

Research Methods and Technology Special Communication

Cite this article: Hanson HA, Leiser CL, Bandoli G, Pollock BH, Karagas MR, Armstrong D, Dozier A, Weiskopf NG, Monaghan M, Davis AM, Eckstrom E, Weng C, Tobin JN, Kaskel F, Schleiss MR, Szilagyi P, Dykes C, Cooper D, and Barkin SL. Charting the life course: Emerging opportunities to advance scientific approaches using life course research. *Journal of Clinical and Translational Science* 5: e9, 1–8. doi: [10.1017/cts.2020.492](https://doi.org/10.1017/cts.2020.492)

Received: 17 March 2020

Revised: 4 June 2020

Accepted: 4 June 2020

Keywords:

Translational life course research; complexity science; life course research priorities; life course methods; life course and lifespan; cytomegalovirus


Address for correspondence:

H. A. Hanson, PhD, MS, Department of Surgery and Population Sciences, Huntsman Cancer Institute at the University of Utah, Salt Lake City, UT 84132, USA.
Email: heidi.hanson@hci.utah.edu

© The Association for Clinical and Translational Science 2020. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.



Charting the life course: Emerging opportunities to advance scientific approaches using life course research

Heidi A. Hanson^{1,2} , Claire L. Leiser^{1,3}, Gretchen Bandoli⁴, Brad H. Pollock⁵, Margaret R. Karagas⁶, Daniel Armstrong⁷, Ann Dozier⁸, Nicole G. Weiskopf⁹, Maureen Monaghan^{10,11}, Ann M. Davis^{12,13}, Elizabeth Eckstrom¹⁴, Chunhua Weng¹⁵, Jonathan N. Tobin^{16,17}, Frederick Kaskel¹⁸, Mark R. Schleiss¹⁹, Peter Szilagyi²⁰, Carrie Dykes²¹, Dan Cooper²² and Shari L. Barkin²³

¹Population Sciences, Huntsman Cancer Institute at the University of Utah, Salt Lake City, UT, USA; ²Department of Surgery, University of Utah, Salt Lake City, UT, USA; ³Department of Epidemiology, University of Washington, Seattle, WA, USA; ⁴Department of Pediatrics, University of California San Diego, La Jolla, CA, USA; ⁵Department of Public Health Sciences, School of Medicine, University of California Davis, Davis, CA, USA; ⁶Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA; ⁷Department of Pediatrics, University of Miami Miller School of Medicine, Miami, FL, USA; ⁸Public Health Sciences, University of Rochester, Rochester, NY, USA; ⁹Department of Biomedical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA; ¹⁰Children's National Hospital, Washington, DC, USA; ¹¹George Washington University School of Medicine, Washington, DC, USA; ¹²University of Kansas Medical Center, Kansas City, KS, USA; ¹³Center for Children's Healthy Lifestyles & Nutrition, Kansas City, MO, USA; ¹⁴Division of General Internal Medicine & Geriatrics and Oregon Clinical & Translational Research Institute, Oregon Health & Science University, Portland, OR, USA; ¹⁵Department of Biomedical Informatics, Columbia University, New York, NY, USA; ¹⁶Center for Clinical and Translational Science, The Rockefeller University, New York, NY, USA; ¹⁷Clinical Directors Network (CDN), New York, NY, USA; ¹⁸Department of Pediatrics, Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, NY, USA; ¹⁹University of Minnesota Clinical and Translational Science Institute, Minneapolis, MN, USA; ²⁰Department of Pediatrics, UCLA Mattel Children's Hospital, University of California at Los Angeles, Los Angeles, CA, USA; ²¹Clinical and Translational Science Institute, University of Rochester, Rochester, NY, USA; ²²Institute for Clinical and Translational Science and Department of Pediatrics, University of California at Irvine, Irvine, CA, USA and ²³Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, USA

Abstract

Life course research embraces the complexity of health and disease development, tackling the extensive interactions between genetics and environment. This interdisciplinary blueprint, or theoretical framework, offers a structure for research ideas and specifies relationships between related factors. Traditionally, methodological approaches attempt to reduce the complexity of these dynamic interactions and decompose health into component parts, ignoring the complex reciprocal interaction of factors that shape health over time. New methods that match the epistemological foundation of the life course framework are needed to fully explore adaptive, multi-level, and reciprocal interactions between individuals and their environment. The focus of this article is to (1) delineate the differences between lifespan and life course research, (2) articulate the importance of complex systems science as a methodological framework in the life course research toolbox to guide our research questions, (3) raise key questions that can be asked within the clinical and translational science domain utilizing this framework, and (4) provide recommendations for life course research implementation, charting the way forward. Recent advances in computational analytics, computer science, and data collection could be used to approximate, measure, and analyze the intertwining and dynamic nature of genetic and environmental factors involved in health development.

Introduction

Fifteen years after mapping the human genome, an individual's health is clearly shaped not only by genomics, but also by a complex web of factors, from molecular to societal, that varies over time. Genetics are fixed at conception, though gene expression is dynamic and "genetics" of health and disease are complex and develop across the lifespan. While our understanding of the genetics of disease evolves, another medical "revolution" has been occurring. During the past several years, generation of massive quantities of data has become commonplace and excitement surrounding "big data" as well as machine learning (ML) for medicine is palpable. However, the explosion of these two fields has outpaced the evolution of theoretical and conceptual frameworks that should underlie research in these arenas. Studies that are not grounded in solid frameworks may lack reproducibility and biological plausibility. We propose

that the marriage of a “life course” theoretical framework with complexity science associated with big data applications has paradigm-shifting potential for understanding health development across the lifespan. It also offers potentially significant translational science impact.

The Barker Hypothesis is often cited as an early use of the life course research lens in chronic disease epidemiology [1]. The landmark Dutch Famine natural experiment used a large historic cohort database, where the exogenous insult was lack of nutrition in a defined population with available health outcomes data [2]. The study demonstrated a strong association with poor maternal nutrition and later offspring risk for obesity and diabetes, but the outcome depended on the gestational timing of the exposure this general phenomenon of early life exposures (ELEs) affecting later disease manifestation has since evolved to the more encompassing Life Course Health Development (LCHD) framework [3]. LCHD provides a theoretical framework for embracing complexity in health research by integrating a developmental perspective into the concept of health and describing how *health development* is applicable to both individuals and populations. The model is built on the recognition that although early determinants, exposures, and influences may not become phenotypically evident for years or decades, such events can lead to significant adverse later life health consequences. Health risks and disease conditions evolve over time, coexist with positive health states, and can be mitigated or exacerbated by social, physical, and biologic contexts.

Life course research is often conflated with lifespan research; however, the ideas are distinct and related to our recommendations for moving beyond theory to implementation. The focus of this article is to (1) delineate differences between lifespan and life course research, (2) articulate the importance of complex systems science as a methodological framework in the life course research toolbox to guide our research questions, (3) raise key questions that can be asked within the clinical and translational science domain utilizing this framework, and (4) provide recommendations for life course research implementation, charting the way forward.

Differentiating Lifespan from Life Course

Confusion often surrounds the terms *lifespan* and *life course*. *Lifespan* is a measure of longevity reflecting underlying biologic aging of an individual that occurs for everyone. *Life course* encompasses lifespan and is influenced by the interaction of contextual factors over time that affect health and development and vary among individuals (see Table 1 for a list of definitions). Furthermore, exposures, whether physiologic or sociologic, can have differential impact on health outcomes based on the following:

- a. Dose: The amount of endogenous or exogenous exposure (such as stress or air pollution).
- b. Duration: Short- or long-term accumulation of physical and psychosocial exposures throughout the life course on health and development from gamete to grave.
- c. Timing: The timing of exposure in relation to human development, recognizing that some periods of development allow for more plasticity and adaptation than other periods.

Matching Theory to Methods: Complex Systems Science as a Methodological Framework in the Life Course Research Toolbox to Guide Our Research Questions

A theoretical tenet of life course research is the interdependence between and dynamic nature of factors that shape health. Recent

Table 1. Core definitions

Term	Definition
Lifespan	A measure of longevity reflecting the underlying chronologic biologic aging of an individual that occurs for everyone.
Life course	The interaction of contextual factors over time that affect health and development and varies among individuals.
Framework	A conceptualization structure providing guidance to the researcher as research questions are developed and theories are tested.
Theoretical framework	A framework that has been derived from tested theories and is generally accepted. The set of concepts drawn from the theory can be thought of as the guide to build and support a study and should define the epistemological, methodological, and analytical approach to the problem.
Methodological framework	Methodological guide linking the theoretical framework to the appropriate methodological tools.
Complexity science	The study of complex systems and problems that are dynamic, unpredictable, and multidimensional. These patterns emerge over time as a result of systems that are interconnected [57–59].

advances in high-performance computing create possibilities for the combination of data across a vast number of sources to make sense of the complex, interdependent factors that shape health. As a result of these technological advances, as with genomic medicine two decades ago, the potential for ML-based methods to solve complex disease questions has emerged [4,5]. While ML can identify important patterns that may not be extractable by the human mind, theoretical and methodological grounding is necessary to avoid misattributing causation to spurious associations. A hybrid approach is needed, one that connects theory to methodological frameworks to allow for scientifically rigorous and biologically plausible inquiries.

A Life Course Approach Influences New Hypotheses

Below we highlight three case examples illustrating the complex nature of human health and disease and the importance of couching questions related to the development of immunity, type 2 diabetes (T2DM), and Alzheimer’s disease (AD) within the *LCHD theoretical framework* using *methodological tools from complexity science*. This demonstrates how using a life course lens within complexity science frameworks can lead to generation of new scientific hypotheses and associated interventions that consider the high dimensionality, nonlinearity, and heterogeneity of health and disease.

Case 1: The Development of Immunity and Impact of Ubiquitous Persistent Infection

Early infancy and childhood represent windows where fundamental processes prime the immune system and development of the microbiome occurs. For example, early life acquisition of human cytomegalovirus (CMV), a ubiquitous childhood infection, can lead to *enhanced* responsiveness to vaccines including those for influenza

[6,7]. Conversely, later in the life course, ongoing CMV infection may promote immunosenescence, paradoxically conferring *decreased* vaccine responsiveness and increased susceptibility to other infections [8]. Thus, the timing of exposure to and acquisition of CMV has potentially a profound impact on immune system priming and function, and functional immunologic outcomes can clearly evolve over the life course.

However, sole examination of timing of CMV infection does not fully describe its broad influence over the immune network, which is reactive and adaptive to both heritable and nonheritable influences [8]. A study by Brodin *et al.* used monozygotic identical twins discordant for CMV serostatus (i.e., one twin infected, the other uninfected) to demonstrate substantial differences in CD8+ cell effector function, gamma-delta T-cell populations, and differences in both cytokine levels and overall responsiveness to IL-10 and IL-6. These parameters vary with age; generally heritable factors have limited influence and nonheritable factors increase influence as cumulative environmental exposures shape the microbiome over the life course. In fact, responses to influenza vaccine were found to be entirely influenced by nonheritable factors, indicating that interactions between the microbiome and infections are complex and driven by a combination of inherited and nutritional or pathogenic factors, which vary based on timing of exposure during vulnerable windows of development. Shifts in the microbiome, in turn, affect immune system health at large, which continually adapts to environmental circumstances [9]. Social patterns of immunity and infection may also give rise to differential trajectories of exposure and illness within populations. The example of CMV infection is instructive, given that there are striking disparities in congenital infection and early life seroprevalence depending upon race, ethnicity, and socioeconomic status [10,11].

Because the development and influence of the immune system are dependent on the timing of exposure to infections and long-term influence on immune system functioning, a life course lens leads us to different research questions that could shed light on prevention strategies and improve health across the lifespan. Research questions should be framed historically, as occurring within a specific time context, and holistically, focusing on genetic changes over time and multiple step processes to disease initiation rather than reducing the question to cause and effect of a single variable.

Case 2. Emerging Adolescent and Adult Type 2 Diabetes Mellitus (T2DM)

Individuals may be genetically predisposed to the development of T2DM, with rates of heritability ranging from 20% to 80%. To date, over 150 genetic variations that increase risk of T2DM diagnosis have been identified [12]. While genetic predisposition is an important contributor, it cannot fully explain the overall risk for the development of T2DM. T2DM is a heterogeneous phenotype [13] that involves mechanisms at multiple levels (individual, family, community, policy) that interact dynamically, consistent with the LCHD framework. A complex set of factors throughout the life course contributes to risk of disease in adulthood, some having additive or multiplicative interactive effects and others, such as interactions between transcription/translation regulation and metabolic networks, affecting risk through feedback loops [14].

Dose, duration, and timing of exposures may also have important roles between environment and T2DM risk. The level and length of time spent in unhealthy environments (i.e., dose and duration) may be important factors to consider when linking possible exposures to T2DM. The developmental period (i.e., timing)

of exposure is also an important consideration. During the prenatal period, higher maternal prepregnancy BMI, maternal hyperglycemia, and exposures to endocrine disrupting compounds are associated with later T2DM risk [15,16]. In newborns, low infant birth weight and associated rapid catch-up growth in the first 6 months of life increase the risk for T2DM in adolescence and young adulthood [17]. Furthermore, the infant gut microbiota, influenced by pre-, peri-, and early postnatal factors, may also affect risk associated with adult-onset disorders such as T2DM [18]. Childhood and young adult exposures such as poverty, adverse childhood experiences, depression, and engagement in risk behaviors (e.g., smoking, alcohol use) all further influence the risk for the development of T2DM in middle adulthood, and account for some racial and ethnic disparities observed [19].

Selecting tools from complexity science allows for holistic approaches to understanding the associations and can be used to assess the relative contributions of factors that could be altered to benefit health. For example, a recent study by Frisard *et al.* used the Women's Health Initiative Clinical Trials Cohort and Observational Study Cohort, both longitudinal cohorts, to investigate the association between antidepressant use over time and T2DM risk [20]. The authors used marginal structural models (MSMs) to allow for predictor variables such as antidepressant use and BMI to vary across an average of 7.6 years of follow-up. A handful of other studies investigating T2DM risk used similar approaches to allow for variation in measures, such as socioeconomic status, over time [21,22]. Other approaches have incorporated multiple types of genetic and dietary information. Network analyses have been used to integrate genetic, genomic, and functional data to explore the mechanisms predisposing individuals to T2DM [23,24]. Random forests have been used to predict metabolic syndrome status based on dietary and genetic data [25].

The current literature lacks in-depth studies that combine a spectrum of data ranging from genes to environment and utilize multilevel models and methods that embrace the complexity of T2DM. Novel research questions will likely require new methods that allow for the incorporation of time and complex interactions. For example, investigating the relationship between neighborhood, diet and exercise, differential DNA methylation over time, and resulting changes in fasting glucose and insulin metabolism in individuals transitioning from normal to pre-T2DM and T2DM requires a methodological framework that can model the complexity of the relationships without ignoring changes over time. By looking through a life course lens, drawing from existing tools and the creation of new tools, identified risk factors at the individual, parental, and environmental level along with the timing and interaction of the exposures could be used to direct targeted T2DM interventions.

Case 3. Cognitive Impairment/AD in Older Ages

Cognitive decline is an example of a complex phenotype that may have multiple interrelated risk factors across the lifespan, with genetics and environment playing an important role in shaping risk. Early life risk factors, including learning disabilities, educational attainment, impaired body growth and development, and lower childhood socioeconomic status all impact adult brain structure and function, and when present, increase vulnerability to the development of cognitive impairment/ AD in later life [26–30]. While AD is a major cause of dementia in Western societies, it develops asymptotically decades before diagnosis. Genes play a role, but by themselves rarely predict when or if AD will emerge [31], suggesting the need to consider the interactive role of environment.

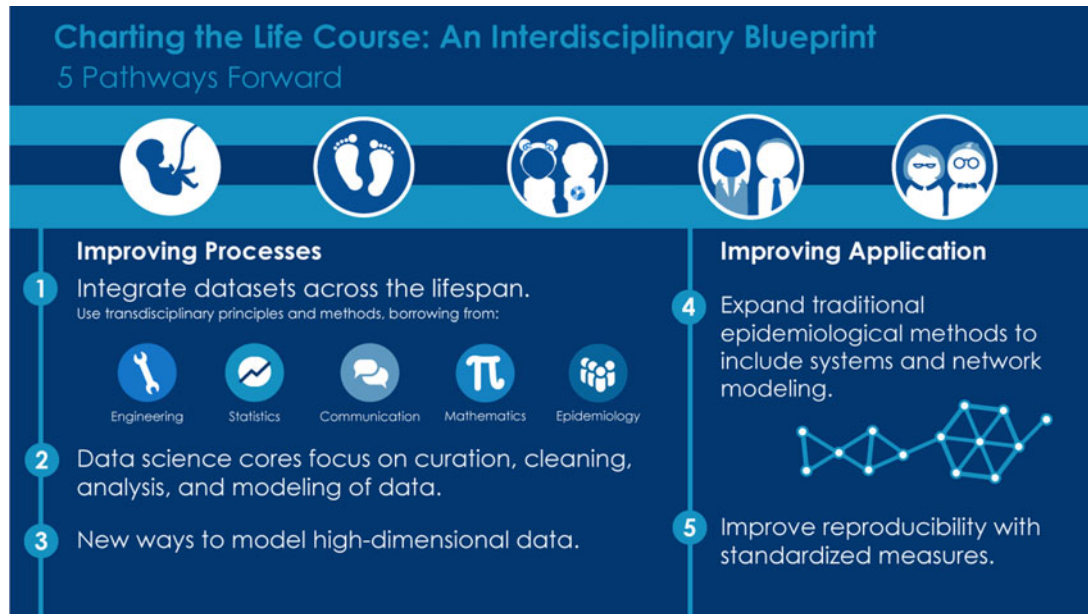


Fig. 1. Five recommended pathways for moving transnational life course research forward.

Some gene–environment interactions may be expressed in epigenetic markers; several epigenetic candidates demonstrate an association with dementia later in life [32]. This is referred to as the Latent Early Life Associated Regulation model which suggests epigenetic markers acquired from gestation to older age could modify gene expression that may induce or accelerate AD [33]. These nonheritable factors include infections, stress, nutrition, adiposity, and exposure to toxins (such as pesticides) which may have differing effects on risk of AD development depending on the degree and timing of the exposure, as well as the interaction of these factors [34].

While exposures such as these in early childhood can alter brain structure, predisposing the brain to increased risk for learning and memory dysfunction throughout life, exposure at other times in life can also affect brain health and risk for neurodegeneration. Chronic medical conditions, such as hypertension and T2DM in adolescence or adulthood, are associated with later cognitive impairment. This may be the result of defects in insulin signaling and consequent activation of neuroinflammatory pathways. However, the manifestation of cognitive impairment could be affected by control of the underlying condition, with cumulative effects of dysregulation over time exacerbating the onset and timing of cognitive impairment [35].

Moreover, it is not just early in life that these various exposures affect whether and when cognitive impairment will emerge, but also during middle age when there are many physiological changes. One such physiologic change includes an alternation in hormone levels. As women are 50% more likely to develop dementia sometime in their lives compared with men, multiple studies have turned to examining estrogen levels and total reproductive years to understand these associations [36]. However, specific mediators of neural injury have not yet been elucidated [31]. Women with pregnancy-related complications, such as preeclampsia, are also at increased risk of developing later life cognitive decline [37]. Using a complexity science lens would inform new hypotheses to look not only at estrogen levels, fertility, and pregnancy outcomes over one's lifespan, but examine the interactions of exposures such as infections, medications, chronic illnesses (such as T2DM), early

childhood brain development, and social deprivation. By examining interactions, timing of these interactions, dose of exposure, and duration over time, different types of interventions would emerge.

Another example of how this lifecourse framework could change the timing and type of interventions is the relatively new discovery that the gut microbiome mediates regulation of brain homeostasis, potentially resulting in the alteration of microglial activation and brain function in middle age. Middle-aged mice who received prebiotic diets had reversal of stress-induced immune priming and reduction in aging-related monocyte infiltration, modulating the peripheral immune response and altering neuroinflammation in middle age [38]. This could redirect our thinking – it is not only the *intervention type*, such as providing prebiotics, but the *intervention timing*, during middle age, that needs to be considered.

Taken together, the complex web of family history, genetic and epigenetic influences, environment and social context, chronic medical conditions, and microbiome influence require a life course approach to truly help us understand how to prevent or delay the onset and improve treatment of cognitive impairment and AD in older age [39,40].

Charting the Life Course Recommendations

We are not the first to note a need for a complexity science approach to the epidemiology of health and disease [41–43]; however, we think new technological innovations can now enable this approach to improve population health using a life course theoretical framework. We offer recommendations to build the necessary databases (process recommendations) that would allow for application of complexity science to generate and test new hypotheses (application recommendations). These five recommendations are summarized in Fig. 1. *Process recommendations* include the following: (1) creating systematic processes for longitudinal integration of datasets across the lifespan representing multiple levels of exposures from physiologic to sociologic; (2) utilizing data science core resources to prepare and package integrated datasets to make them accessible for researchers to generate and test new hypotheses; and (3) developing and validating ways to model high dimensional

data. *Application recommendations* include the following: (4) promoting and applying existing methods that fit a complexity science methodological framework; and (5) developing analytical methods that can capture multiple dimensions of time (timing, dose, and duration). Below, we elucidate next steps to act upon these recommendations.

Integrate Datasets Across the Lifespan

Prioritization for the creation of datasets that allow for investigations of the interplay between factors at the molecular level to the organ level within a single individual, the study of individual interactions with the environment, and the elucidation of how these factors vary between people and over time is necessary to move life course research forward. Large amounts of data from individuals in different contexts and across time will allow us to explore gene-environment interactions or family member to family member influences in ways that were impossible before.

It is becoming increasingly common for studies to link multiple large datasets to allow for a more in-depth examination of disease risk. For example, the Utah Population Database (UPDB) is an immense genealogical dataset that has been record-linked to many statewide datasets (including the Utah Cancer Registry and statewide medical claims from 1996 to present), with annual updates. The full dataset contains nearly 10 M people, 4 M of whom have 3–18 generations of genealogy available, and infrastructure that links distinct records for a specific person allows the UPDB to create a depiction of the life history of an individual based on medical and administrative data [44,45]. UPDB links EHRs data to government datasets, such as birth and death certificates, census record, geospatial data, and social determinants of health. Creating a “tapestry” of information sources allows for a wealth of measurement on factors influencing health and disease, identification of sensitive windows of exposure, and targets for prevention [46]. EHRs linked to government datasets, such as birth and death certificates, census records, geospatial data, and social determinants of health, as well as other large data (e.g., school, child welfare, and employment records).

Utilize Data Science Core Resources

As the amount of patient-related data linked to external sources increases, data science cores focused on the curation, cleaning, analysis, and modeling of data need developed. Data science builds upon principles and methods from across multiple disciplines, including ethics, engineering, statistics, communication, mathematics, and epidemiology. Organizing teams including experts from across these disciplines would allow for the development of innovative methods and increase interpretability of large amounts of complex data. In addition to data curation and methods development, these teams should be focused on developing ways to communicate results from the scientific community to the lay population, as well as standards for data sharing and for improving algorithm portability (such as the model adopted by the www.ohdsi.org) to enable a wider range of analysis on these curated datasets.

Develop New Ways to Model High Dimensional Data

In order to fully leverage data integrated from a vast number of sources, new ways to model high dimensional data collected serially need to be developed. Combining data across multiple sources and utilizing ML (variable selection) techniques can help us develop more accurate predictive models by identifying large-

scale properties of interacting variables and enabling targeted approaches by clustering patients with similar characteristics or phenotypes. This hypothesis-generating approach will allow us to expand on traditional hypothesis-driven approaches and find innovative ways to improve population health through pattern discovery and disease prediction. ML identifies patterns that explain the data available in both supervised (trained on an outcome) or unsupervised (trained to find a pattern) approaches. Pattern discovery algorithms are generally unsupervised, and can be used to refine our definition of phenotype [47, 47–50]. Algorithms are developed to seek out variables and combinations of variables to predict outcomes. ML differs from traditional regression methods because it can handle large amounts of combinations of data to examine interactions and nonlinear assumptions [51]. However, algorithms can also overfit models and lead to false conclusions. Consequently, data generated from ML approaches need validation from different populations. Moreover, outcomes are only as good as the data input and therefore can lead to biases reflected in the datasets themselves [52].

Expand Traditional Epidemiological Methods to Include Systems and Network Modeling

Complex systems assume that the functional form is nonlinear, distributions are nonnormal, individuals are heterogeneous, and health reflects multilevel structures, dynamic temporality, and interaction between factors. Traditional epidemiological methods need to expand to include systems modeling, network analyses, and machine learning. Complex systems analyses model the dynamic interactions between factors at the same level (e.g., cells, individuals) and across levels (e.g., tissue, neighborhood). Agent-based models are a commonly used approach for complex systems analyses and have been used in chronic disease epidemiology [53], health disparities [54], and health behaviors research [55] to explore dynamically complex processes and strengthen understanding of causal processes. Network analyses have been very useful for unraveling the complexity of genetic pathways and their interactions in chronic diseases. Phylogeography, a form of network analysis, is a method that can be used to study spatial viral disease distribution and would be useful for COVID-19-related studies [56]. ML methods take a more agnostic approach to epidemiological discovery by learning the models that best fit the data. ML methods, therefore, can play an important role in discovering unknown predictors of disease as well as refining phenotypic definitions of disease. Pattern identification, dimension reduction, and phenotypic definitions can be used to hone the search for causal mechanisms [45,47]. Methods embracing complexity science have the potential to identify factors having the largest impact on disease risk, the importance of duration and dose of exposure, periods across the lifespan that might be most amenable to intervention, the dynamic nature of risk, and moderating factors that should be considered.

Improve Reproducibility with Standardized Measures

As we move forward in identifying these measures, standard ways of operationalizing measures would improve reproducibility. Progress in recommendations 1–4 will be for naught without high-quality, semantically interoperable, and structured data. The penultimate solution would be a single American healthcare system, allowing for standardized records collection and storage. Until that goal can be achieved, investigators will have to continue to derive

Table 2. Examples of existing methods that can be used to investigate health across the lifespan using multiple interactions, time as a dimension, system science-based approaches, and computational approaches

	Challenges	Potential Solutions
Process Recommendations	Integrate Datasets Across the Lifespan	
	– Data storage	Close collaboration with computer science and data science colleagues.
	– Fragmented healthcare leads to disparate data types and storage	Development of novel ways to combine historical data that are flexible and scalable. Identify ways to quantify error when combining data from multiple sources.
	– Ethical aspects and need for data oversight committees	Data repositories should have an oversight committee that includes representatives from the community. Data storage and sharing ethics should be a common topic of conversation for staff involved in curating, maintaining, and distributing the data.
	– Data need to be standardized, curated, and high quality	Development of rigorous and sharable standards for standardizing and curating data. Quantify error related a dataset or measure.
	– Data ownership	Recognition that individuals are owners of their data. Outward facing programs that garner community involvement and knowledge about the datasets being used for health research. Discussion about how to give back to the community. Encourage discussion with the community about knowingly contributing data that can lead to medical innovations.
Utilize Data Science Core Resources	– Scalability	Data science core resources need to be able to serve a broad range of research questions across multiple disciplines. Processes should be developed that are general enough to allow them to be far-reaching.
	– Need to have a culture of data sharing	Academic health centers should openly enable, encourage, and reward data sharing.
	– Complying with rules for data access	Create easily accessible documents that clearly state data use and reuse rules. Educate data users, trainees, and investigators on responsible data use and reuse practices.
Develop New Ways to Model High Dimensional Data	– Requires transdisciplinary teams that may have differences in communication	Organize workshops, symposia, and meetings that regularly bring together scientists from multiple disciplines. Provide educational opportunities for investigators to learn how to present their research in lay language to make it accessible for all audiences.
	– Novel methods may be deemed high risk	Develop NIH initiatives, like Physical Sciences in Oncology (PSO), that bring together scientists from across a wide range of disciplines to address major questions and barriers in research.
	– Models are only as good as the data	Set benchmarks for data processing and transparency high.
Application Recommendations	Expand Traditional Epidemiological Methods to Include Systems and Network Modeling	
	– Requires training and a shift in epidemiological thinking	Develop easily accessible synchronous and asynchronous learning opportunities for investigators.
	– Models rest on assumptions about the distribution, sparsity, magnitude of effect, relationship between actors, and others that are carefully considered.	Create minimum standards for publishing complexity and machine learning results in health care journals. Create guidelines for systematic and rigorous research.
	– Algorithms and simulations are not substitutes for formal statistical modeling	Selection of models should be based on the epistemological foundation of the question. Methods should be combined with more traditional models to fully understand problems. Limitations should be clearly stated.
Improve Reproducibility with Standardized Measures	– Institutional specific software and data storage procedures	Develop standardized, extensible data schemas with well-defined entities, attributes, and metadata.
	– High variation in data types and structure across institutions.	Document data architecture and create templates to ensure standardized collection of data

innovative solutions to overcoming the fragmented nature of data collection and storage.

Conclusions

Health and disease evolve over time through complex interactions. New technology allows for the development of new methods and measures to effectively conduct life course research, applying a life course framework drawing from tools – from complexity science methodology to both generating and testing novel hypotheses. These recommendations are not without limitations (Table 2); however, the benefits outweigh the limitations. The field will benefit from a consistent approach to generate verifiable big data and a systematic approach to identify key methods and measures to maximize utility of that data. New statistical analytic approaches to handle the complexity of time, dose and duration of exposures, and context need to be developed and tested. These recommendations begin to chart the way forward in life course research to further develop the field and move it from theory to application.

Acknowledgements. The authors would like to thank Karen Bandeen Roche, William Hay, and Rashmi Gopal-Srivastava for their valuable feedback. They would also like to thank Brenna Kelly for designing Fig. 1.

This publication was made possible by the following CTSA grants from the National Center for Advancing Translational Science (NCATS), National Institutes of Health (UL1TR001067, UL1TR002001, UL1TR001873, UL1TR001442, UL1TR001876, UL1TR002369, UL1TR001414, U01TR002004, UL1TR002494, UL1TR001086, UL1TR001866). Heidi A. Hanson is also partially supported by the National Institutes of Health K07 Award 1K07CA230150-01. Nicole G. Weiskopf is partially supported by the National Library of Medicine K01 Award LM012738. Brad H. Pollock is partially supported by NIH UL1TR000002. Daniel Armstrong is partially supported by Agency for Community Living 90DDUC0031-03. Dan Cooper is partially supported by NIH R01HL110163 and P01HD048721. M Schleiss is partially supported by R01 HD079918. F. Kaskel is partially supported by NIHT32 NIDDK 5T32DK 007110-46. C Weng is also supported by the National Library of Medicine R01 LM009886 project. G Bandoli is also partially supported by the National Institutes of Health K01 Award K01 AA027811.

Disclosures. The authors declare no potential conflicts of interest.

References

- Barker DJP. Fetal origins of coronary heart disease. *BMJ* 1995; **311**(6998): 171–174.
- Roseboom TJ, Meulen JHPvd, Osmond C, Barker DJP, Ravelli ACJ, Bleker OP. Adult survival after prenatal exposure to the Dutch famine 1944–45. *Paediatric and Perinatal Epidemiology* 2001; **15**(3): 220–225.
- Halfon N, Forrest CB, Lerner RM, Faustman EM. *Handbook of Life Course Health Development*. Cham, Switzerland: Springer International Publishing, 2017.
- Kermany DS, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 2018; **172**(5): 1122–1131.e1129.
- Bennett TD, et al. Data science for child health. *The Journal of Pediatrics* 2019; **208**: 12–22.
- Furman D, et al. Cytomegalovirus infection enhances the immune response to influenza. *Science Translational Medicine* 2015; **7**(281): 281ra243–281ra243.
- Castrucci MR. Factors affecting immune responses to the influenza vaccine. *Hum Vaccin Immunother* 2018; **14**(3): 637–646.
- Brodin P, et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell* 2015; **160**(1–2): 37–47.
- Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell* 2014; **157**(1): 121–141.
- Fowler KB, et al. Racial and ethnic differences in the prevalence of congenital cytomegalovirus infection. *The Journal of Pediatrics* 2018; **200**: 196–201.e191.
- Manicklal S, Emery VC, Lazzarotto T, Boppana SB, Gupta RK. The “silent” global burden of congenital cytomegalovirus. *Clinical Microbiology Reviews* 2013; **26**(1): 86–102.
- Prasad RB, Groop L. Genetics of type 2 diabetes – Pitfalls and possibilities. *Genes* 2015; **6**(1): 87–123.
- Tuomi T, Santoro N, Caprio S, Cai M, Weng J, Groop L. The many faces of diabetes: a disease with increasing heterogeneity. *The Lancet* 2014; **383**(9922): 1084–1094.
- DenBraver NR, Lakerveld J, Rutters F, Schoonmade LJ, Brug J, Beulens JWJ. Built environmental characteristics and diabetes: a systematic review and meta-analysis. *BMC Medicine* 2018; **16**: 12.
- Hivert MF, Vassy JL, Meigs JB. Susceptibility to type 2 diabetes mellitus – From genes to prevention. *Nature Reviews Endocrinology* 2014; **10**(4): 198–205.
- Giulivo M, Lopez de Alda M, Capri E, Barceló D. Human exposure to endocrine disrupting compounds: their role in reproductive systems, metabolic syndrome and breast cancer: a review. *Environmental Research* 2016; **151**: 251–264.
- Eriksson JG, Kajantie E, Lampl M, Osmond C. Trajectories of body mass index amongst children who develop type 2 diabetes as adults. *Journal of Internal Medicine* 2015; **278**(2): 219–226.
- Tamburini S, Shen N, Wu HC, Clemente JC. The microbiome in early life: implications for health outcomes. *Nature Medicine* 2016; **22**(7): 713–722.
- Bancks MP, Kershaw K, Carson AP, Gordon-Larsen P, Schreiner PJ, Carnethon MR. Association of modifiable risk factors in young adulthood with racial disparity in incident type 2 diabetes during middle adulthood. *JAMA – Journal of the American Medical Association*. 2017; **318**(24): 2457–2465.
- Frisard C, et al. Marginal structural models for the estimation of the risk of Diabetes Mellitus in the presence of elevated depressive symptoms and antidepressant medication use in the Women’s Health Initiative observational and clinical trial cohorts. *BMC Endocrine Disorders* 2015; **15**: 56–56.
- Farmer RE, et al. Metformin use and risk of cancer in patients with type 2 diabetes: a cohort study of primary care records using inverse probability weighting of marginal structural models. *International Journal of Epidemiology* 2019; **48**(2): 527–537.
- Nandi A, Glymour MM, Kawachi I, VanderWeele TJ. Using marginal structural models to estimate the direct effect of adverse childhood social conditions on onset of heart disease, diabetes, and stroke. *Epidemiology* 2012; **23**(2): 223–232.
- Fernández-Tajes J, et al. Developing a network view of type 2 diabetes risk pathways through integration of genetic, genomic and functional data. *Genome Medicine* 2019; **11**(1): 19.
- Liu J, et al. An integrative cross-omics analysis of DNA methylation sites of glucose and insulin homeostasis. *Nature Communications* 2019; **10**(1): 2581.
- Szabo de Edelenyi F, et al. Prediction of the metabolic syndrome status based on dietary and genetic parameters, using Random Forest. *Genes & Nutrition* 2008; **3**(3): 173–176.
- Whalley LJ, Dick FD, McNeill G. A life-course approach to the aetiology of late-onset dementias. *The Lancet Neurology* 2006; **5**(1): 87–96.
- Evans DA, et al. Education and other measures of socioeconomic status and risk of incident Alzheimer Disease in a defined population of older persons. *Archives of Neurology*. 1997; **54**(11): 1399–1405.
- Russ TC, Kivimäki M, Starr JM, Stamatakis E, Batty GD. Height in relation to dementia death: individual participant meta-analysis of 18 UK prospective cohort studies. *British Journal of Psychiatry* 2014; **205**(5): 348–354.
- Wang X-J, Xu W, Li J-Q, Cao X-P, Tan L, Yu J-T. Early-life risk factors for dementia and cognitive impairment in later life: a systematic review and meta-analysis. *Journal of Alzheimer’s Disease* 2019; **67**: 221–229.
- McGurn B, Deary IJ, Starr JM. Childhood cognitive ability and risk of late-onset Alzheimer and vascular dementia. *Neurology* 2008; **71**(14): 1051–1056.
- Seifan A, Schelke M, Obeng-Aduasare Y, Isaacson R. Early life epidemiology of Alzheimer’s Disease: a critical review. *Neuroepidemiology* 2015; **45**(4): 237–254.
- Sanchez-Mut JV, Gräff J. Epigenetic alterations in Alzheimer’s Disease. *Frontiers in Behavioral Neuroscience* 2015; **9**: 347–347.

33. **Lahiri DK, Zawia NH, Greig NH, Sambamurti K, Maloney B.** Early-life events may trigger biochemical pathways for Alzheimer's disease: the "LEARN" model. *Biogerontology* 2008; **9**(6): 375–379.
34. **Miller DB, O'Callaghan JP.** Do early-life insults contribute to the late-life development of Parkinson and Alzheimer diseases? *Metabolism: Clinical and Experimental* 2008; **57**(2): S44–S49.
35. **Zilliox LA, Chadrasekaran K, Kwan JY, Russell JW.** Diabetes and cognitive impairment. *Current Diabetes Reports* 2016; **16**(9): 87.
36. **Gilsanz P, Lee C, Corrada MM, Kawas CH, Quesenberry CP, Whitmer RA.** Reproductive period and risk of dementia in a diverse cohort of health care members. *Neurology* 2019; **92**(17): e2005–e2014.
37. **Fields JA, et al.** Preeclampsia and cognitive impairment later in life. *American Journal of Obstetrics and Gynecology* 2017; **217**(1): 74.e71–74.e11.
38. **Boehme M, et al.** Mid-life microbiota crises: middle age is associated with pervasive neuroimmune alterations that are reversed by targeting the gut microbiome. *Molecular Psychiatry* 2019. doi: [10.1038/s41380-019-0425-1](https://doi.org/10.1038/s41380-019-0425-1)
39. **Thein MS, Igbineweka NE, Thein SL.** Sickle cell disease in the older adult. *Pathology* 2017; **49**(1): 1–9.
40. **Lin L, Zheng LJ, Zhang LJ.** Neuroinflammation, gut microbiome, and Alzheimer's disease. *Molecular Neurobiology* 2018; **55**(11): 8243–8250.
41. **Galea S, Riddle M, Kaplan GA.** Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology* 2010; **39**(1): 97–106.
42. **Luke DA, Stamatakis KA.** Systems science methods in public health: dynamics, networks, and agents. *Annual Review of Public Health* 2012; **33**(1): 357–376.
43. **Roux AVD.** Complex systems thinking and current impasses in health disparities research. *American Journal of Public Health* 2011; **101**(9): 1627–1634.
44. **Skolnick M BL, Dintelman S, Mineau GP.** A computerized family history database system. *Sociology and Social Research* 1979; **63**: 506–523.
45. **Hanson HA, et al.** Harnessing population pedigree data and machine learning methods to identify patterns of familial bladder cancer risk. *Cancer Epidemiology, Biomarkers & Prevention* 2020; **29**(5).
46. **Corn M.** Archiving the phenome: clinical records deserve long-term preservation. *Journal of the American Medical Informatics Association* 2009; **16**(1): 1–6.
47. **Hanson HA, et al.** Family study designs informed by tumor heterogeneity and multi-cancer pleiotropies: the power of the Utah population database. *Cancer Epidemiology, Biomarkers & Prevention* 2020; **29**(4).
48. **Sweatt AJ, et al.** Discovery of distinct immune phenotypes using machine learning in pulmonary arterial hypertension. *Circulation Research* 2019; **124**(6): 904–919.
49. **Zhou S-M, et al.** Defining disease phenotypes in primary care electronic health records by a machine learning approach: a case study in identifying Rheumatoid Arthritis. *PLoS One* 2016; **11**(5): e0154515–e0154515.
50. **Beaulieu-Jones BK, Greene CS.** Semi-supervised learning of the electronic health record for phenotype stratification. *Journal of Biomedical Informatics* 2016; **64**: 168–178.
51. **Mullainathan S, Spiess J.** Machine learning: an applied econometric approach. *Journal of Economic Perspectives* 2017; **31**(2): 87–106.
52. **Obermeyer Z, Emanuel EJ.** Predicting the future – Big data, machine learning, and clinical medicine. *New England Journal of Medicine* 2016; **375**(13): 1216–1219.
53. **Nianogo RA, Arah OA.** Agent-based modeling of noncommunicable diseases: a systematic review. *American Journal of Public Health* 2015; **105**(3): e20–e31.
54. **Langellier BA.** An agent-based simulation of persistent inequalities in health behavior: understanding the interdependent roles of segregation, clustering, and social influence. *SSM Popul Health* 2016; **2**: 757–769.
55. **Galea S, Hall C, Kaplan GA.** Social epidemiology and complex system dynamic modelling as applied to health behaviour and drug use research. *International Journal of Drug Policy* 2009; **20**(3): 209–216.
56. **Forster P, Forster L, Renfrew C, Forster M.** Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences* 2020; **117**(17): 9241–9243.
57. **Benham-Hutchins M, Clancy TR.** Social networks as embedded complex adaptive systems. *The Journal of Nursing Administration* 2010; **40**(9): 352–356.
58. **Paley J, Eva G.** Complexity theory as an approach to explanation in health-care: a critical discussion. *International Journal of Nursing Studies* 2011; **48**(2): 269–279.
59. **Miles A.** Complexity in medicine and healthcare: people and systems, theory and practice. *Journal of Evaluation in Clinical Practice* 2009; **15**(3): 409–410.