Application of Disease Etiology and Natural History to Prevention in Primary Health Care: A Discourse

Franklin White

Department of Community Health and Epidemiology, Dalhousie University, Halifax, NS, Canada; Pacific Health and Development Sciences, Victoria, BC, Canada

Highlights of the Study

• Etiology and natural history are keys to identifying intervention opportunities.
• Prevention is defined at three levels: primary, secondary, and tertiary.
• Strategies include health promotion, reducing risk factors, case finding, screening, addressing functional health, preventing errors, and integrated models.
• Health literacy and language are critical considerations for public communication and patient-centered care.

Keywords
Prevention · Etiology · Natural history · Primary health care · Health literacy

Abstract
The principles of etiology and natural history of disease are essential to recognizing opportunities for prevention across the illness spectrum. They have a bearing on how illness is experienced, how differently it can be perceived at the time of first contact with the health system, and how it may appear at later stages. Opportunities for prevention arise at every stage in the process, and three main levels are described: primary, secondary, and tertiary. Prevention strategies include health promotion focused on determinants, clinical prevention to reduce modifiable risk factors, case finding, screening, and addressing functional outcomes relevant to quality of life; the importance of preventing errors is also recognized. The distinction between incidence effects and treatment effects of prevention is explored. This review also examines the differing roles of language in health science and public communication, aspects of disease classification, related issues in patient-centered care, the prevention paradox, and integrated models of disease prevention.

Introduction
This review portrays the etiology and natural history of diseases as essential to primary health care (PHC) in promoting health and preventing disease. Although written for health professionals most likely to encounter individuals early in the process of seeking help for illness or injury, the principles are applicable at any stage of a person’s illness.
While no disease model is free from theoretical and practical limitations, especially considering the vast range of circumstances that affect health, etiology and natural history provide a robust framework that serves well to conceptualize illness entities, and ways of intervening in them. Alternative clinical, social, and environmental paradigms that can meet critical scrutiny should also be considered [1]. Using all applicable frameworks, practitioners must respond to the needs of people as they present with illness.

Disease itself has enduring value as a sustainable construct; the most obvious benefit is in making a diagnosis. According to a view expressed in a BMJ debate: “… acknowledgement by the medical profession that a patient’s condition has a name and is a legitimate illness is immensely reassuring and enabling… Crudely handled, medicalization can perpetuate disability and exclusion. But used constructively and appropriately it is the first step towards recovery” [2].

It is self-evident that not all diagnosed conditions will reach full recovery, but (at the very least) any person given a diagnosis will want to know its prognosis (the likely course and possible outcomes). Prognoses can vary from recovery to death, and a range of intermediate outcomes which may impact on function and quality of life. To respond intelligently to this expectation ideally requires that the natural history of the disease is understood, and that intervention efficacies are known.

Detailed knowledge about prognosis (with and without intervention) comes mainly from epidemiological studies that quantify the likelihood of possible outcomes, while efficacy estimates generally require the conduct of randomized controlled trials, and may vary with the characteristics of patients selected for study. For potentially fatal conditions, mortality estimates are calculated, for example the case-fatality rate (proportion of cases of a disease that are fatal within a specified period of time), which is often used for acute infectious diseases (IDs), and the 5-year survival rate (proportion of people alive 5 years after diagnosis, or a designated intervention), often used for cancer.

Despite prognostic implications, there exist more nuanced interpretations. The potential for stigma attached to particular disease labels is long recognized. A diagnosis in itself can adversely influence clinical management, if driven more by the label than by a patient’s actual needs [3]. Particular diagnoses may affect lifestyle, relationships, and the ability to earn a living. Aided by the internet, people are taking more control over their health, and sometimes perceive the medicalization of some conditions as unwarranted.

Recognizing such variations in perception may help practitioners to contextualize how different people interpret seemingly similar afflictions differently, and to understand why some are resistant to adopting approaches offered by the formal health system, and to appreciate the role of some alternative models and practices [4]. Even so, the disease focus remains legitimately at the core of much medical and public health practice.

The Importance of Language and Health Literacy

Before enlarging on aspects of diagnosis, etiology, natural history, and disease prevention, the role of language must be recognized, as it is linked to our cognitive ability to think, learn, reason, and communicate. Without common understanding, people may arrive at varying interpretations of what is written, read, or said; when this happens, it creates confusion across the spectrum from professional to public communication. Core PHC skills therefore include listening, asking questions, diagnosing, and recommending a course of action. Precision in language is critical for rigorous thought and clarity in expressing an approach to patient care and in providing public health information [5]. However, this linguistic principle does not stop there: rigorous use of language is never sufficient, as it must be integrated with effective knowledge translation to achieve language more readily understood by particular patients and the public at large, people mostly not steeped in biomedical science and technology. This is not a new idea: “…physicians absorb… diagnostic frameworks and population-based guidelines, and translate them… to the level of a single person whose illness is but one piece of life and whose profile never quite matches the one in the textbook” [6]. Also, among health care recipients, health literacy is a variable commodity: “…health literacy includes listening, speaking, and conceptual knowledge that make it possible to understand health interactions, forms, and instructions… health care environments have cultures of their own, ways of doing things, and uses of language that are different than what average persons experience in their day to day lives… The greater the difference between our lived experiences and those of others, the more likely our frames of reference will be different. Therein lies the potential for misunderstanding” [7].

The transition to a more “patient-centered approach” to clinical management owes much to the emergence of family medicine as a discipline that makes continuing care its special responsibility. Among the exponents was Ian McWhinney (1926–2012), a founder of the specialty when it was launched about half a century ago. He wrote:
“An understanding of the meaning of the illness for the patient should be as important for the physician as reaching a clinical diagnosis” [8].

Linguistic principles apply at all levels from local to international. Communicating more effectively in local languages can help implement public health strategies to support sick people and prevent transmission. For example, in the Democratic Republic of the Congo, the technical term “suspected case” meant “criminal” in the local Nande language, initially impeding efforts to manage an Ebola virus disease outbreak [9]. Public perceptions of newly medicalized disorders with accompanying diagnostic terms reveal that use of medical language in communication can bias how a condition is perceived: if it is a “disease,” how serious, and how common or rare is it [10]? The choice of language can influence a patient’s understanding, their decision to seek care, and whether to comply with prevention or treatment. Patient decision making, or self-triage [11], thus has implications for medical care. Over-emphasizing diagnosis, as distinct from what the illness means to the patient, therefore requires attention. A rational middle ground exists: neither indiscriminate acceptance of medicalization nor criticizing new terms just because they are medicalized can be justified [12].

“Primary Care” or “Primary Health Care”?
The terms “primary care” (PC) and “primary health care” are sometimes used interchangeably, but are not identical frameworks. Because they are mutually supportive and have a similar relationship to the natural history of disease, both are relevant here. PC is a profession-centered concept based on a clinical role: “Health or medical care that begins at time of first contact between a physician or other health professional and a person seeking advice or treatment for an illness or an injury” [13]. Described by some ethicists as “the traditional medical model” [14], this is an inaccurate label. It is not exclusive to medicine: it applies to many other kinds of service provider. Neither is it inherently “traditional” (taken to imply “adhering to customs that are respected because they are time-honored or integral to a certain culture or history”) [15]. To the contrary, although “time-honored,” the PC model is used because it works. Evidence shows that it helps to prevent illness and death, and (in contrast to specialty care) is associated with more equitable distribution of health in populations [16].

An important focus of PC is clinical decision making: any act of diagnosis based on presenting symptoms and signs that leads to a decision regarding prognosis, treatment, referral, or counseling. The task is often not straightforward: the meaning of symptoms can vary greatly between this first health system contact and the referral context; for example, the diagnostic spectrum for cough presenting in a community clinic will differ from that among patients referred to a specialized chest clinic. Similarly, persons referred with severe headache to a neurology clinic are more likely to have a brain tumor than the much larger number presenting with a similar complaint in an emergency clinic [17, 18].

Furthermore, among the reasons specialists are more likely to arrive at a valid diagnosis is that their case load has been triaged by clinicians in PC: preliminary testing has often been done, and the condition is often further advanced and easier to diagnose (its “pre-test probability” has been elevated). This should come as no surprise: it is predicted by Bayes’ theorem, a mathematical formula for applying conditional probability [19]. Bayes’ theorem is now introduced early in medical school and in standard texts, e.g., Chapter 3 in *Harrison’s Principles of Internal Medicine* [20]. Even so, adherence to Bayes’ principles is largely absent in clinical practice – low probability diseases are still tested for causing unneeded cost and risk, and high probability diseases often ignored when a single negative test result is returned [21].

The broader concept of PHC derives from a societal perspective more aligned with the role of organized public health systems: “Essential health care made accessible at a cost that a country can afford, with methods that are practical, scientifically sound and socially acceptable. Everyone should have access to it and be involved in it, as should other sectors of society beyond health. It should include community participation and education on prevalent health problems, health promotion and disease prevention, provision of adequate food and nutrition, safe water, basic sanitation, maternal and child health care, family planning, prevention and control of endemic diseases, immunization against vaccine-preventable diseases, appropriate treatment of common diseases and injuries, and provision of essential drugs” [1].

According to the World Health Organization (WHO), PHC “provides whole-person care for health needs throughout the lifespan, not just for a set of specific diseases… it is rooted in a commitment to social justice and equity and in recognition of the fundamental right to the highest attainable standard of health, echoing Article 25 of the Universal Declaration on Human Rights” [22]. For this review, it is relevant to highlight the phrase “not just for a set of specific diseases.” Yes, PHC is more than this, but WHO’s language explicitly includes disease within PHC, just as it encompasses PC as clinically defined.
The Language of Disease and Illness

The term “disease” does not convey the same meaning for everyone; it varies with role, context, and perception. Those directly affected experience it as “illness.” Informal caregivers (families, friends, others) become aware at an early stage. Self-care and mutual support may be sufficient to navigate the process, but may also be critical, depending on whether the formal health system enables access. From a purely clinical perspective, “illness” has been oversimplified as “the subjective sensation of experiencing a diseased state” [13]; it is more complex when viewed from the patient perspective: “it is possible for an individual to have a disease, yet be unaware of it and act accordingly; it is also possible for people to feel and/or act sick without showing evidence of any objectively verifiable disease. In the former instance there is no illness, though there may be disease. In the latter case there is certainly illness” [23].

The term “disease” has ancient roots in the English language, literally “lack of ease” or “uneasiness.” In the 21st century, it came to be defined as any departure from good health or from normal physiological and/or psychological function; this definition encompasses “disorders” but must not be misconstrued as referring to natural processes, such as normal childbirth, menopause, sexual preference, aspects of aging, and bereavement, which require distinct recognition. Nor should it be applied to signs or symptoms per se, for example a rash, a lump, elevated temperature, cough, pain, nausea, or weakness; although these may be expressions of underlying disease, until adequately investigated they do not in themselves constitute disease. In biomedical terminology, disease is defined by clinical, pathological, and epidemiological criteria that enable systematic study and application. At present, four major categories are commonly recognized: injuries, IDs, non-communicable diseases (NCDs), and mental and behavioral disorders, each category comprising many conditions.

Diseases pass through stages: susceptibility, pathological onset, pre-symptomatic, clinical, then resolution. Each may be modifiable by intervention, such as prevention, treatment, and rehabilitation, as well as by self-care and social adjustments. Depending on disease type and severity, outcomes vary from recovery to death, with intermediate outcomes such as impairment (a physical or mental defect at the level of a body system or organ), disability (temporary or long-term reduction of one’s capacity to function), or handicap (reduction of one’s capacity to fulfill a social role as a consequence of an impairment, disability, inadequate training for the role, or other circumstances) [4].

The first International List of Causes of Death, was adopted by the International Statistical Institute in 1893. Now known as the International Classification of Diseases (ICD), and published under WHO auspices since 1948, this goes through successive revisions. For example, WHO member states will use ICD-11 starting from 2022 [24]. With advances in knowledge and experience, disease taxonomy continues to evolve [25].

Reflecting ongoing research and development, the formal integration of psychosocial elements within the biomedical model gave rise in 1980 to a supplementary classification of Impairments, Disabilities, and Handicaps to ICD-9. This subsequently led to widespread adoption of an International Classification of Functioning, Disability and Health (ICF), which was launched in 2001 [26, 27]. The ICF is applicable to functional entities as outcomes or as starting points for clinical and public health intervention at the individual or population level. While this was a significant step forward [28] because abilities and participation are context dependent and related to quality of life, there are ongoing efforts to improve definitions and measurement [29, 30]. As famously stated by Greenwood: “The scientific purist, who will wait for medical statistics until they are nosologically exact, is no wiser than Horace’s rustic waiting for the river to flow away” [31].

Etiology

The term “etiology” means the science of causes; from a scientific perspective, all diseases must have causes. A cause is something that produces an effect; in epidemiology it is customary to distinguish necessary cause, sufficient cause, proximal cause, and distal cause. A necessary cause is one without which a condition cannot occur. Sufficient cause is defined as a set of minimal conditions and events that inevitably produce health, disease, and injury. A proximal cause is an immediate precipitating factor; a distal cause is more remote. These concepts are embedded within epidemiology [32], the discipline that studies the distribution and determinants of health-related states or events in specified populations, including diseases, causes of death, behaviors, responses to intervention/non-intervention, and the provision and use of health services.

It is now recognized that virtually all diseases have multifactorial causation; in other words, varying combinations of causes are required to produce the effect [33]. This gives rise to a composite framework of health and illness: tissues and organs operating at biological level, perception and experience at psychological level, and attribution of meaning at social level [34]; integrating these
elements is critical to understanding the clinical picture. While it was once argued that some IDs, genetic disorders, and traumatic injuries could be considered unifactorial, this was only ever true to the extent that the necessary cause was a defined microbiological agent, a defective gene, or a singular event such as an explosion. In causation as now understood, numerous factors play roles, and some may hold potential for prevention. In 21st century medicine, illness is viewed as a continuum that may flow in either direction, including reversibility in many conditions.

Factors relevant to multifactorial causation include the following [32]. Predisposing factors: those that prepare, sensitize, condition, or otherwise create a state of susceptibility so that the host tends to react in a deleterious fashion to a disease agent, personal interaction, environmental stimulus, or specific incentive. These factors are “necessary” but rarely “sufficient” to cause the phenomenon under study. Enabling factors: those that facilitate the manifestation of disease, disability, ill-health, or use of services or conversely those that facilitate recovery from illness, maintenance or enhancement of health status, or more appropriate use of health services. These factors may be “necessary” but are rarely “sufficient” to cause the phenomenon under study. Precipitating factors: those associated with onset of a disease, illness, accident, behavioral response, or course of action; usually one factor is more important or more obviously recognizable than others if several are involved and one may often be regarded as “necessary.” Reinforcing factors: those tending to perpetuate or aggravate the presence of a disease, disability, impairment, attitude, pattern of behavior, or course of action. They may tend to be repetitive, recurrent, or persistent and may or may not necessarily be the same or similar to those categorized as predisposing, enabling, or precipitating. Risk factors: an aspect of personal behavior or lifestyle, and environmental exposure, or an inborn or inherited characteristic that, on the basis of epidemiological evidence, is known to be associated with health-related condition(s) considered important to prevent. There are several types of risk factor: risk marker; determinant; modifiable risk factor; non-modifiable risk factor.

An understanding of multifactorial causation in itself may be applied to the design of prevention strategies, such as the Precede-Proceed model [35]; this employs an analysis of predisposing, reinforcing, and enabling factors in the design of behavioral and environmental interventions across the health spectrum. Similarly, actions on modifiable risk factors is a key to clinical prevention strategies, and is also utilized by public health organizations within health promotion strategies. In operational terms, etiology and natural history are interdependent, offering the potential to identify and integrate potential clinical and public health interventions if properly supported by relevant health systems development.

**Natural History of Disease**

The natural history of disease refers to the progression of a disease process in an individual over time, in the absence of treatment [36]. Hippocrates (c. 460–375 BC) was among the first to regard disease as a natural rather than a supernatural phenomenon, encouraging physicians to look for causes using objective observation and deductive reasoning [37]. In modern times, what is known about the natural history of any disease is constructed largely from observations of affected persons followed over time. Ideally this requires studies of defined cohorts which commence with the onset of the condition; however, although such studies have been carried out for numerous diseases, rigor can be difficult to achieve especially for chronic conditions with insidious onset that increasingly dominate the global burden. For example, inception cohorts for multiple sclerosis are virtually impossible to recruit because clinical onset is often discordant with biological onset and disease duration may exceed that of an investigator’s career, even their lifespan [38].

Viewed historically, much of what we know about the natural history of most diseases has been pieced together pragmatically, from astute clinical observations as well as formal studies of defined phases of a condition, whether this be studies of etiology, prevention and treatment efficacy, or of prognostic outcomes. Thereby, most often from multiple sources, a composite picture of a dynamic disease process (its natural history) is constructed. Relevant examples exist for all disease categories: injuries, e.g., meniscal tears [39], childhood trauma [40]; IDs, e.g., trachoma [41], human papilloma virus (HPV) [42]; NCDs, e.g., type 2 diabetes [43], Crohn’s disease [44]; mental and behavioral disorders, e.g., schizophrenia [45], bipolar disorders [46]. Such knowledge is continually reorganized in light of emerging research and experience, while management of emerging diseases, e.g., COVID-19, is rendered particularly challenging because so little is known to support initial interventions.

There is an honored tradition for such observational enquiry in the field of general practice that has strengthened with the emergence of family medicine as a specialty. Among its heroes is William Pickles, who wrote a classic...
entitled *Epidemiology in Country Practice*. His scientific contribution was to document, at a level not previously achieved, the development and evolution of ID epidemics. He did this by astute observation, charting, and excellent record keeping. In 1939 he wrote the following: “Let me recommend… particularly to those… entering country practice, this observation of the natural history of epidemic diseases… We… practitioners are in a position to supply facts from our observations of nature, and it is, I feel most strongly, our plain duty to make use of this unique opportunity” [47]. Stemming from such eclectic origins, the discipline of clinical epidemiology has become steadily formalized, contributing much to the elucidation of etiology and natural history of diseases, and to our existing and growing knowledge about disease prevention.

As illustrated in Figure 1 (a generic model), the passage of a person through time is represented by the thick line moving left to right. Thus, a healthy person may enter a subclinical process which in turn may cross a threshold to become recognized as clinical disease. The model shows how the person may emerge from the clinical situation with a range of possible outcomes, from death to partial or full recovery. This dynamic sequence holds for virtually all conditions: for IDs, transition from subclinical to clinical disease reflects the incubation period (time interval between invasion by an infectious agent and the onset of symptoms or signs), while for cancer (for example) it is the latent period or interval between exposure and manifestations (a complex multistep process where initial cell changes may become irreversible with progression to detectable neoplasia).

For individuals who cross the clinical threshold to become “sick,” treatment becomes the intervention priority. Nonetheless, opportunities for preventive intervention exist at every stage in the natural history. Treatment in itself may be preventive, not of the disease as it occurred but of its consequences in terms of chronicity or functional outcome; or it may be indirectly preventive such as the treatment of individuals with a disease transmissible from person to person, which may in turn reduce the rate of secondary transmission, e.g., individuals on antiretroviral treatment who have an undetectable virus load cannot transmit HIV to others; or it may result in a quarantine decision as in the current COVID-19 pandemic for which (at the time of writing) no fully effective drug or vaccine has been identified.

Any depiction of the natural history of disease has limitations. For example, risk factors and determinants shown under the primary prevention column (Fig. 1) may also function under subsequent columns dealing with secondary and tertiary prevention. Also, while the diagram emphasizes prevention in the PHC context, more specific diagrams may be applicable for diseases for which the intervention focus is on progression factors.

**Levels of Prevention**

Prevention refers to any intervention intended to stop something from happening. In PHC, this includes poli-
cies and actions to reduce the incidence and/or prevalence of disease, disability, and premature death, to reduce the prevalence of disease precursors and risk factors in the population, and, if none of these are feasible, to slow its progress and reduce associated disability and social impacts. This concept is usefully classified within three major levels: primary, secondary, and tertiary. In epidemiological terms, primary prevention aims to reduce disease incidence, secondary prevention aims to reduce disease prevalence by shortening its duration, and tertiary prevention aims to reduce the number and/or impact of complications. Two other levels are recognized in some contexts: “primordial” and “quaternary.” Primordial prevention aspires to establish and maintain conditions that minimize hazards to health [48] (as distinct from dealing with risk factors already present); however, because the aim of such efforts (in epidemiological terms) is to reduce disease incidence, it is appropriate to view primordial as a form of primary prevention. Quaternary prevention is defined by the World Organization of Family Doctors (WONCA) as actions taken to identify and protect patients at risk of over-medicalization and to promote ethically acceptable interventions [49]; this may be considered a form of tertiary prevention as already defined.

While PHC practitioners are generally well versed in aspects of health promotion and prevention, the extent to which such measures are incorporated in their practices, based on evidence, offers room for improvement [50]. For example, population health might be enhanced by capturing missed screening opportunities, as long as unnecessary diagnostic tests are avoided, given the potential harm that inappropriate testing can give rise to [51].

**Primary Prevention**

Primary prevention initiatives vary with the needs of individuals and populations through the life course, and include such measures as fortification of staples with minerals or vitamins, childhood immunization, smoking prevention, contraceptive counseling, nutrition guidelines, and eliminating contaminants in air, food and drinking water. Primary prevention aims to reduce disease incidence by intervening on modifiable risk factors or other preventive measures, e.g., health promotion, that influence the likelihood of disease outcomes. For a classic example, providing clean water, sanitation, and hygiene education has been shown to reduce the incidence of diarrheal diseases [52].

Incidence effects are well illustrated by mass vaccination against childhood infections, which leads typically to reductions in disease commensurate with the extent of coverage. Reducing the incidence in turn leads to reductions in rates of related complications and mortality which may also be considered an “incidence effect,” keeping in mind the time lag between incidence and the occurrence of related outcomes. For example, rubella vaccination leads not only to a reduction in its incidence, but also to a secondary reduction in congenital rubella syndrome (CRS); a core rationale for rubella vaccination is to protect pregnant women from this serious outcome in their offspring (CRS outcomes include congenital heart disease, hearing impairment, cataracts, and developmental delay) [53].

Illustrating a more delayed incidence effect is the use of vaccine to prevent infection by strains of HPV associated with long-term risk of cervical cancer. The International Agency for Research on Cancer recognized HPV as the “necessary” cause in the natural history of cervical cancer and HPV types 16 and 18 as carcinogenic agents based on strong and consistent associations between infection and disease [54]. Since 2007, HPV vaccination has been implemented in many countries; strong evidence of protective efficacy against premalignant and cancerous lesions emerged [55]. Based on such evidence, in 2018 the WHO launched a global goal of eliminating cervical cancer [56].

Incidence effects can also be reversed, as observed in the current upsurge of measles in many countries where vaccination levels have been allowed to slip [57]; this is largely a result of anti-vaccination groups disseminating misinformation. Comparable events have occurred before due to failure to promote scientifically reliable public information. In Britain, a reduction in pertussis vaccination in 1974 was followed by an epidemic of over 100,000 cases of pertussis and 36 deaths by 1978. In Japan during the same period, vaccination rates fell precipitously leading to a jump in pertussis from 393 cases and no deaths in 1974 to 13,000 cases and 41 deaths in 1979. In Sweden, the incidence rate per 100,000 children aged 0–6 years increased from 700 in 1981 to 3,200 in 1985 [58].

Considered within primary prevention, “primordial prevention” refers to measures that address underlying determinants of health that may require policy shifts and related actions. In the language of etiology, this addresses distal causes, while in terms of natural history it targets the underlying conditions that promote disease onset. As primary prevention it acts mostly to reduce incidence, but does so through more complex chains of causation such that its impact is generally less immediate and more diffuse. However, all of this can be considered within the...
operational domain of health promotion, including healthy public policy [59, 60]. In many respects, therefore, the distinction between primordial prevention and health promotion lies more in its professional constituency than its operational approach; thus, the term is encountered in some medical specialties, while the term “health promotion” is more widely understood across public health and PHC. The term “primordial prevention” is obtuse to the public at large, while the term “health promotion” offers an intuitive meaning to virtually everyone.

“Health promotion” itself is a broader concept that impinges on all levels of prevention, from primary to tertiary: “it consists of policies and processes that enable people to increase control over and improve their health. These address the needs of the population as a whole in the context of their daily lives, rather than focusing on people at risk of specific diseases, and are directed towards action on the determinants of health” [13]. Much of the current focus of health promotion involves a concern for health inequities, and acting on the underlying social determinants, e.g., power, money, and resources that give rise to them, e.g., deficits in human rights, literacy, gender equality, opportunity, and related needs for a “health in all policies” approach [61]. Keep in mind that, while social determinants are powerful, other determinants (biological, environmental, health systems) also influence health and also offer vitally important avenues for intervention.

Responsibility for primary prevention is mostly shared among front-line practitioners and public health staff working with communities. It is supported by regulatory agencies which set standards for air, water, and food quality, and offer specialized functions relating to radiation protection, hazardous products and wastes, and the regulation of vaccines, drugs, and devices. Similar measures apply to workplace hazards, their removal, reduction, or amelioration, e.g., air exchange, temperature control, ventilation, protective clothing and equipment, adequate lighting, and other best practices, thereby comprising a central role within the discipline of occupational health. They apply in the home: safe handling of food, building standards, child safety measures, and so on. At times of threat from epidemic disease and other community exposures, e.g., toxic chemicals, fire, floods, and heatwaves, primary prevention requires timely decisions and prompt actions [4]. There is no better example at present than the COVID-19 pandemic, where timely action to promote social distancing has been the most critical step in reducing transmission.

Secondary Prevention
This refers to early detection and prompt intervention to control disease and minimize disability. The “iceberg phenomenon” is relevant here, referring to a common situation where only a relatively small proportion of cases, the “tip of the iceberg,” comes to the attention of the health care system. The “submerged portion” includes disease not medically attended, attended but not accurately diagnosed, and diagnosed but not reported [62]; it may include inapparent infections (subclinical and incubating cases and carriers, which are significant in the spread of IDs). The proportion of missed cases varies with disease severity, especially during the early phases of the natural history when prevention is more likely to be effective. For examples, the submerged portion for type 2 diabetes is about 50% in many developed countries; for psychiatric disorders it may be as high as 80%. Corresponding estimates are higher for less developed countries. For some conditions such as HIV/AIDS and cancer of the cervix, the size of the submerged portion has decreased with improved case finding and screening methods.

The epidemiological outcome of effective secondary prevention can generally be considered a “treatment effect” – there is no incidence effect, except for reduced secondary transmission of some IDs. The potential for arresting epidemic spread underlies such preventive practices as contact-tracing, quarantine, and active treatment.

Secondary prevention measures fall mostly within two categories: case finding and screening programs. Such practices often include elements of health promotion, for example, in child health clinics, in addition to immunization (primary prevention), children may undergo anthropometry: parents of children not meeting height and weight norms may then be counseled on aspects of nutrition or referred for assessment. In clinical practice, opportunistic case finding during routine or periodic examination may yield previously undetected cases of chronic conditions such as hypertension, diabetes, and cancer, thereby leading to earlier intervention than would otherwise occur. In public health practice case finding is essential to communicable disease control through “contact tracing” individuals who have had close contact with a diagnosed case of conditions such as tuberculosis, sexually transmitted infections, COVID-19; it is also employed in investigating foodborne illness to identify those who may be at risk [63].

While case finding can be considered a normal clinical activity when carried out on the professional judgment of
a clinician in consultation with a patient, formal screening programs require a more rigorous decision-making process due to ethical, logistical, and natural history considerations. Screening programs are a complex health systems enterprise that may entail significant infrastructure investment in aspects such as laboratory support, information systems, institutional technologies, and quality assurance. This being so, we must clearly define what is meant by “screening.” Screening is defined as presumptive identification of unrecognized disease or defect by the application of tests, examinations or other procedures which can be applied rapidly. A screening test is not intended to be diagnostic: it sorts out apparently well subjects who probably have a disease from those who probably do not. A decision to offer a screening program must be justified in accordance with decision-making rules that take into account disease severity and prevalence in a given setting, that the proposed screening test has good performance characteristics, e.g., acceptable error rates and predictive values, the feasibility and acceptability of test procedures including cost and resource implications, attributes such as age, sex, family history, and whether risk factors are present, and evidence of intervention effectiveness. Screening programs vary with geographic jurisdiction depending on epidemiologically defined need and the availability of technical and financial resources. Some are offered to entire cohorts, e.g., antenatal, early childhood, occupational groups, while others may be offered to individuals at particular risk, e.g., families with a history of particular genetic conditions.

Among the key principles laid out by Wilson and Jungner [64] in 1968, in a highly cited document that has stood the test of time, is that “the natural history of the condition, including development from latent to declared disease, should be adequately understood.” Related to this, it has become an ethical requirement of screening initiatives that presumptive identification of disease will lead to improved prognosis. Although meeting this overall expectation is a challenging exercise on several fronts, many screening programs do achieve this status: for example, screening for high blood pressure, cervical cancer, and mammography for breast cancer. This duly noted, principles for justifying a screening program are not always respected by advocates of particular practices for which the evidence base may be flawed, incomplete, or controversial; common examples include breast self-examination [65] and PSA testing for prostate cancer [66].

In assessing screening initiatives, the “appearance” of an improved prognosis can arise as a result of bias (systematic error). The three most common are now described. Selection bias: error due to systematic differences in characteristics between those who take part in screening and those who do not: if those differences are associated with a better outcome, then the apparent “improvement” may be erroneously attributed to the screening program. Lead-time bias: over-estimation of survival time, due to the backward shift in the starting point in measuring survival, i.e., early diagnosis does not necessarily result in improved prognosis. Length bias: selection of disproportionate numbers of long-duration cases in one group (more likely to show up at any point in time, especially if screened) but not in another, i.e., unscreened people include those with all durations [4].

Tertiary Prevention

This consists of measures aimed at mitigating the impact of long-term disease and disability by: eliminating or reducing impairment, disability, and handicap, minimizing suffering, optimizing function and quality of life, and maximizing the potential years of life. Tertiary prevention targets both the clinical and outcome stages of a disease. It is implemented in symptomatic patients and aims to reduce the severity of disease impacts including associated functional sequelae, e.g., diabetic foot care education [67] and cardiac rehabilitation in post-myocardial infarction patients [68]. It envisions the preventive act in terms of achieving not only a better clinical outcome, but also in enhancing functional outcomes and restoring social roles. For example, special care dentistry has emerged as a discipline to facilitate oral care for people with intellectual and developmental disabilities (IDD) for whom conventional dental care presents challenges [69]; in fact, the Precede-Proceed model has been applied to an oral health strategy for adults with IDD [70]. Thus, tertiary prevention aims to encompass broad societal outcomes, e.g., reducing stigma in health facilities [71], harm reduction for addictions [72], and wheelchair-friendly building designs [73]. More needs to be done in all societies to address such needs in order to improve the quality of life for everyone affected by functional challenges; while not “treatment” in the usual medical sense, the necessary actions at policy level require advocacy and leadership from those on the front lines of medicine and public health.

Considering quaternary prevention as a subcategory of tertiary prevention, much of its rationale has emerged from over-medicalization of the elderly [49, 74]. However, such a perspective is relevant to all patients: people may suffer harm from medical interventions from conception, through childhood, and indeed during their entire life course. In this broader context, the approach is
described as “action taken to protect individuals from medical interventions that are likely to cause more harm than good” [75]. Examples include overtreatment and over-diagnosis, and actions include protection from unnecessary and ethically questionable examinations and interventions. Although the motivation behind this movement is to promote best practices, additional impetus comes from related medical malpractice claims where a particular conclusion is of interest: factors that predict that a patient will resort to litigation include a prior poor relationship with the clinician and a perception that the patient is not being kept informed [76].

Some observers approach these concerns from a systems perspective, focusing on the inevitability of medical errors. By recognizing that such events occur and learning from them, an effective approach to resolving them requires promoting a culture that recognizes safety challenges and implements viable solutions rather than harboring a culture of blame, shame, and punishment. The Joint Commission (a US accreditation body) thereby recognizes two major error types: (1) errors of omission as a result of actions not taken e.g., not strapping a patient into a wheelchair, (2) errors of commission due to wrong actions taken e.g., administering a medication to which a patient has a known allergy; stemming from this a typology of errors has been developed along with prevention modalities [77]. Because errors typically occur from the convergence of many contributing factors, this builds upon the principle of multifactorial causation, where the analysis of modifiable factors may serve as a basis for preventive strategies.

**Discussion**

As stated at the outset, etiology and natural history provide a robust framework that serves well to conceptualize illness entities, and ways of intervening in them. Viewing all knowledge as theory, the goal of science is the development of better theories [78], and research in turn is driven by uncertainty surrounding accepted theories. Thus, the biomedical concept of disease has grown to include psychosocial elements critical to functional health. The existence of initially competing paradigms did not require that we negate a general theoretical framework that itself grew out of ancient wisdom [37] and has been continually modernized. In the 21st century, this evolving model remains essential in defining and analyzing illnesses, and ways of intervening in them.

This duly stated, what has been reviewed here is by no means a unifying theory of disease. The potential for prevention so far explored has mostly to do with manipulating external environments, while future challenges may increasingly turn to human evolutionary biology for a more complete understanding. It unlikely that we will ever prevent all disease because our capacity to change all aspects of external environments is limited, as is our capacity to improve the evolved design of the human organism itself [79].

Sound preventive approaches tend to build upon one another. In fact, the natural history model is compatible with other models now in common use, such as Haddon’s matrix as applied to the design of injury prevention, the Precede-Proceed model that guides health education and promotion initiatives, and the social ecological model through which strategies are designed, implemented and assessed to ensure that preventive actions are reinforced at all levels from individual to family, community, and society at large.

Despite the transformation of the model, some bioethicists portray a limited perspective: “…as medicine focuses on changing individual’s bodies to reduce suffering, its increasing influence steals attention away from changing the social structures and expectations that can produce such suffering in the first place” [80]. Although the motivations behind such a view are no doubt intended to be constructive, especially by encouraging attention to the determinants of health, such assertions reveal a lack of insight or even awareness of how operational paradigms have already shifted. A basic mischaracterization is to reduce medicine to a focus “on changing individual’s bodies.” As discussed earlier, it is the integration across the biopsychosocial spectrum that completes the clinical picture; while more needs to be achieved in this direction, this particular assertion is simply archaic, perhaps even a “straw man” – creating a position that is easy to refute, then attributing that position to the medical profession [81].

More fundamentally, it is incorrect to portray medical and public health or health policy interventions as a zerosum game (a situation in which gains to one group or individual can occur only at the expense of losses elsewhere), and to do so reflects a lack of understanding of what PHC is actually about. Just because resources are allocated to dealing with disease at the level of individuals does not mean that other disciplines (including public health) are thereby impeded from addressing environmental and social determinants. Although more needs to be done to address determinants, even with the most optimal performance of health promotion and prevention efforts, people will still need medical care; it is not therefore a simplistic choice between one or the other ap-
proach, but getting the balance right in both developed and developing countries [82, 83]. In support of this, WHO’s Health in All Policies framework emphasizes the accountability of policymakers for health impacts at all policy levels, including an emphasis on the consequences of public policies on health systems, determinants of health, and well-being [84]. This in turn contributes to sustainable development.

A core principle underlying population health is that a large number of people at small risk may generate more cases of disease than a small number at high risk [85]. Related to this is the “prevention paradox”: a preventive measure that brings large benefits to the community may offer little to most participating individuals. For example, to prevent one vehicle accident death, thousands of people must wear seatbelts. Applying this principle to cardiovascular disease in East Asia, a reduction of just 3% in average blood pressure (as might be achieved by sustained reductions in dietary sodium or caloric intake) is estimated to reduce the incidence of disease (largely among clinically defined non-hypertensive persons) almost as much as would hypertensive therapy of all hypertensive persons in the population [86]. PC medicine, especially in the form of family practice, offers opportunities to improve continuity of care as part of a life-course approach to health, where integrated approaches become more feasible than with less coordinated and episodic medical care. It enables people to remain active in the workforce, helps to sustain household income, and encourages people to invest in their own and their children’s health and well-being [87].

Integrated approaches may be scaled up to society as a whole, perhaps best illustrated by the North Karelia Project. Half a century ago, extremely high cardiovascular mortality in North Karelia (a province in Finland), was of great concern to its population. In response, the North Karelia project was launched by the Finnish government in 1972. Health policy initiatives included tobacco control legislation and food industry participation (to reduce dietary fat and salt), supported by health education and promotion, while PC physicians and nurses organized around detection and intervention on modifiable risk factors at the individual level. After an initial 5 years, this strategy was extended to Finland as a whole. The main aims were to reduce extremely high serum cholesterol, blood pressure, and smoking levels with lifestyle changes and improved drug treatment, especially for hypertension. Major declines were observed for serum cholesterol, blood pressure, and smoking. Coronary mortality reduced in the middle age population by 84% over 4 decades (1972–2014). About two thirds of the mortality decline was explained by risk factor changes and one third by improved treatments. In the contemporary global situation of burgeoning NCDs, the North Karelia experience offered a powerfully motivating lesson [88] and similar initiatives were subsequently initiated in some 30 countries of Western Europe and the Americas, and several elsewhere in the world, collectively referred to as the CINDI-CARMEN-INTERHEALTH network [89].

Success in creating integrated health promotion and disease prevention systems requires vision, leadership and management skills, and training strategies to support front-line staff. The process should be guided by evidence and supported by monitoring and evaluation. To address disease burdens at population level takes vision and long-term commitment, measured in decades. All such initiatives are works in progress, and beyond the scope of this review to discuss further; an early review was published by the WHO [90].

Conclusion

Virtually all diseases are the result of interactions across our internal and external environments and behaviors that may influence the process at any stage from susceptibility and exposure risk to clinical outcome. While prevention owes much of its historical success to the conceptual and applied knowledge that continues to flow from the biomedical sciences, as health systems move to address more adequately the spectrum of disorders now dominating the global burden of disease, we must look beyond disaggregated risk factors and determinants to embrace how these interrelate throughout the life course, including biological experiences of early life, as well as cultural and social influences. The scientific approach to investigating any condition that affects people’s health starts with formulating theories, moves to gathering and analyzing data, then to inferences and interpretations, and closes with knowledge translation to specific audiences. Constructing and deconstructing, knowledge advances in small increments, and so it is with our understanding of the natural history of disease.

Acknowledgements

I thank Debra Nanan for critiquing the manuscript. This review is dedicated to the memory of our colleague, the late John Murray Last (1926–2019) whose epidemiology and public health reference texts are used around the world. His light continues to shine.
Conflict of Interest Statement

The author has no conflicts of interest to declare.

Author Contributions

Aside from the references cited, this review is otherwise entirely my own work.

Funding Sources

There was no applicable funding.

References

1. White F. Primary health care and public health: foundations of universal health systems. Med Princ Pract. 2015;24(2):103–16.
2. Wessey S. What do you think is a non-disease? Pros and cons of medicalisation. BMJ. 2002 Apr;324(7342):912.
3. Bedson J, McCarrison R, Croft P. Labelling chronic illness in primary care: a good or a bad thing? Br J Gen Pract. 2004 Dec;54(509):932–8.
4. White F, Stallones L, Last JM. Integrated approaches to disease prevention and control. In: Global public health – ecological foundations. New York: Oxford University Press; 2013.
5. Sarnat HB. Semantics do matter! Precision in scientific communication in pediatric neurology. J Child Neurol. 2007 Nov;22(11):1245–51.
6. Martin D, Pollack K, Woolard RF. What would an Ian McWhinney health care system look like? Can Fam Physician. 2014 Jan;60(1):17–9.
7. Partida Y. Language barriers and the patient encounter. Virtual Mentor. 2007 Aug;9(8):566–71.
8. McWhinney IR. Why we need a new clinical method. Scand J Prim Health Care. 1993 Mar;11(1):3–7.
9. Marzotto M. We won’t achieve zero Ebola cases in the Congo until we get the language right. The New Humanitarian. 2019 Decem-

10. Young ME, Norman GR, Humphreys KR. The role of medical language in changing public perceptions of illness. PLoS One. 2008;3(12):e3875.
11. Cooper AA, Humphreys KR. The uncertainty is killing me: self-triage decision making and information. Electronic J Appl Psychol. 2008;4(1):1–6.
12. Kaczmarek E. How to distinguish medicalization from over-medicalization? Med Health Care Philos. 2019 Mar;22(1):119–28.
13. Last JM. A dictionary of public health. New York: Oxford University Press; 2007.
14. Callahan D, Parnes E. The ends of medicine: shaping new goals. Bull N Y Acad Med. 1995;72(1):95–117.
15. Stack Exchange. English language and usage. https://english.stackexchange.com/questions/166803/conventional-vs-traditional (accessed October 18, 2019).
16. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q. 2005;83(3):457–80.
17. White F, Nan D. Clinical decision making Part I: errors of commission and omission. J Pak Med Assoc. 2003 Apr;53(4):157–9.
18. Nanan D, White F. Clinical Decision Making Part II: why diagnostic procedures vary in performance. J Pak Med Assoc. 2003;53(5):194–7.
19. White F, Nan D: A Bayesian approach to treatment efficacy. Ann R Coll Physicians Surg Can. 1999;32(2):111.
20. Mark DB. Decision-making in clinical medicine. In: Fauci AS, Braunwald E, Kasper DL, et al, editors. Harrison’s principles of internal medicine. 17th ed. New York: McGraw-Hill Professional; 2008.
21. Medow MA, Lucey CR. A qualitative approach to Bayes’ theorem. Evid Based Med. 2011 Dec;16(6):163–7.
22. World Health Organization. Primary health care. Fact sheet. Geneva: WHO; 2019. https://www.who.int/news-room/fact-sheets/detail/primary-health-care Accessed 16 October 2019.
23. Idler EL. Definitions of health and illness and medical sociology. Soc Sci Med. Med Psychol Med Sociol. 1979 Nov;13A(6):723–31.
24. World Health Organization. Classifications. International classification of diseases 11th revision (ICD-11). Geneva: WHO; 2018.
25. Sturmberg JP, Martin CM. Diagnosis - the limiting focus of taxonomy. J Eval Clin Pract. 2016 Feb;22(1):103–11.
26. World Health Organization. International classification of impairment, disability and handicap. Geneva: WHO; 1980.
27. World Health Organization. International classification of functioning, disability and health. Geneva: WHO; 2001.
28. Jones RB. Impairment, disability and handicap—old fashioned concepts? J Med Ethics. 2001 Dec;27(6):377–9.
29. Barbotte E, Guillemin F, Chau N; Lorhandi-
