

Review

Cancer Cluster Investigations: Review of the Past and Proposals for the Future

Michael Goodman^{1,†}, **Judy S. LaKind**^{2,3,4,†,*}, **Jerald A. Fagliano**^{5,†}, **Timothy L. Lash**^{1,†},
Joseph L. Wiemels^{6,†}, **Deborah M. Winn**^{7,†}, **Chirag Patel**^{8,†}, **Juliet Van Eenwyk**^{9,†},
Betsy A. Kohler^{10,†}, **Enrique F. Schisterman**^{11,†}, **Paul Albert**^{11,†} and **Donald R. Mattison**^{12,13,†}

¹ Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Road, Atlanta, GA 30322, USA; E-Mails: mgoodm2@emory.edu (M.G.);

timothy.lee.lash@emory.edu (T.L.L.)

² LaKind Associates, LLC, 106 Oakdale Avenue, Catonsville, MD 21228, USA

³ Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Howard Hall Suite 200, 660 W. Redwood Street, Baltimore, MD 21201, USA

⁴ Department of Pediatrics, College of Medicine, Milton S. Hershey Medical Center, Pennsylvania State University, 500 University Drive, Hershey, PA 17033, USA

⁵ Division of Epidemiology, Environmental and Occupational Health, New Jersey Department of Health, P.O. Box 369, Trenton, NJ 08625, USA; E-Mail: Jerald.Fagliano@doh.state.nj.us

⁶ Division of Cancer Epidemiology, Department of Epidemiology & Biostatistics, School of Medicine, University of California, Helen Diller Family Cancer Research Building, HD 274 1450 3rd Street, San Francisco, CA 94158, USA; E-Mail: Joe.Wiemels@ucsf.edu

⁷ Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, 9609 Medical Center Drive, Bethesda, MD 20892, USA; E-Mail: WINNDE@MAIL.NIH.GOV

⁸ School of Medicine, Stanford University, 1265 Welch Road, Stanford, CA 94305, USA; E-Mail: cjp@stanford.edu

⁹ Washington State Department of Health, P.O. Box 47812, Olympia, WA 98504, USA; E-Mail: juliet.vaneenwyk@doh.wa.gov

¹⁰ North American Association of Central Cancer Registries, Inc., 2121 W. White Oaks Drive, Suite B, Springfield, IL 62704, USA; E-Mail: bkohler@naaccr.org

¹¹ Division of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, 6100 Executive Blvd., Bethesda, MD 20892, USA; E-Mails: schistee@mail.nih.gov (E.F.S.); albertp@mail.nih.gov (P.A.)

¹² Risk Sciences International, 325 Dalhousie Street, Ottawa, ON K1N 7G2, Canada

¹³ McLaughlin Centre for Population Health Risk Assessment, University of Ottawa, 325 Dalhousie Street, Ottawa, ON K1N 7G2, Canada; E-Mail: mattisond@risksciences.com

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: lakindassoc@comcast.net;
Tel./Fax: +1-410-788-8639.

*Received: 27 November 2013; in revised form: 13 January 2014 / Accepted: 20 January 2014 /
Published: 28 January 2014*

Abstract: Residential clusters of non-communicable diseases are a source of enduring public concern, and at times, controversy. Many clusters reported to public health agencies by concerned citizens are accompanied by expectations that investigations will uncover a cause of disease. While goals, methods and conclusions of cluster studies are debated in the scientific literature and popular press, investigations of reported residential clusters rarely provide definitive answers about disease etiology. Further, it is inherently difficult to study a cluster for diseases with complex etiology and long latency (e.g., most cancers). Regardless, cluster investigations remain an important function of local, state and federal public health agencies. Challenges limiting the ability of cluster investigations to uncover causes for disease include the need to consider long latency, low statistical power of most analyses, uncertain definitions of cluster boundaries and population of interest, and in- and out-migration. A multi-disciplinary Workshop was held to discuss innovative and/or under-explored approaches to investigate cancer clusters. Several potentially fruitful paths forward are described, including modern methods of reconstructing residential history, improved approaches to analyzing spatial data, improved utilization of electronic data sources, advances using biomarkers of carcinogenesis, novel concepts for grouping cases, investigations of infectious etiology of cancer, and “omics” approaches.

Keywords: cancer; cluster investigations; cancer biomarkers; case grouping; leukemia; exposome; infection

1. Introduction

Residential clusters of non-communicable diseases are a source of enduring public concern, and at times, controversy [1–3]. Compared to clusters in which cases are linked by common occupation such as working with asbestos in a cluster of mesothelioma [4], or share an unusual risk factor such as prenatal exposure to diethylstilbestrol in a cluster of clear cell carcinoma of the vagina [5], clusters that appear to arise in a given geographic area or in a given community are particularly difficult to study.

Descriptions of non-occupational geographic clusters of cancer (primarily leukemia) can be found in the literature as far back as the beginning of the 20th century [6] and published systematic reviews of this issue span nearly 40 years [7,8]. Other diseases that have been reported to cluster in space and

time include birth defects [9,10], autism [11–13], multiple sclerosis [14,15], amyotrophic lateral sclerosis [16,17] and suicide [18,19]. While a wide array of health outcomes have been reported to cluster, what sets cancer clusters—and especially pediatric cancer clusters—apart are the frequency with which they are reported and the existence of population-based cancer registries to readily and accurately identify cases in a defined geographical area. We therefore focus on cancer clusters in this paper, although much of the content would apply in equal measure to clusters of other diseases.

The published recommendations on how to conduct cluster investigations have remained largely unchanged over the last three decades. In 1981, Aldrich [20] proposed starting with a definition of the potential cluster event, followed by the determination of the population at risk, and then an assessment of whether further study is warranted. Once a full study is deemed necessary, the investigators would consider developing a study questionnaire aimed at testing “the battery of reported theories related to the specific disease etiology” [20].

In 1989, a National Conference on Clustering of Health Events summarized the preceding twenty years of experience of cluster investigations and discussed specific methodological features of such investigations [21–25]. In 1990, the Centers for Disease Control and Prevention (CDC) issued their Guidelines for Investigating Clusters of Health Events [26]. The CDC guidelines outlined a four-stage approach, which was similar to that proposed by Aldrich [20], and included the following components: initial contact and response (Stage 1); an assessment to confirm existence of a cluster (Stage 2); an evaluation of feasibility of a full scale epidemiologic study (Stage 3); and, if warranted, a formal etiologic investigation (Stage 4).

In 2007, CDC issued an addendum to the 1990 guidelines by specifically addressing investigations of cancer clusters [27]. Among the criteria used to justify the move from initial to more complex stages of investigation were: a statistical excess of a single type of cancer; a rare cancer type; a common cancer in an unusual age group; or suspected exposure to a known carcinogenic agent with sufficient elapsed time since exposure [27]. The most recent CDC guidelines for cancer cluster investigations were issued in late 2013, and continued to recommend the previously adopted four-stage approach. In addition, the 2013 guidelines highlighted data sources and statistical techniques that could be used in cancer cluster investigations, and described possible approaches for developing effective communication strategies. The stated goals of these updated guidelines were “to provide needed decision support to public health agencies in order to promote sound public health approaches, facilitate transparency and build community trust” [28].

While goals, methods and conclusions of cluster studies are debated in both the scientific literature [8,25,29–31] and the popular press [32–35], investigations of reported residential clusters rarely provide definitive answers about disease etiology [8,36–40]. Further, it is inherently difficult to study a cluster for a disease with complex etiology and long latency such as most cancers. Despite this difficulty, evaluation of clusters remains an important function of local, state and federal public health agencies. Early and timely involvement of public health agencies is critical because a poor initial response can result in missed opportunities for an investigation and education and may increase the level of uncertainty and concern in a community, potentially resulting in the need to later expend additional public health resources.

Several recent reviews argued that progress in cluster research may require fundamental, rather than incremental, changes in methodology, and have recommended the development and testing of novel or

previously understudied hypotheses [8,41–44]. The current communication summarizes deliberations of the multi-disciplinary two-day workshop “*Advancing Cancer Cluster Assessments: Starting the Dialogue*” held in April 2013 with the goal of advancing the search for new approaches to studying this issue. The workshop included researchers with specific relevant expertise in epidemiology, biostatistics, informatics, exposure science, clinical medicine, disease surveillance, and risk communication. Workshop participants came from a variety of settings, including federal and state public health agencies, academic and government research organizations, and the private sector. Although several participants had first-hand involvement in cluster investigations, the workshop did not focus on findings from previous studies, but rather used past experience to identify key issues that need to be considered in future cluster investigations. The results of the workshop discussions are described here. We first review definitions and goals associated with cancer cluster investigations, then describe investigation-related challenges, and finally describe novel or under-explored approaches that could potentially be added to the arsenal of current approaches for investigating clusters. It is the hope of the workshop participants that this communication will prompt those involved in various aspects of cancer cluster investigation (representatives of the community, health agencies, and academic research institutions) to consider new ways of thinking about this long-standing problem.

2. What is a Cancer Cluster and What are the Goals of Investigating Clusters?

In its 1990 guidelines, the CDC defined a cluster as “...an unusual aggregation, real or perceived, of health events that are grouped together in time and space and that are reported to a health agency” [26]. CDC later sharpened the definition, in the context of cancer investigations, as “...a greater-than-expected number of cancer cases that occurs within a group of people in a geographic area over a defined period of time” [27,45]. This re-definition focuses on the cluster as a statistical excess in a specified population, geographic area, and time period, and is not dependent on its perception, reporting, or existence of a common cause.

Many clusters reported to the public health agencies by concerned citizens are accompanied by an expectation that an investigation will uncover a specific environmental cause of disease in the affected community [3,30,37,46]. By this measure, with few exceptions, cancer cluster investigations have not been successful [8]. Public health authorities and researchers acknowledge that cluster investigations rarely find statistical associations between local factors and disease incidence, and further that these investigations cannot demonstrate causality [8,27,31,37].

However, while understanding the role of known or perhaps novel risk factors is an objective of cluster investigations, it may not be the only objective. Even if following an investigation the etiology of disease remains unclear, the report of a cluster by the community and the proposed link to a possible cause can sometimes bring to light public health, environmental, social or other problems that should and could be mitigated even if not directly related to the community-reported concern. Neutra [31] emphasized that, as “part of good, empathetic public health practice”, health agencies need to have trained staff to promptly respond to concerns about potential clusters, including assessment of disease occurrence as well as environmental factors of concern to the community. The 1990 CDC Guidelines [26] noted that “reports of clusters cannot be ignored,” and public health agencies should adopt a leadership role in responding to concerns that “maintains community relations...without

excessively depleting resources". The intention of CDC's guidelines and many states' cluster response protocols is to screen and prioritize reports to limit investigations to those most likely to produce meaningful results [27,37]. Similarly, Condon *et al.* [29] noted that health agencies have a responsibility to the public to respond to community concerns, and that interactions in the course of an investigation provide opportunities to educate an engaged group of citizens on the frequency, etiology, and prevention of cancer, as well as on exposure issues of concern. Further, without this engagement, health agencies might miss the rare instances where cancer cluster investigations using current methodologies might be productive. This engagement allows health agencies to address environmental issues or other locally important cancer-related factors, such as screening [29].

Thus, cancer cluster investigations may best be seen as the fulfillment of a health agency's general mission to protect and improve health, rather than as a basic research program in the environmental etiology of cancer. However, in terms of advancing the basic (as opposed to applied) science of cancer etiology and prevention, researchers will remain interested in exploring clusters in terms of causality and those types of basic science explorations will most often fall outside the scope of health agency activities.

3. Cancer Cluster Investigation Challenges

While a wide array of health conditions aggregate in space and time, cancer clusters present several unique challenges for the affected community and for health agencies and researchers. These challenges, which drive the need for continued thought on novel approaches for investigating cancer clusters, are briefly described here:

Timing of disease development: Most malignancies have induction periods measured in decades. Exceptions to this are cancers in infants and children (where by definition the induction period cannot be longer than months or several years), leukemias arising from radiation and chemotherapy treatments for certain cancers [47], and cancers in immunosuppressed organ transplant recipients [48,49]. This long induction period presents a particular challenge in investigations of residential cancer clusters [50], because even though current address is routinely collected in cancer registries, the complete residential history is usually not available. True geographic clusters may need to be defined by the co-localization of individuals many years prior to the cancer diagnoses.

Defining a "case" for inclusion: Case identification and classification present additional problems in cancer cluster investigations. A reported cluster may comprise individuals with a very rare and histologically distinctive cancer such as glioblastoma multiforme [51]. However, most reports of cancer clusters include cases presenting with cancers of different organs that are not known to have a common etiology or common genetic basis. Attempting to determine a common underlying cause in this type of situation will likely produce a misleading result or no conclusive result. Further, even cancers that arise from the same organ and have the same International Classification of Disease (ICD) code (e.g., acute lymphoblastic leukemias) may represent different molecular types and have different etiologic mechanisms and should therefore not be viewed as a single group of cases [52].

Problem of small numbers: Sparsely populated geographic regions often experience wide year to year fluctuations in the number of cancer cases, leading to unstable estimates of cancer incidence. This impedes researchers' ability to establish presence or absence of a cluster [37]. Small numbers of cases complicate the implementation of case control studies aimed at testing causal hypotheses,

because these studies tend to lack statistical power and often produce measures of associations that are too imprecise to allow meaningful conclusions.

Defining boundaries and cluster area populations: The boundaries of perceived clusters are often based on social or neighborhood networks involving known cases rather than on the more relevant boundaries dictated by exposures of interest [53]. This misspecification may limit our ability to identify a cluster and to understand its etiology [54]. The result can be either failure to detect a true cluster (due to exclusion of potentially relevant cases) or observing a cluster where none exists (by excluding exposed disease-free individuals).

Migration: Due to the long induction period between exposure to a carcinogen and development of disease, some exposed members of a population may no longer be living in the community where a cluster develops, resulting in under-counting of cases. Conversely, a case contributing to the overall cancer cluster may have been exposed to a carcinogen from an earlier exposure in a different geographic region resulting in over-counting of cases. In either situation, population movement in or out of a community may result in misclassification of exposure [55]. The effect of migration on cluster investigations may be particularly pronounced and difficult to assess if migrants and those who do not change residence differ with respect to socioeconomic, exposure, demographic or health-related characteristics [56].

Challenges related to cancer registries: Population-based cancer registries are the best source of data for measuring cancer burden in a geographic area and over time [57]. Cancer registries in the USA are certified annually by the North American Association of Central Cancer Registries based on the completeness, timeliness, and accuracy of data, which has contributed to highly standardized and reliable data. While these registries are fundamental to understanding the distribution of cancer in time and space, they do not currently contain all of the information necessary for investigating cancer clusters (e.g., residential history). As information reported to the registries comes exclusively from medical records, most data on personal behavioral risk factors or environmental exposures are not captured. Complete ascertainment of cancer cases can take up to two years from the date of diagnosis, due to local reporting laws and the complexity of the data [58]. For this reason, using registry data to confirm a reported excess of cancer cases can delay confirmation for up to two years.

4. Proposed Novel or Under-explored Approaches for Investigating Cancer Clusters

The previously noted lack of success in identifying environmental risk factors through investigations of residential clusters indicates a need for fundamentally novel—rather than incrementally improved—approaches. Several novel or under-explored potentially productive approaches are described, each in different stages of development by academic researchers and/or health agencies. Each approach has advantages as well as obstacles to its implementation. We recognize that adoption of new tools will likely require additional resources that may not be currently available to public health agencies or academic researchers and that obtaining necessary resources may require concerted and combined efforts of state/federal health agencies and academic research institutions. However, the addition of one or more of these tools to the current armamentarium may help in advancing our ability to detect clusters and improve our understanding of etiology of disease clusters.

4.1. Rapid Case Ascertainment

As mentioned previously, an important barrier to cluster investigations is the time lag between diagnosis and complete enumeration of cases in cancer registries, potentially resulting in a missed possible cluster that would only be detected when all reporting for that time period is complete. This time lag could be avoided or minimized through rapid case ascertainment (RCA) methods whereby initial information about newly diagnosed cases is obtained through expedient transmission of pathology reports [59]. The modern RCA systems such as ePath collect electronic pathology reports and notify registry personnel or eligible researchers about new cancer cases with very little delay thereby allowing continuous assessment of cancer occurrence [60]. More recently developed approaches take advantage of the ePath technology by using natural language and knowledge-based processing to identify relevant tumor information in free text pathology reports. Software performing these tasks is currently being tested at several cancer registries [61]. Although modern RCA methods could play an important role in cluster investigations, their full integration into day-to-day cancer surveillance practices will likely take several years. Meanwhile, improving timeliness and completeness of cancer registration should be emphasized [60,62] by utilizing informatics technology including RCA methods and matching with relevant and evolving medical record databases.

4.2. Reconstructing Residential History

As nearly all cancers have protracted latency, current address, which is readily ascertained from the registry data, may be less important than residential history. Until recently, this presented a nearly insurmountable methodological limitation of cluster investigations, which had to use interviews to account for residential mobility. In recent years, however, residential history data have become increasingly available through population directories. Many of these directories are commercially available and could be used to construct residential histories during a cluster investigation.

One recent study assessed the accuracy of residential histories in a population directory from LexisNexis, Inc. (Miamisburg, Ohio, USA). The analysis compared residential histories recorded in the LexisNexis directory to information collected from written surveys in a case-control study of bladder cancer in Michigan. The lifetime addresses obtained from LexisNexis and those reported in the surveys matched for 71.5% of participants [63]. The authors concluded that while higher accuracy is desirable, the availability of residential history from population directories such as LexisNexis represents a “vast improvement over the assumption of immobile individuals currently used in many spatial and spatiotemporal studies”.

4.3. Application of Spatial Statistics

Traditional approaches of working through the steps of cluster investigations [28] involve assessing rates for administrative geographical units such as ZIP codes or census tracts. An alternative approach is to examine clustering of disease in time and space untethered to pre-defined geographic units. This methodology was first suggested more than two decades ago [64], but computational and data management barriers at that time were formidable. Modern computer technology, however, enabled rapid developments in geospatial statistics and the practical applications of new methods of

identifying and investigating clusters of cancer and other diseases. A number of currently available global clustering statistical tests are aimed at evaluating presence or absence of “hot spots of disease” on the map [64–68]. These tests, all based on the null hypothesis of “spatial randomness” have been reviewed in detail previously [69]; most were found to perform well, but depend heavily on the underlying assumptions. An example of a practical application of these global clustering tests is the recent analysis of brain cancer mortality in the USA, which demonstrated that brain cancers were more common in parts of Arkansas, Mississippi and Oklahoma, but found no specific localized clusters [70].

The use of tests of spatial randomness without an a priori expectation may be viewed as an advantage because performance of a test can be assessed in on its own merit; or a disadvantage because, as noted in a review by Kulldorff *et al.*, the findings of statistical analyses “may or may not correspond to true and interesting geographic patterns of the disease” [69]. Lawson proposed addressing this issue by applying a Bayesian approach that first incorporates a priori distribution for the study area and time of interest based on a pre-existing concern, and then performs a statistical assessment using one of the clustering tests [71,72].

As statistical methods for evaluating spatial patterns of health conditions and risk factors continue to develop [73]; their refinement presents a number of practical challenges. For example, it is important to keep mind that enhanced granularity of spatial data may require new ways of protecting confidentiality [74].

4.4. Continuous Monitoring of Registry Data

Although cluster analysis is not used in all registries, it could potentially be incorporated into routine practice assuming sufficient staff training and allocation of resources. In terms of feasibility, some state-based cancer registries use software such as SaTScan™ to verify community-reported cancer clusters and to find hot spots of late stage disease and other indications of screening and treatment need for selected cancers. Conducting constant monitoring could enable these agencies to quickly detect and investigate cancer clusters regardless of whether community members also report the same cluster and to perform descriptive epidemiologic studies that identify geographic aggregation of certain malignancies (see for example, [75]). Proactive scanning, even on a daily basis, is commonplace in influenza surveillance or in monitoring of asthma attacks, *i.e.*, conditions that are common, and develop relatively quickly following rapid changes in the environment. By contrast, monthly or even yearly proactive scanning presents a much greater challenge in cancer surveillance because the true changes in cancer occurrence are relatively slow and because small numbers of cases observed in a limited geographic area tend to produce incidence estimates that are unstable and difficult to interpret.

Continuous monitoring is not without limitations. One issue is the potential obligation for public health agencies to investigate and communicate findings of all software-identified cancer clusters. This obligation may overwhelm sparse public health resources at some agencies. A second issue is the need to verify whether data mining methods are up to the task of cancer cluster identification (e.g., can data mining be used to address the previously mentioned lack of historical residential data?). Lastly, spatial uncertainty must be addressed. Spatial uncertainty is the lack of, or the error in, knowledge about geographical position such as patients’ addresses that include P.O. boxes or rural routes (these are known to mischaracterize geographic location). It is also unclear how that uncertainty

affects any association between environmental exposures and disease [76–78]. Further, research on how to visually display the extent of uncertainty is needed [79]. Although geographic information systems are becoming increasingly sophisticated in terms of addressing this issue, more research is needed to improve statistical methods and spatial data collection and quality control [80]. Thus, before state health departments embark on proactive monitoring for cancer, researchers need to verify that this approach has utility, given issues of latency and mobility, multiple comparisons, and temporal instability caused in part by small numbers. Issues associated with potential harm due to false positives as well as communication and ethical issues must also be evaluated.

4.5. Improved Utilization of Electronic Data Sources

The linking of cancer cluster information with other forms of now rapidly digitalized health data [81] such as the electronic health record [82], population characteristics, and health care resources [83–85] in real-time could be helpful in pinpointing potential causal agents.

A key concern is patient privacy; individuals must usually provide consent before such a linkage could occur. Alternatively, linkages could be performed with de-identified personal and geographic data. Current technical and practical barriers that would need to be resolved include incompatible data sets, lack of data standards, and data quality/integrity concerns. In order to resolve conflicts between data sets, researchers could utilize tools currently employed in software engineering to document digital processes (e.g., modifications in data formats and structure) and to track and ensure data integrity when consolidating multiple data sources.

Because reports of residential cancer cluster investigations emanate from a residential network, another novel opportunity would involve harnessing the social network for data gathering on exposures and lifestyles. Development of a common interview, which could potentially be administered over the internet or mobile devices, would allow its rapid dissemination to members of the residential network. This interview could be customized to examine topics of particular relevance to each cluster, including the environmental issues of greatest concern. Linking the common interview questions across multiple potential cancer clusters may identify commonalities that would be missed when each is evaluated alone. This strategy would also help address the common problem of sparse data, which often plagues residential-based cancer cluster investigations. Members of the affected residential network could be enlisted to aid with customization of the common interview and with data collection (e.g., by linking residents to the interview and/or by directly collecting data via mobile devices).

Information about clusters may be indirectly ascertained from digital social networks, which could shed light on individuals' lifestyles and behavior from their interactions on these digital networks [86]. Kosinski *et al.* [87] demonstrated how preferences captured in Facebook (“likes”) predict behaviors of clusters of a social network, such as alcohol intake, smoking status, and narcotic use. For example, the more an individual “likes” to drink alcohol, the higher the probability that that individual’s social network also prefers that behavior. Hurdles that must be overcome in order to use this type of approach include obtaining consent from entire networks (as a coherent whole) and addressing the proprietary nature of these data. Information derived from these services needs to be validated for its utility in public health surveillance.

4.6. Advances Using Biomarkers of Carcinogenesis

Traditional case-control studies of cancer clusters are problematic in part because of small sample sizes and the inability to control for confounders [88]. However, novel study designs that take into account new appreciations of the biological or “natural” history of cancer as a disease may help facilitate future investigations. Cancer is understood to evolve over a period of years or decades, with each new characteristic induced by multiple genetic and epigenetic changes. This concept originated with the recognition of the stepwise morphologic and genetic evolution of colon cancer [89]. Now many cancer sites have been described in exquisite genomic and epigenomic detail, with documented sequential progression of disruptions in the normal cell physiology [90]. This progression can be highly variable. Most age-related epithelial cancers have long latencies with more than five genetic mutations required. By contrast, childhood and therapy-related cancers (such as those that are related to the *MLL* gene) require only a small number of genetic changes and may have latencies of only several months [91,92]. Interestingly, any given population will harbor some persons carrying pre-cancerous cells; indeed, all individuals harbor some mutations that can contribute to cancer if more mutations occur [93,94]. Space-time clusters of cancer are likely to be related to causal factors that put an entire community at risk, but only impact cancer incidence in those at-risk individuals that have precancerous cells at the verge of becoming tumorigenic. Thus, a cause of the cluster may trigger the disease in only a small number of cases even though many individuals were exposed. This consideration underscores the need for alternative endpoints that are associated with increased risk of cancer (i.e., biomarkers of risk), but are detectable prior to tumor occurrence [95]. The relevant biomarkers available as endpoints can reflect genetic, epigenetic or RNA related changes as well tissue-based differences in protein levels.

The use of biomarkers as endpoints has been advocated for cancer prevention trials [96], but could also apply to evaluation of clusters or effects of environmental exposures. Before these biomarkers are used in population studies, however, they need to be validated against clinically meaningful outcomes to avoid misinterpretation of results.

4.7. Novel Concepts for Grouping Cases

With a trove of longitudinal clinical examinations and measures increasingly available in the health record, clinical characteristics of cases to be included in clusters could be better defined. For example, routine characterization of myeloid leukemias has evolved from the French-American-British Classification, which relied primarily on morphologic features, to the 2001 World Health Organization (WHO) classification which recommended cytogenetic assessment, to the 2008 WHO classification, which combines morphologic, cytogenetic, and molecular analyses [97]. Most recent clinical recommendations for the management of acute myeloid leukemia (AML) in children and adolescents indicate that an AML diagnostic workup should include at a minimum “morphology with cytochemistry, immunophenotyping, karyotyping, FISH (fluorescent *in situ* hybridization), and specific molecular genetics in the bone marrow” [98]. This example indicates that information on molecular characteristics of tumors is becoming increasingly available in the medical records and therefore can be used in cancer cluster investigations. The current SEER (Surveillance, Epidemiology and

End Results) coding system (ICD-O-3) includes the most relevant cytogenetic and morphologic criteria and simply adopting this coding scheme will help to incorporate the most pertinent and specific diagnostic details in a systematic fashion [99].

It is becoming increasingly clear that many cancer types can be subdivided into entities based on molecular characteristics that may have distinct etiologies, prognoses, and responses to therapies [100]. Molecular markers of tumors are increasingly being incorporated in routine practice to establish cancer progress or guide treatment; cancer registries are beginning to find ways to capture these data, as well [100]. Molecular information could be collected from medical records of individuals within putative cancer clusters and be used to classify cases into more homogenous subgroups for analysis; this has the potential to be useful in uncovering etiologic factors that are relevant to only certain of the cancer subtypes. For example, triple negative breast cancer has some shared and some different risk factors compared to other forms of breast cancer [101].

In addition, biomarkers can be used in cancer cluster investigations to identify tumors with similar molecular characteristics that may share a common cause. Our current method of classifying cancer by primary site (e.g., organ) and/or broad histological type may insufficient for understanding cancer etiology. Cancer cells may share common characteristics regardless of cancer site [102] and common cellular pathways for growth and survival exist across multiple tissues. These characteristics include rapid cell growth, resistance to apoptotic signals, uncoupling of differentiation and cell division, and maintenance of the ends of chromosomes (telomeres). An example is the mutation of *TP53* or *RAS* genes which are mutated across cancers of the lung, colon, pancreas, blood, skin and other sites [103]. These common mutations in disparate cancers may have similar causes, for instance nucleophilic chemicals or aflatoxin [104]. Another example would be *IDH* mutations, common in leukemia, brain cancer, and cartilaginous tumors, and related to broad epigenetic patterning [105].

Therefore, it is possible that for cancer cluster investigations, cancers should be reclassified according to subtype within a major cancer type as well as according to their carcinogenesis features such as presence of mutations or epigenetic changes as opposed to location or appearance. For some cancers (e.g., pediatric leukemias), this type of data may already be available in medical records. For other types of cancers, data are currently being collected only for research purposes.

4.8. Infection and Cancer Clusters: An Example of Pediatric Leukemia

Pediatric leukemia is a disease known to involve genetic aberrations that occur during distinct time periods: the first aberrations occur during pregnancy (prenatally) and subsequent aberrations occur postnatally [106]. Leukemia incidence is, at least to some extent, calendar time-dependent, although not unequivocally seasonal [107] and thus leukemia clusters are likely an expression of postnatal causal events which have impacted communities at about the same time. It is hypothesized that such causal events are likely to be infectious [108–110]. For example, flu epidemics are often followed by transient increases in leukemia rates [111]. Further, a widely publicized leukemia cluster in Niles, IL was reported to be “accompanied by the parallel appearance of rheumatic-like illness” in the same community, suggesting a common infectious etiology [112].

A more recent example of a potential infectious cause of leukemia is found in the description of the Fallon, NV cluster, which affected children from 2 to 19 years of age and included a range of

common childhood leukemia diagnoses [113]. All leukemia cases occurred in the space of three years and most were restricted to one year [43]. With such a disparate age range and leukemia subtype diagnoses, the cluster is unlikely to be linked to cancer “initiating” events that occur prenatally [106]. The initiating mutations occurring earlier in the children’s life may have dissimilar causes and identities, leading to different subtypes of leukemia at different ages, despite disease diagnoses being tightly clustered in time. The epidemic appearance of the cluster only makes sense as a clustering of “secondary genetic events” precipitated by a new environmental stimulant such as infection, one that might have been introduced to the community from the town’s transient military population [43].

Similarly, an apparent cluster of seven cases of childhood acute lymphoblastic leukemia (ALL), which occurred over a four-week period in Milan, Lombardy, Italy, was associated with an outbreak of the AH1N1 influenza virus which occurred several weeks prior to the diagnoses [114]. The authors note that this is “compatible with the “delayed infection” hypothesis for childhood ALL in which an abnormal immune or inflammatory response to a common infection promotes ALL in susceptible individuals”.

Infection is not the only potential cause for time-dependent clustering, as shown by other examples of leukemia clusters that may have been incited by chemical stimuli [115,116]. However, infection remains a viable theory in leukemia clustering (e.g., “population mixing” theories [117]), and the role of infection in leukemia and other cancers is currently under exploration using sequencing and discovery methods similar to that described for new emerging viral illnesses [118].

Considering that cancer clusters (if related to a common a cause) are likely to be a response to a proximate (in time) change in the environment and also are likely to be a rare response to a common factor, cluster investigations should focus on the identification of factors that have impacted the community at large rather than just the individuals who contracted cancer. The likelihood of success for this type investigation would be increased if it were performed immediately upon identification of a cluster. Such an investigation can compare a community with other communities that have not experienced similar health outcomes, and focus on agents that factor into the known etiology of specific cancer types. For example, Steinmaus *et al.* [119] examined the Fallon, NV cluster in this fashion by comparing the Fallon community with other communities of similar size in different locations that held military bases.

For leukemia, infectious stimuli can be explored by reviewing hospital records and registry data to search for unusual co-occurrences of related health events prior to or concurrent to the cluster. Biological samples can be retrieved from cancer cases and community members (tumor and constitutive material) to test for specific hypotheses (infectious agents), or in the absence of specific tests more exploratory profiling of chemical and infectious exposures. Academic or industry laboratories that could help support such efforts should be recruited at early stages if possible.

4.9. “Omics” Approaches

To deal with the complexity of multiple exposure factors that are difficult to study using existing methods, researchers have developed the concept of the exposome, which describes the “totality of environmental exposures” an individual encounters from birth to death [120]. The exposome concept was introduced as an analog to the genome, which encapsulates almost all of the hereditary information of an individual and consists of 3 billion chemical bases that encode about 20,000 genes.

Genomic technologies are already used to examine clustering of disease. For example, Palacios *et al.* [121] discovered a novel pathogen in a cluster of patients who developed encephalopathies shortly after a solid organ transplant from a single donor. By applying high-throughput sequencing technology of samples from deceased patients, the investigators were able to isolate genetic material of the causal virus amongst a complex mixture of host microflora without any *a priori* knowledge of the infectious agent. Like the genome, studies of the exposome may be designed to query various combinations of environmental factors. Such studies may be possible after ascertainment of a “baseline” or “reference” exposome from population-based biomarker surveillance data [122,123]. Unlike the genome, however, the technology needed to ascertain an individual’s exposome is still in the conceptual stage.

5. Conclusions

In this communication, we reviewed the challenges associated with successfully identifying community cancer clusters and their causes and described scientific advances—in various stages of maturity—that could potentially be harnessed to improve our ability to conduct community cancer cluster investigations in a way that might lead to a better understanding of cancer etiology. Following are key conclusions and recommendations:

1. The challenges to understanding why cancers may cluster in time and space were first enumerated several decades ago, but still limit investigations today.
2. While understanding the role of known or perhaps novel risk factors is an objective of cluster investigations, health agencies have a responsibility to the public to respond to community concerns. Interactions during a cluster investigation provide opportunities to bring to light a public health, environmental, social or other health problem as well as to educate an engaged group of citizens on the frequency, etiology, and prevention of cancer, as well as on exposure issues of concern.
3. Advances in our understanding of cancer development and cause, coupled with new methods of spatial statistics and novel technologies, present opportunities for examining cancer clusters in novel ways and may lead to greater success in identifying cancer clusters and understanding cancer cluster etiology.
4. Technological advances may also improve the collection of information on residential history and population characteristics.
5. Biological advances can improve the use of biomarkers for understanding cancer etiology, for identifying and defining cases, and considering under-explored possible causes of cancer clusters such as infection.

The advances described here, including those that are in the early stages of development, will require a commitment of resources in order to bring these various approaches to fruition. While cluster investigations serve several purposes, public health protection related to cancer cluster investigations will ultimately derive from fundamentally improved methods for investigating those clusters.

Acknowledgments

The Steering Committee, which was responsible for development of workshop goals and direction, included Michael Goodman, Rollins School of Public Health, Emory University, Betsy Kohler, North American Association of Central Cancer Registries, Judy S. LaKind, LaKind Associates LLC., School of Medicine, University of Maryland, College of Medicine, Pennsylvania State University, Donald R. Mattison, Risk Sciences International, University of Ottawa, Enrique F. Schisterman, Eunice Kennedy Shriver National Institute of Child Health and Human Development, and Kimberly Wise, American Chemistry Council. The authors thank Li Zhou of the National Cancer Institute, for commenting on some statistical aspects of this paper. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of RFHEE or the National Institutes of Health.

Author Contributions

All authors participated in the workshop and contributed equally to this work.

Conflicts of Interest

This workshop was supported by a grant from the Research Foundation for Health and Environmental Effects (RFHEE). RFHEE was not involved in the design, collection, management, analysis, or interpretation of the information in the manuscript; or in the preparation or approval of the manuscript. Workshop participants or their affiliated organizations received an honorarium (except Jerald Fagliano, Deborah Winn, Juliet Van Eenwyk, Enrique Schisterman, Paul Albert) and travel support (except Jerald Fagliano, Deborah Winn, Enrique Schisterman, Paul Albert, Betsy Kohler). Judy LaKind received support for Workshop development; Judy LaKind and Michael Goodman consult to governmental and private sectors. Joseph Wiemels is supported by the following grants: NIEHS/EPA P01ES018172, NCI R01CA155461, NIEHS R01ES09137. Chirag Patel is supported by the following grant: NHLBI T32HL007034. Michael Goodman receives salary support from the National Cancer Institute's contract N01 PC35135. Donald Mattison is an employee of Risk Sciences International and the McLaughlin Centre for Population Health and consults to government and private sectors, and conducts research with government, private sectors and independently. No other competing interests are declared.

References

1. McCoy, H.V.; Trapido, E.J.; McCoy, C.B.; Strickman-Stein, N.; Engel, S.; Brown, I. Community activism relating to a cluster of breast cancer. *J. Commun. Health* **1992**, *17*, 27–36.
2. United States Congress. ATSDR: Problems in the Past, Potential for the Future? Hearing before the Subcommittee on Investigations and Oversight, 2009. Available online: <http://www.gpo.gov/fdsys/pkg/CHRG-111hrg47718/html/CHRG-111hrg47718.htm> (accessed on 6 July 2013).
3. Winn, D.M. Science and society: The Long Island Breast Cancer Study Project. *Nat. Rev. Cancer* **2005**, *5*, 986–994.
4. Otte, K.E.; Sigsgaard, T.I.; Kjaerulff, J. Malignant mesothelioma: Clustering in a family producing asbestos cement in their home. *Br. J. Ind. Med.* **1990**, *47*, 10–13.

5. Herbst, A.L.; Ulfelder, H.; Poskanzer, D.C. Adenocarcinoma of the vagina—Association of maternal stilbestrol therapy with tumor appearance in young women. *N. Engl. J. Med.* **1971**, *284*, 878–881.
6. Boyle, P.; Walker, A.M.; Alexander, F.E. Methods for investigating localized clustering of disease. Historical aspects of leukaemia clusters. *IARC Sci. Publ.* **1996**, *135*, 1–20.
7. Caldwell, G.G.; Heath, C.W., Jr. Case clustering in cancer. *Southern Med. J.* **1976**, *69*, 1598–1602.
8. Goodman, M.; Naiman, J.S.; Goodman, D.; LaKind, J.S. Cancer clusters in the USA: What do the last twenty years of state and federal investigations tell us? *Crit. Rev. Toxicol.* **2012**, *42*, 474–490.
9. De Wals, P. Investigation of clusters of adverse reproductive outcomes, an overview. *Eur J. Epidemiol.* **1999**, *15*, 871–875.
10. Elliott, L.; Loomis, D.; Lottritz, L.; Slotnick, R.N.; Oki, E.; Todd, R. Case-control study of a gastroschisis cluster in Nevada. *Arch. Pediatr. Adolesc. Med.* **2009**, *163*, 1000–1006.
11. Bertrand, J.; Mars, A.; Boyle, C.; Bove, F.; Yeargin-Allsopp, M.; Decoufle, P. Prevalence of autism in a United States population: The Brick Township, New Jersey, investigation. *Pediatrics* **2001**, *108*, 1155–1161.
12. Gee, A. Californian autism clusters leave researchers baffled. *Lancet* **2010**, *376*, 1451–1452.
13. Matsuishi, T.; Shiotsuki, Y.; Yoshimura, K.; Shoji, H.; Imuta, F.; Yamashita, F. High prevalence of infantile autism in Kurume City, Japan. *J. Child Neurol.* **1987**, *2*, 268–271.
14. Nicoletti, A.; Iorio, S.; Reggio, E.; Tarantello, R.; Liberto, A.; Iorio, F.; Patti, F.; Reggio, A. A possible spatial and temporal cluster of multiple sclerosis in the town of Linguaglossa, Sicily. *J. Neurol.* **2005**, *252*, 921–925.
15. Schiffer, R.B.; McDermott, M.P.; Copley, C. A multiple sclerosis cluster associated with a small, north-central Illinois community. *Arch. Environ. Health* **2001**, *56*, 389–395.
16. Proctor, S.P.; Feldman, R.G.; Wolf, P.A.; Brent, B.; Wartenberg, D. A perceived cluster of amyotrophic lateral sclerosis cases in a Massachusetts community. *Neuroepidemiology* **1992**, *11*, 277–281.
17. Sienko, D.G.; Davis, J.P.; Taylor, J.A.; Brooks, B.R. Amyotrophic lateral sclerosis. A case-control study following detection of a cluster in a small Wisconsin community. *Arch. Neurol.* **1990**, *47*, 38–41.
18. Fowler, K.A.; Crosby, A.E.; Parks, S.E.; Ivey, A.Z.; Silverman, P.R. Epidemiological investigation of a youth suicide cluster: Delaware 2012. *Del. Med. J.* **2013**, *85*, 15–19.
19. Robertson, L.; Skegg, K.; Poore, M.; Williams, S.; Taylor, B. An adolescent suicide cluster and the possible role of electronic communication technology. *Crisis* **2012**, *33*, 239–245.
20. Aldrich, T.E. A procedure for investigating cancer cluster reports. *Med. Hypotheses* **1981**, *7*, 809–817.
21. Anonymous. Proceedings of the National Conference on Clustering of Health Events, Atlanta, GA, USA, 16–17 February 1989.
22. Caldwell, G.G. Twenty-two years of cancer cluster investigations at the Centers for Disease Control. *Amer. J. Epidemiol.* **1990**, *132*, S43–S47.
23. Goodman, R.A.; Buehler, J.W.; Koplan, J.P. The epidemiologic field investigation: Science and judgment in public health practice. *Amer. J. Epidemiol.* **1990**, *132*, 9–16.

24. Ross, A.; Davis, S. Point pattern analysis of the spatial proximity of residences prior to diagnosis of persons with Hodgkin's disease. *Amer. J. Epidemiol.* **1990**, *132*, S53–S62.
25. Rothman, K.J. A sobering start for the cluster busters' conference. *Amer. J. Epidemiol.* **1987**, *132*, S6–S13.
26. CDC. Guidelines for investigating clusters of health events. *MMWR* **1990**, *39*, 1–23.
27. Kingsley, B.S.; Schmeichel, K.L.; Rubin, C.H. An update on cancer cluster activities at the Centers for Disease Control and Prevention. *Environ. Health Perspect.* **2007**, *115*, 165–171.
28. CDC. Investigating Suspected Cancer Clusters and Responding to Community Concerns: Guidelines from CDC and the Council of State and Territorial Epidemiologists. Recommendations and Reports. Available online: http://www.cdc.gov/mmwr/preview/mmwrhtml/rr6208a1.htm?s_cid=rr6208a1_e (accessed on 26 September 2013).
29. Condon, S.K.; Sullivan, J.; Nettoreba, B. Letter to the editor. Response to “Cancer clusters in the USA: What do the last twenty years of state and federal investigations tell us?” *Crit. Rev. Toxicol.* **2013**, *43*, 73–74.
30. Goodman, M.; Naiman, J.S.; Goodman, D.; LaKind, J.S. Response to Condon *et al.* Comments on “Cancer clusters in the USA: What do the last twenty years of state and federal investigations tell us?” *Crit. Rev. Toxicol.* **2013**, *43*, 75–76.
31. Neutra, R.R. Counterpoint from a cluster buster. *Amer. J. Epidemiol.* **1990**, *132*, 1–8.
32. Fagin, D. *Toms River: A Story of Science and Salvation*; Bantam Books: New York, NY, USA, 2013.
33. Gawande, A. The cancer-cluster myth. *The New Yorker*, **1998**, February, 34–37.
34. Harr, J. *A Civil Action*; Vintage Books: New York, NY, USA, 1995.
35. Johnson, G. Cancer Cluster or Chance? The Link between Environmental Contaminants and Cancer is Surprisingly Weak, if not Imaginary. Available online: http://www.slate.com/articles/health_and_science/medical_examiner/2013/03/cancer_cluster_in_toms_river_new_jersey_the_link_to_a_superfund_site_is.html (accessed on 24 September 2013).
36. Aldrich, T.; Sinks, T. Things to know and do about cancer clusters. *Cancer Invest.* **2002**, *20*, 810–816.
37. Thun, M.J.; Sinks, T. Understanding cancer clusters. *Ca-A Cancer J. Clin.* **2004**, *54*, 273–280.
38. Trumbo, C.W. Public requests for cancer cluster investigations: A survey of state health departments. *Amer. J. Public Health* **2000**, *90*, 1300–1302.
39. Wartenberg, D.; Greenberg, M. Solving the cluster puzzle: Clues to follow and pitfalls to avoid. *Stat. Med.* **1993**, *12*, 1763–1770.
40. Williams, L.J.; Honein, M.A.; Rasmussen, S.A. Methods for a public health response to birth defects clusters. *Teratology* **2002**, *66*, S50–S58.
41. Berman, D.W.; Cox, L.A., Jr.; Popken, D.A. A cautionary tale: The characteristics of two-dimensional distributions and their effects on epidemiological studies employing an ecological design. *Crit. Rev. Toxicol.* **2013**, *1*, S1–S25.
42. Cox, L.A., Jr.; Popken, D.A.; Berman, D.W. Causal vs. spurious spatial exposure-response associations in health risk analysis. *Crit. Rev. Toxicol.* **2013**, *43*, 26–38.
43. Francis, S.S.; Selvin, S.; Yang, W.; Buffler, P.A.; Wiemels, J.L. Unusual space-time patterning of the Fallon, Nevada leukemia cluster: Evidence of an infectious etiology. *Chem. Biol. Interact.* **2012**, *196*, 102–109.

44. Wakefield, J.; Kim, A. A Bayesian model for cluster detection. *Biostatistics* **2013**, *14*, 752–765.
45. CDC. Cancer Clusters. National Center for Environmental Health. 2012. Available online: <http://www.cdc.gov/nceh/clusters/> (accessed on 18 September 2013).
46. Navarro, K.; Janssen, S.; Nordbrock, T.; Solomon, G. *Health Alert: Disease Clusters Spotlight the Need to Protect People from Toxic Chemicals*; Natural Resources Defense Council: New York, NY, USA, 2011.
47. Curtis, R.E.; Boice, J.D., Jr.; Stovall, M.; Bernstein, L.; Greenberg, R.S.; Flannery, J.T.; Schwartz, A.G.; Weyer, P.; Moloney, W.C.; Hoover, R.N. Risk of leukemia after chemotherapy and radiation treatment for breast cancer. *N. Engl. J. Med.* **1992**, *326*, 1745–1751.
48. Chapman, J.R.; Webster, A.C.; Wong, G. Cancer in the transplant recipient. *Cold Spring Harb. Perspect. Med.* **2013**, *3*, doi:10.1101/cshperspect.a015677.
49. Engels, E.A.; Pfeiffer, R.M.; Fraumeni, J.F., Jr.; Kasiske, B.L.; Israni, A.K.; Snyder, J.J.; Wolfe, R.A.; Goodrich, N.P.; Bayakly, A.R.; Clarke, C.A.; *et al.* Spectrum of cancer risk among USA solid organ transplant recipients. *JAMA* **2011**, *306*, 1891–1901.
50. Jacques, G.M.; Meliker, J.; Kaufmann, A. In search of induction and latency periods: Space-time interaction accounting for residential mobility, risk factors and covariates. *Int. J. Health Geogr.* **2007**, *6*, doi:10.1186/1476-072X-6-35.
51. Division of Epidemiologic Studies, Illinois Department of Public Health. *Incidence of Glioblastoma in Zip Code 60453 of Oak Lawn (Cook County), Illinois 1993–1997*; Illinois Department of Public Health, Springfield, IL, USA, 2000.
52. Pui, C.H.; Relling, M.V.; Downing, J.R. Acute lymphoblastic leukemia. *N. Engl. J. Med.* **2004**, *350*, 1535–1548.
53. Bender, A.; Williams, A.N.; Bushhouse, S. Statistical anatomy of a brain cancer cluster—Stillwater, Minnesota. *Disease Control Newsletter Minnesota Department of Health* **1995**, *23*, 4–7.
54. Olsen, S.F.; Martuzzi, M.; Elliott, P. Cluster analysis and disease mapping—Why, when, and how? A step by step guide. *BMJ* **1996**, *313*, 863–866.
55. Nuckols, J.; Airola, M.; Colt, J.; Johnson, A.; Schwenn, M.; Waddell, R.; Karagas, M.; Silverman, D.; Ward, M.H. The impact of residential mobility on exposure assessment in cancer epidemiology. *Epidemiology* **2009**, *20*, S259–S260.
56. Pearce, J.; Boyle, P. Is the urban excess in lung cancer in Scotland explained by patterns of smoking? *Soc. Sci. Med.* **2005**, *60*, 2833–2843.
57. Hankey, B.F.; Ries, L.A.; Edwards, B.K. The surveillance, epidemiology, and end results program: A national resource. *Cancer Epidemiol. Biomarker Prev.* **1999**, *8*, 1117–1121.
58. Clegg, L.X.; Feuer, E.J.; Midthune, D.N.; Fay, M.P.; Hankey, B.F. Impact of reporting delay and reporting error on cancer incidence rates and trends. *J. Natl. Cancer Inst.* **2002**, *94*, 1537–1545.
59. Lipscomb, J.; Gotay, C.C.; Snyder, C. *Outcomes assessment in cancer: Measures, methods and applications*; Cambridge University Press: Cambridge, UK, 2005.
60. CDC, NPCR. NPCR-AERRO ePath Reporting Activities. Available online: <http://www.cdc.gov/cancer/npcr/informatics/aerro/activities/epath.htm> (accessed on 12 September 2013).

61. Cernile, G.; Goodman, M.; Ward, K. Automated Cancer Data Extraction and Rapid Case Ascertainment from Text Based Electronic Pathology Reports. In Proceedings of American Medical Informatics Association Joint Summits on Translational Science, San Francisco, CA, USA, 19–23 March 2012.
62. CDC, NPCR. Meaningful Use of Electronic Health Records. Available online: http://www.cdc.gov/cancer/npcr/meaningful_use.htm (accessed on 12 September 2013).
63. Jacquez, G.M.; Slotnick, M.J.; Meliker, J.R.; AvRuskin, G.; Copeland, G.; Nriagu, J. Accuracy of commercially available residential histories for epidemiologic studies. *Amer. J. Epidemiol.* **2011**, *173*, 236–243.
64. Turnbull, B.W.; Iwano, E.J.; Burnett, W.S.; Howe, H.L.; Clark, L.C. Monitoring for clusters of disease: Application to leukemia incidence in upstate New York. *Amer. J. Epidemiol.* **1990**, *132*, S136–S143.
65. Besag, J.; Newell, J. The detection of clusters in rare diseases. *J. Roy. Statist. Soc. Ser. A Stat.* **1991**, *154*, 143–155.
66. Tango, T. A class of tests for detecting “general” and “focused” clustering of rare diseases. *Stat. Med.* **1995**, *14*, 2323–2334.
67. Swartz, J.B. An entropy-based algorithm for detecting clusters of cases and controls and its comparison with a method using nearest neighbours. *Health Place* **1998**, *4*, 67–77.
68. Whittemore, A.S.; Friend, N.; Brown, B.W.; Holly E.A. A test to detect clusters of disease. *Biometrika* **1987**, *74*, 631–635.
69. Kulldorff, M.; Song, C.; Gregorio, D.; Samociuk, H.; DeChello, L. Cancer map patterns: Are they random or not? *Amer. J. Prev. Med.* **2006**, *30*, S37–S49.
70. Fang, Z.; Kulldorff, M.; Gregorio, D.I. Brain cancer mortality in the United States, 1986 to 1995: A geographic analysis. *Neuro. Oncol.* **2004**, *6*, 179–187.
71. Lawson, A.B. Disease cluster detection: A critique and a Bayesian proposal. *Stat. Med.* **2006**, *25*, 897–916.
72. Lawson, A.B. Commentary: Assessment of chance should be central in investigation of cancer clusters. *Int. J. Epidemiol.* **2013**, *42*, 448–449.
73. Moraga, P.; Kulldorff, M. Detection of spatial variations in temporal trends with a quadratic function. *Stat. Methods Med. Res.* **2013**, in press.
74. Jones, S.G.; Kulldorff, M. Influence of spatial resolution on space-time disease cluster detection. *PLoS One* **2012**, *7*, doi:10.1371/journal.pone.0048036.
75. Wagner, S.E.; Bauer, S.E.; Bayakly, A.R.; Vena, J.E. Prostate cancer incidence and tumor severity in Georgia: Descriptive epidemiology, racial disparity, and geographic trends. *Cancer Cause. Control* **2013**, *24*, 153–166.
76. Beale, L.; Abellan, J.J.; Hodgson, S.; Jarup, L. Methodologic issues and approaches to spatial epidemiology. *Environ. Health Perspect.* **2008**, *116*, 1105–1110.
77. Goovaerts, P. Accounting for rate instability and spatial patterns in the boundary analysis of cancer mortality maps. *Environ. Ecol. Stat.* **2008**, *15*, 421–446.
78. Thew, S.L.; Sutcliffe, A.; de Bruijn, O.; McNaught, J.; Procter, R.; Jarvis, P.; Buchan, I. Supporting creativity and appreciation of uncertainty in exploring geo-coded public health data. *Methods Inform. Med.* **2011**, *50*, 158–165.

79. MacEachren, A.M.; Robinson, A.; Hopper, S.; Gardner, S.; Murray, R.; Gahegan, M.; Hetzler, E. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartogr. Geogr. Inform. Sci.* **2005**, *32*, 139–160.
80. Jacquez, G.M. A research agenda: Does geocoding positional error matter in health GIS studies? *Spat. Spatiotemporal Epidemiol.* **2012**, *3*, 7–16.
81. Brownstein, J.S.; Freifeld, C.C.; Madoff, L.C. Digital disease detection—Harnessing the web for public health surveillance. *N. Engl. J. Med.* **2009**, *360*, 2153–2157.
82. Jensen, P.B.; Jensen, L.J.; Brunak, S. Mining electronic health records: Towards better research applications and clinical care. *Nat. Rev. Genet.* **2012**, *13*, 395–405.
83. Lian, M.; Struthers, J.; Schootman, M. Comparing GIS-based measures in access to mammography and their validity in predicting neighborhood risk of late-stage breast cancer. *PLoS One* **2012**, *7*, doi:10.1371/journal.pone.0043000.
84. Richardson, D.B.; Volkow, N.D.; Kwan, M.P.; Kaplan, R.M.; Goodchild, M.F.; Croyle, R.T. Spatial turn in health research. *Science* **2013**, *339*, 1390–1392.
85. Wangia, V.; Shireman, T.I. A review of geographic variation and geographic information systems (GIS) applications in prescription drug use research. *Res. Soc. Admin. Pharm.* **2013**, doi:10.1016/j.sapharm.2012.11.006.
86. Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; Barabasi, A.L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; *et al.* Social science. Computational social science. *Science* **2009**, *323*, 721–723.
87. Kosinski, M.; Stillwell, D.; Graepel, T. Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5802–5805.
88. Rothman, K.J. Clustering of disease. *Amer. J. Public Health* **1987**, *77*, 13–15.
89. Vogelstein, B.; Fearon, E.R.; Hamilton, S.R.; Kern, S.E.; Preisinger, A.C.; Leppert, M.; Nakamura, Y.; White, R.; Smits, A.M.; Bos, J.L. Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **1988**, *319*, 525–532.
90. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674.
91. Abdulwahab, A.; Sykes, J.; Kamel-Reid, S.; Chang, H.; Brandwein, J.M. Therapy-related acute lymphoblastic leukemia is more frequent than previously recognized and has a poor prognosis. *Cancer* **2012**, *118*, 3962–3967.
92. Sam, T.N.; Kersey, J.H.; Linabery, A.M.; Johnson, K.J.; Heerema, N.A.; Hilden, J.M.; Davies, S.M.; Reaman, G.H.; Ross, J.A. MLL gene rearrangements in infant leukemia vary with age at diagnosis and selected demographic factors: A Children’s Oncology Group (COG) study. *Pediatr. Blood Cancer* **2012**, *58*, 836–839.
93. Kennedy, S.R.; Loeb, L.A.; Herr, A.J. Somatic mutations in aging, cancer and neurodegeneration. *Mech. Age. Dev.* **2012**, *133*, 118–126.
94. Beckman, K.B.; Ames, B.N. Oxidative decay of DNA. *J. Biol. Chem.* **1997**, *272*, 19633–19636.
95. Goodman, M.; Bostick, R.M.; Kucuk, O.; Jones, D.P. Clinical trials of antioxidants as cancer prevention agents: Past, present, and future. *Free Radical Biol. Med.* **2011**, *51*, 1068–1084.
96. Janakiram, N.B.; Rao, C.V. Molecular markers and targets for colorectal cancer prevention. *Acta Pharmacol. Sin.* **2008**, *29*, 1–20.

97. Davis, K.L.; Marina, N.; Arber, D.A.; Ma, L.; Cherry, A.; Dahl, G.V.; Heerema-McKenney, A. Pediatric acute myeloid leukemia as classified using 2008 WHO criteria: A single-center experience. *Amer. J. Clin. Pathol.* **2013**, *139*, 818–825.
98. Creutzig, U.; van den Heuvel-Eibrink, M.M.; Gibson, B.; Dworzak, M.N.; Adachi, S.; de Bont, E.; Harbott, J.; Hasle, H.; Johnston, D.; Kinoshita, A.; *et al.* Diagnosis and management of acute myeloid leukemia in children and adolescents: Recommendations from an international expert panel. *Blood* **2012**, *120*, 3187–3205.
99. SEER Program Quality Control Section. ICD-0-3 SEER Site/Histology Validation List, 5 December 2012. Available online: <http://seer.cancer.gov/icd-o-3/sitetype.icdo3.d20121205.pdf> (accessed on 18 September 2013).
100. McDermott, U.; Downing, J.R.; Stratton, M.R. Genomics and the continuum of cancer care. *N Engl. J. Med.* **2011**, *364*, 340–350.
101. Boyle, P. Triple-negative breast cancer: Epidemiological considerations and recommendations. *Ann. Oncol.* **2012**, *23*, 7–12.
102. Hanahan, D.; Weinberg, R.A. The hallmarks of cancer. *Cell* **2000**, *100*, 57–70.
103. Barbacid, M. Ras oncogenes: Their role in neoplasia. *Eur. J. Clin. Invest.* **1990**, *20*, 225–235.
104. Pfeifer, G.P.; Besaratinia, A. Mutational spectra of human cancer. *Hum. Genet.* **2009**, *125*, 493–506.
105. Schaap, F.G.; French, P.J.; Bovee, J.V. Mutations in the isocitrate dehydrogenase genes IDH1 and IDH2 in tumors. *Adv. Anat. Pathol.* **2013**, *20*, 32–38.
106. Wiemels, J. Chromosomal translocations in childhood leukemia: Natural history, mechanisms, and epidemiology. *J. Natl. Cancer Inst. Monogr.* **2008**, 87–90.
107. Gao, F.; Chia, K.S.; Machin, D. On the evidence for seasonal variation in the onset of acute lymphoblastic leukemia (ALL). *Leuk. Res.* **2007**, *31*, 1327–1338.
108. Eden, T. Aetiology of childhood leukaemia. *Cancer Treat. Rev.* **2010**, *36*, 286–297.
109. Greaves, M. Infection, immune responses and the aetiology of childhood leukaemia. *Nat. Rev. Cancer* **2006**, *6*, 193–203.
110. Infante-Rivard, C. Chemical risk factors and childhood leukaemia: A review of recent studies. *Radiat. Prot. Dosim.* **2008**, *132*, 220–227.
111. Kroll, M.E.; Draper, G.J.; Stiller, C.A.; Murphy, M.F. Childhood leukemia incidence in Britain, 1974–2000: Time trends and possible relation to influenza epidemics. *J. Natl. Cancer Inst.* **2006**, *98*, 417–420.
112. Heath, C.W., Jr.; Hasterlik, R.J. Leukemia among children in a suburban community. *Amer. J. Med.* **1963**, *34*, 796–812.
113. Steinberg, K.K.; Relling, M.V.; Gallagher, M.L.; Greene, C.N.; Rubin, C.S.; French, D.; Holmes, A.K.; Carroll, W.L.; Koontz, D.A.; Sampson, E.J.; *et al.* Genetic studies of a cluster of acute lymphoblastic leukemia cases in Churchill County, Nevada. *Environ. Health Perspect.* **2007**, *115*, 158–164.
114. Cazzaniga, G.; Bisanti, L.A.; Palmi, C.; Randi, G.; Pregliasco, F.; Deandrea, S.; *et al.* A Childhood Leukaemia Cluster in Milan: Possible Role of Pandemic AH1N1 Swine Flu Virus. In Proceedings of American Society of Hematology Annual Meeting, Atlanta, GA, USA, 8–11 December 2012.

115. Costas, K.; Knorr, R.S.; Condon, S.K. A case-control study of childhood leukemia in Woburn, Massachusetts: The relationship between leukemia incidence and exposure to public drinking water. *Sci. Total Environ.* **2002**, *300*, 23–35.
116. Maslia, M.L.; Reyes, J.J.; Gillig, R.E.; Sautner, J.B.; Fagliano, J.A.; Aral, M.M. Public health partnerships addressing childhood cancer investigations: Case study of Toms River, Dover Township, New Jersey, USA. *Int. J. Hyg. Environ. Health* **2005**, *208*, 45–54.
117. Kinlen, L.J. An examination, with a meta-analysis, of studies of childhood leukaemia in relation to population mixing. *Br. J. Cancer* **2012**, *107*, 1163–1168.
118. Chiu, C.Y. Viral pathogen discovery. *Curr. Opin. Microbiol.* **2013**, *16*, 468–478.
119. Steinmaus, C.; Lu, M.; Todd, R.L.; Smith, A.H. Probability estimates for the unique childhood leukemia cluster in Fallon, Nevada, and risks near other US military aviation facilities. *Environ. Health Perspect.* **2004**, *112*, 766–771.
120. Wild, C.P.; Scalbert, A.; Herceg, Z. Measuring the exposome: A powerful basis for evaluating environmental exposures and cancer risk. *Environ. Mol. Mutagen.* **2013**, *54*, doi:10.1002/em.21777.
121. Palacios, G.; Druce, J.; Du, L.; Tran, T.; Birch, C.; Briese, T.; Conlan, S.; Quan, P.L.; Hui, J.; Marshall, J.; *et al.* A new arenavirus in a cluster of fatal transplant-associated diseases. *N. Engl. J. Med.* **2008**, *358*, 991–998.
122. Patel, C.J.; Bhattacharya, J.; Butte, A.J. An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS One* **2010**, *5*, doi:10.1371/journal.pone.0010746.
123. Patel, C.J.; Chen, R.; Kodama, K.; Ioannidis, J.P.; Butte, A.J. Systematic identification of interaction effects between genome- and environment-wide associations in type 2 diabetes mellitus. *Hum. Genet.* **2013**, *132*, 495–508.