicine, Keele University, Keele, Staffordshire, UK
e-mail: p.r.croft@keele.ac.uk

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_12

effects of raised serum urate) and his likely response to treatment (immediate symptom relief by anti-inflammatory drugs and future prevention of gout attacks by allopurinol). This is why diagnosis represents the pinnacle of the textbook model of clinical art and science. Getting it right determines what treatment needs to be given. In this traditional model, prognosis (predicting what will happen in the future) follows from the diagnosis and treatment.

But this model is changing. One reason is because patient outcomes vary. Prognosis can differ substantially between people with similar diagnoses and treatment. Clinicians have known this for centuries, and part of their skill lies in communicating this uncertainty to patients, for example, when reassuring the patient that an apparently alarming diagnosis (cancer or arthritis) can have a benign course [2].

The science of clinical medicine has also changed to take account of this variability, for example, by disease-staging or identifying novel biomarkers. The aim is to characterise patients more precisely according to their likely future outcomes and response to treatment ('personalised medicine') [3]. Data on patient outcomes has become the arbiter of useful diagnosis by addressing the question: 'Will patients benefit in the future as a result of this new subgroup or biomarker?' The population sciences of epidemiology and statistics have underpinned this shift to prognosis as a driver of modern diagnostic process [4].

Prognostic information provides a framework for characterising variability even in people with an apparently uniform, biologically rooted diagnosis like gout [5, 6]. Serum urate levels predict long-term outcomes (renal compromise, cardiovascular disease) regardless of gout symptoms [7]. Obesity increases the risk of acute gout attacks [8]. Genetic markers are linked to both elevated urate levels and acute gout attacks [9]. Gout sufferers in some ethnic groups have poorer outcomes than others, raising questions about equitable access to care and underlying mechanisms behind the differences [10].

This book as a whole concerns the problems and challenges faced by patients who, in contrast to Fred, have a disease that does not fit available diagnostic categories and who may have been failed by current diagnosis and diagnostic process. Could prognosis and prognostic research help fill the gap?

From a clinical perspective, such patients fall into two groups. First are those who, like Fred, present a clinical picture consistent with pathological disruption of an internal organ system or systems, such as an immunological disease. This is why the patient sits in the consulting rooms of a relevant expert specialist. But their condition, unlike Fred's, fits no criteria for known diseases or available treatment. They have neither diagnostic label nor clear prognosis. They are excluded from treatment trials and, because of small patient numbers, have no relevant outcome data. They are individuals with a biological 'disease with no name', adrift in a measured world that currently excludes them. Many chapters in this book offer insights, perspectives and new evidence designed to help such individuals.

This chapter, however, concerns a second group – patients with symptoms but no obvious disease and no clinical picture of biological abnormality, their conditions often labelled by their symptoms: low back pain, widespread pain and chronic

fatigue. This is unsettling because symptoms are supposed to point to underlying pathology and diagnosable disease that can explain the illness and indicate treatment.

In the first group of patients, the clinician sees evidence of deeper pathology but finds it fits no known disease classification. The second group poses a different problem: the possibility that there is no useful diagnosis to make despite patients' distressing, disabling symptoms.

This chapter considers whether prognosis and prognostic science can offer an alternative to the diagnostic process for this second group as a practical way of framing their problem and guiding safe effective healthcare. There are four sections, each discussing a challenge posed by such patients.

Challenge 1: The Benefits and Hazards of Searching for a Diagnosis

> It's all very well you telling me I shouldn't Xray patients with back pain, but I remember the man with simple back pain whom I decided to Xray. It showed spinal metastases from his undiagnosed prostate cancer. *Mike, an experienced senior UK family doctor attending a guideline seminar.*
>
> If they had told me at the first consultation that there might not be a diagnosis to make, it would have been far better. *Tracy, a 25-year-old woman, attending a UK public meeting on pain, whose forearm pain had been investigated inconclusively by many hospital departments over a 2-year period.*

Pathological abnormalities, such as Fred's raised urate, provide targets for mechanism-based treatments. Conditions such as low back pain (LBP] threaten this model. LBP looks similar to gout on paper – chronic, long-term, with intermittent acutely painful episodes or flares. It affects one-third of adults in any 1 year, one in ten of whom develop persistent disabling pain, making it globally the single most prevalent cause of years lived with disability [11]. Some important spinal pathologies that can cause LBP, such as cancer and spinal cord compression, need urgent diagnosis. However, such 'red flag' diagnoses are rare in the absence of other symptoms: between 1 and 10 per 1000 primary care LBP consulters annually [12]. But this is common enough in an average clinician's working life for doctors like Mike to resist guideline advice against routine imaging of the spine in uncomplicated LBP [13, 14]. The evidence for that guideline advice derives from studies showing that such imaging leads to a rise in the number of interventions but no added improvement in patient outcomes [15, 16]. Most patients with LBP in primary care do not have an important or useful diagnosis to be made that would alter their outcome.

This is the diagnostic balancing act for modern primary care: the hazard of missing serious treatable conditions ('underdiagnosis') versus the dangers and costs of pursuing and treating conditions that pose no threat to health or survival ('overdiagnosis') [17, 18]. Prognostic evidence about likely future outcomes, with and without diagnostic procedures and labelling, helps restrain diagnosis for diagnosis' sake [5]. A growing number of publications are investigating the point on the spectrum of a

condition when the harms of searching for and making a diagnosis, and of the resulting treatment, outweigh possible benefits – such as identifying mild cases of conditions that overlap with normal variability (e.g. attention-deficit/hyperactivity disorder [19]) or disease types that, untreated, have no adverse outcomes (e.g. some melanomas [20]). Overdiagnosis may result in different degrees of waste and costs in contrasting healthcare systems (e.g. USA versus UK), but, as Tracy's quote illustrates, patients in any healthcare system do not necessarily appreciate unbridled diagnostic enthusiasm for its own sake.

Challenge 2: How to Classify Patients with Symptoms that Do Not Have a Diagnosis?

> It is a most excellent thing for a physician to cultivate prognosis …… predicting and foretelling……
>
> *Hippocrates. Fifth-Century BCE*

For centuries before a clinical science emerged to underpin diagnosis and effective treatment, prognosis was the consummate clinical skill – the art of prediction based on cumulative knowledge of patients past and present. Modern prognosis research now provides numerical estimates of the likelihood of future outcomes. Four types of prognostic study provide the architecture for this science [3]. Each study type contributes evidence to guide care for patients with a symptom like LBP in the absence of an important or useful diagnosis.

Type I studies classify such patients according to 'what is likely to happen in the future', given current knowledge, treatments, healthcare systems and social attitudes. In one study, patients presenting in primary care with LBP were followed for 12 months and grouped into those who did and did not develop persistent disabling pain [21]. Type 2 prognosis studies then ask what factors at baseline predict future outcomes. In the case of LBP, these include physical factors such as leg pain or widespread pain, and psychosocial factors such as anxiety about work capacity [22]. These factors may be preventable influences on low back pain progression, but they also enable further subgrouping of patients by risk of future outcomes. Type 3 studies then combine multiple prognostic factors into a single risk score for individual patients. An example is the STarT Back tool, a self-complete questionnaire that provides a score which, after red flag diagnoses are ruled out, places patients with LBP in one of three strata of predicted 12-month risk of disabling pain: low, medium and high [22]. Patients receive interventions matched to their risk stratum – e.g. exercise for those at low risk, physical therapy for the medium risk and more complex pain management for the high risk. In a UK trial of LBP patients randomised to care that did or did not use the STarT Back prognostic tool, the screened group was better off than controls 12 months afterwards, in terms of pain and disability and return-to-work rates, and had lower healthcare costs because unnecessary referrals were avoided in the low-risk group [23]. Type 4 studies explore whether responses to treatment differ by baseline prognosis [3].

'Prognostic classification with a purpose' – describing, grouping and managing people according to their risk of future outcomes – has here replaced obsession with diagnostic classification for its own sake and provides an alternative framework for managing patients with LBP. Such prognostic stratification must link with effective care or avoidance of low-value care in order to be useful, just as diagnostic classifications must do, and this may need changes in the healthcare system [24].

Challenge 3: Explaining Symptoms When There Is No Diagnosis – The Example of Chronic Pain

> Someone believed me, and understood what I was suffering. Here was a diagnosis at last.... recognition and the possibility of a proper treatment or even a cure [1]

The example of successful stratified care for LBP leaves unanswered how to explain a recurrent disabling symptom such as chronic LBP when there is no diagnosable pathology or label to go with it. And labels are important. For many patients with chronic widespread pain, the arrival of the diagnostic term 'fibromyalgia' was greeted with relief, as the quotes above illustrate, because it validated patients' distress and the pursuit of resources to help manage their work loss and disability. Yet it has not led to discovery of a distinctive pathology or specific treatments for the condition. And there is an alternative – to explain the pain itself.

The crucial insight for this emerged from neurophysiology. The old view of pain as a fixed response in the brain to a source of injury or disease, such that once the source was removed or treated, the pain would disappear, was replaced by the notion of pain as a flexible memory laid down across the whole nervous system that can outlive its original cause and be modified, crucially, over time by the whole biopsychosocial environment of the patient. Such explanations have been incorporated by well-informed articulate primary care and musculoskeletal professionals into their conversations with patients. Evidence suggests providing such neurophysiological explanations may contribute, albeit modestly, to improved patient prognosis [25]. Prognostic studies support the content of such explanations by identifying the factors that predict future progression of pain from acute to chronic. These embrace the biophysical (severity and extent of initial pain and injury), the psychological (emotional and cognitive factors) and the social (work status and relationships, compensation systems) [11]. Interventions such as multidisciplinary rehabilitation that address all these components in combination are effective in improving outcomes for patients with chronic LBP [26].

The irony is that chronic pain, this complex multifaceted long-term problem, has now achieved the status of a diagnosis in the latest International Classification of Diseases (ICD-11) [27]. Symptoms and their prognosis underpin this new

[1] Quotes from patients with chronic unexplained widespread pain on receiving the diagnosis of fibromyalgia.

diagnostic group, which can be applied in the presence or absence of concurrent diagnoses (e.g. rheumatoid arthritis or inflammatory bowel disease) that might be the original source of pain. One-third of people with inflammatory bowel disease (Crohn's or ulcerative colitis), for example, still report chronic pain after their gut inflammation has remitted on biologic treatments [28]; one-third of people with new onset rheumatoid arthritis whose inflammatory markers have responded to treatment report refractory pain 2 years later [29].

The new ICD category will hopefully encourage implementation of effective chronic pain management among such patients, whilst the 'symptom' label means no one is excluded – people are defined by what they complain of, not by the presence or absence of underlying diagnosable pathology. The new category builds on an understanding of biological mechanisms of pain perception and modulation and links to mechanism-targeted interventions, which range from novel molecular neurological targets to social and psychological interventions that help people to work around their pain. But it does not solve all the problems faced by patients who have 'symptoms with no diagnosis'.

Challenge 4: How Do We Care for the Sick Individual in a Measured World?

> What shall I do with all this immensity in a measured world?[2]

Application of healthcare data to personalised medicine, via machine learning and artificial intelligence, promises to improve estimates of an individual's likely future outcome and response to treatment. Yet this risk prediction is still based on group-level measurements.

Such evidence informs policy decisions for groups, and individuals may benefit as part of a group. For example, risk estimates of poor outcomes associated with older age and comorbidity if people were to contract Covid-19 [30] were used in prognostic models that helped to drive UK policies for shielding vulnerable groups during the pandemic and prioritising them for vaccination. Yet for the individual worrying about returning to the workplace, uncertainty remained, however precise the estimates of their risk. The future for the individual cannot be known for certain.

In an international report on diagnosis, the authors emphasised how caring consultations and conversations with patients, in addition to good prognostic evidence, remain critical to resolving the balancing act between under- and overdiagnosis. These latter, the report argues, are not in opposition but are 'sides of the same coin, unified by the need for a more thoughtful, caring and conservative approach to diagnosis' [31].

[2] Marina Tsvetayeva, Selected Poems, translated by Elaine Feinstein, Second edition. Oxford: Oxford University Press, 1981.

Such caring may include 'gut instinct' when it represents a clinician's long-time knowledge of their patient (a person with LBP perhaps, in their 70s, who rarely complains but today is not their usual self) and reflects Bayesian thinking ('Is this person more likely than the average person with LBP to have a serious cause, and make further investigation more likely to deliver useful information?'). But caring can also align with diagnostic restraint, if the patient consulting with LBP today reveals they have a relative who recently died from metastatic prostate cancer and it is this worry they want to discuss with a doctor who has time to help them weigh up if medical investigations are really the best way to resolve their anxiety.

Planning the Future

The search for mechanisms underlying 'symptoms with no diagnosis' must continue. So too must the search for new ways to integrate the many influences on prognosis in people with chronic symptoms such as LBP – from the personal (e.g. obesity) to the contextual (e.g. no access to high value care), the social (e.g. social inequality) to the biological (e.g. injury) – and hence create combined biopsychosocial models attractive to everyone. But we need also to classify people in ways which (mechanism or no mechanism) help to relieve their symptoms and enable them to live active lives despite symptoms. Prognosis and prognostic research offer frameworks for this, but they are still rooted in population science that is informative, policy-driving and useful, but not finally able to say with certainty what will happen to that particular individual in the consulting room.

Shared management and treatment decisions must, therefore, be guided by clinical experience and patient perceptions as well as evidence from prognostic studies. This concerns the difference between 'risk' and 'uncertainty'. Risk is the empirical estimate of what will happen in future to the group in which the patient is classified. Shared decision-making between patient and clinician is informed by those estimates, but must also incorporate the uncertainty of what will actually happen in the future to that particular individual. This is old-style prognostication, drawing on the values and experience of both clinician and patient, but invigorated and informed by evidence from prognosis research.

References

1. Del Mar C, Doust J, Glasziou P. Clinical thinking. Oxford: BMJ Books, Blackwell Publishing Ltd; 2006.
2. Chauffard A. Medical prognosis: its methods, its evolution, its limitations. Br Med J. 1913;2:286–90.
3. Riley RD, et al. Prognosis research in healthcare. Concepts, methods, and impact. Oxford: Oxford University Press; 2019.
4. Armstrong D. Diagnosis: from classification to prediction. Soc Sci Med. 2019;237:112444. https://doi.org/10.1016/j.socscimed.2019.112444.

5. Croft P, et al. The science of clinical practice: disease diagnosis or patient prognosis? Evidence about "what is likely to happen" should shape clinical practice. BMC Med. 2015;13:20. https://doi.org/10.1186/s12916-014-0265-4.
6. Rothenbacher D, et al. Frequency and risk factors of gout flares in a large population-based cohort of incident gout. Rheumatology. 2011;50:973–81.
7. Gaubert M, et al. Hyperuricemia and hypertension, coronary artery disease, kidney disease: from concept to practice. Int J Mol Sci. 2020;21:4066. https://doi.org/10.3390/ijms21114066.
8. Roddy E, Choi HK. Epidemiology of gout. Rheum Dis Clin N Am. 2014;40:155–75.
9. Dalbeth N, et al. Lancet. 2016;388:2039–52.
10. Dalbeth N, et al. The experience and impact of gout in Māori and Pacific people: a prospective observational study. Clin Rheumatol. 2013;32:247–51.
11. Hartvigsen J, et al. Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. Lancet. 2018;391:2356–67.
12. Henschke N, et al. Prevalence of and screening for serious spinal pathology in patients presenting to primary care settings with acute low back pain. Arthritis Rheum. 2009;60:3072–80.
13. Chou R, et al. Clinical Guidelines Committee of the American College of Physicians. Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. Ann Intern Med. 2011;154:181–9. https://doi.org/10.7326/0003-4819-154-3-201102010-00008.
14. National Institute for Health and Care Excellence. Low back pain and sciatica in over 16s: assessment and management. NICE Guideline No. 59. London: National Institute for Health and Care Excellence (UK); 2020. ISBN-13: 978-1-4731-2186-7.
15. Jarvik JG, et al. Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. JAMA. 2003;289:2810–8.
16. Jarvik JG, et al. Association of early imaging for back pain with clinical outcomes in older adults. JAMA. 2015;313:1143–53.
17. Moynihan R, et al. Preventing overdiagnosis: how to stop harming the healthy. BMJ. 2012;344:e3502.
18. Treadwell J, McCartney M. Overdiagnosis and overtreatment: generalists - it's time for a grassroots revolution. Br J Gen Pract. 2016;66:116–7.
19. Kazda L, et al. Overdiagnosis of attention-deficit/hyperactivity disorder in children and adolescents: a systematic scoping review. JAMA Netw Open. 2021;4:e215335. https://doi.org/10.1001/jamanetworkopen.2021.5335.
20. Welch HG, et al. The rapid rise in cutaneous melanoma diagnoses. N Engl J Med. 2021;384:72–9. https://doi.org/10.1056/NEJMsb2019760.
21. Dunn KM, et al. Characterizing the course of low back pain: a latent class analysis. Am J Epidemiol. 2006;163:754–61.
22. Hill JC, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum. 2008;59:632–41.
23. Hill JC, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet. 2011;378:1560–71.
24. Cherkin D, et al. Effect of low back pain risk-stratification strategy on patient outcomes and care processes: the MATCH randomized trial in primary care. J Gen Intern Med. 2018;33:1324–36.
25. Tegner H, et al. Neurophysiological pain education for patients with chronic low back pain: a systematic review and meta-analysis. Clin J Pain. 2018;34:778–86.
26. Foster NE, Anema JR, Cherkin D, et al. Lancet Low Back Pain Series Working Group. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet. 2018;391:2368–83.
27. Treede RD, et al. Chronic pain as a symptom or a disease: the IASP classification of chronic pain for the international classification of diseases (ICD-11). Pain. 2019;160:19–27.

28. Keohane J, et al. Irritable bowel syndrome-type symptoms in patients with inflammatory bowel disease: a real association or reflection of occult inflammation? Am J Gastroenterol. 2010;105:1789–94.
29. Olofsson T, et al. Pain over 2 years after start of biological versus conventional combination treatment in early rheumatoid arthritis: results from the randomised controlled SWEFOT trial. Arthritis Care Res. 2020; https://doi.org/10.1002/acr.24264.
30. Williamson EJ, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020;584:430–6.
31. Schiff GD, et al. Ten principles for more conservative, care-full diagnosis. Ann Intern Med. 2018;169:643–5.

When the Illness Has No Name: Focus on Clinical Trials in Systemic Lupus Erythematosus

13

Richard Furie

Chapter

There is little question that *uncertainty* is an unwelcomed intruder into our systemic lupus erythematosus (SLE) clinical trials. Particularly vulnerable are studies investigating the effects of experimental drugs on *extra-renal* manifestations, such as those within the mucocutaneous and musculoskeletal domains. Despite study protocols' requirements for enrolled patients to satisfy SLE classification criteria [1–3] as well as specific thresholds of disease activity, several items comprising both the classification criteria and disease activity instruments [4, 5] are sufficiently subjective to varied interpretations by investigators. It is this subjectivity that increases the uncertainty of the outcomes of SLE clinical trials, promotes placebo responses, and jeopardizes proper drug development. The steps taken over the last two decades to enhance objectivity in SLE clinical trials have resulted in improved trial designs and outcomes.

Do All Subjects Enrolled in SLE Clinical Trials Have SLE?

My first encounter with a clinical trial that made me ponder these questions was the phase 2 belimumab study [6]. Belimumab [7] is a monoclonal antibody that targets and antagonizes a key growth factor for B cells, known as B lymphocyte stimulator, or BLyS (also known as B-cell-activating factor, or BAFF) [8]. Since SLE is the prototypic autoimmune disease with clear evidence of B lymphocyte hyperactivity, it was scientifically sound to approach the treatment of SLE with a drug that

R. Furie (✉)
Zucker School of Medicine at Hofstra/Northwell, Division of Rheumatology, Northwell Health, Great Neck, NY, USA
e-mail: rfurie@northwell.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_13

subdues B lymphocytes. The phase 1 study, which established safety and observed a pharmacodynamic effect, served as the foundation for the phase 2 study [9]. The phase 2 study was a typical *extra-renal* SLE study that investigated whether belimumab could reduce disease activity. Eligibility required a diagnosis of SLE as well as a certain degree of clinical activity based upon a commonly used SLE disease activity instrument, known as the SLE Disease Activity Index [5].

The study failed to achieve its endpoint of reduction in disease activity. Pertinent to this discussion is the fact that 28% of the enrolled patients at entry into the study were seronegative for antinuclear antibodies (ANA) or double-stranded DNA (DNA) antibodies. This finding was not so surprising since the classification criteria for SLE used at the time did not require ANA or DNA antibody positivity for SLE classification. The protocol required a history of such autoantibodies but not necessarily at entry into the study. What was unsettling was the relatively high frequency of seronegativity despite the high degree of clinical activity at study entry. Did these seronegative patients truly have SLE or perhaps they had SLE but their clinical activity at the time might have been overstated?

Do Classification Criteria Ensure the Purity of the Study Enrollment?

Of the 4 criteria (out of a possible 11) that had to be met to be classified as SLE, arthritis and rash have been two of the more common manifestations present in patients entering SLE clinical trials. Delving into some of the clinical nuances of SLE raises the following questions. Did patients have inflammatory arthritis or just arthralgias? Did patients have cutaneous lupus in the malar distribution or was it rosacea? There is no easy way to answer these questions as patients in clinical trials may have been diagnosed in the distant past and in many situations diagnosed by a different physician. Despite the existence of several different SLE classifications and definitions of each criterion, they all consist of objective and subjective criteria. Thus, the determination of an SLE diagnosis is ultimately in the hands of the clinician.

Do SLE Study Subjects Actually Have Their Stated Disease Activity?

Although lupus nephritis and *extra-renal* clinical trials have the same goals of reducing disease activity, their designs are dissimilar owing largely to treatment algorithms that are quite different. Lupus nephritis trials enroll patients with proliferative nephritis in need of induction therapy with an immunosuppressive or cytotoxic agent. With the requirement of a recent kidney biopsy demonstrating active International Society of Nephrology/Renal Pathology Society [10] class III or IV lupus nephritis and proteinuria above a specific threshold, the validity of the diagnosis and the activity of the disease are rarely called into question. A tissue

diagnosis and a laboratory result are the ultimate proof. Confirmatory central pathology reading is sometimes incorporated in the protocol.

The measures included in lupus nephritis trials that determine eligibility and outcomes contrast sharply with those incorporated into *extra-renal* SLE studies with respect to their degrees of objectivity. A typical *extra-renal* SLE study requires patients with disease activity that exceeds a particular threshold based on the SLEDAI or BILAG, commonly used disease activity instruments. A rather conventional SLEDAI requirement for entry is 6 points, which represents moderate disease activity. Six points can be attained with arthritis (4 points) and rash (2 points) or a combination of oral ulcers, alopecia, and rash (each 2 points). With rather large incentives for both investigators and patients, patients may be entering studies with overstated clinical activity or disease manifestations not attributed to SLE. SLEDAI criteria that are commonly present at the screening visit where evidence of the disease manifestation is not required include components of the musculoskeletal (arthritis, myositis) and mucocutaneous domains (rash, alopecia, oral ulcers). As previously mentioned, is arthritis being checked on the SLEDAI form because of true synovitis or do they have tender joints on a non-inflammatory basis. Likewise, is the rash a lupus rash or one of its many mimics? Additional examples include oral ulcers and alopecia. Is the oral ulcer a classic palatal ulcer or is it an aphthous ulcer? Is alopecia being scored because the patient noted hair in the shower drain or is it from active SLE?

The inclusion of a patient with overstated clinical activity enrolls an uninformative patient who is as likely to respond to placebo as the experimental drug. Let's return to the phase 2 belimumab trial where 28% of the cohort were serologically inactive at baseline. Compared to the serologically active group, the serologically inactive group had a lower baseline SLEDAI score, a higher frequency of oral ulcers, and many other characteristics that favored lower disease activity. Similar occurrences were observed in other SLE drug development programs. Although the phase 2 epratuzumab (antibody to CD22, a protein on B lymphocytes) study enrolled patients based on BILAG scores, 7% of the entire cohort entered with lupus headache at baseline, a feature "worth" 8 points on SLEDAI. Kalunian and colleagues noted in the two tabalumab (monoclonal antibody directed against BLyS) phase 3 *extra-renal* studies that alopecia was present at baseline in 61% and 55% of the patients, and oral ulcers were present in 30% and 35% of patients [11]. Response rates of alopecia and oral ulcers in the placebo groups were approximately 80% and 45%, respectively. With such high placebo response rates, it is very difficult, if next to impossible, for an experimental therapy to prove itself.

Are Study Subjects Taking Their Medications?

The success of a clinical trial is not only determined by the investigators' enrollment of informative patients. It is vital to select patients who are reliable. Clinicians constantly face non-compliance and non-adherence issues with their SLE patients, but such behavior is not acceptable in a study. Missing data is the nemesis of any

clinical trialist. While attendance at study visits is easily tracked, medication adherence beyond optional pill counts is not. Measuring blood concentrations of hydroxychloroquine, non-adherence was first emphasized by Costedoat-Chalumeau et al. [12]. SLE clinical trials until recently have not measured concentrations of some of the standard of care background medications. Therefore, the lupus research community is unaware of the impact of this important issue on study outcomes.

What Steps Can Be Taken to Promote Certainty in SLE Clinical Trials?

Ensure Patients Have SLE

The most recent SLE classification criteria, the 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus, requires an ANA of at least 1/80 titer on HEp-2 cells or an equivalent positive test at least once; otherwise, the patient is considered not to have SLE [1]. While these criteria have excellent sensitivity and specificity, they do not add to the SLE clinical trial cause. Positive ANAs are encountered quite frequently in the general healthy population. Although definitions exist for the various criteria, fulfillment of some of the clinical domains and criteria still involves subjective assessments. Thus, classification criteria do not guarantee the diagnosis, but they certainly enrich the clinical trial population and provide a level of comfort regarding the composition of the trial cohort. Overriding any issues raised by the classification conundrum is the requirement for satisfaction of additional entry criteria related to disease activity. Therefore, I feel the more important need is the enrichment of clinical trials with informative patients.

Enrich the Study Population with Informative Patients

Although many factors contribute to the success of a clinical trial, the one that rises to the top is the enrichment of informative patients, that is, the enrollment of patients with bona fide activity related to SLE that is potentially reversible.

Since the phase 2 belimumab study, all SLE clinical trials have required a positive ANA or DNA antibody at screening in order for patients to enroll in the study. A couple of sponsors have gone a step further and eliminated ANA from the requirement recognizing the lack of specificity of this test. In some protocols, antibodies to Sm have been an option to satisfy the autoantibody requirement. While more restrictive entry criteria will create a more informative cohort, recruitment into studies will no doubt be compromised.

In more recent years, some trials that require a SLEDAI threshold be exceeded for enrollment have restricted the “soft” SLEDAI items from counting toward the screening SLEDAI requirement. Alopecia (worth 2 points), oral ulcers (worth 2 points), and lupus headache (worth 8 points) are some of the SLEDAI items that

have been targeted for exclusion in order to reduce background "noise." In addition, it is commonplace to see a requirement for a clinical SLEDAI of 4 points to be present at both the screening and baseline visits. A clinical SLEDAI consists of clinical items only; laboratory items are excluded from the point totals. This ensures that clinical activity is still present just prior to receipt of the experimental therapy. Another benefit of incorporating the clinical SLEDAI is the demonstration that reductions in disease activity with the SLEDAI are not the sole result of improvement in laboratory tests; rather, a true clinical improvement has taken place.

While the SLEDAI and BILAG provide opportunities to evaluate specific organ domains, "extra-renal" trials now incorporate additional organ-specific instruments. To quantify cutaneous activity and damage, the Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI) is generally included in clinical trials [13]. For studies focusing on effects of experimental therapies on skin lesions in SLE, a minimum amount of CLASI activity is typically required for entry. By raising the minimum requirement for a CLASI score at entry, the experimental drug is given a true test, and the placebo effect is reduced. Despite the complexity and expense of photography, it is sometimes incorporated in early proof of concept trials to confirm the degree and extent of cutaneous involvement. Adapted from rheumatoid arthritis trial designs, SLE clinical trials now include assessments of swollen and tender joints beyond the musculoskeletal measures incorporated into SLEDAI and BILAG. In contrast to SLEDAI and BILAG, joint scores afford the ability to apply thresholds for entry into studies as well as to better quantify changes over the course of the study. Similar to cutaneous disease, the goal with arthritis is to reduce the placebo response by requiring greater activity.

A post-hoc analysis of the two phase 3 belimumab studies, BLISS-52 and BLISS-76 [14, 15], led to the term "high disease activity" by van Vollenhoven et al. [16]. He and his co-authors demonstrated in the combined phase 3 dataset that a subset of patients with DNA antibodies and low complements at baseline, accounting for approximately 50% of the original cohort, had an enhanced effect when treated with belimumab compared to placebo treatment. Although this was a post-hoc analysis, other sponsors recognized the potential benefit of utilizing this observation prospectively. As a result, several trial designs appeared where high disease activity patients with SLEDAI scores above the typical value of 6 were required. Another approach utilized in the anifrolumab (a monoclonal antibody to the type I interferon receptor that inhibits the type I interferon pathway) program was to incorporate both SLEDAI and BILAG requirements at entry [17]. All of these strategies were intended to increase disease activity among those patients screening for study participation. These efforts would ultimately reduce the placebo response rate and increase the effect size.

Adjudication, now routine in SLE clinical trials, has undoubtedly improved our trial design methodology. It can be performed at any step in the study – from entry to the final outcome assessment. When performed at entry, inappropriate and uninformative patients for the study may be excluded.

Enrichment of a study cohort based on a molecular signature or trait that associates with an increased response rate has been the dream of investigators. However,

predictive biomarkers have remained elusive. In the anifrolumab phase 2 and 3 studies [17–19], patients were classified as interferon-high or interferon-low based on the presence or absence of an interferon signature. The effect sizes were greater in the interferon-high groups in all three studies, owing to higher placebo responses in the interferon-low groups. Such findings support identifying a molecular signature that may enhance effect sizes. A similar observation of heightened effect in those with an Aiolos (a nuclear transcription factor crucial to inflammatory cell development) signature was made in the phase 2 iberdomide (a modulator of cereblon and promoter of Aiolos degradation) SLE study reported by Merrill et al. at the American College of Rheumatology meeting in 2020.

Promote Compliance and Adherence

Easier said than done is to enforce patient visit compliance and medication adherence. Participants in clinical trials are selected, in part, based on their reliability. While they may appear for their study visits, investigators and coordinators have never been assured patients were taking their study or background medications except if the medication was administered parenterally during the research visit. Pill counts are quite antiquated, but we are now witnessing the integration of drug level determinations, such as hydroxychloroquine assays, at screening. Those who allege taking the medication yet have no evidence of drug in their blood are potentially excluded from the study.

Conclusions

One might expect clinical trials to be pure without any need to discuss uncertainty. However, this has not been the case. SLE clinical trials have been plagued by enrollment of uninformative patients, high placebo responses, and countless failed studies. However, lessons learned over the last two decades have resulted in refinements in clinical trial design and many more successful trials. The lupus community will be rewarded with many more therapies, which will no doubt improve outcomes for those with SLE.

References

1. Aringer M, et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for systemic lupus erythematosus. Arthritis Rheum. 2019;71(9):1400–12.
2. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1997;40(9):1725.
3. Petri M, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum. 2012;64(8):2677–86.

4. Symmons DP, et al. Development and assessment of a computerized index of clinical disease activity in systemic lupus erythematosus. Members of the British Isles Lupus Assessment Group (BILAG). Q J Med. 1988;69(259):927–37.
5. Bombardier C, et al. Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. Arthritis Rheum. 1992;35(6):630–40.
6. Wallace DJ, et al. A phase II, randomized, double-blind, placebo-controlled, dose-ranging study of belimumab in patients with active systemic lupus erythematosus. Arthritis Rheum. 2009;61(9):1168–78.
7. Baker KP, et al. Generation and characterization of LymphoStat-B, a human monoclonal antibody that antagonizes the bioactivities of B lymphocyte stimulator. Arthritis Rheum. 2003;48(11):3253–65.
8. Moore PA, et al. BLyS: member of the tumor necrosis factor family and B lymphocyte stimulator. Science. 1999;285(5425):260–3.
9. Furie R, et al. Biologic activity and safety of belimumab, a neutralizing anti-B-lymphocyte stimulator (BLyS) monoclonal antibody: a phase I trial in patients with systemic lupus erythematosus. Arthritis Res Ther. 2008;10(5):R109.
10. Weening JJ, et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. Kidney Int. 2004;65(2):521–30.
11. Kalunian KC, et al. Clinical trial parameters that influence outcomes in lupus trials that use the systemic lupus erythematosus responder index. Rheumatology. 2018;57(1):125–33.
12. Costedoat-Chalumeau N, et al. Very low blood hydroxychloroquine concentration as an objective marker of poor adherence to treatment of systemic lupus erythematosus. Ann Rheum Dis. 2007;66(6):821–4.
13. Albrecht J, et al. The CLASI (Cutaneous Lupus Erythematosus Disease Area and Severity Index): an outcome instrument for cutaneous lupus erythematosus. J Invest Dermatol. 2005;125(5):889–94.
14. Navarra SV, et al. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. Lancet. 2011;377(9767):721–31.
15. Furie R, et al. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. Arthritis Rheum. 2011;63(12):3918–30.
16. van Vollenhoven RF, et al. Belimumab in the treatment of systemic lupus erythematosus: high disease activity predictors of response. Ann Rheum Dis. 2012;71(8):1343–9.
17. Furie R, et al. Anifrolumab, an anti-interferon-alpha receptor monoclonal antibody, in moderate-to-severe systemic lupus erythematosus. Arthritis Rheum. 2017;69(2):376–86.
18. Furie RA, et al. Type I interferon inhibitor anifrolumab in active systemic lupus erythematosus (TULIP-1): a randomised, controlled, phase 3 trial. Lancet Rheumatol. 2019;1(4).
19. Morand EF, et al. Trial of Anifrolumab in active systemic lupus erythematosus. N Engl J Med. 2020;382(3):211–21.

The Epidemiology of Systemic Lupus Erythematosus

14

S. Sam Lim

Epidemiology

Epidemiology is the science that studies the patterns, causes, and effects of health and disease conditions in defined populations. The father of epidemiology is widely regarded as John Snow. Born in 1813 as the son of a laborer, he went on to receive his medical degree from the University of London and achieved acclaim as a pioneer anesthetist. Early in his life, the first cholera pandemic had ravaged Europe. In 1831, the next wave had reached Britain and resulted in 30,000 deaths in just the first year. At the time, the cause of cholera was unknown but different theories abounded, including blaming miasmas or "vapors" from disintegrating materials from swamps or contagion. Dr. Snow had suspected cholera was spread by contaminated water, which initially did not gain much traction within the scientific community. When a severe outbreak occurred in a neighborhood near his home, he had a chance to prove his theory. In a systematic and detailed fashion, he interviewed the families of those who had fallen ill and mapped the cases. Noting that the concentration of deaths centered around users of the Broad Street water pump, he eventually convinced the local authorities to test his theory by removing the pump handle and rendering it inaccessible. The spread of cholera cases significantly dropped, and hence, the science of epidemiology was born.

Epidemiology has also advanced our understanding across a broad spectrum of disciplines, including chronic autoimmune conditions. Understanding who has these conditions will establish the foundation to better address other related and important questions, including how best to focus drug development, understand its

S. S. Lim (✉)
Department of Medicine, Division of Rheumatology, Emory University School of Medicine, Atlanta, GA, USA

Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA
e-mail: sslim@emory.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_14

impact, and direct efforts to where the needs are the greatest. Unlike infectious disease epidemiology, whereby a single vector, or agent, carries and transmits a disease, autoimmune diseases may not be as easily defined and introduces other challenges. These issues came to a head during an epidemiologic investigation of systemic lupus erythematosus (SLE) in the early 2000s. SLE is the prototype of a systemic autoimmune disease. There is tremendous amount of clinical variability, potentially involving any organ system in the body, as well as severity, ranging from mild to life-threatening illness.

As interest and research activity increased in SLE, the fundamental question of "How many people have SLE?" arose. Our charge from the Centers for Disease Control and Prevention (CDC) was to create a registry in order to identify and count everyone with SLE within a well-defined geographic location in a specific time period [1]. We chose the two central counties within metropolitan Atlanta, Fulton and DeKalb, in the years 2002–2004. It was important to define a relatively large population (around 1.5 million people) with a significant proportion of Black individuals, who are at higher risk for developing SLE. In partnering with our state health department, we were considered a public health surveillance project, which allowed us unprecedented access to medical records and databases without having to obtain consent from individuals, which would have been impractical for such a large project. We then embarked on finding individuals who may have SLE and then confirming their diagnosis.

Identifying Individuals Who May Have SLE

There is tremendous diversity and variability of findings in SLE and a lack of definitive and accurate diagnostic testing. Therefore, the "gold standard" for the diagnosis is based on the judgment of an experienced clinician who recognizes the various ways autoimmune features may present in SLE while excluding other causes. Since it was not possible to have such a large number of individuals evaluated by a rheumatologist, the primary experts in diagnosing SLE, we had to narrow our list to those who possibly may have SLE and then review their medical records to confirm the diagnosis.

Given that the diagnosis of SLE comes from physicians, we sought centralized lists from physician encounters in hospitals and clinics and other databases, including those from Medicaid and the Veterans Administration Hospital. Diagnoses are not directly captured by the medical system. Rather, codes are used, in part, to represent diagnoses but are often driven by other considerations. The International Classification of Diseases (ICD) is the coding system maintained by the World Health Organization that is routinely used in medical encounters in the USA and other countries. There are ICD codes related to SLE and associated conditions, which have been shown to be present in most cases of diagnosed SLE. However, these codes can also be used when the diagnosis has not been made. In the evaluation of someone who may have SLE, certain specialized tests may be needed. In order for some of those tests to be paid for by insurance companies, a test may need to be linked to an appropriate ICD codes in order for them to be reimbursed,

regardless of whether the diagnosis is final or not. And as is the case when there are high volumes of data initially entered manually, there will be a certain proportion entered in error. The remainder are from situations in which SLE may have been in consideration but not yet confirmed.

We also sought individuals who have had tests suggestive of SLE. Underlying the heterogeneous clinical features of SLE is the production of autoantibodies, the result of an immune reaction against one's own tissues. We are able to measure many of these autoantibodies through routinely available tests, which suggests that we can easily screen for this condition in a laboratory database. And indeed, it is true that some of these autoantibodies are quite good in this regard, being nearly entirely present in individuals with SLE. However, some are also present in the general population and may never lead to disease or are present in other chronic medical conditions. Others that are found almost exclusively in SLE are present in just a fraction of cases, rendering them less useful as a way of findings the bulk of SLE cases. Similarly, biopsies of skin and kidneys can be suggestive but not diagnostic on its own. We utilized these tens of thousands of administrative codes and laboratory/pathology tests to comprise a list of people potentially with SLE.

Defining Individuals with SLE

Classification criteria have been established for which variable approaches underlying clinical judgment cannot be applied uniformly across large numbers of individuals. Classification criteria provide a consistent way of defining individuals for epidemiologic and other studies. We utilized the 1997 Update of the 1982 Revised American College of Rheumatology (ACR) Criteria for Systemic Lupus Erythematosus, which are comprised of 11 criteria: (1) malar rash, (2) discoid rash, (3) photosensitivity, (4) oral ulcers, (5) arthritis, (6) serositis, (7) renal disorder, (8) neurologic disorder, (9) hematologic disorder, (10) immunologic disorder, and (11) antinuclear antibodies. In order to be classified as having SLE, the presence of four or more criteria is required at any time while all other reasonable diagnoses are excluded. Underlying several of the criterion are specific subdefinitions. For example, hematologic disorder is defined by hemolytic anemia, low white blood cells (leukopenia), low lymphocyte count (lymphopenia), and/or low platelets (thrombocytopenia). Again, these must be attributed to SLE and not from other causes.

With our list of potential SLE patients, we entered facilities where nearly all SLE patients would eventually seek care and be diagnosed (hospitals, rheumatologists, nephrologists, and dermatologists) in and around our catchment area and pulled all available medical records for these individuals. Highly trained medical abstractors reviewed them to identify elements of the ACR Criteria for SLE, which were found in varied forms and locations. For example, the clinical finding of a malar rash, an inflammatory skin lesion found on the cheeks of the face often triggered by sun exposure, can be located in the physical examination portion of the medical record. It can also be found in the narrative or past medical history as part of the reported historical features of SLE. Similarly, laboratory and pathology results can be documented in their original reports or referenced in the narrative or past medical history.

There are also different levels of certainty with respect to the source, which we also captured. The narrative in the medical record may state that a history of malar rash was reported by another rheumatologist who attributed it to SLE perhaps by a faxed letter or phone conversation. This may lend a higher degree of confidence for some, coming from a specialized physician, but tempered if lacking additional source documentation. Other documentation may include patient reported features, such as a suggestive history of a facial rash after sun exposure or even the use of the specific term "malar rash" itself, though without clarity as to how they were defining or conceptualizing it. Laboratory and pathology results can similarly be mentioned in the narrative or reported by the patient but not supported by documentation.

Real-World Experiences

On paper, our original charge appeared to be a standard epidemiologic exercise. But in reality, it highlighted the real-world challenges and nuances of capturing those diagnosed with this condition. The uncertainty in how one arrives at the diagnosis of SLE may seemingly be addressed through the use of a standardized definition in the ACR Criteria. Even so, we were left to struggle with what to do with a significant number of individuals who did not meet the strict definition of having four or more criteria. As a homage to those near the cut-off point, where we could have easily missed or not have had access to a single criterion, we also report in our publications a secondary case definition of those with three ACR criteria with a final diagnosis of SLE by their treating, board-certified rheumatologist. To date, this has passed muster by peer reviewers who tend to be strict in the application of definitions and criteria. However, as of this writing, we have only analyzed and reported features with physicians as the source, which is deemed the highest level of accuracy in scientific circles. So, what to make of the large numbers of features potentially consistent with ACR SLE Criteria reported by patients alone? In our database, there are significant numbers of individuals who do not meet three or at least four ACR criteria but have been consistently tagged with administrative codes associated with SLE from various sources throughout the healthcare system. Many also have been described as having a final diagnosis of SLE by their treating physicians, including rheumatologists, and have been treated as such by having been prescribed immunosuppressive medications. Do these individuals "count"?

How Many Have SLE? It Depends

The CDC funded other SLE epidemiology registries in different parts of the country using similar methodology in order to obtain incidence and prevalence estimates from different racial/ethnic groups. In addition to our registry in Atlanta, others were located in Detroit, New York City, San Francisco, and in the Indian Health Service. Estimates from these sites were then pooled together to come up with the total number of individuals in the USA with SLE. In 2020, we published our final

results in a manuscript titled "Prevalence of Systemic Lupus Erythematosus in the United States: Estimates from a Meta-Analysis of the Centers for Disease Control and Prevention National Lupus Registries" in the journal *Arthritis & Rheumatology* [2]. Our conclusion was that in 2018, 204,295 individuals in the USA fulfilled the ACR classification criteria for SLE. The part of the conclusion critical for interpreting the final number is that these were individuals fulfilling established criteria but given our methodology, a certain number of individuals might not have been counted when they should have been. However, rarely is this appreciated. Instead, the take-home message of most in the scientific and lay communities has been that the CDC determined that fewer than 205,000 individuals have SLE. This total number ran significantly counter to what others had thought. Given prescription and coding data, the pharmaceutical industry projected a potentially larger pool of eligible patients for therapeutics they were developing. Patient advocacy groups, such as the Lupus Foundation of America, have been touting figures as high as 1.5 million from survey data of individuals who self-reported the diagnosis.

Which numbers/projections are correct? The answer is: it depends. It depends on your perspective and what information you're trying to relay, which in turn informs the approach and definitions best used. The CDC-supported registries used a case definition anchored on the ACR Criteria for SLE. These criteria seek to identify clinical and laboratory features that are as accurately attributed to SLE as possible and not another cause, short of individual assessments by a study rheumatologist. As such, the cases identified can be more confidently associated with the biological and autoimmune processes of SLE, which may be helpful, for example, when determining if rates of SLE are changing over time due to certain environmental exposures. From a patient advocacy standpoint, as our observations have supported, there are clearly many more individuals than classified by the ACR Criteria that carry the diagnosis of SLE or believe that they have features consistent with SLE. For these individuals, their symptoms and the impact on their function and quality of life are no better or worse when considering how many criteria they may or may not have. With respect to the pharmaceutical industry, the opportunity to treat with an immunomodulatory agent appropriate for SLE is not limited to those defined by the ACR Criteria but exists across this spectrum. The healthcare system ultimately seeks to better understand the association of these administrative codes with healthcare utilization, costs, and outcomes regardless of the criteria. As in any classic Venn diagram, there may an area of overlap that is shared among all of them significant areas are the circles themselves and the larger shared areas they have with one or two other circles. Each space, shared or unshared, large or small, tells a different and equally important story. Yet, they all coexist in the SLE universe.

Future Directions

Having been directly responsible for the analytical decisions made from our registry and the conclusions drawn from them, I have been made culpable in front of different audiences, each with their own perspective and priorities. I have learned to avoid

the urge to defend our results. They are what they are, according to the approach and definitions we used. Rather, the challenge is to help people understand the issues addressed in this article so that they can best interpret the results for their purposes.

It is nearing a time when another epidemiologic study of SLE on a population level is due. It is difficult to interpret a prevalence rate in isolation and more informative knowing how rates and characteristics change over time. And with the advent of the wide use of the electronic medical record, big data, and artificial intelligence, we will undoubtedly be able to approach these questions with more efficiency and power than our approach that utilized mostly manual review of paper medical records. However, until our understanding of the immunobiology of SLE leads us to more accurate and commercially available testing, the issues raised in this article will remain considerable. Therefore, future studies must recognize the need for flexibility of definitions that was apparent to us early in the development of our registry and acknowledge that even the best laid plans will have its shortcomings. After the completion of our registry activities, we obtained permission to contact individuals that met ACR Criteria for SLE and offer them the opportunity to enroll in a study whereby they would be followed over time through surveys. Thankfully, the vast majority have agreed to participate. But to our surprise, there were several individuals who adamantly refused and even denied having had SLE, even though we documented many lupus codes, ACR criteria, laboratory results, and physician notes with SLE as a diagnosis. Again, we were reminded that perspective matters.

Back to John Snow

John Snow died without ever knowing exactly what caused cholera, which took the combined efforts of several people across generations. One hundred and thirty years before John Snow, Antonie van Leeuwenhoek was a pioneer in microscopy and rendered the first drawing of bacterium. Dr. Snow never learned that his contemporary, Filippo Pacini, a physician and anatomist, was the first to discover *Vibrio cholerae*, the bacterium that causes cholera. Louis Pasteur, known as the father of modern microbiology, only a year after John Snow had developed strong evidence against the miasma theory. And finally, it was not until 26 years after John Snow that Robert Koch, a bacteriologist, rediscovered, isolated, and first cultured *Vibrio cholerae*. This is a good reminder that our efforts and varied perspectives are not in vain. They are the patchwork that will one day lead us toward more clarity.

References

1. Lim SS, Bayakly AR, Helmick CG, Gordon C, Easley KA, Drenkard C. The incidence and prevalence of systemic lupus erythematosus, 2002–2004: The Georgia Lupus Registry. Arthritis Rheum (Hoboken, NJ). 2014;66(2):357–68.
2. Izmirly PM, Wan I, Sahl S, Buyon JP, Belmont MH, Salmon JE, et al. The incidence and prevalence of systemic lupus erythematosus in New York County (Manhattan), New York: the Manhattan Lupus Surveillance Program. Arthritis Rheum. 2017;69(10):2006–17.

Part IV

Diagnoses

Managing and Tolerating Diagnostic Uncertainty

15

Paul K. J. Han

Introduction

My point of view is that of a behavioral researcher, but one who has also wrestled with diagnostic uncertainty as a clinician, patient, and family member. My primary interest lies in the human experience of uncertainty, and the strategies clinicians and patients use to cope with it. From my perspective, diagnostic uncertainty is one particular instance of a more general, complex experience that pervades all of health care, has diverse psychological effects, and is managed in numerous ways. The problem I will address in this chapter, however, is that the medical management of diagnostic uncertainty by clinicians has historically been confined to the pursuit of diagnostic knowledge and certainty. Although this pursuit is a necessary part of the management of diagnostic uncertainty, it is not sufficient to meet all of our needs as clinicians and patients. When the illness has no name, we need not only to attain greater knowledge but to paradoxically maintain uncertainty and enhance our capacity to tolerate it. The management of diagnostic uncertainty is thus a much broader endeavor than normally acknowledged, and in this chapter, I will attempt to outline what this endeavor entails.

Conceptualizing Diagnostic Uncertainty

The first task in understanding the management of diagnostic uncertainty is to establish a coherent working definition of the problem being managed. Diagnostic uncertainty has two conceptual elements: diagnosis and uncertainty. Diagnosis, a paramount concern of medicine as an applied scientific endeavor, is an act of

P. K. J. Han (✉)
Division of Cancer Control and Population Sciences, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
e-mail: paul.han@nih.gov

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_15

classification involving the assignment of some pre-defined, socially constructed conceptual category to a given health problem [1]. This classificatory act serves several instrumental functions. It motivates, enables, and justifies clinical action, satisfies fundamental psychological needs for order and certainty, reinforces particular cultural values and beliefs, and legitimates various social roles, rights, privileges, and arrangements. The correctness or accuracy of any given diagnosis depends on the degree to which it effectively serves these clinical, psychological, cultural, and social functions.

Uncertainty, the other key element of diagnostic uncertainty, is a mental state consisting of the conscious, metacognitive awareness of ignorance about some aspect of the natural world [2]. As such, uncertainty is not synonymous with mere ignorance, or lack of knowledge. Rather, uncertainty is itself a distinct form of knowledge—specifically, a reflective, higher-order knowledge of one's lack of knowledge, without which one would simply be unconsciously ignorant (unaware that one does not know).

In medicine, uncertainty arises from three principal *sources*—probability, ambiguity, and complexity—which constitute symbolic or informational manifestations of deeper root causes of ignorance. Probability manifests the indeterminacy of health outcomes due to inherent randomness in all natural events; ambiguity manifests the indeterminability of health outcomes due to missing or incomplete evidence; and complexity manifests the intractability of health outcomes due to the multiplicity, heterogeneity, or conditional nature of their causes, attributes, or manifestations [2]. Uncertainty in medicine also pertains to various *issues*: (1) scientific (encompassing the diagnosis, prognosis, causes, and treatment of a given health problem), (2) practical (encompassing the structures and processes of health care), and (3) personal (encompassing psychosocial, moral, and existential concerns). Uncertainty in medicine also has different *loci*; it is variably located in the minds of different people—clinicians, patients, family members, and other stakeholders. While some people may be uncertain (consciously aware of their ignorance about some important health-related issue), others are simply ignorant.

Integrating the meaning of its two constituent elements, diagnosis and uncertainty, yields a coherent working definition of diagnostic uncertainty as the *conscious, metacognitive awareness of ignorance about the appropriate conceptual category for a given health problem*. In contrast to other types of medical uncertainty, diagnostic uncertainty pertains specifically to the substantive issue of what existing biomedical category to assign to the problem at hand or how to assign it. Yet diagnostic uncertainty ultimately originates from the same sources as all other medical uncertainties—probability, ambiguity, and complexity—and is variably located in the minds of clinicians, patients, and other persons.

Managing Diagnostic Uncertainty

Like the medical problems it pertains to, diagnostic uncertainty itself needs to be managed, and the potential approaches are curative and palliative in nature and intent. Fig. 15.1 outlines the range of specific strategies clinicians use to manage

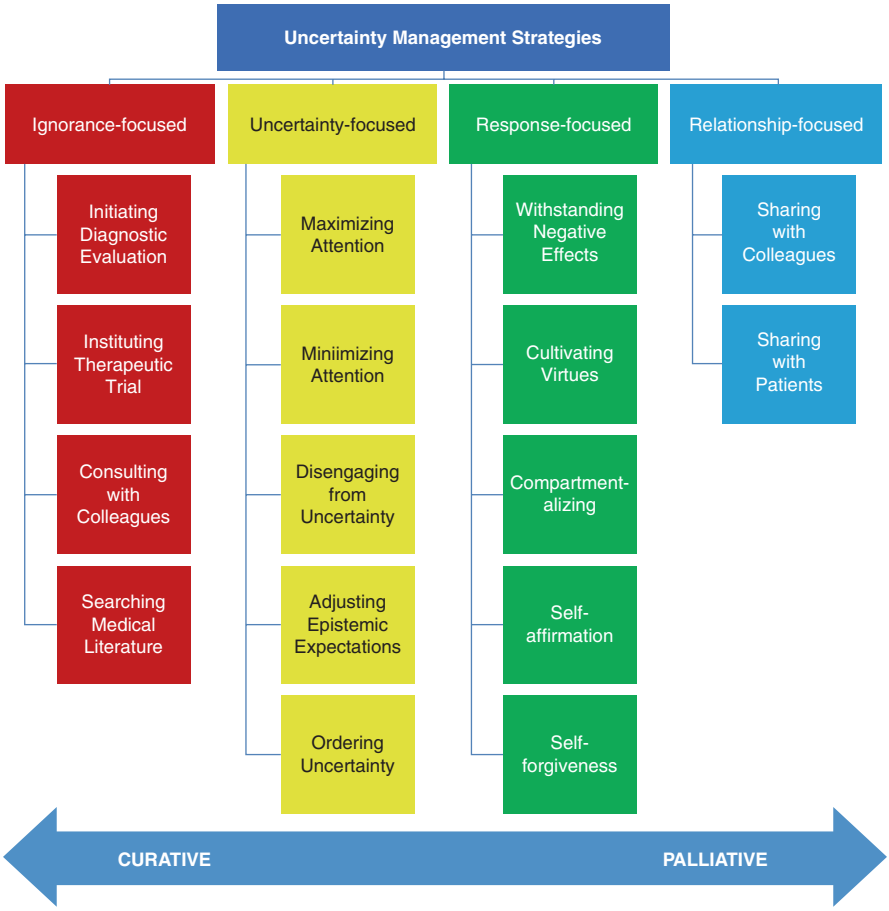

Fig. 15.1 Diagnostic uncertainty management strategies: conceptual taxonomy

diagnostic as well as other types of uncertainty in medicine. I will now briefly discuss these strategies and make the case that the effective management of diagnostic uncertainty requires greater attention to palliative strategies.

Ignorance- and Uncertainty-Focused Strategies

If diagnostic uncertainty represents the conscious awareness of ignorance about the appropriate conceptual category for a given health problem, then it follows that diagnostic uncertainty can be managed by addressing either (1) diagnostic ignorance or (2) the conscious awareness of this ignorance [3]. Historically, the medical management of diagnostic uncertainty has focused almost exclusively on these two strategies, and reducing diagnostic ignorance has been the primary approach. The signature activities of clinical care—taking a medical history, performing a physical examination, obtaining laboratory, and imaging studies—are all efforts to cure diagnostic uncertainty by eliminating key knowledge gaps about the appropriate

conceptual category for the health problem at hand. Other ignorance-focused strategies for managing diagnostic uncertainty also include initiating therapeutic trials, seeking professional consultation, and searching the medical literature. These various activities ultimately reduce diagnostic uncertainty by decreasing clinicians' and patients' level of ignorance.

Uncertainty-focused strategies for managing diagnostic uncertainty are also curative in intent; however, they aim to reduce not the ignorance that is the object of uncertainty, but the conscious, metacognitive awareness that is also necessary for uncertainty to exist. One dominant strategy consists of *minimizing attention* to diagnostic ignorance by ignoring what is unknown or restricting the scope of one's attention to some diagnostic possibilities and not others (e.g., "rheumatologic" *vs*. "oncologic" diagnoses, or physical *vs*. psychological manifestations of illness). The converse strategy, *maximizing attention*, consists of maintaining a high index of suspicion for particular diagnoses; a primary example is the common approach of treating particular symptoms or signs (e.g., chest pain, breast mass) as diagnostic of serious conditions (e.g., myocardial infarction, breast cancer) "until proven otherwise." Adopting such an epistemic posture paradoxically reduces uncertainty by focusing the diagnostic evaluation and restricting alternative possibilities. *Disengaging from uncertainty* is another strategy that entails emotionally distancing oneself from one's uncertainty in various ways such as transferring responsibility of managing uncertainty to others. Another common strategy for managing diagnostic uncertainty, *adjusting epistemic expectations,* involves altering one's expectations and demands for knowledge and certainty; examples include accepting lower levels of specificity for a given diagnostic test, in exchange for higher levels of sensitivity (or vice versa). *Ordering uncertainty* is a final uncertainty-focused strategy that entails imposing some logical structure or process that makes it more manageable; examples include instituting diagnostic "pathways" or algorithms that provide clinicians and patients with a tangible action plan. The common feature of all of these activities is that they ultimately reduce diagnostic uncertainty simply by decreasing clinicians' and patients' *awareness* of their ignorance, as opposed to their level of ignorance per se.

The Necessity of Diagnostic Uncertainty

Ignorance- and uncertainty-focused strategies alone, however, are ultimately insufficient for managing diagnostic uncertainty. A principal reason is that although diagnostic ignorance may be reducible, it is not completely curable. The question of what specific conceptual category is most appropriate for a given illness in an individual patient is fundamentally indeterminate due to randomness and variability in disease processes, fundamentally indeterminable due to the finite nature of empirical data and the shortcomings of all diagnostic measures and tools, and fundamentally intractable due to the multifarious causes and protean manifestations of disease in individuals. Diagnostic ignorance is thus both logically and practically ineliminable: no matter how clear a given diagnosis might appear, some amount of diagnostic uncertainty will always remain and require further management.

Yet diagnostic uncertainty is not only logically and practically ineliminable, but also medically and morally necessary. It enables clinicians to avoid premature closure in establishing a diagnosis, and to remain open to alternative diagnostic possibilities. Diagnostic uncertainty can thus ensure the well-being of patients; the metacognitive capacity to entertain the possibility that a given diagnosis is inaccurate can prevent clinicians from initiating or maintaining medical interventions that are non-beneficial or even harmful. Diagnostic uncertainty also enables clinicians to avoid reifying existing medical categories that are merely conceptual, and to see each patient as an individual suffering person rather than a mere instantiation of some universal pathophysiologic entity. Diagnostic uncertainty can thus promote patient-centered health care; the metacognitive capacity to view diagnoses as imperfect heuristic abstractions rather than "real things" can prevent clinicians from stigmatizing patients with specific diagnoses or marginalizing persons with medically unexplained symptoms [4]. It can help clinicians respect the epistemic authority of patients whose illnesses do not fit into existing conceptual boxes, and to be more receptive to their perspectives and needs [5].

The necessity of diagnostic uncertainty means that ignorance- or uncertainty-focused management strategies alone will never be adequate. Diagnostic uncertainty is incurable and also serves essential, adaptive functions; efforts to manage diagnostic uncertainty thus need to focus not simply on eliminating but on maintaining and accepting it. The problem, however, is that uncertainty has psychologically aversive effects. A large body of social science research has documented how uncertainty arising from various sources—probability, ambiguity, and complexity—provokes negative cognitive, emotional, and behavioral responses including perceptions of vulnerability, fear and anxiety, and avoidance of decision making. These psychological responses comprise distinct syndromes that have been assigned their own diagnostic labels: risk aversion, ambiguity aversion, and complexity aversion [6]. The existence of these aversive responses complicates any effort to accept and maintain diagnostic uncertainty in medicine. It raises the need to palliate diagnostic uncertainty—that is, to somehow ameliorate its negative psychological effects.

Response- and Relationship-Focused Strategies

Historically, however, the palliation of diagnostic uncertainty has not been a primary focus of clinical care, and approaches to the task have not been formalized in medical practice or education. Consequently, the various strategies that clinicians and patients use to palliate diagnostic and other important uncertainties are mostly enacted in an informal and non-deliberative manner. Nevertheless, these palliative strategies can be explicitly categorized according to their main targets: (1) psychological responses to uncertainty and (2) interpersonal relationships [3].

Response-focused uncertainty management strategies include *withstanding negative effects* of uncertainty, including vulnerability, fear, and indecision. Conceptually simple but practically difficult, this strategy entails stoically enduring the suffering that uncertainty produces. Arguably all clinicians and patients enact this strategy to some extent; however, differences in individuals' tolerance of uncertainty make

some people less capable of enduring it, and more susceptible to "burnout" and other negative psychological sequelae [7]. Another response-focused uncertainty management strategy, *cultivating virtues,* involves upholding moral ideals (e.g., industriousness, thoroughness) that represent meaningful goals even if perfect knowledge is unattainable. A common example is adherence to the goal of "leaving no stone unturned," a strategy that enables many clinicians and patients to focus on the process—rather than the outcome—of pursuing diagnostic knowledge. *Compartmentalizing psychological responses* involves uncoupling and sequestering various cognitive, emotional, or behavioral responses to uncertainty in order to limit their deleterious effects; the "detached concern" often practiced by clinicians is a prime example of this strategy. *Self-affirmation*, another important response-focused strategy, entails acknowledging one's core values and strengths; examples for clinicians include altruism or commitment to the welfare of others, as well as intellectual rigor. Affirming such values promotes a sense of self-integrity that can protect people against negative consequences of stressors such as uncertainty. *Self-forgiveness* is a final response-focused strategy that involves absolving oneself from guilt and blame and caring less about the negative evaluation of others. For clinicians, this strategy is particularly important yet challenging to enact due to the certainty-focused professional norms of medicine and the ever-present threat of malpractice litigation.

Relationship-focused strategies for managing diagnostic uncertainty resemble response-focused strategies in their palliative nature and intent, but differ in targeting interpersonal relationships. *Sharing with colleagues* and *sharing with patients* are important relationship-focused strategies that palliate negative effects of uncertainty in at least two ways. First, they promote the mutual exchange of experiences and concerns among individuals, which fosters feelings of camaraderie, emotional support, and trust and prevents feelings of social and existential isolation. Diagnostic uncertainty, like most significant threats in life, is more tolerable when it is faced together rather than alone. Second, sharing uncertainty establishes joint responsibility—professional, legal, and moral—for its management, which lessens the cognitive, emotional, and decisional burdens of the task for any given individual.

Managing Uncertainty: The Meaning and Necessity of Tolerance

I have briefly outlined some of the strategies clinicians and patients use to manage diagnostic uncertainty in medicine. I have argued that historically, medicine has been preoccupied with ignorance- and uncertainty-focused efforts aimed at the curative goal of increasing knowledge and certainty about the appropriate conceptual category for a patient's illness. Consequently, medicine has devoted very little formal attention to response- and relationship-focused strategies aimed at the palliative goal of ameliorating the negative psychological effects of uncertainty. Yet this imbalanced focus, I have argued, ignores both the incurable nature of uncertainty and its psychological and moral necessity for clinicians and patients. There is much

more to managing diagnostic uncertainty than acquiring the knowledge required to make a diagnosis—as unquestionably important as this task is. While clinicians and patients wait for diagnostic evaluations to be completed, they must cope—if only temporarily—with the aversive awareness of their own ignorance. Even after a diagnosis is made, clinicians must have the presence of mind to realize that all diagnoses are fallible constructs that, in spite of their usefulness, can monopolize medical attention and prevent patients from being understood and treated as persons. And when no definitive diagnosis is forthcoming—when the illness has no name—clinicians must help patients find some way of living without a biomedical category, outside of conventional conceptual boundaries.

For all of these reasons, medicine's efforts to cure diagnostic uncertainty need to be supplemented with equal efforts to maintain and palliate it. When the illness has no name, the task of managing diagnostic uncertainty necessarily becomes less informational and more emotional and relational in nature; accordingly, clinicians need to devote greater attention to response- and relationship-focused management strategies. Importantly, such heightened attention does not preclude the continued pursuit of diagnostic knowledge; as with all other medical problems, palliative and curative interventions for diagnostic uncertainty are not mutually exclusive. Palliative interventions should supplement—not supplant—clinicians' efforts to cure as much diagnostic ignorance as they can.

The ultimate goal of all these efforts is to increase the uncertainty tolerance of both patients and clinicians. Here I am not referring to "uncertainty tolerance" in the merely descriptive sense I have defined elsewhere as "the set of negative and positive psychological responses—cognitive, emotional, and behavioral—provoked by the conscious awareness of ignorance about particular aspects of the world" [7]. Rather, I am referring to uncertainty tolerance in a normative sense: as a moral goal for health care. Uncertainty tolerance in this normative, moral sense consists of an *adaptive balance* in people's varied responses to uncertainty [6]. Exactly what specific psychological responses make up this balance has no single answer: it depends on the individual and situation, and the task of every clinician and patient is to work together to find it. This effort, furthermore, ultimately requires clinicians and patients to cultivate specific moral virtues and psychological capacities: the humility to acknowledge the fallibility of one's knowledge, the flexibility to adjust one's diverse responses to uncertainty, and the courage to move forward in spite of one's ignorance [6].

The critical unanswered question is exactly how to cultivate these virtues and ultimately increase the capacity of clinicians and patients to tolerate diagnostic and other important uncertainties in medicine. I have explored some potential answers to this question, but much more conceptual and empirical research is needed. I have simply tried to make the point that the management of diagnostic uncertainty is a task that entails more than pursuing knowledge and acquiring information. It is a broader, more complex endeavor that entails accepting uncertainty, minimizing its potential harms, and maximizing its potential benefits. Effectively managing diagnostic uncertainty in this larger sense thus requires expanding the repertoire of strategies used by clinicians and patients. More fundamentally, however, I believe it also requires shifting medicine toward a new paradigm that treats medical uncertainty

not as a mere deficit state equivalent to ignorance, but as a form of knowledge to be preserved and palliated. It remains for further research to determine how to achieve this broader paradigm shift.

References

1. Jutel A. Sociology of diagnosis: a preliminary review. Sociol Health Illn. 2009;31(2):278–99.
2. Han PKJ, et al. Varieties of uncertainty in health care: a conceptual taxonomy. Med Decis Mak. 2011;31(6):828–38.
3. Han PKJ, et al. How physicians manage medical uncertainty: a qualitative study and conceptual taxonomy. Med Decis Mak. 2021;41(3):275–91.
4. Atkins CGK. My imaginary illness: a journey into uncertainty and prejudice in medical diagnosis. Ithaca: Cornell University Press; 2010.
5. O'Leary D. Why bioethics should be concerned with medically unexplained symptoms. Am J Bioeth. 2018;18(5):6–15.
6. Han PKJ. Uncertainty in medicine: a framework for tolerance. New York: Oxford University Press; 2021.
7. Hillen MA, et al. Tolerance of uncertainty: conceptual analysis, integrative model, and implications for healthcare. Soc Sci Med. 2017;180:62–75.

Is There a Textbook for Non-textbook Patients?

16

Jillian Rose

Is There a Textbook for Non-textbook Patients?

To understand the lived experiences of patients with illnesses with no name, we wanted to hear from the patients directly. Several patients, who expressed concern about living with the illness, were suggested by their medical care team. For this workshop presentation, we invited 30 patients of diverse backgrounds to participate in 30–60-minute semi-structured interviews to explore the patients' perspective about living with the illness and about its impact on their lives. Fifteen patients participated. The interviews addressed what patients wanted clinicians to consider and how can they share advice for other patients. The interviews were transcribed, and a thematic analysis was conducted by two separate individuals, who then finalized the key themes together.

Participants' mean age was 37 years old. Approximately half identified as White/Caucasian and half as minority women (African American, Asian American, Latina, and others).

Several key themes that emerged included delays in diagnoses; not being believed by doctors and not being believed by family members; women who are not taken seriously; invisible symptoms; power struggles with doctors; treatments that don't work; and mental health. Patients asked their physicians to listen/show empathy discussing and addressing mental health. Patients' advice to other patients is to find the right doctor, listen to your body, and find the right support team.

Overall, the interviews make clear that diagnostic uncertainty strongly impacts the physical and emotional health of patients; it impedes patients' ability to plan for

The author is grateful for the enthusiastic, patient, and collaborative assistance of BellaRose Virginia Smith.

J. Rose (✉)
Hospital for Special Surgery, New York, NY, USA
e-mail: rosej@hss.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_16

their immediate future. The patient perspective also made clear that listening and empathy shown by clinicians are needed to provide a sense of normalcy. When present, it is lifesaving.

The following sections list the themes identified by the authors, introduced by a patient's verbatim response.

Theme: Delays in Diagnosis

> I have only even seen chronic illness portrayed as a huge loss.
>
> This tragic loss. This awful thing that kind of erases a future.

Most participants described delays in being given a diagnosis—8 to 10 years before from symptoms to diagnosis. Several had been misdiagnosed many times. Participants who indicated delays in diagnoses reported that clinicians did not listen to them and used strict categories and criteria to diagnose them. Two participants reported being diagnosed within a few months of having symptoms. Both attributed this to their doctors' being good listeners. They were curious, asked questions, and were good investigators.

Theme: Not Believed by Doctors

Almost all participants shared that they experienced not being believed by a doctor. Participants especially highlighted that because they did not fit into a diagnostic criterion they were dismissed and not taken seriously and often told it's all in your head because what they were describing did not show up in their lab results.

Theme: Not Believed by Family Members/Family in Denial

Some participants shared that they had very supportive families; others shared that their family members were in denial and wanted to ignore their symptoms due to fear and the uncertainty of the illness. Most participants shared due to a lack of a definitive diagnosis they were often not taken seriously by family members.

Theme: Women Symptoms Are Not Taken Seriously

Concerns around not being believed due to one's gender, culture, or age were identified as an issue for over half of participants. Some participants reported experiencing bias and discrimination during their medical encounter and reported concerns around their symptoms being dismissed or being referred to as hormonal. Most participants who shared these concerns indicated that these experiences were not uncommon.

Theme: Symptoms

All participants reported pain and fatigue. Participants described in detail how it felt to live with pain throughout their bodies constantly, as well as how debilitating fatigue was. The symptoms made planning their lives challenging due to the unpredictable nature of the illness. Many patients described how betrayed they felt by their bodies and the many negative ways that their symptoms impacted their quality of life.

Theme: Power Struggle with Doctors

Several participants expressed concerns about a power struggle they felt with their doctors, where the doctors had all the decision-making power. However, a quarter of participants shared that they had experienced shared decision-making throughout their care experiences.

Theme: Treatments Don't Work

Participants expressed concerns about many of the treatments they were prescribed not working. However, many reported following through with those treatment because they felt guilty reporting that they were not experiencing any benefit from the medications. Some shared how hard their doctors worked to try to help them feel better; however, the medications just don't work so I just don't say how I feel; I don't want my doctor to be discouraged.

Theme: Mental Health

All participants reported struggling with depression, anxiety, and/or isolation. They identified the emotional aspect of the illness as the most challenging aspect they had to cope with. However, almost all participants reported not being asked by their doctor about this mental health and how they were coping with the illness. Two participants reported that their doctors were very hands-on with assessing their emotional health and recommending therapy and other holistic practices to better cope with pain.

Theme: Advice to Doctors

- *Listen/show empathy to patients.* See them as real people and partners in care.
- *Discuss and address mental health concerns.* All participants shared the need for doctors to ask about emotional health and to normalize discussing and addressing it. Participants frequently reported that they don't usually bring up their mental health because they think that it's not something that doctors care about, although it is often what they are most challenged by.
- *Be open to alternative treatments, particularly when what you're doing doesn't work and/or the risk/benefit ratio is very low.* Participants also encouraged doctors to explore other methods of treatment that may be outside the box.

Theme: Advice to Patients

- *Find the right doctor.* Participants frequently shared that patients should find the right doctor for them.
- *Be open and honest.* They advised other patients to be open with their doctors and to cultivate good communication. Participants shared that they were their doctor's most important source in figuring out what's happening with their bodies, so it's important to speak up.
- *Listen to your body.* All participants advise other patients to listen to their bodies, especially when they need to rest. Participants frequently expressed concerns around balancing work, family, appointments, and other life demands. However, they wanted to share with other patients that listening to what your body needs, how your body is feeling, and resting appropriately were all important in managing the illness and having a good quality of life.
- *Find the right support team.* Some participants discussed their resilience in coping with the illness for several years, the credited support groups and other informal peer support, and family support networks with their ability to manage and cope with their illness.

The main points are summarized above. Table 16.1 provides more specific details and quotes, some of which are emotional and explicit, others of which introduce topics physicians rarely consider.

Table 16.1 The table identifies the types of questions asked of participants, themes that emerged through structured analysis of the interviews, and a few examples of verbatim patient quotations on the impact of living with the illness with no name

Question	Theme	Patient quotation
How long did it take to receive a diagnosis after your first symptom?	8–10 years	One doctor just looked at my record and didn't want to even do bloodwork; he said I just didn't fit into the diagnostic category Doctors need better criteria for diagnosing; it all just seems so rigid and out of date I had to see over six doctors in 1 year for someone to take me seriously
What were barriers to care?	Doctors not listening, strict diagnostic categories	Maybe I'm not describing what I 'm feeling correctly, but I know what I'm feeling. I know it
How can mediators help?	Listen, be curious, be a good investigator	Even though doctors do not know the actual diagnosis, we're okay with it as long as we know you are trying to figure it out with us

Table 16.1 (continued)

Question	Theme	Patient quotation
Did you feel like you were believed by your doctors? By your family?	Not believed by doctors; symptoms are not real/it's in your head; not taken seriously; it's in your head; family in denial	Is there a textbook for non-textbook patients? Everything looks okay on paper, except I feel like I'm dying a little bit everyday inside It wasn't like I was faking it; I would see my joints swell up, my feet swell up, and my hands swell up. I'm not making this up. This is happening I couldn't give them a name so they thought I was making it up I was told you are too young to be on disability. My coworkers think I am just taking advantage of the system Advice to patients: Diagnosing this illness is hard and information is limited, so doctors have to listen and care about what patients are saying People know cancer and MS but don't know what it means to be "lupus-like" or "maybe RA" or "fibro" so they just think it's in your head because they don't understand it; I don't understand it most days
Do you think your gender or culture had anything to do with you being believed?	Yes, women's symptoms are not taken seriously; we are viewed as uneducated; young patients are not believed	Unfortunately, women of color deal with discrimination and patronizing doctors, even as a PhD I faced racist and religious microaggressions from doctors; my gender identity later became an issue as well Doctors are conditioned to not take women's feelings seriously, which can impact whether or not they are believed and receive care I felt dismissed as crazy/hormonal because of gender I experienced bias from healthcare workers because of my gender and background Because I don't use the specific medical terminology to describe my symptoms, my illness is still real and I need help

(continued)

Table 16.1 (continued)

Question	Theme	Patient quotation
What is the most challenging part of having this illness?	Symptoms, pain, fatigue, power struggle with doctors, treatments don't work, mental health challenges, depression, anxiety, isolation	The pain feels like some kind of poison in my body It's hard for people to believe that I'm in excruciating pain 24/7 and every single decision I make has to take my illness into consideration There is such a "I'm the doctor, you have to listen to me" attitude, even when they're wrong My old rheumatologist would put me down for not responding to the medication My doctors are in denial about the medications working, so I just use my traditional medicines to treat my condition You don't want to be a burden so you don't tell the truth about the medications not working Doctors work hard to put the treatment plan together, and they have so much hope, but I feel guilty to say I am still fatigued and I am still in pain. It just doesn't work so I just don't say how I feel; I don't want my doctor to be discouraged I had to ask my doctor, "can you please tell me I'm not crazy? Tell me that I'm not alone?" If only the doctors would have asked, I would have gladly shared. But I was too embarrassed to say I was depressed I was dying from depression, not my symptoms, but my doctor never asked me how I was coping mentally. I could cope with the physical symptoms, but I lost so much of me mentally I was depressed, but I had to go on my own to look for help Doctors don't understand the physical, mental, and emotional pain patients endure, because of their illnesses and its consequences You feel lonely. You feel alone even when if you are surrounded

Table 16.1 (continued)

Question	Theme	Patient quotation
What would you like doctors to consider when treating patients with your illness?	Listen to patients. Show empathy, explore other methods of therapy, discuss and address mental health concerns, use plain language to explain research and treatment	I had to leave that rheumatologist; I wasn't getting better; I think if she listened to what I had to say, she would have made me feel better I wanted to tell her what is going on inside of me, but she did not give me the opportunity to tell her; she said I can tell her next time, she had a plan See us as a person. Don't see us as just like a "patient." see us like a person coming to you because we are feeling so bad that we just can't live this way any longer If you put me on opioids, don't treat me like a drug addict when I come to you; treat me with empathy; I am in pain My doctor encouraged me to try other therapies; acupuncture and meditation especially helped with the pain Doctors find other therapies unconventional, but must understand that if it make a patient feel better, it should be looked into I don't want a quick solution or a magic pill; help me to better cope with this I want to participate in clinical trials, but it's never explained well Most of the time, I had no clue what the doctor was saying; after 100 tests, all I understood was your ANA was negative so I still couldn't get a diagnoses
What advice do you have for patients?	Be open with your doctor, share your symptoms, seek help for mental health concerns, listen to your body, find the right doctor, find a good support system, rest	I often feel like I am complaining, but it's important for us to share our symptoms I had to change doctors several times, until I found a doctor who was open and listened to what was important to me Talking with a therapist can help with coping We need to stop pretending that everything is fine, because we don't want our doctors to feel like they failed us Just learning to listen to my body made a big difference to me Having a support system was key in not feeling alone; I think doctors should recommend more support groups If I don't rest, my body takes it from me and then I go into a flare

(continued)

Table 16.1 (continued)

Question	Theme	Patient quotation
Is there anything else you would like to share?	Insurance barriers, explaining medications better, diet and nutrition are key, unable to explain illness, reluctant to take medication, struggled with accepting illness, want better ways to communicate symptoms, need for more information	No real diagnosis in 20 years, but the symptoms are consistent Doctors should create multiple ways to test patients to provide more concrete answers for patients like me Doctors could create a symptom chart to follow when assessing patients Being diagnosed with a chronic illness is difficult, not just because of the physical symptoms but also the emotional and betrayal of your own body

The Changing Role of Uncertainty in Physician-Patient Relationships

17

Andrew Schafer

A number of years ago, when I was chief of medicine at a prominent medical school, I used one of my weekly meetings with our internal medicine residents in training to give them a surprise multiple choice quiz. At the end, I asked them to put their names on their answer sheets before submitting them. Over the years, I had become increasingly disturbed by repeatedly finding specific and definitive but poorly documented and misleading diagnoses of medical problems in the medical records of many patients. For example, if a patient had chronic kidney disease and also happened to have mild diabetes, the diagnosis in the chart might be "diabetic nephropathy" – even though there was not a shred of objective evidence that diabetes caused the kidney damage. The only thing accomplished by stating this specific diagnosis in the records was to provide physicians who would later care for that patient with false reassurance that a firm diagnosis had been already made and therefore there was no further need to be concerned about it. I would have written something like "chronic kidney disease of unknown etiology, possibly (or probably) diabetic nephropathy." So, my quiz to the residents consisted of about a dozen questions based on very brief case histories. An example was a previously healthy middle-aged man who was found to be anemic. For diagnosis, the options were (not in this order) (A) nutritional anemia; (B) erythroid hypoplasia; (C) anemia due to occult gastrointestinal bleeding; (D) anemia of unknown etiology; or (E) anemia of chronic disease. Although more about it is known today, at that time "anemia of chronic disease" was a waste can term for multiple miscellaneous disorders associated with anemia. In this case, there was not even a major chronic disease in the patient to justify the diagnosis of (E).

None of the 15 or so residents chose answer (D), but many did choose answer (E). All the other cases and questions were constructed similarly, including an

A. Schafer (✉)
Department of Medicine, Weill Cornell Medical College, New York, NY, USA

New York-Presbyterian Hospital/Weill Cornell, New York, NY, USA
e-mail: ais2007@med.cornell.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_17

option to respond to each with "_____ of unknown etiology" or some variant thereof. Not a single resident chose that option for any question, even though that would have been the truthful answer for each one. Why? Were the residents overconfident about their knowledge base? Was it lack of nosological humility [1], i.e., knowing what one doesn't know? Did they think that I would be critical of them if they did not have a specific answer? Or were they all already enculturated by their teachers to be uncertainty-aversive in medicine, assuming that uncertainty might be a stigma of weakness or ignorance?

Concepts

Within the definition of *uncertainty* in medicine, I shall include the term *ambiguity*. The latter more specifically refers to the physician confronting decisions that require a difficult choice of available options [2]. In her essay on "Training for Uncertainty," the renowned medical sociologist, Renée Fox, distinguished between two different forms of uncertainty confronted by physicians and medical trainees. The first derives from the doctor's limited mastery of available knowledge. The second is the result of limitations in current medical knowledge [3]. It is sometimes difficult to distinguish between personal ignorance or ineptitude and the limitations of present-day knowledge. One of my colorful but haughty professors in medical school was fond of responding to difficult questions with: "I don't know. And when I say that, I mean that the answer to that question is *not known*."

Uncertainty and ambiguity have permeated medical practice since antiquity [4]. Today, with the advent of specialization and subspecialization, they may be more pervasive in some fields than others. At one extreme might be primary care physicians, who must deal with not only ambiguous new symptoms related by a patient but also the need to integrate or distinguish those symptoms from each other and from many other illnesses the patient might already have, and psychiatrists, who still only have an extremely limited repertoire of objective tests with which to make an accurate diagnosis. On the other side might be, for example, oncologists whose patients initially present to them with a biopsy-proven cancer diagnosis already made. But by no means does this exempt oncologists from uncertainty as they do further molecular testing that yields results of unknown significance, face ambiguity about treatment choice in many cases, and have to deal with the uncertain nature of treatment toxicity and complications such patients may have. Although awareness of the physician's own uncertainties and limitations is imperative to being a good doctor, it is true that taking it to the extreme is maladaptive and can paralyze decision-making and action.

The Evolution of Physician-Patient Relationship

How physicians have recognized and managed medical uncertainty has evolved throughout history as a function of the medical knowledge available, the level of understanding of both physicians and patients regarding the nature of human

disease, and cultural and societal norms and mores at the time and place. In the mid-twentieth century, Thomas Szasz, a distinguished but controversial psychiatrist, classified the doctor-patient relationship into three models. The activity-passivity model described the physician doing something to a patient who is unable to respond or assumes a fully passive role. In the guidance-cooperation model, the doctor tells the patient what to do and the patient cooperates and obeys. In the mutual participation model, the doctor helps the patient to help himself, and the patient becomes a participant in the partnership [4]. The first two models are paternalistic relationships. The term paternalism is not necessarily a pejorative one. Until contemporary times, physicians fully rooted in the Hippocratic tradition considered paternalism to be only a hardline form of beneficence, even if it was at the expense of the patient's autonomy [5]. But most physicians today have learned that the patient's own values are important determinants of what is best for him. This notion was codified in 1960 when the first code of patients' rights was written.

The Role of Uncertainty in the Physician-Patient Relationship

It is my opinion that the type of a physician-patient relationship that is established in each individual case is determined to a large extent by the physician's comfort and willingness to share with his patient the uncertainties that influence her thinking and recommendations. What is the risk? The patient may indeed leave the physician's practice for care elsewhere when he mistakes the doctor's expression of uncertainty for hesitation, indecisiveness, or lack of sound knowledge and good judgment. In many situations, the physician might be fully aware of her uncertainties but just does not disclose them when she is speaking with the patient. In other, more pernicious cases where patients are not made aware of uncertainties, the physician may have become so encrusted in her own sense of uncompromising certainty and infallibility that she has become resistant to new knowledge that might contradict her entrenched beliefs acquired through personal experience and perceived wisdom. Regarding this latter group, a recent study found that physicians with high but fragile self-confidence (as measured by a narcissistic personality inventory standard scale} respond to ego threats by expressing an even greater sense of self-perceived invulnerability [6].

The traditional authoritarian relationship that has governed interactions between physicians and their patients until not so long ago has had its roots in the shamans of antiquity, dating back to 4000 BC, who still practice shamanism today in some parts of the world, projecting an image of omnipotence by possessing privileged links that enable them to commune with the spirits and gods. Authoritarian physicians have donned the mask of infallibility as a means of gaining and maintaining professional control of the patient [7]. Further, a physician's ability to preserve her own power over the patient in the doctor-patient relationship has depended largely on her ability to control the patient's uncertainty [8]. Perhaps the peak of physician authoritarianism was reached in the late nineteenth and early twentieth centuries in response to the revolution of new knowledge that was developing at that time about human disease, like the germ theory, the cell theory, and advances in physiology of

the circulation and heart, and most importantly the advent of rational treatment options. This led to the emergence of the "expert physician," devoid of humility and endowed with hubris [9].

So, have physicians been authoritarian by nature? Probably not. But Ludmerer Ludmir et al. conducted an interesting analysis of the autocratic tendencies of rulers of countries who were also physicians, using an established scale ranging from −10 to +10, in which the former represented fully autocratic governments and the latter fully democratic ones. Of 1254 rulers (prime ministers, presidents, etc.), 32 were also physicians, like Bashar al-Assad, the ophthalmologist-dictator of Syria. Physician rulers had significantly lower mean scores than non-physicians [10].

Despite the tradition of patients grumbling about doctors, and satires depicting medicine as quackery, extortion, and parasitism, doctors have acquired almost divine status throughout the centuries [11]. Doctors have been seen as Godlike or at least more-than-human, much as the shamans have been [12]. It is written in Ecclesiasticus to "Honour a physician with the honour due unto him for the uses which ye may have of him: for the Lord hath created him. For the most High cometh healing, and he shall receive honour of the king [13]."

Nonetheless, "authority" in medicine does not have to always have a malefic connotation. In fact, in 1957, the Scottish polymath scholar, T.T. Paterson, defined medical authority and called it "Aesculapian authority," describing it as benevolent, valuable, and magnanimous. Three components of Aesculapian authority were defined. First is sapiential authority, which is derived from the expertise and wisdom of the physician from dedicated study and experience, not from any position the physician might occupy. The second is moral authority, rooted in the Hippocratic Oath, by which doctors do what is right and what is good for the individual patient. The third is charismatic authority, by which doctors are not expected to be reasonable at all times; they must have the opportunity to be arbitrary when necessary since life and death themselves are largely arbitrary [14].

Today, the attitudes of physicians regarding uncertainty in medicine and how that affects the doctor-patient relationship appear to be shifting again, particularly in Western civilizations. Moira Stewart's pioneering work has developed the concept of "patient-centered" care [15]. Exponentially increasing use of the Internet by patients, especially in more affluent countries of the West, has begun to level the playing field of up-to-date medical information between patients and their doctors. Egalitarianism is replacing authoritarianism in the physician-patient relationship in these societies, incorporating mutual participation, respect, and shared decision-making as information is communicated in ways that maximize its understanding [16, 17].

Looking to the Future

So, does the concept of uncertainty in medicine have anything to do with the issue of physician-patient relationships? Quite a lot, I believe. With the revolution in genetic and molecular medical science in the twenty-first century, epitomized by the

Human Genome Project that was completed in 2003, perhaps the greatest feat of inward exploration in history, by mapping, decoding, and sequencing all the genes of the human body, it was fully expected that the large gap in our knowledge of human disease would diminish and even disappear. But the opposite has occurred. The massive amount of data these advances have generated has resulted in even more uncertainty. While gradually more and more of the new information will be successfully translated to target the development of new treatments for previously untreatable diseases, the vast majority of discovered information will remain uninterpretable. In fact, it is even today exposing the existence of diseases that were heretofore not even known about. When a clinician today orders blood or tissue testing for next generation sequencing (NGS), hoping to identify a precise molecular cause of a patient's medical problem, the test may prove useless. Even more problematically, along with the gene mutations the clinician is looking for, she is apt to find other genomic variants, the significance of which for the patient is completely unknown. The same is true with other "omics" testing, like proteomics, transcriptomics, epigenomics, and metabolomics. Abnormalities found within this massively expanding, seemingly infinite amount of new structured and unstructured data [18] in any given patient will only widen the uncertainty chasm in medical knowledge. It is the price we pay for important discoveries and progress in medical science. The more we look, the more we will find. So, what is the information-overloaded physician supposed to communicate to her patient about all the unknowns about his condition?

The answer is that it will require a seismic paradigm shift in the physician-patient relationship. Today, the confluence of astonishing growth in computer capacity, with zettabytes of information immediately accessible on Internet, suggests that the application of artificial intelligence (AI) to medical practice is within reach and likely inevitable [19]. How will this affect the physician-patient relationship? The adoption of robust AI into clinical practice will undoubtedly bring with it a new level of humility among physicians, their more ready acceptance of uncertainty in medicine, and recognition of their human limitations. These projections may evoke dystopian visions of robots replacing physicians in the future of health care. At the same time, however, the data input required to feed the AI system is extremely complex and prone to serious errors even now. And the rate of technological advances to deal with these limitations will likely be greatly exceeded by the rate of new information acquisition. More importantly, the output of AI information, which is already at risk for generating erroneous and even nonsensical answers, will continue as the complexity of input increases. Finally, the good physician's personal experience and intuition, the nuances of human behavior she can observe and sensibly incorporate into diagnosis and treatment, and the incalculable benefits of human-to-human interaction in healing or failure to heal can never be reproduced formulaically and algorithmically by a robot. Therefore, in summary, the physician-patient relationship will inevitably continue to change, perhaps even radically, but uncertainty in medicine can never disappear until we have attained absolute and permanent perfection in the practice of medicine, the impossible.

References

1. Fins JJ. "When a diagnosis has no name" workshop; 2021.
2. Iannello P, et al. Ambiguity and uncertainty tolerance, need for cognition, and their association with stress. A study among Italian practicing physicians. Med Edu Online. 2017;22(1):1270009. https://doi.org/10.1080/10872981.2016.1270009.
3. Fox RC. Training for uncertainty. In: Merton RK, Reader G, Kendall P, editors. The student-physician: introductory studies in the sociology of medical education. Cambridge, MA: Harvard university Press; 1957. p. 207–41.
4. Szasz TS, et al. The doctor-patient relationship and its historical context. Am J Psychiatry. 1958;115:522–8.
5. Weiss GB. Paternalism modernized. J Med Ethics. 1985;11:184–7.
6. Alexander GC, et al. Brief report: physician narcissism, ego threats, and confidence in the face of uncertainty. J Appl Soc Psychol. 2010;40:947–55.
7. Katz J. Why doctors don't disclose uncertainty. Hast Cent Rep. 1984;14:35–44.
8. Waitzkin H, Stoeckle JD. The communication of information about illness. Clinical, sociological, and methodological considerations. Adv Psychosom Med. 1972;8:180–215.
9. Cushing A. History of the doctor-patient relationship. In: Brown J, Noble LM, editors. Clinical communication in medicine. John Wiley & Sons, Ltd. p. 2–16.
10. Ludmir EB, et al. The physician as dictator. Lancet. 2017;390(10099):1023. https://doi.org/10.1016/S0140-6736(17)32147-5.
11. Osmond H. God and the doctor. N Engl J Med. 1980;302:555–8.
12. Goranson A, et al. Doctors are seen as godlike: moral typecasting in medicine. Soc Sci Med. 2020;258:113008.
13. Ecclesiasticus. 38:1–25. King James Bible. https://www.bibleopening.com/passage?query=sirach. 38:1–25.
14. Siegler M, Osmond H. Aesculapian authority. Hastings Cent Rep. 1973;1:41–52, including Paterson TT. Notes on Aesculapian authority. Unpublished manuscript. 1957, p. 7.
15. Stewart M, et al. Patient-centered medicine: transforming the clinical method. Thousand Oaks: Sage; 1995.
16. Kaba R, Sooriakumaran P. The evolution of the doctor-patient relationship. Int J Surg. 2007;5:57–65.
17. Byrne PS, Long BEL. Doctors talking to patients. A professional evolution. J Am Med Assoc. 1996;275:152–6.
18. Noorbakhsh-Sabet N, et al. Artificial intelligence transforms the future of healthcare. Am J Med. 2019;132:795–801.
19. Wachter R. The digital doctor: Hope, hype, and harm at the Dawn of Medicine's computer age. New York: McGraw Hill; 2017. p. 94–5.

Syndromes in Search of a Name: Disorders of Consciousness, Neuroethics, and Nosological Humility

18

Joseph J. Fins

Introduction: Terry's Story

In 1993 Angilee Wallis got a call from her son's nursing home. Terry, who was in the permanent vegetative state following a car accident in 1984, was "not right." Evidently, overnight Terry's roommate, an elderly man with advanced dementia, got tangled up in his sheets, asphyxiated himself, and died. Terry was seemingly disturbed by what happened even though that would appear impossible given the fact that he carried a "diagnosis" of permanent unconsciousness. Nonetheless, the nurse's aide caring for Terry sensed that something was wrong, that he was in distress. Using her mother's intuition, she called Mrs. Wallis so she could come to the home and comfort her son.

For my book, *Rights Come to Mind: Brain Injury, Ethics, and the Struggle for Consciousness*, I interviewed Mrs. Wallis and she told me about that fateful day:

> …One of the aides called me from work one morning and told me she was not supposed to do that but… that man had passed away that night, and that it had bothered Terry…I needed to be down there…[when she arrived] Terry was lying there with his eyes open wide, he would not go to sleep, I mean he was making no noise at the time. But I stayed there with him most all the day until he finally went to sleep. So I don't know what he saw, but I know he saw something. And I know it had, *now,* I knew then it had to be something that was really bad [1].

J. J. Fins (✉)
Division of Medical Ethics, Weill Medical College of Cornell University, Weill Cornell Medical College, New York, NY, USA

Adjunct Faculty, The Rockefeller University, New York, NY, USA

Visiting Professor of Law and Solomon Center Distinguished Scholar in Medicine, Bioethics and the Law, Yale Law School, New Haven, CT, USA
e-mail: jjfins@med.cornell.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_18

The operative word in Mrs. Wallis's recounting is *now* for by the time I spoke with her Terry's condition had undergone a dramatic change. In 2003, 19 years after his car accident, he started to speak for the first time since he had fallen into a coma and then the vegetative state [2]. His words, first "Mom" and then "Pepsi," garnered international attention and were described as a miracle awakening in the international media [1]. Over the ensuing weeks, he developed greater fluency, but he was like Rip Van Winkle, stuck in time. For him, it was still 1984 and Ronald Reagan was president, but Mrs. Wallis had her son back. He knew who she was and as importantly who he was.

It was this appreciation of self and others that gave Mrs. Wallis pause when she retold the story of Terry's reaction to his roommate's death. Now that he had started to talk and tell her he was there, she understood in retrospect *why* he had been so upset by the overnight events. Despite the diagnosis he carried, she wondered, perhaps he hadn't been unconscious at all, but rather aware and unable to communicate. Perhaps he hadn't been in the vegetative state as all of his doctors thought. It left her with a sense of horror that Terry had had to experience the trauma of seeing his roommate die and bear witness to that death without voice. Had it not been for his nurse's aide's moral intuition, his distress would have gone unnoticed and unattended.

Mrs. Wallis's newfound perspective was more than a function of her son talking after 19 years. It was that science had caught up to Terry's narrative. It now had a way to explain what had happened and give a name to his condition. In 2002, an expert panel published the diagnostic criteria for the Minimally Conscious State (MCS) a state of liminal consciousness in which patients have intention, attention, and memory [3]. Patients may say their name, look up when you come into the room, or reach for a cup. The challenge is that these behaviors are episodic and intermittent and not reproducible. And when they are not demonstrating these behaviors, MCS patients look as if they are vegetative. Biologically MCS patients are also distinct from vegetative ones. Patients in MCS have intact neural networks in contrast to the disintegration of vegetative patients whose brains are unable to work as a unit [4, 5].

The advent of the minimally conscious state was a step in the evolution of neuroscience's description of disorders of consciousness that continues to this day. As such this process of nosologic refinement provides an historical analog to rheumatologic conditions about which this volume is primarily concerned. But it is more than a lesson in the evolving history of diagnostics. It is also a lesson in *nosologic humility* [6]. Simply put in 1993, when Terry was shaken by his roommate's death, neither he nor neuroscience had words to describe his reaction. He was voiceless, and neurology would have to wait another decade for the creation of a diagnostic category for patients who appeared vegetative but in factor harbored consciousness.

There is a lesson in this disconnect in what we think we know and what is actually understood. In retrospect it is apparent that Terry wasn't vegetative in 1993. But that was his diagnosis because there were no other options. But, if he wasn't vegetative, then what was he? Even if someone pointed to the inconsistencies of his presentation, the power of Kuhnian forces would squelch any violation of the prevailing

paradigm and assert a kind of certitude which would be unwarranted [7]. And yet that is precisely the challenge of novel thinking in the face of diagnostic ignorance.

Such circumstances call for humility rather than hubris if we hope to advance knowledge and diagnostic thinking. As the great anatomist Oliver Wendell Holmes, Sr. – father of the jurist and Harvard anatomy professor – asserted in an 1862 volume entitled *Border Lines of Knowledge in Some Provinces of Medical Science*, "The best part of knowledge is that which teaches where knowledge leaves off and ignorance begins. Nothing so clearly separates a vulgar from a superior mind, than the confusion in the first between the little that it truly knows, on the one hand, and what it half knows and what it thinks it knows, on the other." Observing that "Science is the topography of ignorance," Holmes reminds us, "That which is true of every subject is especially true of that branch of knowledge which deals with living beings… [8]."

The Origins of the Vegetative State

What was true for Holmes in the mid-nineteenth century has been true for the last half century with respect to disorders of consciousness. The story begins with the introduction of the persistent vegetative state in 1972 in a landmark article in *The Lancet* by the Scottish neurosurgeon Bryan Jennett, the originator of the Glasgow Coma Scales, and Fred Plum, the American neurologist who first described the locked-in-state [9]. Given that this volume speaks to the question of *when the illness has no name*, it is fitting that Jennett and Plum entitled their article "The Persistent Vegetative State after Brain" adding the subtitle: "A Syndrome in Search of a Name."

Fred Plum was known as an exceedingly precise wordsmith and skilled editor (he was chief editor of the *Archives of Neurology* and the founding editor of the *Annals of Neurology*) [10]. With this provenance the essay on the vegetative state becomes especially instructive for those seeking to name novel conditions. First, there is the subtitle. Why syndrome and not a diagnosis in search of a name? A syndrome is a symptom complex which describes a specific condition but for which a clear causal explanation remains unknown [11]. In contrast a diagnosed disease is marked by a clear cause, as in a pneumonia caused by a specific pathogen. The vegetative state was properly described as a syndrome because many etiologies (anoxia, trauma) can cause it and because the mechanisms of the physiologic and anatomic derangements leading to the vegetative state were unknown and themselves diverse. Jennett and Plum's careful parsing of syndrome versus diagnosis points to nomenclature that can serve as an intermediate way of giving a name to a condition that is in the *process of becoming* a formal disease. To name a syndrome, a disease prematurely can be as problematic as not naming a condition at all.

Jennett and Plum understood the vegetative state as one of wakeful unresponsiveness in which autonomic functions were preserved and the eyes were open but there was no awareness of self, others, or the environment. They postulated that this state of unconsciousness represented the isolated function of the brain stem in the

absence of higher cortical function. In naming their syndrome, they looked to history and philosophy for a predicate that illustrated this clinical phenomenon and nodded towards a putative mechanism. To that end they invoked Aristotle's *De Anima*, translated as "on the soul" [12].

Aristotle, who as both a philosopher and a botanist sought to systematize nature, conceived of two biological faculties, the vegetative and the animalic. The former is nutritive and foundational and is modeled on the plant. The second, seen in animals and humans, is one of sensation, movement, and thought. This hierarchy worked well for Jennett and Plum who saw ascending and dependent faculties from the brain stem, responsible for vegetative and autonomic function, on to the cortex which allowed for higher cortical functions [12].

Once they settled on vegetative, Jennett and Plum needed a temporal descriptor for their syndrome. In their careful choice of "persistent," we see their conceptual prudence and an acknowledgment of the limits of their knowledge. They wanted to convey a syndrome of long duration, but they neither knew how long it would last nor whether it would last forever. Thus, instead of simply giving the condition some sort of temporal marker, they *explained* their choice:

> Certainly we are concerned to identify an irrevocable state, although the criteria needed to establish that prediction reliably have still to be confirmed. Until then "persistent" is safer than "permanent" or "irreversible"; but "prolonged" is not strong enough, and unless it is quantified it is meaningless [9]."

This passage explains the rationale for persistent and their temporal prudence. Based on available data, they could not say the condition was permanent. So given this contingency and the possibility of evolving new knowledge, they viewed persistent as safer than permanent as a temporal modifier. In 1994, the Multi-Society Task Force published a two part consensus statement in the *New England Journal of Medicine* adding "permanent" as a second descriptor to vegetative states that persisted for 3 months after anoxic and 12 months after traumatic brain injury [13]. This reflected new knowledge of both the epidemiology of the condition and an understanding of how differing etiologies (anoxia versus trauma) influenced the natural history of the condition. The neurologist, James L. Bernat, in an interview for *Rights Come to Mind* explained the Task Force's thinking:

> The adjective "persistent" refers only to a condition of the past and continuing disability with an uncertain future, whereas "permanent" implies irreversibility. Persistent vegetative state is a diagnosis; permanent vegetative state is a prognosis [1].

Bernat's comments are notable for elevating Jennett and Plum's syndrome to a diagnosis and further distinguishing a diagnosis from a prognosis, which here is how *stable* a diagnosis will be over time.

The new millennium brought with it a new diagnostic category with the Minimally Conscious State. As noted, the MCS criteria were published in 2002 [3]. Notably, the subtitle of that paper was "Definition and Diagnostic Criteria." Like the 1994 Multi-Society Task Force Report on the Vegetative State, the early syndromic

rumblings of Jennett and Plum had been replaced with language about diagnosis and fixed criteria which helped to distinguish this condition from the vegetative state.

From Phenotype to Mechanism

In her volume *Making Medical Knowledge,* Miriam Solomon writes that there are multiple ways of knowing and creating new medical knowledge [14]. Despite the favored placed of evidence-based medicine in clinical trials, understanding mechanisms of illness is important for classification and to the development of new therapeutics [14, 15]. In the context of brain injury, understanding mechanisms of injury and recovery have been catalyzed by functional neuroimaging during the first two decades of the twenty-first century.

Before the advent of neuroimaging, diagnostic assessment was conducted at the bedside based on neurological signs and behavioral assessment tools. By giving an *interior view* of the brain, neuroimaging opened up the possibility of additional refinement of diagnostic categories and the identification of discordances between behaviors seen at the bedside and activity within the brain.

In 2019 I focused on these discordances at a conference on personalized medicine at the Pontifical Academy of Sciences in 2019 [16, 17]. In my lecture, I analogized the behaviors seen at the bedside and the circuitry visualized on functional neuroimaging to the phenotype/genotype distinction so important in molecular medicine and cancer biology. Like the peas in Mendel's garden, what appears the same phenotypically may have significantly different underlying biologies. This has been true for malignancies, with differing genetic arrays, that respond differently to therapeutic agents based on their molecular vulnerabilities. The susceptibility of acute promyelocytic leukemia to ARA-C comes to mind as an exemplar of why underlying biology is therapeutically dispositive [18].

The same phenomenon is informing the understanding and diagnostic classification of disorders of consciousness and hinges on the presence of covert consciousness, as seen on neuroimaging but obscured on the clinical exam. This was first identified in a 2006 paper in *Science* by Adrian Owen and colleagues [19]. These investigators demonstrated the ability of a patient clinically diagnosed as being in the vegetative state as being able to follow volitional commands when placed in a scanner. Using functional magnetic resonance imaging (*f*MRI), the patient was asked to imagine playing tennis, walking about her house, and linguistically distinguish similarly sounding words with different meanings. When she imagined doing these tasks, she activated motor, spatial, and linguistic regions of the brain associated with the *performance* of those tasks.

In a paper I coauthored with Nicholas Schiff, we noted that this was evidence of *non-behavioral MCS* in which a patient *behaviorally* thought to be in the vegetative state could not be in the vegetative state because of command followings [20]. If the vegetative state is a state of wakeful *unresponsiveness*, activations on the *f*MRI indicated responsiveness and volitional activity not seen in patients who are unconscious.

A follow-on study by Monti et al. published in the *New England Journal of Medicine* used volitional command following to create a communication channel for a patient who was otherwise thought to be in the vegetative state but was in fact in MCS, albeit without motor output [21]. That patient was able to answer yes/no questions by imagining tennis imagery for *yes* and spatial navigation tasks for *no*. While this patient under further scrutiny did ultimately demonstrate subtle motor output, that finding would not likely have been elucidated without the evidence of volitional command following on *f*MRI.

Studies such as these, as well as a parallel study conducted at Weill Cornell [22], were pivotal because they opened up a way of classifying patients beyond their bedside phenotype, suggesting what my colleague Nicholas Schiff has described as cognitive motor dissociation (CMD) [23]. CMD, or the discordance between volitional cortical activity indicative of consciousness in the absence of correlative behavioral output, is a phenomenon which traverses the range of disorders of consciousness.

More recently evidence of CMD, and covert consciousness, has been identified acutely in the Intensive Care Unit with evidence that some patients thought to be in coma – classically defined as an eyes-closed state of unresponsiveness – having the ability to covertly demonstrating awareness on *f*MRI [24, 25] or EEG [26]. Of note patients with CMD at the start of their brain injury had better outcomes a year later than those who did not. There are normative implications to this finding beyond the prognostic relevance of CMD in the ICU [27]. When families appreciate that their loved ones *are there* in a way that may not be readily apparent at the bedside, they are more reflective about decisions to withhold or withdraw life-sustaining therapy. This becomes an important corrective to empirical studies that have shown that 70% of deaths following acute traumatic brain injury are accounted for by decisions to withdraw life-sustaining therapy [28].

Naming Wisely

Despite the overwhelming importance of covert consciousness, and its demonstration of the perils of relying solely on phenotype for establishing a diagnosis of consciousness, in 2010 European investigators proposed the *unresponsive wakefulness syndrome* (UWS) as a new name for the vegetative state [29]. Perhaps motivated by a sense that the term was pejorative, despite overwhelming scholarly evidence [12], other articulations of a vegetative nervous system [30], and Plum's own assertion to the contrary [31], UWS has unfortunately gained currency.

While the advancement of this name could reflect an unawareness of the etymological origins of the vegetative state dating back to antiquity, there is a deeper lack of awareness manifest in the moniker of UWS to describe the vegetative state. As such UWS presents a cautionary tale for all who would seek to name the nameless conditions and syndromes that have captured our attention. To be blunt, a bad name may be worse than no name at all.

Proponents who seek to rename the vegetative state as the UWS have made two critical mistakes. First there is an editorial concern. The phrase "unresponsive wakefulness" bears a resemblance to what Jennett and Plum actually said in their *Lancet* paper. There, describing the vegetative state, they noted that "it seems wakefulness without awareness" [9]. So, to start there is the question of originality and authorial provenance. Simply stated, Jennett and Plum said it first.

But there is a more substantive concern in Jennett and Plum cautions about the vegetative state – "that it *seems* wakefulness without awareness" [9]. And here is the brilliance of their logical deductions: They knew what they could not know. In 1972 they had no way to peer into the brain and discern function. So they hedged and asserted that vegetative patients seemed to be awake and unaware, but they could not be sure. It might be possible, and so they were hesitant to assert definitively that the observed wakefulness of the vegetative state was invariably associated with a lack of awareness. Their hedge was a virtuosic application of logic: what one cannot truly demonstrate one cannot truly know. Their caution held out the possibility of what we would, decades hence, call covert consciousness or cognitive motor dissociation [32].

Jennett and Plum's careful parsing was forward looking in 1972 in contrast to the introduction of UWS in 2010. Why would anyone want to describe a condition as wakeful and unresponsive now that we have examples of patients who *seem* wakeful and unresponsive at the bedside and yet demonstrate responsiveness on functional neuroimaging? Given this, it makes little sense to have an unresponsive wakefulness syndrome. Patients could be behaviorally unresponsive and not be in a vegetative state if they had covert consciousness. This real possibility points to the ideological risks inherent in naming (or renaming) a condition [33]. We have to be careful with our choices and name wisely.

The Place of Time

A central consideration with all names is their place of time. Is a condition acute or chronic, is it self-limited or of long duration? Or, as in the case of by Jennett and Plum, should the vegetative state be described as persistent or permanent? Everything in medicine is governed by the clock [34], from the cell cycle to visiting hours, immunization schedules, and how we structure clinical specialties (think the life cycle of pediatrics to geriatrics).

We need to be especially careful about our invocation of time. Indeed when we name conditions, we will be judged over time. In this regard, history has been kind to Jennett and Plum. The careful reader will recall how they decided to settle upon "persistent" [9] to describe the vegetative state in 1972 and that a permanent category was added in 1994 [13]. More recently – affirming the temporal prudence of Jennett and Plum's original rejection of "permanent" – experts from the American Academy of Neurology (ANA), the American Congress of Rehabilitation Medicine (ACRM), and the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDDLR) [35] made the judgment that a "permanent"

category was no longer justified. Based on a systematic evidence-based review, they determined as many as 20% of patients designated as in the permanent vegetative state might recover consciousness. Because of this finding, they redesignated the permanent vegetative state as the *chronic* vegetative state [36].

In an essay, coauthored with the neurologist James L. Bernat, accompanying the ANA, ACRM, and NIDDLR evidence-based review and practice guideline, I explored the ethical, palliative, and legal implications of this redesignation. We argued that 20% overestimated the number of recoveries as it did not account for the high rate of misdiagnosis of covert consciousness [37]. Nonetheless, the redesignation had normative significance because the predicate for the right to die, established in *Quinlan*, *Cruzan*, and *Schiavo*, was the *permanence* of the vegetative state. It became a proxy for unbreachable futility.

To make a general point about time and naming: the moral significance we attach to diagnoses – their meaning for patients and families and the choices they make to treat or to withhold care – are often predicated upon perceptions of a future. When we, as physicians, cannot adequately characterize that future, the time course and trajectory of our patients' diseases, these life-altering choices become even more difficult.

The Greeks had two conceptions of time – *chronos* and *kairos* – that speak to the challenge of making significant choices absent temporal clarity. *Chronos* is chronological time of the sort distinguished by adjectives like persistent, permanent, or chronic. *Kairos*, on the other hand, is the deeper significance of these designations and what they mean for ethical and clinical choices. *Kairos* speaks to the timeliness of our choices. The deep challenge for us as we construct a nosology when we are trying to ascertain chronological biomarkers of illness is that the clock does not stop. Life continues and choices need to be made. The problem, as one scholar put it, was that for the Greeks *chronos* must precede *kairos* [38]. Facts are the predicate for value choices. But this is not always possible as exemplified by medicine's grappling with the life and death choices posed by the COVID-19 pandemic [39] or whenever we are confronted by epistemic uncertainty [40].

What's in a Name?

In naming things, we create order and structures of knowledge that first improve the identification of conditions and then lead to novel therapeutics. Osler said it best in an essay called "The Leaven of Science" published in his *Aequanimitas*: "The determination of structure with a view to the discovery of function has been the foundation of progress" [41]. I would assert that those structures can be anatomic, such as the anatomic circuitry underlying disorders of consciousness or structures of knowledge that create the framework for imagining how progress can be made. Both start with our typologies of knowledge and are dependent upon the grammar and syntax of the diagnostic language we seek to construct.

While the progress made in developing therapeutic strategies for disorders of consciousness is beyond the scope of this essay, suffice it to say that the elucidation

of covert consciousness and the *mesocircuit* underlying consciousness [42] – itself named to reflect the central role played by the thalamus as a hub linking the brain stem and the cortex – has led to efforts to use deep brain stimulation (DBS) to modulate the thalamus to integrate cortical function. As we reported in *Nature* [43], with stimulation of the bilateral interlaminar nuclei of the thalamus, a MCS patient who was only able to communicate by episodically moving his eyes had improved cognitively mediated behaviors, limb control, and the ability to take oral feeding. With DBS he was able to say the first 16 words for the Pledge of Allegiance, tell his mother he loved her, and voice preferences. It was *agency ex machina* [44], the restoration of voice via a neuroprosthetic, an effort conceivable and made possible by structures of knowledge created by the evolving nomenclature which has classified disorders of consciousness since Jennett and Plum first advanced the persistent vegetative state decades ago [9].

New Nosologies, Pragmatism, and Disability Rights

As a physician ethicist, I have sought to draw upon the American Pragmatic Tradition, specifically John Dewey's theory of inquiry [45], to develop a method of moral problem-solving that I have called *clinical pragmatism* [45]. Clinical pragmatism seeks to blend ethical theory and practice in the service of the real, or what Dewey would call the construction of the good [46]. As I approached the problems posed by disorders of consciousness, or more generally diseases in search of a name, I have sought to utilize pragmatic philosophical approaches to understand scientific advances and the value choices that progress can prompt [47–49].

Dewey was keenly sensitive to the implications of progress. In 1938, he wrote an essay, entitled "Common Sense and Scientific Inquiry," in which he wrote of the synergy between science and philosophy and how "Inventions of new agencies and instruments create new ends; they create new consequence which stir men to form new purposes" [50]. It is an apt allusion both to the purpose of pragmatism, also known as instrumentalism, and the place that technology plays in science [51]. In the case of disorders of consciousness, the advent of functional neuroimaging and the discovery of covert consciousness have created new ends and purposes.

One of those purposes is to appreciate issues of inequity that have historically afflicted patients with disorders of consciousness. As I have documented in *Rights Come to Mind* [1], after brilliant acute care that saves lives, patients with severe brain injury often are relegated to nursing homes. There they receive what has been euphemistically described as custodial care, without adequate medical treatment or the rehabilitative services that could restore function and ameliorate distress. Central to this has been the restoration of functional communication to those whose consciousness has been silenced by injury through efforts like our use of deep brain stimulation in the minimally conscious state.

The point is that once one identifies – and names – the problem of covert consciousness, one can no longer look away. It becomes a moral imperative to give

voice to the voiceless. In the landmark *Obergefell* decision that legalized gay marriage, Justice Anthony Kennedy spoke of how new insights can prompt a realization of inequities that need to be rectified. In a stirring passage for the majority, he opined that, "…[N]ew insights and societal understandings can reveal unjustified inequalities within fundamental institutions that once passed as unnoticed and unchallenged" [52].

In the context of this anthology, one could assert that *if you name it you own it*. It becomes important to follow through and ensure health equity and access to care. This has been the focus of my work advocating for the needs of patients and families touched by severe brain injury through the prism of a disability rights framework [1, 53, 54] which seeks to achieve societal inclusion under the aegis of the Americans with Disabilities Act [55–57].

Coda

Two years after the onset of Karen Ann Quinlan's coma, Jennett and Plum proposed an outcome study of 1000 comatose patients [58]. Their ambitious plans were reported upon in a 1977 *New York Times* article by Lawrence K. Altman [59]. In the wake of *Quinlan*, Jennett and Plum hoped to better understand the nature of the condition and provide families with better prognostic information so they could make more informed choices. Jennett told Altman that, "We know surprisingly little about the process of recovery" [59]. For his part, Plum hoped their project would help begin to articulate, "the scientific basis of tomorrow's medical ethics" [59]. That ethic of care began 50 years ago when Jennett and Plum first described the persistent vegetative state.

And it all began with a name…

Acknowledgments The author is grateful to Dr. Michael Lockshin for his visionary leadership in formulating the *When the Illness Has No Name* conference and the Hospital for Special Surgery for the opportunity to contribute to this important endeavor. Dr. Fins also acknowledges the support of the Weill Cornell Medical College's Clinical & Translational Science Center [UL1-RR024966] which helped to support his explorations in ethics and translational medicine.

Dedication This essay is dedicated to the memory and legacy of Dr. Fred Plum.

References

1. Fins JJ. Rights come to mind: brain injury, ethics and the struggle for consciousness. New York: Cambridge University Press; 2015.
2. Schiff ND, Fins JJ. Hope for "Comatose" patients. Cerebrum 2003;5(4):7–24. Reprinted as: Schiff ND and Fins JJ. Hope for "Comatose" Patients. *A* Cerebrum Classic in The Dana Foundation's Cerebrum 2007, Read CA, editor. New York: Dana Press, 2007. p. 185–203.
3. Giacino JT, et al. The minimally conscious state: definition and diagnostic criteria. Neurology. 2002;58(3):349–53.

4. Laureys S, et al. Cortical processing of noxious somatosensory stimuli in the persistent vegetative state. NeuroImage. 2002;17(2):732–41.
5. Schiff ND, et al. fMRI reveals large scale network activation in minimally conscious patients. Neurology. 2005;64:514–23.
6. Fins JJ. The Jeremiah Metzger lecture: disorders of consciousness and the normative uncertainty of an emerging nosology. Trans Am Clin Climatol Assoc. 2020;131:235–69.
7. Kuhn TS. The structure of scientific revolutions. 2nd ed. Chicago: The university of Chicago Press; 1972.
8. Holmes OW. Border lines of knowledge in some provinces of medical knowledge. Boston: Ticknor and Fields; 1862.
9. Jennett B, Plum F. Persistent vegetative state after brain damage. A syndrome in search of a name. Lancet. 1972;1(7753):734–7.
10. Altman LK. Fred Plum, Neurologist who helped coin 'Persistent Vegetative State,' dies at 86. The New York Times. June 12, 2010. Accessed on 10 January 2010 at: https://www.nytimes.com/2010/06/13/health/13plum.html.
11. Calvo F, et al. Diagnoses, syndromes, and diseases: a knowledge representation problem. AMIA Ann Symp Proc. 2003;2003:802.
12. Adams ZM, Fins JJ. The historical origins of the vegetative state: received wisdom and the utility of the text. J Hist Neurosci. 2017;26(2):140–53.
13. Multi-Society Task Force on PVS. Medical aspects of the persistent vegetative state. Parts I and II. N Engl J Med. 1994;330:1499–508. 1572–9
14. Solomon M. Making medical knowledge. New York: Oxford University Press; 2014.
15. Fins JJ. A review of *Making Medical Knowledge* by Miriam Solomon. New York: Oxford University Press; 2014. Notre Dame Philosophical Reviews. March 7, 2016. http://www.ndpr.nd.edu/news/65137-making-medical-knowledge/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+PhilosophicalReviews%2FNews+(Notre+Dame+Philosophical+Reviews).
16. Personalized Medicine and Disorders of Consciousness: An alternate convergence of knowledge towards a New Clinical Nosology. Presented at "Revolution of Personalized Medicine." Pontificia Academia Scientiarum. The Vatican, Vatican City. April 8, 2019.
17. Fins JJ. Personalized Medicine and Disorders of Consciousness: an alternate convergence of knowledge towards a New Clinical Nosology. In: Barilan YM, Brusa M, Ciechanover A, editors. Can precision medicine be personal; can personalized medicine be precise? Oxford: Oxford University Press; 2021. In Press. (Expected, 2021).
18. Coombs CC, et al. Acute promyelocytic leukemia: where did we start, we are we now, where are we now, and the future. Blood Cancer J. 2015;5:e304.
19. Owen AM, Coleman MR, Boly M, et al. Willful modulation of brain activity in disorders of consciousness. Detecting awareness in the vegetative state. Science. 2006;313(5792):1402.
20. Fins JJ, Schiff ND. Shades of gray: new insights from the vegetative state. Hast Cent Rep. 2006;36(6):8.
21. Monti MM, et al. Willful modulation of brain activity in disorders of consciousness. N Engl J Med. 2010;362(7):579–89.
22. Bardin JC, et al. Dissociations between behavioural and functional magnetic resonance imaging-based evaluations of cognitive function. Brain. 2011;134(3):769–82.
23. Schiff ND. Cognitive motor dissociation following severe brain injuries. JAMA Neurol. 2015;72(12):1413–5.
24. Edlow BL, et al. Early detection of consciousness in patients with acute severe traumatic brain injury. Brain. 2017;140(9):2399–414.
25. Edlow BL, Fins JJ. Assessment of covert consciousness in the intensive care unit: clinical and ethical considerations. J Head Trauma Rehabil. 2018;33(6):424–34.
26. Claassen JR. Detection of brain activation in unresponsive patients with acute brain injury. N Engl J Med. 2019;380(26):2497–505.

27. Fins JJ. Disorders of consciousness in clinical practice: ethical, legal and policy considerations. In: Posner JP, Saper CB, Claassen J, Schiff ND, editors. Plum and Posner's diagnosis of stupor and coma. 5th ed. New York: Oxford University Press; 2019. p. 449–77.
28. Turgeon AF, et al. Mortality associated with withdrawal of life-sustaining therapy for patients with severe traumatic brain injury: a Canadian multicentre cohort study. CMAJ. 2011;183(14):1581–8.
29. Laureys S, et al. European Task Force on Disorders of Consciousness. Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. BMC Med. 2010;8:68.
30. Timme W, Davis TK, Riley HA, editors. The vegetative nervous system; an investigation of the most recent answers. Proceedings of the Association of Nervous and Mental Diseases Volume IX. Baltimore: Williams & Wilkins; 1928.
31. Plum F, et al. Coordinated expression in chronically unconscious persons. Philos Trans Biol Sci. 1998;353:1929–33. https://doi.org/10.1098/rstb.1998.0345.
32. Fins JJ. Disorders of consciousness in clinical practice: ethical, legal and policy considerations. In: Posner JP, Saper CB, Claussen J, Schiff ND, editors. Plum and Posner's diagnosis of stupor and coma. 5th ed. New York: Oxford University Press; 2019. p. 449–77.
33. Fins JJ, Plum F. Neurological diagnosis is more than a state of mind: diagnostic clarity and impaired consciousness. Arch Neurol. 2004;61(9):1354–5.
34. Fins JJ. My time in medicine. Perspect Biol Med. 2017;60(1):19–32.
35. Giacino JT, et al. Comprehensive systematic review update summary: disorders of consciousness: report of the guideline development, dissemination, and implementation Subcommittee of the American Academy of neurology; the American Congress of Rehabilitation Medicine; and the National Institute on Disability, Independent Living, and Rehabilitation Research. Neurology. 2018;91(10):461–70. Published simultaneously in Arch Phys Med Rehabil 2018;99(9):1710–9.
36. Giacino JT, et al. Practice guideline: disorders of consciousness. Neurology. 2018;91(10):450–60. Simultaneously published in Archives of Physical Medicine and Rehabilitation 2018;99(9):1710–9.
37. Fins JJ, Bernat JL. Ethical, palliative, and policy considerations in disorders of consciousness. Neurology. 2018;91(10):471–5. Simultaneously published in Archives of Physical Medicine and Rehabilitation. 2018;99(9):1927–31.
38. Smith JE. Time, Times and the 'Right Time': "Chronos" and "Kairos". Monist. 1969;53(1):1–13.
39. Fins JJ. COVID-19 through time. Issues Sci Technol. 2021;37(3):73–8. https://issues.org/covid-19-through-time-medical-knowledge-fins/.
40. Fins JJ. Time to Death: *Chronos, Kairos* and the "Longest Distance between Two Poles". In: Fine R, Levinson J, editors. The pursuit of life: the promise and challenge of palliative care. University Park: The Pennsylvania State University Press; 2021. In Press.
41. Osler WO. The leaven of science. In: Aequanimitas. London: H.K. Lewis; 1904.
42. Schiff ND. Recovery of consciousness after brain injury: a mesocircuit hypothesis. Trends Neurosci. 2010;33(1):1–9.
43. Schiff ND, et al. Behavioral improvements with thalamic stimulation after severe traumatic brain injury. Nature. 2007;448(7153):600–3.
44. Fins JJ, et al. In pursuit of *agency ex Machina*: expanding the map in severe brain injury. Am J Bioeth Neuroreth. 2021;12(2–3):200–2. https://doi.org/10.1080/21507740.2021.1904050.
45. Fins JJ, et al. Clinical pragmatism: a method of moral problem solving. Kennedy Inst Ethics J. 1997;7(2):129–45.
46. Fins JJ, et al. Clinical pragmatism: bridging theory and practice. Kennedy Inst Ethics J. 1998;8(1):39–44.
47. Fins JJ. Clinical pragmatism and the care of brain injured patients: towards a palliative neuroethics for disorders of consciousness. Prog Brain Res. 2005;150:565–82.
48. Fins JJ. Neuroethics and disorders of consciousness: a pragmatic approach to neuro-palliative care. In: Laureys S, Osseries O, Tononi G, editors. The neurology of consciousness, cognitive neuroscience and neuropathology. 2nd ed. Amsterdam: Elsevier; 2015. p. 241–55.

49. Fins JJ. Towards a pragmatic neuroethics in theory and practice. In: Racine E, Aspler J, editors. The debate about Neuroethics: perspectives on the field's development, focus, and future. Berlin: Springer; 2017. p. 45–65.
50. Dewey J. Common sense and scientific inquiry. In: Boydston JA, editor. Logic. The Theory of Inquiry. Collected in John Dewey, The Later Works, 1925–1953. Volume 12:1938. Carbondale and Edwardsville: Southern Illinois University Press; 1991. p. 66–85.
51. Fins JJ. Neuroethics and the lure of technology. In: Illes J, Sahakian BJ, editors. Epilogue, the Oxford handbook of neuroethics. New York: Oxford University Press; 2011. p. 895–908.
52. *Obergefell v. Hodges*, 135 S. Ct. 2584 (2015).
53. Fins JJ. Giving voice to consciousness: Neuroethics, human rights and the indispensability of neuroscience. The Society for Neuroscience David Kopf Lecture on Neuroethics. Camb Q Healthc Ethics. 2016;25(4):583–99.
54. Fins JJ. The civil right we don't think about. Sunday Review. The New York Times. August 27, 2017. SR-10.
55. Americans with Disabilities Act of 1990. Public Law 101–336. 108th Congress, 2nd session (July 26, 1990).
56. Wright MS, Varsava N, Ramirez J, Edwards K, Guevremont N, Ezer T, Fins JJ. Justice and severe brain injury: legal remedies for a marginalized population. Fla State Univ Law Rev. 2018;45:313–82.
57. Fins JJ, et al. Disorders of consciousness and disability law. Mayo Clin Proc. 2020;95(8):1732–9.
58. Fins JJ. A once and future clinical Neuroethics: a history of what was and what might be. J Clin Ethics. 2019;30(1):27–34.
59. Altman LK. Coma study seeks ways to predict recoveries. The New York Times, April 14, 1977. Accessed 10 January 2020 at: https://www.nytimes.com/1977/04/14/archives/coma-study-seeks-ways-to-predict-recoveries.html.

Reflections on the Conference by a Physician-Patient

19

Jerome Groopman

My Interest in this Conference

The issues covered in the conference are of interest to me not only as a hematologist-oncologist who has engaged in laboratory research, conducted clinical trials, and cared for the people with complex maladies but also as a patient who experienced the vicissitudes of diagnostic uncertainty.

The importance of a doctor in providing a patient with a coherent understanding of his or her illness and choice of treatment options became vividly apparent during my training as a hematology fellow at the University of California, Los Angeles. I saw a middle-age woman with bone, dermal, and oral lesions that the pathologist termed consisted of "Langerhans cells." Delving deeply into the literature, I found a multiplicity of diagnostic schemes for these disorders with scant insight into their pathogenesis. In 1981, I published an article in the Annals of Internal Medicine that began with an epigraph meant to reflect the state of the field:

> There is an ancient Chinese classification of animals into 13 categories, including those belonging to the emperor, tame animals, four-footed animals, those resembling flies, embalmed animals, mythologic animals, and those not included in the foregoing classes [1].

We proposed to move diagnosis from morphology to a more pathophysiological footing, specifically the cell of origin, in this context monocyte-macrophage, whether reactive to an external stimulus causing proliferation, a manifestation of a lipid storage disorder, or driven autonomously as a neoplasm.In my field of hematology-oncology, the revolution in molecular biology with advances in DNA technology soon allowed for pathological diagnoses based on histology to move to the level of genetics and clonality. Indeed, now classical terms like "non-Hodgkin's lymphoma" can be subdivided into more meaningful categories based not only on

J. Groopman (✉)
Medicine Harvard Medical School, Experimental Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_19

cell of origin but also importantly on identifying oncogenes and tumor suppressor genes as drivers of the neoplasm. This informs prognosis as well as choice of treatment, particularly the so-called targeted therapies which are selected based on the genetics. Other technologies also have importantly advanced refinements in diagnosis. Imaging, particularly PET scanning, is now regularly employed in assessing lymphoma and other malignancies. As hematologists and oncologists, we stand on a much more solid scientific foundation to explain to our patients the nature of their disorder and what to expect with regard to current treatment; we also are designing clinical trials in a more rational way, grouping patients based on molecular diagnoses.

So, for example, histiocytoses are now classified based on a deeper understanding of both the cell of origin of the lesion and its pathogenesis. The classification from the Histiocyte Society currently in use divides the disorders into five categories [2].

- The "L" (Langerhans) group – this includes LCH, indeterminate cell histiocytosis, Erdheim-Chester disease (ECD), mixed LCH/ECD, and extracutaneous juvenile xanthogranuloma.
- The "C" (cutaneous and mucocutaneous) group – this encompasses a range of disorders localized to the skin and/or mucosa surfaces that do not meet diagnostic criteria for LCH, including juvenile xanthogranuloma, adult xanthogranuloma, and cutaneous Rosai-Dorfman disease.
- The "R" (Rosai-Dorfman disease) group – this includes Rosai-Dorfman disease and miscellaneous non-cutaneous histiocytoses that do not meet diagnostic criteria for LCH.
- The "M" (malignant histiocytoses) group – this includes primary malignant histiocytoses involving the skin, lymph nodes, digestive system, central nervous system, and other areas and also encompasses malignant histiocytoses secondary to other diseases (e.g., follicular lymphoma, lymphocytic leukemia, hairy cell leukemia, acute lymphoblastic leukemia).
- The "H" group (hemophagocytic lymphohistiocytosis) group – this includes primary hemophagocytic lymphohistiocytosis (HLH) and macrophage activation syndromes due to Mendelian inherited conditions and secondary HLH due to infection, malignancy, rheumatologic syndromes, iatrogenic immune suppression or activation, or other conditions.

Note that despite the advances in diagnosis of these disorders, there is still uncertainty. This is clear in some of the categories above, where disorders appear to be shoehorned (the "R" group) by default.

Human biology is complex. Not only do we deal with genetics but also we deal with epigenetics, the modification of gene expression by environmental and other factors. Thus, as several conference speakers emphasized, diagnosticians seeking genetic determinants of disease still must deal with probabilities. Despite pathophysiological refinements in categories, there can remain stark differences in the manifestations of a specific disorder and its evolution or prognosis in any individual

over time. Stephen J. Gould, the Harvard biologist given a diagnosis of incurable and rapidly fatal mesothelioma, wrote an essay titled "The Median isn't the Message." [3] (Gould lived more than a decade and died of another cancer.) And, as articulated by several speakers, since for the ill individual a diagnosis is a story, the spectrum of outcomes for the patient means there is a multiplicity of potential narratives. In the field of oncology heavily based on genetics, this is vividly illustrated by the spectrum of outcomes for a woman with a given BRCA mutation. Some women will have early onset breast cancer and ovarian cancer, while other women within the same family with the same mutation will not develop cancer until a much older age, if ever. So diagnosis does not rigidly dictate prognosis; the median is not everyone's message, despite shared genetics.

Thinking as a Physician

Some 15 years ago, I began to question my thinking as a physician, specifically, why I would make a correct diagnosis and why at times I would miss one. I began searching the literature and found very little to explain diagnostic success and failure. The Institute of Medicine had published a report in 1999 on errors in medicine, but these were system errors, such as mistaking one patient for another and giving the wrong medication, remedied by system solutions like double checking the name on the hospital bracelet [4]. One of the few physician thinkers who had grappled with the issue was Dr. Pat Croskerry, a speaker at the conference, whom I featured in my writing and stands as a seminal researcher in the field of misdiagnosis. Drawing on the work of Daniel Kahneman and Amos Tversky, who defined thinking shortcuts, or heuristics, that appear to be evolutionarily imprinted and beneficial when we make judgments under time pressure and uncertainty, as physicians do, but can also lead us astray, I crafted a simple mnemonic that has served me well in trying to avoid the pitfalls of misdiagnoses: The Three As. These are anchoring, attribution, and availability [5].

"Anchoring" is to seize on the first bit of information that the patient might offer, or the first notable laboratory value or finding on an x-ray, and then to pursue thinking along a singular linear path. In effect, you drop your "cognitive anchor" in one diagnostic harbor rather than keeping an open mind on the horizon. "Attribution" is essentially engaging stereotypes or miscellaneous characteristics that color your thinking and cause you to draw on your biases, thereby closing your mind to other possibilities. One illustrative case came from Dr. Donald Redelmeier at the University of Toronto who saw an elderly patient who had been in the Merchant Marine [5]. He arrived in the emergency room wearing grubby clothes, unshaven, with a whiff of alcohol on his breath and was found to have liver disease. The house staff immediately attributed his hepatic dysfunction to alcoholism, relying on the stereotype of the drunken sailor who is disheveled, the alcoholic who doesn't care for himself and, Redelmeier observed, may trigger a sense of disgust in a doctor. The admitting resident did not take a detailed history or perform a careful and detailed physical exam or laboratory assessment and thus missed the underlying

diagnosis, which was Wilson's disease, an abnormality of copper metabolism. (The patient was not an alcoholic.) Attribution errors are particularly prevalent when it comes to patients with mental health issues, as Dr. Croskerry emphasizes, so that complaints of pain are readily attributed to anxiety or somatization without a sufficient exploration of potential organic causes.

"Availability" refers to how we tend to think of a diagnosis when we have recently seen a similar case, especially when the local medical ecology, so to speak, fosters this. If it's flu season and someone comes in with chills, cough, and dyspnea, we quickly think it's flu, again without crafting a broader differential diagnosis. I learned of one such case of availability from Dr. Harrison Alter when he worked in a clinic on a Native American reservation [5]. An elderly Navajo woman during a flu outbreak presented to the emergency room with a history of feverish feeling and was tachypneic. Dr. Alter told me although she was afebrile with a clear chest x-ray, he concluded her fever was masked because she said she had taken aspirin and the x-ray was clear because it was early in her illness and she might be dehydrated. And, because of availability, he fell prey to another cognitive pitfall, confirmation bias, where you ignore aberrant findings that don't fit your initial conclusion, in this case, an abnormal serum chloride in the elderly woman. It turned out she had salicylate poisoning from taking numerous aspirin when she didn't feel good, causing tachypnea to compensate for acidosis. The irony is that Dr. Alter, during his fellowship, had written a thesis on salicylate intoxication. But when you're in the trenches, rather than the library, thinking under time pressure and conditions of uncertainty, heuristics come to the fore.

How to prevent misdiagnosis has been an endeavor of Dr. Croskerry and others who addressed metacognition. This is not easy in the current medical environment where clinical appointments are being shaved down into minutes and open-ended interviews are being sacrificed for "smart phrases" dropped from electronic records, or patients filling out questionnaires rather than being given time and freedom to talk to their physicians. But metacognition, thinking about how we think, is vital if we're going to prevent misdiagnosis. And the beginning of metacognition is to assimilate the knowledge that Dr. Croskerry presented, and incorporate an understanding of the heuristics that we all necessarily employ, but with an alertness to their benefits and risks, so that we can successfully exploit them and our patients do not suffer from their deficiencies.

Thinking as a Patient

I had two experiences as a patient related to diagnosis. One was catastrophic and the other thankfully not. I will address them in that order.

I was an avid distance runner, even tackling marathons, and in my late 20s, I developed pain in my lower back that radiated into my right leg and big toe. I saw a surgeon, was prescribed NSAIDs, and after 2 months without relief had a simple laminectomy for a "bulging disc" seen on myelogram (CT and MRI were not yet clinically in use). The outcome was fair, still some discomfort, so I took up

swimming instead of running, since the pounding on the pavement seem to exacerbate the pain, despite the surgeon assuring me that the nerve was freed of the disc. About a year and a half after the operation, I stood up from a chair and had excruciating back pain, electric shocks radiating into my buttocks. Over the course of weeks, this didn't relent. It was impossible for me to sit or walk. I consulted orthopedic surgeons and neurosurgeons but didn't receive a clear diagnosis for the disabling back pain. (The myelogram was unrevealing.) I tried to ignore it for a while, swimming a few laps, but that seemed to make it only worse. I became despondent. There had to be a cause. I'd been trained as a scientific physician who believed that there was an etiology to every medical malady and that once that etiology was pinpointed, then a specific rational therapy might remedy it. The idea that there could not be a clear diagnosis was anathema to me [6, 7].

Ultimately, a neurosurgeon and orthopedic surgeon told me that I had "spinal instability" and needed a fusion from the fourth lumbar to the first sacral vertebrae to "stabilize my lower spine," with the promise that this would not only fully relieve my back pain but also return me to full functioning. This was exactly what I wanted to hear: a precise diagnosis, "spinal instability" (spondylolisthesis), and a clear solution (fusion), with a glowing prognosis. And, none of it was true.

I awoke from the surgery in even more pain than preoperatively, unable to move my legs. The surgeons wondered whether I might have hemorrhaged around the nerve roots and said they'd be happy to go in again and explore me. Fortunately, my wife, an endocrinologist with a very different mindset, asserted "No way." So I took painkillers, lay on ice in a body brace, and after several weeks began physical therapy, beginning to walk again using parallel bars submerged in a warm pool. I made slow progress over the course of many months. I was never able to run again but did restart swimming and only many years later greatly benefited from a physical therapy program with a contrarian mindset that overcame my fear avoidance behavior responding to every twinge and muscle spasm [6, 7]. This empirical approach markedly improved my functioning.

So it was with a very personal perspective that I listened to Dr. Croft address the issue of diagnosing low back pain as a primary care physician in the United Kingdom. Importantly, he emphasized ruling out serious conditions that can cause low back pain (I'm aware of misdiagnoses of epidural abscess, e.g., that were written off as muscle strain). But once that important hurdle is scaled, the reality is that we don't understand why, for example, when I stood up from a chair, I was seized with disabling electric shocks in my low back.

As it happened, many years after that event, I was perched on a very low chair, this time sitting Shiva for my mother. A cousin about my age came to offer condolences. He was aware of my clinical history and told me that he had suffered the same attack, but instead of running into the arms of surgeons who promised a fix based on a diagnosis that was, in retrospect, illusory (spinal instability or spondylolisthesis is not a common cause of back pain in an athlete in his 20s), he took some anti-inflammatory medication and stayed in bed for 2 weeks and then slowly mobilized under the guidance of his primary care physician. After a few months, the pain and muscle spasm had largely gone. He had suffered one more attack 2 years later

and sought the same solution, which succeeded. But in contrast to this highly trained academic physician-scientist, who was hell-bent on having a diagnosis, my cousin was satisfied with seeing the low back as a black box and accepted tincture of time, so long as he could receive the kind of reassurance that Dr. Croft articulated, that in the vast majority of patients, the low back pain does not indicate anything serious, that it can be managed conservatively, and that over the course of weeks to months it will largely abate.

To be sure, such words of reassurance are not easily absorbed by a patient who is in pain, vulnerable to illusory solutions like I was. Rather, he is at risk of being misled. So this might be an instance where availability, specifically dramatic stories that are both negative and real, even though they are not presented as statistics with P values, or as graphs, can have a real impact. Indeed, the seeking of narratives vis-à-vis diagnosis was highlighted by a number of speakers. We as human beings are wed to stories. So although I am a research-driven hematologist-oncologist, as a writer, and as a person who suffered this catastrophe, part of which I blame on myself, I've taken it as a mission to inform and educate others through anecdotes and stories: both ad hoc, in conversation, and in the pages of the New Yorker magazine and in my own books. I present my personal history as a cautionary tale. And then I go beyond that, and reference experts in the field, primary care physicians like Richard Deyo, physiatrists like James Rainville, and conservative spine surgeons like Eugene Carragee, but I also highlight the financial reality around spine surgery that can drive diagnoses and decision-making to the detriment of patients [6].

When people hear stories, they can dovetail with the kind of prognostic data that Dr. Croft showed. Indeed, one of the most potent tools that we use in hematology-oncology is to have patients with diagnoses like breast cancer speak to survivors in order to obtain a clear picture of what they're facing and also to gain a measure of true hope. Listening to Dr. Croft, I propose that in the face of the diagnostic uncertainty of low back pain, it would be highly beneficial for physicians, with informed consent, to introduce their patients with the condition one or more patients in their practice who also had no clear etiology but nonetheless opted for conservative measures and had a good outcome. Speaking to a person with acute back pain who is currently frightened, desperate, suffering, and providing him or her with a path to a better future may well lessen the anxiety and impulsive decision-making that can be triggered by our diagnostic ignorance.

I learned a great deal, the hard way, from this experience: to be skeptical of glib diagnoses like "spinal instability" without firm evidence. And so a number of years after the catastrophic spinal fusion, I developed inflammation in my right wrist, so severe it was difficult to write or open a jar. I saw a rheumatologist who ruled out an underlying disorder, although as a number of the speakers emphasized, tests for autoimmune conditions are uncertain from lab to lab. But there was no indication of a destructive arthropathy by imaging, including MRI. In fact, no test was unrevealing. I consulted five prominent hand surgeons and got five different diagnoses. The most terrifying experience, to put it bluntly, was when I saw a surgeon who told me that I had a "hyper-reactive synovium" and that his recommendation was to undergo an operation where he would strip the synovium around my wrist [5]. Not being a

rheumatologist, I went home after this appointment and Googled "hyper-reactive synovium." I couldn't come up with a discrete diagnostic category. It seemed to me he was making something up off the top of his head. This surgeon occupied a prestigious position in his field. Another prominent hand surgeon said he wasn't sure what was wrong but he would cut me open and figure it out in the operating room. That diagnostic uncertainty would only be resolved by exploration under the knife was not satisfactory.

The irony was that the fifth consultation was with a young hand surgeon, who, after reviewing my numerous blood tests and sophisticated imaging scans, asked me to do a very simple maneuver: to have a plain x-ray of my wrist at rest and then to grip a wooden stick while my wrist was again x-rayed. While I exerted force on the stick, it became clear that the scaphoid and lunate bones dislocated, indicating that the ligament was lax. This appeared to be the cause of the inflammation. Ultimately, when the ligament was surgically repaired, small cysts were also excavated and grafts implanted in these bones. Afterwards, I was speaking to yet another hand surgeon who opined that if I had 80–85% return of function, it would be an excellent outcome. And that's what happened.

It was a very long journey before a diagnosis was made. But I lived with the debility and the uncertainty because I had learned from the spinal fusion that that was necessary, the alternative being to put myself at grave risk with a blundering operation. One of my mentors in medical school, Dr. Linda Lewis, a neurologist at Columbia-Presbyterian, instructed her students as follows: "Don't just do something, stand there." Her words seem particularly apt in the face of weighing risk and benefit in the face of diagnostic uncertainty.

The Key Lessons

What was the key lesson I distilled from the conference? Uncertainty emerged as a cardinal theme of the conference. To my mind, it should become a central edifice in the building of a medical education and a key element in mindful clinical practice. Alas, it is natural to avoid or ignore uncertainty. It is a deeply uncomfortable state of being. It can appear to undermine our authority as physicians and seemingly reduce our ability to provide a sense of support and confidence to our patients. But I contend this is a self-serving illusion that ultimately is defeating in the patient-physician relationship. Rather, as several speakers argued, we need to recognize the core of uncertainty that exists in diagnosis; affirm it in our clinical trials, academic publications, and teaching of medical students and residents; highlight it in our clinical conferences; and most importantly develop a truthful language with which we can communicate it to our patients and their loved ones. Uncertainty should become an important element in so-called shared decision-making, where the patient and physician partner in the journey of illness, making choices together around the benefits and risks of diagnostic testing and treatment options. To pretend that uncertainty does not exist is to provide false hope and deny the reality of human biology, our genetics, epigenetics, and the diversity of outcomes for any single disorder.

This will not be a simple or easy goal to achieve, particularly in the current environment of electronic records that dislike uncertain diagnostic categories and where insurers seem allergic to reimbursing for time, effort, and tests without a set diagnosis. Our current system is pernicious in many ways, most of all because it constrains diagnostic uncertainty. It inhibits dialogue needed during the patient-physician journey of illness to consider and reconsider diagnosis based on open-ended history taking; among the research physicians, who need to keep an open and agnostic mind with regard to diagnostic uncertainty, the system can inhibit new ideas. How to resist and reform the current system remains an open question but an important one that the conference will help foster.

So, as an academic physician, I left the conference both with a sense of humility and with a sense of hope that important and complex issues around diagnoses were addressed in a rigorous and diverse way. And as a patient who had suffered from the black box of low back pain, lack of an etiology made me vulnerable and impulsive, putting myself into a situation that caused real harm. I was heartened by the repeated focus on how to address uncertainty in a constructive way, aiming to harness modern science to get answers to etiology and pathogenesis but also openly admitting when we still work in ignorance.

References

1. Groopman J, Golde D. The histiocytic disorders: a pathophysiologic analysis. Ann Intern Med. 1981;94:95–107.
2. Emile JF, Abla O, Fraitag S, et al. Revised classification of histiocytoses and neoplasms of the macrophage-dendritic cell lineages. Blood. 2016;127:2672–81.
3. Gould SJ. Bully for brontosaurus: reflections in natural history. W.W. Norton & Company. 1991 [also see, The Median Isn't the Message]. Am Med Associat J Ethics. 2013;15(1):77–81.
4. To err is human. Building a safer health system. The Institutes of Medicine. Washington DC: National Academy Press; 1999.
5. Groopman J. How doctors think. New York: Houghton Mifflin; 2007.
6. Groopman J. A knife in the Back: is surgery the best approach to chronic back pain. The New Yorker. 8 April, 2002.
7. Groopman J. The anatomy of hope: how people prevail in the face of illness. New York: Random House; 2004.

Clinical Ambiguity in the Intelligent Machine Era (*Treats Breaks and Discharges*)

20

D. Douglas Miller

Introduction

Healthcare Megatrends

The title of a children's book about a panda bear, *Eats, Shoots & Leaves*, is an example of lexical ambiguity – a phrase that is open to different semantic interpretations that can create confusion for both humans and intelligent machines [1]. An analogous title for a book parsing US healthcare woes might be *Treats Breaks and Discharges*. But there is no confusion about the root causes of US healthcare sector administrative cost waste – quasi-market forces fostering perverse business incentives and a profit-driven innovation sector inserting new technologies and infusing new drugs into jurisdictions that can no longer afford to meet basic care or more pressing population health needs [2]. And all this well *before* the COVID-19 pandemic!

Concomitant with US healthcare system bloat is the explosion of information on things being generated in the course of daily life and by patients during episodic medical care – projected to exceed 2300 exabytes of big data in the USA in 2020 alone [3]! Burgeoning healthcare data management systems such as electronic medical records (EMRs), ostensibly created to mitigate unnecessary deaths and avert dangerous medical errors, have created an "epidemiology" of unintended negative consequences [4]. Whether such data are generated monitoring life in the womb at its most precarious, or when its benefit to patients' end-of-life outcomes is more dubious, the capacity of human medical professionals to capture, store, and decode the meaning of modern massively complex information now eludes standard statistics and computing methodologies.

D. D. Miller (✉)
Medical College of Georgia, Augusta, GA, USA
e-mail: ddmiller@augusta.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_20

Democratizing Technologies

Enter artificial intelligence (AI), the hottest high technology trend. It is said that a good AI technology can do what healthy human brains do well in 1 s. Nonetheless, today's AI technologies have been shown to augment human performance in several data-dense business and scientific domains, and they are already deeply embedded in several aspects of healthcare: digital imaging, genomic medicine, health insurance coding and reimbursement, and population health management.

Since 2017, and before COVID-19, the US FDA has approved several diagnostic AI technology applications [5]. Some offer direct-to-consumer heart rhythm monitoring, while others guide health professionals in the image-directed acute care of strokes, pulmonary embolus, etc. The future *promise* of AI for augmenting human capabilities is often touted – for reducing medical errors, for restoring time from EMR tasking for more patient-provider interactions [6], and for democratizing AI's benefits into medically underserved and economically disadvantaged populations [7].

One hard lesson learned from technology insertion is that it is never neutral. So, enthusiasts for AI's potential value to healthcare (i.e., productivity, profitability) must be responsibly balanced by patient advocates (i.e., pro-data privacy, anti-data bias) [8]. And current narrow AI applications enjoin unique healthcare risks when machines train on messy medical datasets or when they are overpromoted as being capable of "outthinking cancer." In their patients' interests and before explaining AI model outputs to their patients, healthcare providers must become AI literate to assure that machine-informed medical decisions consider input data provenance. Failure to do so risks wasting precious resources and offering unethical patient care options.

Ambiguity and Common Sense

Due to the increasing complexity of healthcare systems and despite new technologies, modern medical providers are frequently confronted with clinical ambiguities. For example, ambiguity occurs when a medical decision-maker lacks knowledge about probability distributions required to calculate the expected value of information or alternatively is unable to assert credible subjective assumptions about the distributions [9]. Ambiguity in diagnosis and related care delays cost health systems and jurisdictions money and can contribute to medical errors.

Nuances of natural language and sentence structure that humans intuitively understand through commonsense reasoning pose challenges for AI, as evidenced in the training and performance of IBM Watson on the TV gameshow *Jeopardy!* [10]. Despite Watson's intolerance for lexical ambiguities, its neural networks soundly defeated two human champions; Watson's healthcare natural language processing (NLP) applications have since matured.

Ambiguous word expressions and sensory-cognitive tasks continue to baffle even the most intelligent machines. To the point, several global challenges have

Table 20.1 Capacity of AI to emulate different types of human thinking

Name	Year	Creator	Human capacity (AI technology task required)	1st successful attempt
Turing Test	1950	Alan M. Turing	Observable behavior (deception in brief conversation)	Eugene (Goostman) chatbot, 2014
Grand Challenge	2004	DARPA	Autonomous vehicle driving (deep multimodal perception)	Stanley, Stanford U. Racing Team, 2005
Urban Challenge	2007	DARPA	Autonomous vehicle driving (deep multimodal perception)	Boss, Carnegie Mellon U. Team, 2007
Winograd Schema	2012	Hector J. Levesque	Lexical ambiguity resolution (commonsense reasoning)	BERT EMNLP, Google AI (72% acc.), 2019
Spectrum challenge	2014	DARPA	Autonomous radio (navigating wireless RF obstacles)	GatorWings, U. of Florida, 2019
AI Next Challenge	2018	DARPA	Contextual adaptation (abstraction, reasoning, explaining)	Launched >$2B funding of 50 programs
xML Challenge	2018	FICO	Post hoc explanation of black box models (interpretability)	IBM Research, 2019

been launched to judge whether current AI technologies can reproducibly achieve the capacity to mimic human reasoning and common sense for problem-solving (Table 20.1). The elegant solutions to these challenges are highly germane to AI technology applications for decision-making in complex healthcare contexts. The key to real-world applicability of AI rests squarely on how well data used to train intelligent machines represent reality, with abiding awareness of human biases and full disclosure of data provenance [11].

This paper will (1) demonstrate how medical professionals facing clinical ambiguity with partial knowledge can effectively apply *probability science* to complex medical decision-making, (2) discuss translating *data science* lessons learned from rendering machines intelligent in nonmedical sectors and low-validity data environments into healthcare systems, and (3) project the potential for next wave *cognitive computing science* to emulate human expert intuition and to disambiguate difficult diagnoses in unique clinical settings for individual patients.

Probability Science

Knowledge and Decision-Making

Knowledge is defined as a set of conclusions drawn by humans combining evidence (i.e., data) with specified assumptions about unobserved quantities. Knowledge can be created based on available observational and/or experimental evidence. Knowledge creation sorts complexities in the evidence (data) so that medical providers can make credible assumptions and draw logical conclusions that guide complex care.

Medical knowledge has evolved over millennia, as has the clinical methodology of differential diagnosis, an exercise in clinical reasoning that uses knowledge.

Differential diagnosis requires serial cognitive weighting of logical solutions based on knowledge, in order to distinguish a disease or condition from others with a similar clinical presentation, to then reach a conclusion about the most probable diagnosis [12]. Accumulated human knowledge and clinical reasoning skills improve differential diagnosis. For example, accuracy in chest X-ray interpretation is higher among senior radiology consultants than their more junior colleagues [13]. Consultants' knowledge and reasoning expertise is reflected by more efficient eye movements for radiographic pattern recognition during visual search of images.

Medical decision-making is an established but imperfect application of probability science to the interrelated processes of differential diagnosis, testing, and treatment [14]. It requires rigorous specification of knowledge, the care objective, and available decision criteria. Real-world medical providers frequently interface with partial knowledge, incomplete evidence, and best-guess assumptions to help resolve clinical uncertainty. *Partial knowledge* of a patients' health status and treatment response is a pervasive concern in medical decision-making. Care cannot be optimized when a clinician has only partial knowledge of patient health status and treatment response(s). However, with a specified objective and sufficient knowledge of responses to testing and/or treatment, care can be optimized.

Iteration of pre- and post-testing disease likelihoods and thoughtful recalibration is implicit to medical decision-making. The process is often rendered even more unpredictable due to complex interactions among patient covariates plus previous/ongoing treatments which may impact clinical findings, test results, and disease evolution. Each person or patient has covariates (x_1, x_2, … x_n), and each treatment (t_A, t_B, … t_Z) response may vary with these covariates (i.e., demographics, medical histories, prior health status, results of testing, etc.).

Ambiguity occurs in clinical settings where a medical decision-maker lacks knowledge. More specifically, the decision-maker lacks objective probability distributions required to calculate the expected value of information (i.e., unknown decision-relevant quantities) and is also unable to credibly assert subjective distributions (i.e., assumptions) instead [9]. Ambiguity can also come from imprecision associated in drawing inferences from heterogeneous samples of study populations. On the other hand, *uncertainty* describes clinical settings in which a decision-maker places their subjective distribution on unknowns.

The duration and degree to which a medical diagnosis remains ambiguous (or uncertain) to clinicians depend on several interdependent variables: collective care team knowledge and individual provider acumen, availability of testing and effective use of evidence (data), and underlying disease complexity. These factors are subject to human thought patterning effects from prior clinical training and case experiences, individual and systemic biases, and coexisting occult and overt medical conditions (i.e., "when you hear hoof beats, think horses not zebras").

Decision analysis is a mathematically based approach that can be helpful under the circumstances of ambiguity resulting either from a provider's lack of evidence-based knowledge of probability distributions or from an inability to assume subjective distributions. Decision analysis does not prescribe one best plan but instead shows how a preferred plan depends on knowledge, the care objective, and decision

criteria. Decision analysis problems involve a decision-maker (i.e., planner) who must choose a treatment for each person in a population. The first task in decision analysis is to characterize the knowledge of the planner. Testing yields further evidence (data) on health status which may be useful in decision-making in medical practice. Diagnostic testing may ($s = 1$) or may not ($s = 0$) be ordered by clinicians, and the test result may be normal, abnormal, or indeterminate. Testing may be beneficial, neutral, or harmful to patients. All testing carries some degree of risk, but testing can negatively affect patient welfare (W) if it is invasive and cost-ineffective or creates harmful delays in treatment.

A common difficulty contributing to partial knowledge in medical practice are *identification problems*. These problems are encountered when drawing inferences from observational studies of patient responses to feasible testing and treatment options that are designed to maximize W. Mean welfare (W_δ) across a population of patients is determined by the fraction of those in each covariate group that a clinician assigns to each testing and treatment option.

For each possible value of (s,t), if all patients with covariate x were to receive (s,t), then:

$$W_\delta = E\left[y(s,t) | x \right]$$

If all patients with covariate x and test result r were to receive (s,t), then:

$$W_\delta = E\left[y(s,t) | x,\mathrm{r} \right]$$

Optimizing testing and treatment allocation by clinicians maximizes mean welfare (W_δ).

Treatment is a medical care necessity in most disease states. Clinicians dealing with ambiguity and/or uncertainty often decide between aggressive treatment (i.e., an intervention) and active surveillance (i.e., watchful waiting). Because real persons' treatment responses may be unobservable, the result of observational studies wherein treatment selection is related to a treatment response creates a counterfactual ("Why doesn't the gold standard randomized clinical trial drug result work for my patient?"). *Partial identification* of a treatment response combines realistic available evidence (data) with credible assumptions. Using this approach, response boundaries can be established without reaching precise conclusions. When an optimal treatment choice is impossible, decision theory provides reasonable decision criteria.

Data Science

Knowledge Representation

It has been stated that there is no information without knowledge representation. The knowledge representation (KR) theory of intelligent reasoning is inextricably intertwined with data structures [15]. In order to mimic higher human brain

functions, computer scientists use advanced computing to query datasets and data repositories in order to learn patterns called features that reproducibly discriminate data anomalies from data commonalities. KR has diverse roles and consequences in this query-learn process. Implementing ML requires that KR provide a first-order guess about features of a data structure. In typical ML applications, KR is optimized by using well-labeled datasets and digital image "dev" sets to initially train algorithms.

Basic KR technologies include logic, rules, frames, semantic nets, etc. As such, KR is a *surrogate* for internal reasoning by a person or computing program about an action involving an external real-world operation (e.g., bicycle assembly, brain surgery). The only completely accurate (i.e., high fidelity) KR of an object is that object itself. Any surrogate must have a specified *identity* corresponding to the world, and it must have a degree of *fidelity* in relationship to the real-world operation. Surrogates for abstract notions (i.e., actions, beliefs, causality, etc.) are imperfect but inevitable, as are their *inferences* about the real world. Imperfect inferences can be a source of error, which can be minimized by selecting a good representation (i.e., minimizing the inference as the source of error). In the absence of total accuracy, the goal is to balance sources of error against gains.

A *frame* is a stereotyped situation, a structure from memory, like being in a subway line or going to a child's birthday. Humans encountering a new situation or a substantial change in viewpoint adapt the remembered frame and work to fit it into a new reality by changing details (as necessary).

Intelligent machines embody KRs that are either *sanctions* (what can be inferred; conclusions we are permitted to make; largely unconstrained) or *recommendations* (what should be inferred; conclusions that are appropriate to make or "intelligent"). KR is also a medium of intelligent machine expression and communication (often in language) to humans and to other machines about things in the world. In this context, a KR may find it easy or difficult to communicate about useful things (i.e., pragmatic utility) and also have the capacity to be misrepresented or misunderstood (i.e., unintelligibility).

Deconvoluting Data Complexity

Before undertaking AI analytics of high-dimensional datasets (i.e., data dimensions [p] $\gg$ data types [n]), complex data decomposition is often necessary. Principal component analysis (PCA) is a family of preprocessing mathematics that reduces high data dimensionality while best approximating the relevant information in the original raw data matrix (i.e., data manifold) [16]. The covariance among data elements in a data matrix can be represented by vectors (i.e., eigenvectors). Classical PCA seeks solutions to the *eigenstructure* of a covariance matrix; it can be used to solve the problem for relatively clean and Gaussian distributed data (Fig. 20.1).

Robust and dynamic PCA approaches can be applied to solve for best in increasingly complex, messy, and non-Gaussian datasets that arrive sequentially over time in batches that are correlated (i.e., time series), are distributed and stored in multiple

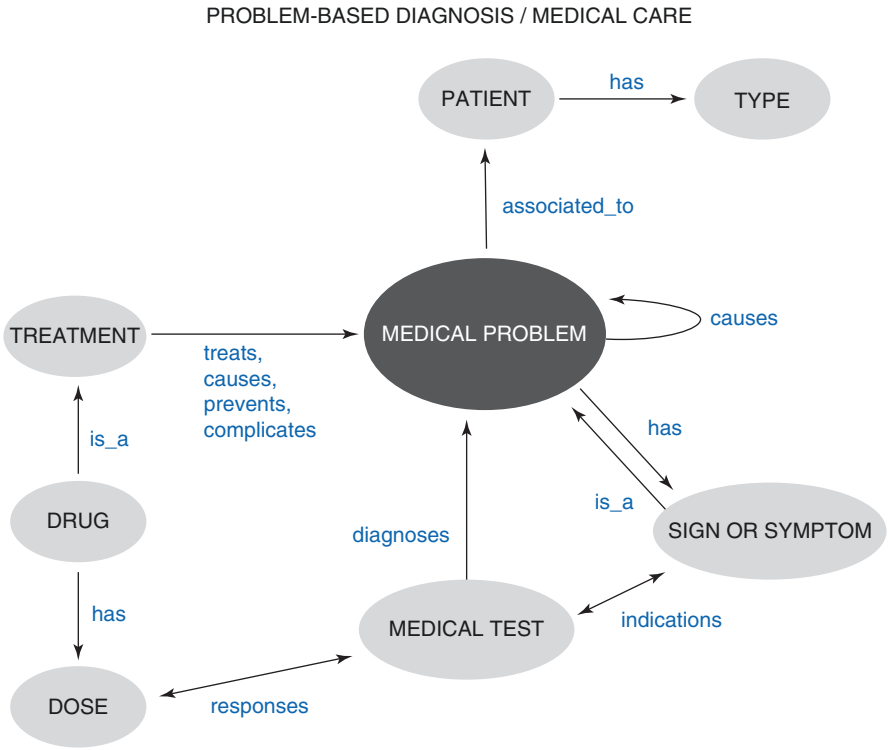

Fig. 20.1 A medical question answering system using NLP technology in a patient with one acute medical problem

locations (i.e., distributed networks), and are comprised of data types that are better represented as a tensor (versus a geometric vector).

By virtue of its capacity to detect sparse outliers within observed digital image data, robust PCA data dimensionality reduction capabilities have been widely applied in the field of computer vision (i.e., facial recognition). Robust PCA has also been validated in digital medical imaging for region-of-interest detection and tracking from under-sampled dynamic (in space and time) 4D magnetic resonance (MR) imaging sequences [17].

Computing Science

Making Machines Intelligent

AI relies on mathematical algorithms rapidly computing solutions in a nodal architecture called a neural network (NN) [18]. Connectivity between NN nodes (i.e., synaptic efficiency) is expressed as weighting patterns (ω_1, ω_2,… ω_x) that can be stored and replicated – so-called features. Once the NN is trained on existing

datasets, otherwise opaque features can be reliably rediscovered in newly presented datasets (i.e., a classifier). Computing efficiency and modeling accuracy for some tasks have gradually improved to levels exceeding that achievable by humans. But by design, the capacity of AI analytics to accurately find a feature never reaches 100%, so that NNs trained on a dataset can be generalizable for learning on new testing datasets.

NNs can also model mathematical functions (f) learned from a dataset (X) to predict a future eventuality (Y) in subsequent data queries (i.e., predictors). Mathematically stated, ML approximates a target function (f) that maps input variables (X) in order to predict an output variable (Y):

$$Y = f(X)$$

As is the case in human brains, intelligent machine knowledge is distributed over a large number of functional units; there is no special analytic subunit for detecting one type of feature. Overall machine predictive accuracy is often compared to human experts' performance using receiver operator characteristic (ROC) curves [19].

The ML analogy to human diagnostic ambiguity is either a low accuracy or irreproducible predictive model that results from KR of data structures that are poorly characterized or of dubious quality (note: machines cannot credibly assert subjective assumptions). The ML analogy to diagnostic uncertainty is predictive model training using data of unclear provenance.

Reasoning and Abstraction

While current first- and second-wave AI technologies have been influenced by cognitive computing advances, they do not possess the human capacity for intelligent reasoning and abstraction. Notwithstanding this, the future potential for AI to reduce human reasoning errors is worthy of consideration.

Human behavior can be influential on intelligent reasoning. The paragon of next wave AI medical technologies would be to adapt to different environments (i.e., contexts) and to augment medical providers' capacity for informed and well-reasoned real-time clinical decision-making behaviors.

Reliability and Transparency

The strength of current diagnostic AI technologies is that they are exceptionally fast and reliable (when compared to human experts) in detecting feature characteristics of common lesions (i.e., cancers, fractures, bleeding, vasculopathy, arrhythmias) on digital medical images (of brains, lungs, breasts, hearts, etc.) such as pathology slides (tumors), medical photographs (skin, retina), and signal sensor tracings

(EKGs, vital signs). However, failure to fully understand ML limitations and/or inability to justify its recommendations to patients without knowing data quality and data provenance is unethical [11].

A further lack of transparency and accountability for some ML predictions – so-called black boxes – reflects these models being too complicated for most humans to comprehend [20]. Whether clinicians are AI savvy or not, they should view such black box ML predictive modeling with appropriate skepticism [21]. However, if the prediction step in AI modeling is separated from the recommendation engine, then actions can be more safely placed into proper context. As is the case in the business, science, and legal sectors, blindly accepting black box modeling can directly contribute to poor medical decisions and to the inefficient use of scarce healthcare resources. Creating new models to mitigate what is happening inside a black box (*explainable* ML or xML) also has inherent risks.

The Digital Image Advantage

While AI is challenged by many things (*q.v.*), it excels at reading digital images because its algorithms are readily trained to accurately recognize the binary features (i.e., edges, shapes, contrast, etc.) of pixels in order to differentiate something that is normal from something that is abnormal. Imaging algorithm training requires well-annotated datasets – of CT scans, MRIs, chest X-rays, retinal images, etc. – without and with the features of actual common medical conditions [17]. Such curated image sets ("dev" sets) are far from ubiquitously available, and their selection introduces the potential for image sampling bias (i.e., underrepresentation of minority cohorts, rare diseases and clinical variants, etc.) that may limit AI applicability in the field.

Conventional ML applications are well established in several digital image-based diagnostic medical fields such as radiology and microscopy – for disease classification, phenotyping, segmentation, and change tracking over time [22]. Computer scientists and clinicians can now apply advanced DL computing on larger and more complex datasets, determining cardiovascular risk factors from retinal photography, converting low-resolution or sparse images to super-resolution microscopy, reading word strings (EMR's, published medical evidence, etc.), and deconvoluting complex matrices (GWAS, immunoassays, proteomics, etc.).

Generative adversarial networks (GANs) were initially developed to test the robustness of discriminative ML and DL algorithms, by deeply characterizing the features in digital images of common objects (i.e., animals, flowers, vehicles, etc.) [23]. GANs can fool discriminator NN's into thinking that a picture with the features of a panda is actually a picture of a gibbon. Virtual patient files created by GANs are used in virtual clinical trials. GANs are also capable of converting real patient MR images to virtual CT images, potentially reducing the radiation burden associated with CT attenuation correction during radiotherapy [17].

Cognitive Human-Intelligent Machine Parallels

Confidence

Everyday life requires that humans acquire and apply an immense amount of knowledge about the world. Both thinking humans and learning machines are designed to *deconvolute* complexity by asking well-informed questions, adding and deleting data in fact arrays, ranking serial thoughts or weighting algorithm performance, and reiterating this process in order to build confidence in either candidate answers or predictive models. The fact that not all data are valid and not all facts prove true adds complexity and reduces confidence.

Whether considering a knowledgeable human's best judgment call or a trained machine's best fit model, intrinsically adaptive learning behaviors are in play. And because much of human working knowledge is subjective and/or intuitive, machine perceptions about information may vary and be difficult to articulate [24].

Information Processing

At any one time, human capacity for consciously processing information is limited to a total of seven information "chunks" (i.e., letters, digits, words) plus or minus two separate bits of information [25, 26]. Human short-term working memory storage capacity is a recall limit of four verbal chunks (i.e., idioms, short sentences), plus or minus one. While these limits vary among individuals and under different external conditions, information processing and working memory ability are critical to human completion of cognitive/mental tasks (i.e., language comprehension, problem-solving, and planning). The brain has multiple storage-specific mechanisms for retaining information, including chunking and other memory rubrics.

For example, language comprehension is a mental task that requires humans to retain ideas from early in a sentence to be combined with ideas later in the sentence. To comprehend an essay, a human must concurrently hold in mind the major premise, the point made in the previous paragraph, a fact, and an opinion presented in the current paragraph – all integrated into a single chunk that permits comprehension before continued reading. Human cognitive processes such as differential diagnosis are influenced by case memories, cumulative experiences, and biases related to knowledge gaps and prior outlier experiences. The AI technology equivalents to text phrase and essay comprehension are natural language processing (NLP) and long short-term memory (LSTM) units, respectively.

Current supervised or unsupervised ML and DL are statistically impressive, reflective of a more processing (compute-intensive) than storage (memory-requisite) technologies. AI is now feasible and more useful to humans due to affordable ultra-fast parallel computing by graphic processing units (GPUs). In 2000, the NVIDIA GeForce 2 chip could compute <50 billion floating-point operations per second (FLOPS). Five years ago, the NVIDIA GTX 1080 card could compute 9 trillion

FLOPS. Today's fastest GPUs can compute >110 trillion FLOPS! This compute power provides real-time processing advantages for reducing data high dimensionality and/or solving dynamic time series data problems.

Learning

Humans are taught *deductive* learning as a means to an understanding of specific concepts from general rules applied to information (i.e., fact arrays). Humans can program intelligent machines for *inductive* learning designed to reflect human problem-solving and emulate deductive reasoning [27]. Striking similarities exist between a machine learning insights derived through algorithm training and professionals developing critical thinking skills during their education and training. For both humans and machines, repeated encounters with varied situations and co-actor behaviors can enable learning.

However, computers and humans work differently with the same body of facts. *Formal* tasks based on fact arrays that are among the most mentally challenging for humans (i.e., timed competitive chess matches) are often the easiest tasks for computers to learn. *Informal* tasks are a struggle for machines – using knowledge to solve actions easy for humans to perform but hard for humans to formally describe [28]. Informal tasks include daily activities that rarely have a clear beginning or end, are likely to be interrupted, are concurrent with other activities, and often need to associate various models and types of information.

Bias

Scientific biases disrupt human cognition and attribution biases disrupt ML. EMR data, health insurance claims, digital device readings, etc. are often generated as a "downstream" consequence of human decisions, which have implicit flaws and biases that are magnified as care becomes more complex and/or compressed (i.e., in higher cognitive load situations). EMR data are potential sources of intrinsic and unrecognized biases in healthcare AI applications. If AI algorithms use EMR or other data generated through a biased process, then the output will reflect that bias [29]. Such AI biases could theoretically be reduced using relatively uniform data sources (i.e., operative vital sign data, emergency department triage data, etc.) that are "upstream" from contaminating clinician judgments and biased human decisions appearing in the medical record as natural language text streams.

Human system-1 thinking failures (i.e., errors due to cognitive biases) decrease with greater knowledge and experience [30]. Both humans and machines can learn from their mistakes resulting from cognitive biases. GANs (q.v.) could also be used to develop more effective system-1 thinking among less experienced medical trainees, exposing them to diverse subtly different patterns of illness presentations (i.e., phenotypes) and to clinical and subclinical disease variants that should signal a fluid change in differential diagnosis thinking [31].

Human system-2 thinking failures (i.e., errors resulting from biases arising from working memory limitations) could theoretically be reduced through the use of AI natural language processing (NLP) technologies that "read" and memorize the peer-reviewed biomedical literature (the evidence) or to rapidly condense years of complex EMR patient information (big data) into organized fact arrays. The human capacity to continuously identify and rapidly reconcile information pattern/feature inconsistencies (i.e., reasoning) greatly enhances human cognitive abilities, exceeding that possible with current NLP technology [32].

Context

Contexts are the shared knowledge domains and environmental settings in which machines and humans interact [33]. AI technologies interdependently affect humans (and vice versa), ideally in a fashion that augments each other's performance. Environmental variables, such as emotional or physical stress, can significantly bias and distort human memory and learning, rendering decision-making sub-optimal [34, 35]. And in the process of validating autonomous machines to perform, AI scientists must improve machines to operate in unfamiliar environments and to function in the face of unanticipated events. Such intelligent machine engineering to confront contexts is not dissimilar from the process of medical learner education for clinical practice.

Highly trained but imperfect human providers and healthcare teams must operate care models and make business decisions in rapidly evolving contexts (i.e., new drugs and medical technologies, new and unevenly applied health policies, etc.). Intelligent machines designed to explain contexts to intuitive humans must employ reasoning, inferences, and/or causality to enhance decision-making.

Intuition and Expertise

A prime characteristic of human cognitive ability, intuition, enables decision-making despite incomplete knowledge. Human intuition is reliant on attention and comprehension of one's surroundings and on the ability to remember events and identify patterns from the environment.

The flaws in human intuitive judgment resulting from heuristics and biases (HB) and the potential for overreliance on expert intuition in naturalistic decision-making (NDM) have been extensively studied [36]. NDM research originally focused on master chess players, who were found to have the capacity to quickly recognize and execute moves based on complex pattern recognition from a repertoire of 50,000 to 100,000 promising lines of game play. In NDM theory, intuition is defined as recognition of patterns stored in memory. Recognition-primed decision (RPD) strategy also takes advantage of tacit knowledge that may be difficult for experts to articulate

and of keen situational awareness in settings such as a building fire, a military action, or a nuclear power plant emergency.

HB is the countervailing school of thought, presenting a skeptical attitude towards expert judgment and expertise. Early research on clinicians' diagnostic judgments showed that simple statistical models and algorithms always performed better due to human informal judgment inconsistency. The capacity for human experts to confidently hold a subjective conviction of their understanding of a clinical case in contextual isolation has been called an "illusion of validity." Even among sophisticated scientists and statisticians, inferior choices are made when relying on intuition over rule-based methodology. It has been shown that humans perform much more poorly than simple algorithms in low-validity environments, such as chaotic or complex clinical situations, when simple cues are often missed. Even in controlled environments, such as personal loan approvals, algorithms have essentially replaced humans; objective performance data trumps subjective impressions of reliability and eliminates human biases from race, gender, etc.

AI exhibits the capacity for unbiased observation (i.e., information processing), persistent cognizance of contexts (situational awareness), and pattern identification between different occurrences (pattern recognition and matching). These very capacities allowed DeepMind computer scientists to train AI to repeatedly win complex games like Go over human champions. One company (Node) has taught an AI to replicate human hunches with generative predictions based on specific past use-case outcomes, but truly generalizable *artificial intuition* awaits.

Limitations of Current Healthcare AI Applications

Intelligibility

Medical providers must be able to understand, validate, edit, and trust an AI model in order to deliver care. But there is often a trade-off between ML model accuracy and intelligibility [37]. The most accurate ML models (i.e., DL, SVMs, boosted trees, etc.) are often not very intelligible to humans, while the most intelligible models (i.e., linear or logistic regression) are often less accurate. This trade-off often limits the accuracy of ML models that can be applied in critical applications such as healthcare. However, reasonable model-to-decider transparency has been achieved for chronic diseases (i.e., diabetes), in critical care (i.e., acute pneumonia), for 30-day hospital readmission risk prediction, etc.

AI can model unexpected insights from complex datasets, extracting features well beyond the capacity of expert clinicians. For example, DL predicted individual nocturnal hypoglycemia with 84% accuracy in a heterogeneous pilot study cohort using a few heartbeats of raw EKG signal recording [38]. By helping medical providers to visualize the key EKG data warning signature, this application overcame the AI intelligibility problem.

Reproducibility

While high technologies like AI have the potential to improve the safety, quality, and ease of care by clinicians, their theoretical benefits have been difficult to reproducibly demonstrate in real-world clinical settings. Whether using commercially available ML tools, or using advanced GANs to create virtual CT scans from actual MR images [17], or using DL for microscopy image reconstruction [22], the potential for introducing artifacts not present in the original training data into AI models is real and of great concern to researchers and clinicians.

Emerging AI technologies could automate work flow processes and thereby enhance decision-making reproducibility in the diagnostic setting, even in contexts that require professional medical judgments. However, there is little research to support this claim. Prior research on new technology insertion into diverse organizations reveals a gap between expectations in theory and applications in practice. When commercially available AI tools were used in the radiology department of a major US academic health center, the decision-making of expert radiologists was measurably slowed [39]. The AI applications introduced an additional (often conflicting) source of opaque diagnostic information that radiologists often later overread with their own findings. The introduction of further diagnostic ambiguity in this setting was related to the interface between a new technology (i.e., AI) and established professional judgment processes (i.e., clinical practice). These fast computing AI tools rendered previously routine professional decision-making tasks nonroutine, adding time-to-diagnosis and creating additional disambiguation work for expert radiologists while likely increasing the costs of care. The potential benefit of AI-augmented diagnostic accuracy and/or reproducibility was mitigated by adding ambiguity to the professional workplace.

In 2018, Google's Automated Retinal Disease Assessment (ARDA) tool was US FDA approved and European Union CE marked for diabetic retinopathy detection. When the tool was subsequently field tested in rural India, where there is a high burden of retinopathy but few ophthalmology providers, DL inferences developed by using hi-res retinal photographic training data could not be readily translated into the undeveloped world due to poorer retinal image quality [40].

Data Quality

The greatest vulnerability of AI applications in healthcare is the quality of the data being entered (largely by humans) into ubiquitous but largely non-interoperable digital data platforms – EMRs and health administrative databases [3, 11]. As noted, while EMRs are complicated information management systems requiring millions of lines of code written by many individual spanning numerous legacy platforms, they are *not* high technologies per se [4]. EMR datasets, originally intended to support health billing functions, are notoriously redundant and holey (i.e., "messy"). EMR alerts and prompts can cause humans to experience automation complacency and contribute to questionable clinical decision-making and/or incorrect treatment

response predictions. EMRs evolved from their humble origins in a sociotechnical context that predated the handheld device/social media era, before the emergence of truly high technologies such as AI, blockchain, internet of things (IoT) devices, 5G enhanced mobile broadband (eMBB), next-generation DNA sequencing (NGS), etc.

Performing AI analytics on messy EMR datasets in the absence of understanding the clinical context predisposes artifacts and biases in predictive models of acute hospitalization outcomes and chronic disease trajectories. For example, the EMR at one leading US academic health center had less <1/5th of hospital progress notes entered manually by clinicians and saw ~50% of medical student, resident, and hospitalist inpatient notes copied and pasted from prior entries [41]. While this common practice makes working with EMRs less onerous, it could perpetuate errors and trigger medical decision-making uncertainties and ambiguities that can permanently contaminate patients' digital healthcare records.

AI application makers promise system-wide solutions for more cost-effective hospital care. Several companies (i.e., Symphony Ayasdi, IBM Watson, Jvion, Medial EarlySign, Pillo Health, Splunk, etc.) are using topological data analysis (TPA) to map persistent homology patterns in patient datasets, to then uncover features with ML. Using TPA and other technologies developed in other sectors, AI tools can query diverse high-dimensional big data sources (i.e., EMRs, corporate performance management (CPM) analytics software platforms, enterprise data warehouses, financial systems), generating unique treatment group clusters for high-cost/high-mortality medical conditions (i.e., pneumonia, sepsis, etc.) [42]. AI tools designed for decision support and clinical care variation management have also been applied to SQL database queries in order to produce customized daily order sets for patients on care paths from the emergency room to admission and through to discharge. But most busy clinicians are not afforded the benefits of data certainty or context clarity in the hurly-burly of EMR-heavy daily medical practice. Training data quality, sourcing bias, and other indurate data problems render the potential for AI technologies to disambiguate clinical uncertainties from EMR's largely aspirational [43].

Lessons Humans Can Learn

From Autonomous Driving Vehicles

Since the first DARPA Urban Challenge in 2007 (see Table 20.1), designing autonomous driving vehicles has tested the limits of AI technologies. Human autonomous system interfaces require a new technology paradigm in which both humans and machines must co-adapt to highly dynamic shared contexts and to each other [44]. In these settings, both are subject to *context variability* influences. Assigning faults or failures to individual humans, autonomous systems, or both demands a fulsome understanding of the complex confluence of *cause and effect* relationships. This impressive but imperfect AI technology has contributed to at least seven human deaths.

The current capacity of on-board and distant electronic multimodal sensors (i.e., camera RGB, LiDAR, radar, ultrasonic, thermal, etc.) to capture and preprocess

single image frame data from both the autonomous vehicle and the dynamic surrounding environment is remarkable. Autonomous vehicle designers are challenged to decide what, when, and how to "fuse" the complex and dynamic information encountered via these multimodal sensors. GPUs and CNNs must robustly extract, classify, and localize features in real time from image frames acquired under varied road and lighting conditions and from other moving vehicles, pedestrians, and cyclists. Explicitly modeling (and propagating) of the uncertainties or the informativeness of each sensing modality is important to the safety of driving an autonomous vehicle. A multimodal object detection network should ideally produce reliable prediction probabilities for object classification (tree or cat) and localization (near and far) [45].

Designers test their digital perception computer vision systems against large-scale realistic annotated image datasets, like the KITTI vision benchmark suite, to reduce uncertainties and biases [46]. But even state-of-the-art (SOTA) visual recognition algorithms used in autonomous driving platforms that rate highly when working on established datasets in the laboratory perform *below* average in the real world. Inference speed variances among fusion network hardware configurations and programming languages are usually self-reported by designers. Without common benchmarks or evaluation metrics, network-to-network predictive uncertainty probabilities remain hard to compare.

Reliability of a multi-model object detection network's uncertainty estimation reflects *robustness*. Robust networks depict higher uncertainty for camera signals acquired during nighttime or adverse weather driving conditions (i.e., "open-world" problems). The reported robustness of sensing modality fusion is generally achieved by a single operation (i.e., addition and an average mean). Newer Bayesian neural network (BNN) approaches to estimating uncertainty assume a prior distribution over the network weights and infer the posterior distribution of weights to extract a better prediction probability.

From Natural Languages and Semantics

Although humans constantly interface with complex dynamic contexts, they most commonly do so though natural language, not multimodal sensors. Natural languages are any language, written or spoken, that has evolved through human use and repetition without conscious planning or premeditation. If humans are confused by natural language usage (i.e., lexical ambiguity), then intelligent machines will also be confused. For example, the human capacity to appreciate sarcasm and nuance in language is lost on machines. Humans and machines share many other natural language ambiguities that limit thinking (cognition) and learning (feature recognition).

Collative semantics (CS) embodies the related concepts of coherence, semantic networks and relations, metonymies, and lexical ambiguity in natural language [47]. CS principles and processes guiding human knowledge creation from and knowledge representation in natural language are applied by its AI technology counterpart – natural language processing (NLP).

Semantic networks are knowledge best understood by humans as a set of related concepts. These networks are cognitively based and consist of arcs and nodes

organized into a taxonomic hierarchy. *Genus* is the name of a (data) class that includes subordinates called *species*. *Differentia* are the properties by which a species is distinguished from another species in the same genus. A vehicle (genus) that carries passengers has differentia from a species like motorbikes. A semantic network (neural or linguistic) is a taxonomy of genus and species in which the nodes are the terms and the links are the arcs between nodes (Fig. 20.3).

Semantic relations are representations of general conceptual relatedness, or coherence. In speech, sematic networks are a common kind of knowledge representation. *Semantic tropes* are computing search engines and analytic software used for chronological grouping of text passages into verbs, adjectives, pronouns, etc. *Metonymies* are the substitution of the name of an attribute for the thing meant, such as a "suit" being substituted for a business executive.

The four CS components used in NLP are sense frames (knowledge representation of word senses), collations (mapping word senses), semantic vectors (mapping with scoring metrics), and screening (rank ordering of semantic relations). These NLP functions quantify and spatially relate the content and meanings within natural language passages.

Coherence is defined as the synergism of knowledge. *Synergism* is defined as the interaction of two discrete agencies to achieve an effect of which none are capable individually. In semantic networks, coherence is based on *inclusion* and *distance*. The small distance between "vehicle" and "car" represents an inclusion (close conceptual relationship). A larger distance between "animal" and "car" is an exclusion (conceptually unrelated). Coherence is the path with the shortest distance between two nodes; numerically it is a measure of conceptual similarity.

From Resolving Lexical Ambiguity

Many words in natural language have a number of possible meanings or word senses (a "bank" of meanings or senses) [48]. The closeness of concepts reflected by inclusion and distance (q.v.) can be distinguished in the process of lexical ambiguity resolution ("lexalytics"). In fact, NLP requires lexical *disambiguation* to determine the exact meaning of a word sense. Resolving words that are spelled the same but have different meanings (i.e., polysemies) requires that NLP consider context.

Non-textual image (visual) information can be an orthogonal source of information for disambiguating word senses [49]. AI models can use a training set of images with associated text (art text-based word sense disambiguation algorithms) to predict word senses related to images from a constrained set of choices. Annotation of a corpus image set (i.e., ImCor, which links images to disambiguated text) can label regions with the most probable word while consider all the choices relevant to the entire image.

Region labeling improves by restricting predicted words to a smaller number of choices such as those known to be in the image caption. An algorithm can "read" a block of text, then answer whether is it related to a matching image, and select the text segments upon which they based their (un-)related classification. Image information can be sufficiently independent from textual-based cues that combining the two sources of information can prove incremental.

From Common Sense Challenges

I can't cut that tree down with that axe. It is too small.

Ambiguity in natural language confuses AI technologies [50]. Since 2011, the annual Winograd Schema Challenge (WSC, Table 20.1) has seen AI research groups from Microsoft, Facebook, the Allen Institute, etc. try to reproduce an *observable behavior* of humans that reflects machines approaching humanlike intelligence. WSC poses a set of multiple choice questions in a particular format of paired sentences that differ only in one or two words and that contain a referential ambiguity that is resolved in opposite directions in the two sentences:

I. The trophy would not fit into the brown suitcase because it was too **big** (*small*). What was too **big** (*small*)?
   Answer 0: The trophy
   Answer 1: The suitcase
II. The town councilors refused to give the demonstrators a permit because they **feared** (*advocated*) violence. Who **feared** (*advocated*) violence?
   Answer 0: The town councilors
   Answer 1: The demonstrators

Common sensed humans solve the two questions readily using their knowledge about the size of objects and spatial reasoning (Question I) and their knowledge about the typical behaviors of demonstrators and interpersonal reasoning (Question II). In 2018, Google AI Language researchers published BERT, the open-source code for empirical methods in natural language processing (EMNLP) AI that achieved 72% SOTA accuracy when tested on WSC problems [51]. BERT dramatically reduced the typically large amount of annotated data required for advanced DL pre-training and reduced training time on a Cloud tensor processing unit (TPU) to 30 minutes and on a single GPU to a few hours. This advance in deeply bidirectional unsupervised language representation (in a multilayered CNN) significantly improved NLP efficiency for question answering and sentiment analysis.

Difficult Questions and Answers

Q: "How should the clinician choose?"

Clinicians frequently estimate the pre-test probability of a patient diagnosis or outcome and use the results of testing to generate a post-test probability. However, more sophisticated medical decision-making relies on accumulated evidence (data) and credible assumptions in order to make estimates of reasonable conclusions that progressively guide patient care. The validity of these estimates and conclusions is usually limited by partial knowledge about the patient's health status and by partial identification of the patient's response to treatment. For both skilled clinicians and intelligent machines, the common reliance on partial knowledge or uncertain information is fraught, limiting both the human confidence and the AI validation necessary for high performance in more complex systems and diverse operational contexts. A lack of knowledge and/or credible assumptions forces diagnostic testing

and treatments to frequently occur under conditions of clinical ambiguity. This and other identification problems render no single unambiguously correct answer to the above question.

A: "I understand why this occurred, and why it did not. I understand why a treatment will succeed, and fail. I know when to trust your care, and that you will make a medical error."

While current narrow AI technologies – handcrafted knowledge and statistical learning – can augment human clinical expert performance using large amounts of high-quality training data, they do not adapt well to changing conditions (contexts) and are often unable to provide human users with clear explanations of their results (intelligibility). In short, narrow AI lacks the capacity to rapidly recognize and adapt to new situations and environments and humanlike communication and reasoning capabilities. Next wave AI promises contextual adaptation and reasoning to model lucid explanations for its recommendations (see the above machine quote), providing clinicians with more than just decision support [32]. These future AI genre *contextual adaptation* systems will learn more from data and better perceive contextual cues, constructing abstraction and reasoning models to explain real-world phenomena in real time (Figs. 20.1, 20.2, and 20.3).

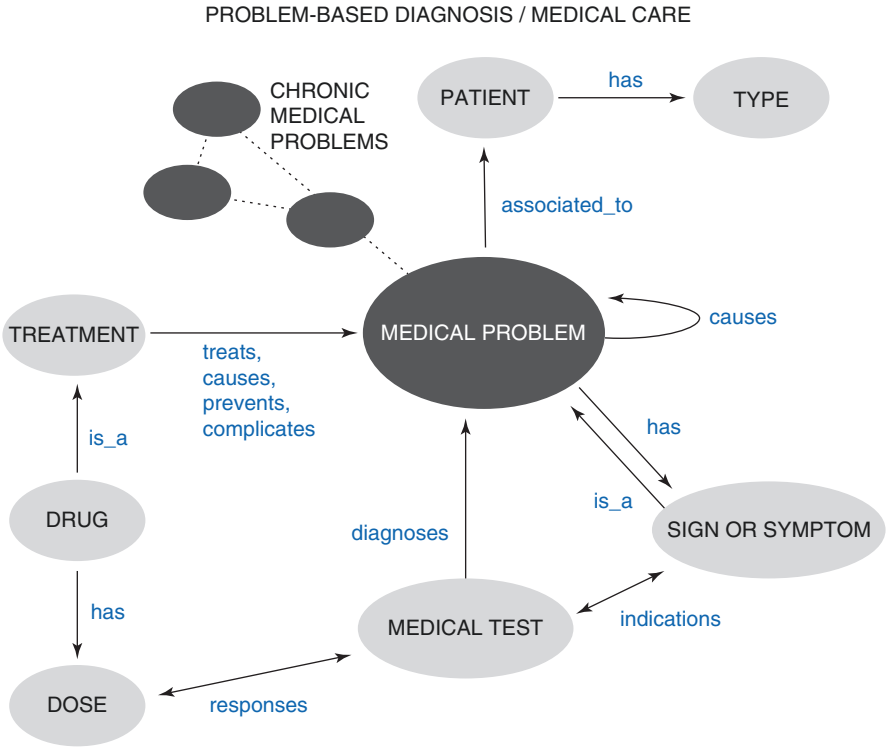

Fig. 20.2 A medical question answering system using NLP technology in a patient with one acute and multiple chronic medical problems

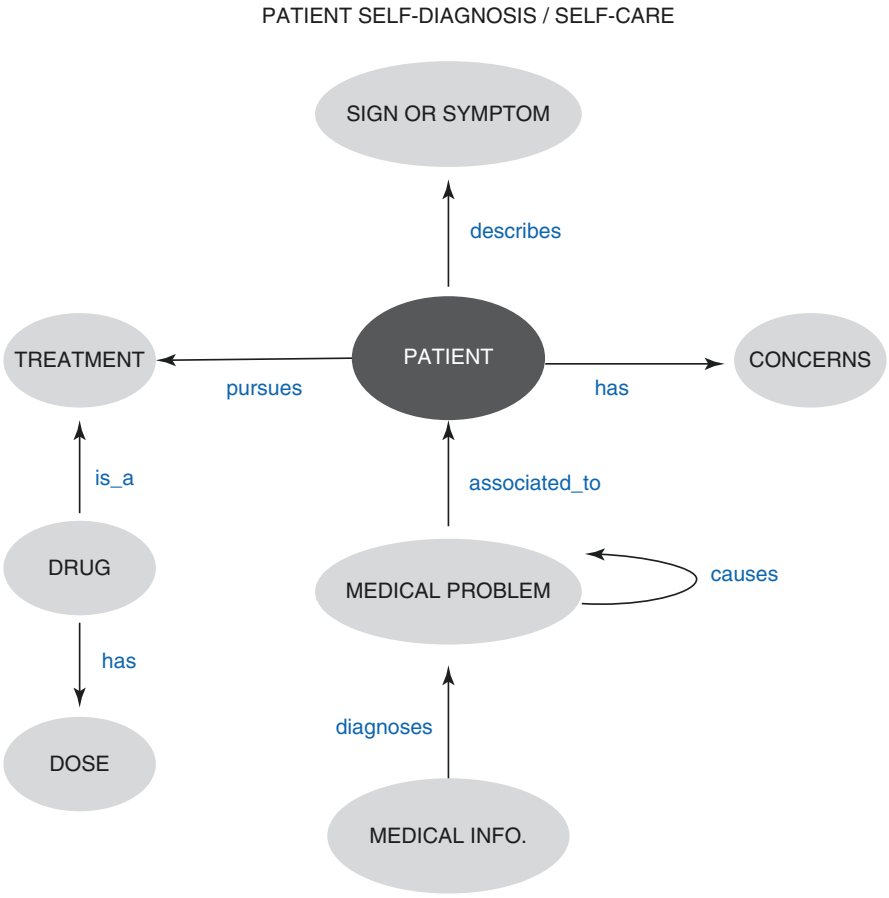

Fig. 20.3 A medical question answering system using NLP technology in a patient pursuing self-diagnosis and/or self-care. The semantic network of self-diagnosis and/or self-care processes differs and is less complex than that involving a healthcare professional

Conclusions

Modern medicine is a deep digital information ocean, teeming with messy data and contextual pollutants, and awash with low-tech data management system inconsistencies (EMRs). Big data are continuously being generated, like waves, from standalone sites and integrated healthcare systems, all of which are being operated by biased humans with varied levels of confidence in their digital information platforms and data management skills. In this setting, the most insidiously dangerous riptide of AI is an overreliance on big datasets for black box predictive modeling.

Trade-offs are implicit whenever a powerful technology (like AI) is inserted into an imperfect system (like healthcare). AI modeling remains highly compute-intensive (requiring iterative trial and error), individually unreliable, and subject to bias from human inputs and from skewed or messy training data. Despite this,

decision support AI is beginning to address healthcare *system* operational complexity, potentially eliminating expensive unnecessary tests, reducing inpatient length of stay, and improving patient outcomes – all key goals of value-based healthcare. However, today's AI applications remain incapable of helping clinicians to disambiguate highly complex *individual* patient diagnostic dilemmas in the real-time clinical moment and may actually make some routine clinical work flows worse.

Deep human knowledge of computing science derived from other AI domains such as autonomous driving vehicles and natural language processing can greatly inform future healthcare AI applications. In order to represent the best interest of both systems of care and individual patients, from the ethical and resource stewardship perspectives, it is important to understand AI's technological underpinnings and core weaknesses *before* attempting to force an AI square peg fix into every healthcare round hole problem.

Primary Considerations When Applying AI Technologies in Healthcare

Concerns about AI

- Technology insertion is never neutral; it often has unintended consequences.
- Adopting "black box" AI models based on suspect data quality or provenance can worsen clinical ambiguities and add to healthcare inefficiencies.
- Today's AI applications remain incapable of disambiguating highly complex *individual* patient diagnostic dilemmas in the real-time clinical moment and may actually make some routine clinical work flows worse.

Using AI

- The challenges of complexity and clinical ambiguity can be addressed by attention to probability, data, and cognitive computing.
- AI technologies may help doctors to disambiguate complex individual patient diagnoses in real time, thereby improving clinical reasoning, mitigating biases, and explaining (or even averting) medical errors.
- Solutions from other data-dense/context-uncertain domains (like autonomous driving vehicles) may be salient to healthcare, where the probability of flawed human reasoning and type 1/2 bias is high.

Duties of the clinician

- Providers must become "AI literate" to assure that machine-informed medical decisions consider input data provenance.
- Providers must explain "black box AI" model outputs to their patients to mitigate wasting resources and offering unethical patient care options.
- Providers must understand AI's technological underpinnings and core weaknesses before assuming that an AI-based approach is correct.

References

1. Truss L. Eats, shoots and leaves. London: Profile Books; 2003.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44–56.
3. Miller DD. The big health data – intelligent machine paradox. Am J Med. 2018;131(11):1272–5.
4. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in healthcare; the nature of patient care information systems-related errors. JAMA. 2004;11(2):104–12.
5. Miller DD. Machine intelligence in cardiovascular medicine. Cardiol Rev. March/April 2020;28(2):54–63.
6. Topol E. Preparing the healthcare workforce to deliver the digital future. National Health Service: London; 2019. https://topol.hee.nhs.uk/
7. Bleicher A. AI will give your doctor superpowers. UCSF Mag. January 2, 2020. https://www.ucsf.edu/magazine/ai-doctor
8. Matheny ME, Whicher D, Thadaney S. Artificial intelligence in health care. A report from the National Academy of Medicine. JAMA. 2020;323(6):509–10.
9. Manski CF. Diagnostic testing and treatment under ambiguity: using decision analysis to inform clinical practice. PNAS (USA). 2013;110(6):2064–9.
10. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? Am J Med. 2018;131:129–32.
11. Miller DD. The medical AI insurgency: what physicians must know about data to practice with intelligent machines. npj Digit Med (Nat). 2019;2:62.
12. Richardson WS, Glasziou P, Polashenski WA, Wilson MC. Evidence about differential diagnosis. BMJ Evid Based Med. 2000;5(6):164–5.
13. Kelly BS, Rainford LA, Darcy SP, Kavanagh EC, Toomey RJ. The development of expertise in radiology: in chest radiograph interpretation, "expert" search pattern may pre-date "expert" levels of diagnostic accuracy for pneumothorax identification. Radiology. 2019;280(1) https://pubs.rsna.org/doi/full/10.1148/radiol.2016150409
14. Bean MA. Probability: the science of uncertainty (with applications to investments, insurance and engineering). Providence: American Mathematical Society; 2001.
15. Davis R, Shrobe H, Szolovits P. What is knowledge representation? AI Mag. 1993;14(1):17–33.
16. Vaswani N, Chi Y, Bouwmans T. Rethinking PCA for modern data sets: theory, algorithms, and applications. Proc IEEE. 2018;106(8):1274–6.
17. Miller DD, Brown EW. How cognitive machines can augment medical imaging. Am J Roentgenol. 2019;212:9–14. https://doi.org/10.2214/AJR.18.19914.
18. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44.
19. Jurczyk T. Gains versus ROC curves. Do you understand the difference? TIBCO Community; 2019. https://community.tibco.com/wiki/gains-vs-roc-curves-do-you-understand-difference?exp=cloud&pref=off
20. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1:206–15.
21. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. JAMA. 2019;319(1):19–20.
22. Deep learning gets scope time. Nat Methods. 2019;16:1195. https://doi.org/10.1038/s41592-019-0670-x.
23. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets; Proceedings of the Neural Information Processing Systems (NIPS) 2014, p. 1–9.
24. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, MA/London: The MIT Press; 2016. www.deeplearningbook.org
25. Miller G. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev. 1956;63:81–97.
26. Cowan N. The magical mystery four: how is working memory capacity limited, and why? Curr Dir Psychol Sci. 2010;19:51–7.

27. Wickelgren I. Ray Kurzweil tells me how to build a brain. Sci Am. December 11, 2012.
28. Kimani S, Gabrielli S, Catarci T, Dix A. Designing for tasks in ubiquitous computing. In: Mostefaoui AK, Maamar Z, Giaglis GM, editors. Advances in ubiquitous computing: future paradigms and directions. Hershey/London: IGI Global Publishing; 2008. p. 172–98.
29. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. JAMA. 2019;322(24):2377–8.
30. Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. Acad Med. 2017;92:23–30.
31. Kahneman D. Thinking, fast and slow. New York: Farrar, Straus, and Giroux; 2011.
32. Launchbury J. A DARPA perspective on artificial intelligence. February 16, 2017. https://www.artificialbrain.xyz/a-darpa-perspective-on-artificial-intelligence/
33. Miller DD, Wood EA. AI, Autonomous machines and human awareness: towards shared human-machine contexts in medicine. Chapter 10. In: Lawless W, Mittu R, Sofge D, editors. Human-machine shared contexts. Amsterdam: Elsevier; 2020.
34. Smith AM, Floerke VA, Thomas AK. Retrieval practice protects memory against acute stress. Science. 2016;345(6315):1046–8. https://doi.org/10.1126/science.aah5067.
35. Maule AJ, Svenson O. Theoretical and empirical approaches to behavioral decision making and their relationship to time constraints. In: Svenson O, Maule AJ, editors. Time pressure and stress in human judgments and decision making. Boston: Springer; 1993. https://doi.org/10.1007/978-1-4757-6846-6_1.
36. Kahneman D, Klein G. Conditions for intuitive expertise: failure to disagree. Am Psychol. 2009;64(6):515–26. https://scholar.google.com/scholar?q=kahneman+klein+2009+American+Psychologist&hl=en&as_sdt=0&as_vis=1&oi=scholart
37. Weld DS. The challenge of crafting intelligible intelligence. 2018. https://arxiv.org/pdf/1803.04263.pdf
38. Porumb M, Stranges S, Pescapè A, Pecchia L. Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events based on ECG. Sci Rep (Nat). 2020;10:170. https://www.nature.com/articles/s41598-019-56927-5
39. Lebovitz S, Lifshitz-Assaf H, Levina N. Doubting the diagnosis: how artificial intelligence increases ambiguity during professional decision making. Soc Sci Res Netw (SSRN). November 4, 2019. https://ssrn.com/abstract=3480593 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3480593
40. Abrams C. Google's effort to prevent blindness shows AI challenges. Wall Street J. January 26, 2019. https://www.wsj.com/amp/articles/googles-effort-to-prevent-blindness-hits-roadblock-11548504004. Accessed 14 Mar 2019.
41. Wang MD, Khanna R, Najafi N. Characterizing the source of text in electronic health record progress notes. JAMA Intern Med. 2017;177(8):1212–3.
42. Miller A. The intrinsically linked future for human and artificial intelligence. J Big Data. 2019;6(38):1–9.
43. Auerbach A, Fihn SD. Discovery, learning, and experimentation with artificial intelligence-based tools at the point-of-Carew – perils and opportunities. JAMA Netw Open. 2021;4(3):e211474. https://doi.org/10.1001/jamanetworkopen.2021.1474.
44. Feng D, Hause-Schutz C, Rosenbaum L, et al. Deep multi-modal object detection and semantic segmentation for autonomous driving: datasets, methods and challenges. arXiv:1902.07830v3[cs.RO]. November 16, 2019.
45. https://www.mathworks.com/help/vision/examples/semantic-segmentation-using-deep-learning.html
46. Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite; 2012. http://www.cvlibs.net/publications/Geiger2012CVPR.pdf
47. Fass D. Collative semantics; 1986. https://www.aclweb.org/anthology/C86-1080.pdf
48. Small SL, Cottrell GW, Tanenhaus MK. Lexical ambiguity resolution: perspective from psycholinguists. Amsterdam: Elsevier; 2013.

49. Barnard K, Johnson M. Word sense disambiguation with pictures. Artif Intell. 2005;167(1–2):13–30. https://doi.org/10.1016/j.artint.2005.04.009.
50. Levesque, HJ, Davis, E, Morgenstern, L. The Winograd schema challenge. Proceedings of the 13th international conference on principles of knowledge representation and reasoning, Association for Advancement of Artificial Intelligence (AAAI); 2012, p. 552–561.
51. Devlin J, Chang M-W. Open sourcing BERT: state-of-the-art pre-training for natural language processing. https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html?m=1

Shame, Name, Give Up the Game? Three Approaches to Uncertainty

21

Vera Wilde

Chapter

What is going on "when the disease has no name"? As a family member and patient, I have observed three main cognitive strategies physicians use to grapple with uncertainty. Some approach it primarily (1) as a threat, others (2) as a classification problem, and others (3) as an ongoing part of health and disease processes. Applying cognitive science insights into why and how uncertainty sometimes threatens physicians underscores the importance of the current movement in rheumatology toward the latter two modes of grappling with it. Both are necessary, because approaching uncertainty as a classification problem has pros and cons.

On one hand, improving classification of cases that are uncertain because they are anomalous, early, and/or mild presentations – as in undifferentiated connective tissue disease (UCTD) or incomplete, latent, preclinical, or prodromal lupus – or simply because they are classically difficult diagnoses like systemic lupus erythematosus (SLE) has the potential to help more patients access needed care sooner, elevating quality of life and functioning and preventing potentially irreversible damage. On the other hand, approaching the problem of uncertainty primarily as a classification problem risks simply shifting the threshold of where patients begin being "in" – recognized as ill and deserving of help – versus "out." The evidence on uncertain cases is insufficient to specify this threshold. Thus, physicians also need to be able to approach uncertainty primarily as part of reality to be accepted rather than (only) a problem to be solved when possible through better classification. This need does not diminish the importance of classification improvements but rather represents a concurrent shift away from approaching uncertainty as a threat.

The most common treatments for UCTD and SLE, low-dose hydroxychloroquine and prednisone, may dampen disease development, as may lifestyle changes

V. Wilde (✉)
Berlin, Germany
e-mail: vera@verawil.de

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4_21

such as avoidance of UV exposure and heat. Low-dose hydroxychloroquine in particular carries very low iatrogenic harm risks, and low-dose prednisone carries relatively low risks while treating most symptomatic autoimmune phenomena. This suggests grappling with uncertainty "when the disease has no name" should include discussion about the possible risks and benefits of these treatment options. In this context, patients should not suffer for years with symptomatic autoimmunity and no appropriate specialized medical treatment offered for it.

But the current focus on diagnostic criteria means precisely that "Organ damage might accrue in a prodromal period prior to a formal diagnosis of SLE being made [1]," although these treatments can help prevent that damage. At the same time, an online survey of 3022 self-reported lupus patients found that most (54.1%) reported having been told there was nothing wrong with them or that their symptoms were psychological [2]." These disconnects seem in part to be products of physicians sometimes approaching uncertainty as threat.

Uncertainty as Threat

"Absence of evidence is not evidence of absence, [3]" and uncertainty in diagnosis is not evidence that a patient presenting with symptoms of unknown etiology is actually well or merely psychologically troubled. Why then would physicians frequently dismiss undiagnosed autoimmune disease sufferers as such?

An empirical answer seems at first to make sense. Depression and anxiety often track with rheumatological diseases including SLE [4] but are also substantially more common than them. Self-reported pain and fatigue as well as frequent infections and other such typical complaints can be associated with mental health as well as rheumatological problems. In light of base rates, Occam's razor would seem to favor the former explanation.

But this reflects a logical fallacy, because mental health and rheumatological problems are not mutually exclusive. To the contrary, the rate of mental health problems in SLE patients is sufficiently high, and mechanisms (such as inflammation) associated with rheumatological and mental health problems sufficiently overlap, such that the base rate of rheumatological disease in depressed and anxious subpopulations is likely to be higher than in general populations. So what some physicians use as a dismissal may well be a clue.

Here is where cognitive bias may come in: Physicians face time pressures that constrain their abilities to be and remain knowledgeable about everything they are supposed to know (information overload), and patients with complex, gradually developing, heterogeneous conditions such as UCTD and SLE present particularly time-intensive diagnostic puzzles. What is a physician seeing many patients a day along with other responsibilities to do?

Perhaps, faced with these pressures, many physicians experience diagnostic uncertainty as the threat of cognitive dissonance. Ironically, desire to help people and see oneself as helping people as a doctor can underpin this mode as belief ("doctors help people") conflicts with behavior ("I don't help this person"). This

conflict can be lessened by rejecting the uncertainty itself ("this patient is [definitely] well or crazy"). This cognitive strategy is time-efficient for the individual physician and protects his or her ego as an expert whose social and legal power comes from knowing about health and disease, insofar as expertise often blurs with certainty.

But it is inefficient (not to mention unpleasant) for the patient, who continues to suffer without appropriate medical help, and society, as patients with unmanaged autoimmunity tend to fare worse than their treated counterparts, with preventable damage causing harm to their families and communities, as well as costing more in healthcare and disability. It is also inefficient for colleagues, who have to deal with the dismissed patients in future interactions. In this way, the cognitive strategy of threat in dealing with diagnostic uncertainty arguably drives physician defection in a collective action problem [5] – a situation in which everyone would be better off cooperating, but enough people choose instead to pursue their immediate self-interest that the behavioral norm undercuts the group's long-term interests. So it benefits everyone – physicians, patients, and society – when physicians are better able to shift cognitive strategies in dealing with diagnostic uncertainty, away from threat and toward classification and process modes.

Taking patients at their word underpins this shift. Questioning patient credibility of self-reported symptoms undermines the possibility of a therapeutic relationship, increasing the likelihood that both doctor and patient will experience an interaction as threatening. This might contribute to a mistrust spiral in which patients trust medicine and science less. The social implications of such spirals can be weighty: Mistrust of the medical profession predicts parental vaccine hesitancy [6]. In this context, changing the way physicians treat patients "when the disease has no name" has the potential to affect public health.

A new set of consensus guidelines setting out clear management options for uncertain diagnoses might help physicians make this shift. The temptation in devising such guidelines, however, is to again deny the central problem of uncertainty – this time by moving the diagnostic threshold. This has the potential to help many people access needed care sooner, preventing harm. But as a cognitive mode for dealing with uncertainty, it also has the potential to perpetuate the patterns of the threat mode by attempting to solve instead of accepting uncertainty. Uncertain disease is likely an unsolvable problem in rheumatology, given the unpredictable and gradual development of heterogeneous immune dysfunctions and other manifestations, the difficult-to-measure nature of typical complaints such as pain and fatigue, and widely reported difficulties in accessing appropriate specialist care early in disease process.

Uncertainty as Classification Problem

Changing the parameters of diagnosable autoimmune disease follows the best available evidence to prevent harm and is thus well worth doing. For example, prodromal lupus patients treated with hydroxychloroquine or prednisone had delayed

classifiable SLE onset, and hydroxychloroquine was also associated with fewer later autoantibody specificities, according to a retrospective study of 130 military personnel [7]. Indeed, this already appears to be standard practice in some places, with the majority of incomplete or potential lupus patients at Brigham and Women's Hospital (66% of 161 patients) [8] and in the Spanish Rheumatology Society Lupus Registry (around 69% of 345 patients) [9] treated with antimalarial medication. These observations still suggest a substantial minority of affected patients (within the universe of identified patients) miss out on effective preventive treatment for unspecified reasons, suggesting room for improvement through updated consensus guidelines.

Relevant literature tends to emphasize these interventions as lupus interventions [10], focusing on the subset of patients with undifferentiated, early, mild, or prodromal disease who go on to develop diagnosable, differentiated diseases including SLE. This suggests researchers may be placing less value on possible quality of life and function improvements for the whole class of symptomatic patients than on preventing disease progression in a more severe subset. But given that undifferentiated disease, too, can disable patients with pain, fatigue, and many of the other same manifestations as diagnosable SLE, this focus might be misplaced from a patient perspective. Given that the difference between undifferentiated and differentiated disease diagnosis can depend entirely on whether patients access the right care at the right time – while disabled and often after having experienced physician dismissal of their complaints – this focus might also be misplaced in a substantial subset of cases that are missed diagnoses (and misdiagnoses). Given that ambiguity, this focus might also hinder research on autoimmunity by assuming distinctions between patient groups that result in part from differences in the timing and quality of physician-patient interactions, and not from differences in the underlying disease.

In that context, approaching uncertainty as a classification problem will tend to create winners and losers among affected patients. An example illustrates this problem: In a recent article in a top subfield journal, Adamichou et al. [11] report testing a machine learning-based model on a sample of patient data from hospital rheumatology clinics. The premise of the tool is that it can assist lupus diagnosis. But by baking in the cognitive distortion of the primacy of the categories of certain lupus and non-lupus patients, its use might keep physicians from learning more about non-lupus, including those for whom diagnosis is uncertain, and keep many of those patients ill and seeking help unsuccessfully. In other words, it seems to assume that diagnostic accuracy is only about quantifiable sensitivity and specificity for a single, binary diagnostic category. Patients with uncertain diagnoses are excluded from the sample, study design, and thinking about the effects of the use of this kind of tool. In this respect, approaching uncertainty as a classification problem is circular: Uncertainty is no longer a problem if all the cases in a given universe are diagnosed.

This kind of research, again, has merit. Improving SLE diagnostic accuracy is important. I used a tool like this one (Stephen Borowitz's Isabel) [12] many years ago in the medical school library to help generate a differential diagnosis that favored SLE when my mom was disabled with a disease without a name. Using that differential to learn more from books, patient support groups, and appropriate

specialists, I eventually presented her case synopsis to a rheumatologist who diagnosed and treated her. Hers was not a corner case. But what about such cases?

Batu et al. [13] report results from a pediatric cohort study on Adamichou et al.'s tool (SLE Risk Probability Index, SLERPI). They find raising the diagnostic threshold in that population decreases sensitivity by around 2% while increasing specificity by around 8%. It is unclear that decreasing sensitivity by a single-digit percentage in order to increase specificity by another single-digit percentage benefits these patients. If clinicians should treat children who score a 7 using this tool the same as they should treat children scoring an 8, raising the threshold to reduce its type II errors while increasing its type I errors would appear to have little practical merit.

In a typical cohort setup, both Adamichou et al. and Bantu et al. use patient data from two groups: those diagnosed with lupus and those diagnosed with other rheumatologic diseases. Maybe 100% diagnosis rates are the norm in the rheumatology departments of the university hospital clinic samples from which these cohorts were drawn. But I have been to many specialists with my mom and been to a fair number myself as a patient over the years. Before her SLE and my UCTD diagnoses, physicians usually dismissed us without diagnosis or treatment. Having headed up a lupus patient support group chapter and also heard similar stories from many female friends with other chronic health problems, I believe these sorts of experiences are common. This suggests that perhaps it is a currently accepted research norm to exclude cases without rheumatologic diagnoses from studies like these. That would make uncertain cases invisible to researchers and physicians reading their work. This reflects one danger of approaching uncertainty as a classification problem: It appears to erase it, but in so doing, it may normalize the dismissal of uncertain cases, promoting the threat approach to dealing with uncertainty when some patients inevitably still fall outside the diagnostic bounds. After all, those patients do not appear to exist in the relevant medical literature.

Uncertainty as Part of Process

Discussions of the problem of uncertain diagnosis in rheumatology often involve the distinction between classification criteria, used to qualify patients for clinical trials, and diagnostic criteria, used to qualify patients for clinical diagnosis and treatment. This distinction assumes an unproven trade-off between privileging internal validity to treat the most severely affected (e.g., SLE patients with lupus nephritis) over external validity to help the most patients (e.g., SLE patients including those misdiagnosed with UCTD) in a notoriously heterogeneous disease group. It thus exemplifies Alvan Feinstein's concern that his evidence-based medicine revolution was hijacked by the "distraction" of quantitative models [14].

Feinstein suggested differentiating between the special collection of data regarded as suitable evidence and practicing evidence-based medicine [15]. Such data, he noted, derive "almost exclusively from randomized trials and meta-analyses," and "do not include many types of treatments or patients seen in clinical practice." Feinstein warned that by calling on doctors to make clinical decisions

based on a highly restricted quality and scope of evidence, such work had an "authoritative aura" which "may lead to major abuses that produce inappropriate guidelines or doctrinaire dogmas for clinical practice." Focus on the boundaries of defined autoimmune diseases using such data leads to precisely such an abuse: failure to name and treat uncertain cases where such treatment may prevent harm and improve quality of life while posing minimal iatrogenesis risks.

Practicing evidence-based medicine, Feinstein went on, means attending to "such cogent clinical features as severity of symptoms, illness, co-morbidity, and other clinical nuances." Filling his prescription means recognizing that applying somewhat exegetical diagnostic criteria with debatable thresholds causes some patients to suffer without medical help, risking irreversible damage from untreated disease progression in addition to living lives curtailed by common problems such as recurrent infections, intermittently disabling pain and fatigue, and the profound isolation of being unable to say what is wrong. Devising a new, more inclusive way of measuring autoimmunity including uncertain cases suggests a new research agenda that could lead to advances in care, quality of life, and communication for many.

Research Agenda

Serum autoimmunity appears to be rising for unknown reasons [16]. This suggests the population of uncertain disease patients may be growing. General population research might include lifestyle and rheumatology questionnaires in an attempt to identify factors that might predispose people to develop practically meaningful symptoms, before clinicians identify disease. This might generate useful insights for diagnosis as well as prevention.

Recent research from the emerging field of nutritional psychiatry establishes that interventions centered on cost-effective nutritional education can lower inflammation and depressive symptoms [17–19]. The relatively low-risk nature of such interventions makes them especially appropriate for studies including patients with uncertain, undifferentiated, and differentiated autoimmune diseases which tend to be characterized by inflammatory processes and are often comorbid with depression and/or anxiety. The importance of diet in lupus has been widely discussed [20], particularly with respect to including "healthy" types of dietary fats [21], probiotics [22], and fresh whole foods (especially fruits and vegetables) [23], caloric restriction or fasting [24], and minimizing ultra-processed foods, particularly free sugars [25]. But to date, no relevant large-scale randomized trials have been conducted. By blocking on patients' different disease states and treatments along with demographics like gender within an approximate equalization rather than true randomization procedure, researchers could account for the increased within-group heterogeneity resulting from including a wider range of potential autoimmune disease sufferers while also assessing the extent to which these findings generalize to traditionally defined rheumatology patients.

Prevention-oriented research including uncertain and undifferentiated cases should also grapple with possible intergenerational effects of autoimmunity, working to identify conditions that may exacerbate or ameliorate them. It appears that in

utero exposures to adverse conditions such as maternal infection and malnutrition can adversely affect offspring, particularly in terms of neurodevelopmental outcomes [26, 27]. Some of the same offspring neurodevelopmental risks (e.g., autism) are associated with maternal autoimmunity [28]. Thus it seems plausible that offering safe, symptomatic treatment for autoimmunity before and during pregnancy might improve offspring outcomes. In the realm of uncertain disease, at what point does the possible cost of iatrogenic harm outweigh the possible benefit of downregulating maternal autoimmunity during crucial developmental windows? Should pregnant women be screened for autoimmunity, just as they are screened for other conditions that might adversely affect fetal development?

And what about the mistrust created or exacerbated by common patient experiences of going years without needed medical help in spite of asking for it? Might accepting uncertainty as part of the process present its own set of challenges in terms of physician credibility to these and other patients? Or can emphasis on upholding the Hippocratic Oath to do no harm trump emphasis on diagnostic accuracy?

As a family member and patient, I think starting from an assumption of patient credibility and accepting uncertainty as part of the process has the potential to enrich therapeutic relationships by reorienting clinical interactions as dialogues in the context of ongoing experiments. But as a scientist, I also understand skepticism toward self-reports, particularly retrospective ones. Maybe there could be more middle ground in practice here, where patients who feel they were not heard or seen in the past have a chance to produce evidence of symptoms that might have occurred in the past. Maybe there should be a different category along the autoimmunity continuum identifying patients who report having met diagnostic criteria unobserved.

By using the scientific method to study the broader universe of cases, this research agenda could help normalize accepting uncertainty as part of the process of diagnosing and treating autoimmune disorders. Just as society has shaped science by demanding certainty, so too can science shape society by accepting uncertainty. The institutional incentive structures of normal medical science, such as grant applications, conference proposals, and academic publication, tend to privilege specificity and staying within narrow parameters that conform to what other people think and are doing [29]. Thus researchers studying diseases like SLE, which often involve long histories of inadequate treatment, underdiagnosis, and eventual progression to more irreversible damage, never seem to meet clinicians who see patients in the process of developing these diseases, before as much recognizable damage is done. Prioritizing mitigating the risks of preventable harm to patients presenting with uncertain autoimmunity might help these experts meet.

Conclusion

There is no perfect diagnostic universe. Maximizing one form of accuracy (sensitivity) usually compromises another (specificity). Similarly, there is no risk-free treatment universe. Maximizing prevention usually risks iatrogenic harm when the treatment involves a traditional medical intervention like a pharmaceutical or surgery.

This implies that diseases that have no names will always be with us. There is no diagnostic approach that banishes them, although there is one that makes them someone else's problem. There is no possible perfect set of guidelines that will prescribe the single best clinical management approach for them, although there are some low-risk treatments for uncertain, mild, or undifferentiated autoimmunity that might be much more broadly appropriate than they are currently applied. Maybe upholding the Hippocratic Oath to do no harm in this context means erring more on the side of believing patients and less on prioritizing diagnostic accuracy in terms of certainty, no matter the cost.

References

1. Barber MRW, et al. Evolving concepts in systemic lupus erythematosus damage assessment. Nat Rev Rheumatol. 2021;17(6):307–8.
2. Daly R, et al. Lupus diagnosis: process and patient experience [abstract]. Arthritis Rheumatol. 2017;69(Suppl 10). https://acrabstracts.org/abstract/lupus-diagnosis-process-and-patient-experience/. Accessed 22 May 2021.
3. Altman DG. Statistics notes: absence of evidence is not evidence of absence. BMJ. 1995;311:485.
4. Figueiredo-Braga M, et al. Depression and anxiety in systemic lupus erythematosus. Medicine (Baltimore). 2018;97(28):e11376.
5. Olson M. The logic of collective action: public goods and the theory of groups, second printing with a new preface and appendix. London: Harvard University Press; 1971.
6. Reuben R, et al. Mistrust of the medical profession and higher disgust sensitivity predict parental vaccine hesitancy. PLoS One. 2020;15(9):e0237755. https://doi.org/10.1371/journal.pone.0237755.
7. James JA, et al. Hydroxychloroquine sulfate treatment is associated with later onset of systemic lupus erythematosus. Lupus. 2007;16:401–9.
8. Al Daabil M, et al. Development of SLE among "potential SLE" patients seen in consultation: long-term follow-up. Int J Clin Pract. 2014;68(12):1508–13.
9. Rúa-Figueroa Í, et al. Comprehensive description of clinical characteristics of a large systemic lupus erythematosus cohort from the Spanish Rheumatology Society Lupus Registry (RELESSER) with emphasis on complete versus incomplete lupus differences. Medicine (Baltimore). 2015;94(1):e267.
10. Bourn R, James JA. Pre-clinical lupus. Curr Opin Rheumatol. 2015;27(5):433.
11. Adamichou C, et al. Lupus or not? SLE Risk Probability Index (SLERPI): a simple, clinician-friendly machine learning-based model to assist the diagnosis of systemic lupus erythematosus. Ann Rheum Dis. 2021;80:758–66. https://doi.org/10.1136/annrheumdis-2020-219069.
12. https://www.healthcareitnews.com/news/decision-help-virginia-hospital.
13. Batu ED, et al. Correspondence on 'Lupus or not? SLE Risk Probability Index (SLERPI): a simple, clinician-friendly machine-learning-based model to assist the diagnosis of systemic lupus erythematosus.' Ann Rheum Dis. 2021:annrheumdis-2021-220246. https://doi.org/10.1136/annrheumdis-2021-220246. Online ahead of print. PMID: 33811035.
14. Feinstein AR. "Clinical judgment" revisited: the distraction of quantitative models. Ann Intern Med. 1994;120:799–805.
15. Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine". Am J Med. 1997;103:529–35.
16. Dinse GE, et al. Increasing prevalence of antinuclear antibodies in the United States. Arthritis Rheum. 2020. https://doi.org/10.1002/art.41214. Epub 2020 Apr 30. PMID: 32266792.

17. Berk M, et al. So depression is an inflammatory disease, but where does the inflammation come from? BMC Med. 2013;11:200. https://doi.org/10.1186/1741-7015-11-200.
18. Jacka FN, et al. A randomised controlled trial of dietary improvement for adults with major depression (the 'SMILES' trial). BMC Med. 2017;15:23. https://doi.org/10.1186/s12916-017-0791-y.
19. Parletta N, et al. A Mediterranean-style dietary intervention supplemented with fish oil improves diet quality and mental health in people with depression: a randomized controlled trial (HELFIMED). Nutr Neurosci. 2019;22(7):474–87. https://doi.org/10.1080/1028415X.2017.1411320. Epub 2017 Dec 7.
20. Constantin M, et al. Significance and impact of dietary factors on systemic lupus erythematosus pathogenesis (review). Exp Ther Med. 2019;17(2):1085–90.
21. Leiba A, et al. Diet and lupus. Lupus. 2001;10(3):246–8. https://doi.org/10.1191/096120301674681790.
22. Mu Q, et al. Control of lupus nephritis by changes of gut microbiota. Microbiome. 2017;5:73. https://doi.org/10.1186/s40168-017-0300-8.
23. Klack K, et al. Diet and nutritional aspects in systemic lupus erythematosus. Rev Bras Reumatol. 2012;52(3):384–408. English, Portuguese.
24. Liu Y, et al. Fasting-induced hypoleptinemia expands functional regulatory T cells in systemic lupus erythematosus. J Immunol. 2012;188(5):2070–3. https://doi.org/10.4049/jimmunol.1102835. Epub 2012 Jan 30. PMID: 22291185; PMCID: PMC3288569.
25. Correa-Rodríguez M, et al. Dietary intake of free sugars is associated with disease activity and dyslipidemia in systemic lupus erythematosus patients. Nutrients. 2020;12(4):1094. https://doi.org/10.3390/nu12041094. PMID: 32326626; PMCID: PMC7231002.
26. Zerbo O, et al. Maternal infection during pregnancy and autism spectrum disorders. J Autism Dev Disord. 2015;45:4015–25. https://doi.org/10.1007/s10803-013-2016-3.
27. Lumey LH, et al. Cohort profile: the Dutch hunger winter families study. Int J Epidemiol. 2007;36(6):1196–204. https://doi.org/10.1093/ije/dym126.
28. Zhu Z, et al. Maternal systemic lupus erythematosus, rheumatoid arthritis, and risk for autism Spectrum disorders in offspring: a meta-analysis. J Autism Dev Disord. 2020;50:2852–9. https://doi.org/10.1007/s10803-020-04400-y.
29. Kolata G. Grant system leads cancer researchers to play it safe. New York Times, June 28, 2009.

Vera Katelyn Wilde obtained her PhD in politics from the University of Virginia before holding postdoctoral positions in psychology at UCLA and government at Harvard, all supported by the National Science Foundation. Her experiences as a family member of a then-undiagnosed lupus patient and later as the head of a lupus patient support group chapter informed her dissertation research on a diagnosis decision support tool

Others Who Spoke

Michael D. Lockshin

Others Who Spoke at the Workshop Concurred with and/or Added New Thoughts on Several Topics

Regarding the question, *What is a diagnosis?*, Peter Grayson and Dan Kastner described how discovery of the genetics of VEXAS created a novel, separable diagnosis, a subset of autoimmune diagnoses, that bridges multiple medical specialties and that demands different physician responses, not necessarily those suggested by the phenotype. Like other speakers, they discussed how emotional conversations about unnamed diagnoses can be, whether or not a new diagnosis is found.

Regarding *The Purposes of Diagnoses*, Susana Serrate-Sztein described how those who fund research use "white papers," written by professional organizations and advocacy groups, to inform scientific review and to help National Institutes of Health staff decide research priorities. Allan Gibofsky pointed out that medications are not "approved" by the Food and Drug Administration (FDA) but instead are given "indications," leaving responsibility of "off-label" use to physicians, insurers, and other stakeholders. It is difficult, he said, to summarize FDA policies regarding diagnoses because the FDA has issued 2633 "guidance" documents, each specific to a medication and a diagnosis, not governed by an overall plan.

In the section on *Diagnostic Uncertainty*, Hardeep Singh discussed the differences between uncertainty and error; the effect of uncertainty on patient stigmatization; the need for incentives to establish priority projects; and quantitating uncertainty (recently published in book form) [1]. He and Jillian Rose discussed the patients' perspectives of how physicians and patients communicate uncertainty. The main messages were having no answer is frustrating, doctors are not willing to discuss uncertainty, and patients are humiliated by not being believed.

Jinoos Yazdany said that, while illness is a journey, an electronic medical record (EMR) is a clinical event based on billing, not on science or patient care. She suggested that artificial intelligence (AI) methods and natural language processing

M. D. Lockshin
Weill Cornell Medicine, Hospital for Special Surgery, New York, NY, USA
e-mail: Lockshinm@hss.edu

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4

(NLP) reviews of EMRs to assign probability of diagnosis rather than binary names would improve consistency use—after the fact, of course, not in real time. She discussed biases in AI. She asked that multiple medical specialties, not just self-appointed experts in silos, provide input to narrative and language searches.

Katherine Liao prioritized phenotypic descriptions; she noted that EMRs have no consistent qualitative and quantitative ways of weighting laboratory reporting. She offered many ways in which NLP can contribute to resolving uncertainty, for instance, by overlaying criteria-based clinical trials with observational megadata to generalize trial results.

Mary Crow concurred that clinical studies of populations require use of clinical phenotypes (*exclusive*) rather than mechanistic descriptions (*inclusive*); the opposite is true for studies of individual patients. Catherine MacLean discussed the choices insurers make, for instance, the conundrum of choosing the preferred option. Is the choice ethical, honest, and valid? Based on good short-term outcome and low cost or on better (but not guaranteed) long-term high cost outcome and very long times to determine cost/benefit.

MacLean, Gibofsky, Lars Noah, David Pisetsky, and others mentioned the possibility of fraud or profit-driven choices in uses of diagnosis names. Daniel Solomon spoke of the inability of journals to adjudicate articles if diagnoses are not named.

In *Discussion and Recommendations*, Pisetsky asked, "Who creates criteria?" then he and Miriam Solomon pointed out the diversity of official organizations, individual authors, self-defined experts who are the definers, and the lack of overarching standards for official uses of criteria.

Others spoke to the positive aspects of "branding" patients with diagnosis names. Branding is useful to reassure that diagnosis-associated rules apply and to add a sense of competence to the conversation. Branding has negative aspects: it can close minds to alternate explanations. Karen Costenbader, noting the time-limited nature of criteria, emphasized that it is important to know who decides which criteria apply, that criteria can be used for harm as well as good, and that doctor-patient combined decision-making is required.

Noah offered this warning about criteria: "Watch what you write. Payers and others will take you seriously." He pointed out that people who are excluded from clinical trials nonetheless find ways to obtain the medications. Although their experiences can be informative, legislation prohibits including them in official reports. He suggested that identifying such patients through public records might make it possible to evaluate people who receive off-label medications and thus approve a drug's efficacy in less restricted populations. Richard Furie and Gibofsky concurred that there is no systematic way of recording adverse events.

Peter Croft pointed out that, in the United Kingdom, because International Classification of Diseases (ICD) codes are not linked to billing, British physicians are more comfortable than Americans with discussing uncertainty. He argues that establishing prognoses is better for society than is providing a diagnosis label. Speaking from the vantage point of a Department of Medicine Chairman, Andrew Schafer noted how the structure of American medicine harms imaginative thinking. In the United States, he said, medical records require that every chart note be

assigned an unambiguous ICD code, regardless of real-time uncertainty. Uncertainty is not taught in medical student and medical resident curricula. Schafer illustrated the point with an anecdote: As a teaching exercise, he provided a brief case scenario of an anemic patient, detailed laboratory tests included; he then asked medical trainees to choose one of five possible diagnoses, four of which listed precise causes. The fifth (and correct) option was "anemia of unknown etiology." Not single resident chose the correct option because of the dishonor associated with uncertainty.

Reference

1. https://urldefense.com/v3/__https://books.google.com/books/about/Diagnosis.html?id=p6AzDwAAQBAJ__;!!M-C4!RcluinDT6nnixTLIGNuQaKAvbFUGb8GbpV_H2nHBMkCRoQ-XNr7CwN-YaLONg6HW$.

Index

M. D. Lockshin et al. (eds.), *Diagnoses Without Names*,
https://doi.org/10.1007/978-3-031-04935-4