

Article

Citation Drift in the Use of a Foundational Low-Dose Risk-Assessment Paper: A Citation Context Analysis of Crump et al. (1976)

Andy Wai Kan Yeung 

Oral and Maxillofacial Radiology, Applied Oral Sciences and Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China; ndyeung@hku.hk

Abstract

Scientific papers can acquire meanings that differ from their original arguments, yet citation counts alone cannot reveal such changes. This study examined how the work of Crump and co-workers in 1976, published as a foundational paper on low-dose linearity in carcinogenic risk assessment, has been cited and represented over time. A cited-reference search of the Web of Science Core Collection was conducted on 1 June 2026 using the paper's title. Of 350 retrieved records, 318 citing papers were included after excluding inaccessible, duplicate, incorrect, non-English, or non-substantive records. Each paper was coded for citation context, stance, accuracy, and accuracy issue using a hybrid deductive–inductive scheme. From 1977 through 2015, 89.4–97.0% of citations remained within broad original-scope contexts, including carcinogenic risk assessment, multistage modeling, low-dose extrapolation, and threshold debate. In 2016–2026, original-scope contexts declined to 38.5% (25/65), while peripheral, generic, or unrelated dose–response uses increased to 52.3% (34/65). Citation stance remained mostly neutral or supportive, but accurate or acceptable citations fell from over 70% before 2016 to 33.8% (22/65) in 2016–2026, while inaccurate or unrelated attributions rose to 53.8% (35/65). These findings illustrate citation drift in the citation history of foundational paper written by Crump and co-workers in 1976 and underscore the need for careful source verification considering that classic references may become shorthand in contemporary scholarly, regulatory, and public-health discussions.

Keywords: citation context analysis; citation accuracy; citation drift; miscitation; scholarly communication; reference verification; foundational papers; low-dose risk assessment; linear no-threshold model



Academic Editor: Andrew Kirby

Received: 10 September 2026

Revised: 20 October 2026

Accepted: 22 October 2026

Published: 26 October 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Low-dose radiation risk assessment often depends on inference beyond the range in which excess risks can be measured directly with high statistical confidence. For stochastic outcomes such as cancer, one influential solution has been to use linear extrapolation from observed higher-dose effects to lower-dose exposure conditions. In radiation protection, this reasoning is commonly expressed through the linear no-threshold (LNT) model, under which each additional increment in ionizing-radiation dose is assumed to contribute some incremental cancer risk, even when the dose is small. The model has therefore functioned not only as a scientific hypothesis about dose response but also as a practical convention for regulation, dose limitation, and protection under uncertainty (ICRP, 2005; UNSCEAR, 2024).

The rationale for low-dose linearity has drawn from several sources, including early radiation-genetics work, postwar expert assessments of radiation hazards, and later quantitative approaches to cancer risk assessment. Muller's mutation experiments in *Drosophila melanogaster* helped support the view that radiation-induced biological effects might occur without a demonstrable threshold (Muller, 1927), and the 1956 BEAR I Report of the U.S. National Academy of Sciences later gave institutional prominence to linear extrapolation from higher-dose evidence to lower-dose exposure conditions (National Research Council, 1956). Nevertheless, the form of the dose–response curve at low doses remains difficult to establish empirically. Biological processes such as DNA repair, adaptive responses, immune surveillance, cell killing, and dose-rate effects have been invoked in support of nonlinear, threshold, or hormetic interpretations, whereas some epidemiological findings have been interpreted as compatible with low-dose linearity (Baldwin & Grantham, 2015; Bosch de Basea et al., 2023; Luckey, 2006). Recent reviews therefore continue to distinguish the practical role of LNT in radiological protection from unresolved uncertainty about the magnitude and shape of low-dose risk (Rühm et al., 2022; Wojcik & Zölzer, 2024).

Within this broader literature, the work of Crump et al. (1976) has been an important but sometimes simplified reference (Crump et al., 1976). The paper did not merely assert that all agents, all diseases, or all dose–response curves must be linear at low doses. Instead, it advanced a more specific theoretical argument: if cancer arises from single cells and an external carcinogenic process acts directly and additively with ongoing background carcinogenic processes, then the additional risk at sufficiently low doses is expected to be approximately linear. The paper also discussed multistage and multi-hit models and emphasized assumptions and limitations, including direct cellular action, background additivity, single-cell origin, and the practical question of how linear “linear” must be for risk-assessment purposes.

The preservation of the original meaning of the cited reference matters, because influential papers may acquire citation for contexts that differ from their original content. A reference may be cited accurately for a specific argument, cited more generally as shorthand for a broader field, or cited inaccurately for claims, methods, or procedures that it did not contain. Conventional citation counts cannot distinguish among these uses. Citation context analysis can therefore provide a more informative view of how a foundational paper is understood, transmitted, simplified, or misrepresented over time. In the past, it was reported that the miscitation rate in scientific literature varied from 19% in geography (Haussmann et al., 2013) to 26% in pediatric orthopedics (Davids et al., 2010). These figures come from individual studies in specific disciplines and do not establish a general range. Meanwhile, Mogull (2017) estimated a lower overall rate of 14.5% in the medical literature after standardizing error definitions across studies and recalculating rates based on quotations actually examined rather than the number of references selected—a methodological correction that yielded lower estimates than previous reviews.

The present study therefore examined the citation history of Crump et al. (1976) using citation context analysis. Specifically, it studied how citing papers have used Crump et al. (1976); whether the cited paper was invoked supportively, neutrally, critically, or in mixed ways; and whether the citation accurately represented the original argument. By focusing on citation context and accuracy rather than citation counts alone, this analysis aimed to clarify how the low-dose linearity argument of Crump et al. (1976) has been incorporated into, and sometimes detached from, the later radiation-protection and risk-assessment literature.

2. Materials and Methods

2.1. Literature Search

The Web of Science Core Collection (WOSCC) literature database was queried on 1 June 2026 to identify citing papers of [Crump et al. \(1976\)](#). Similar to a recent bibliometric work on LNT ([Yeung et al., 2026](#)), WOSCC was chosen because it is a curated multidisciplinary citation database widely used in citation analysis research ([Birkle et al., 2020](#); [Singh et al., 2021](#)) and maintained by Clarivate, the same company that computes Journal Impact Factor. The title of [Crump et al. \(1976\)](#), namely, “Fundamental Carcinogenic Processes and Their Implications for Low Dose Risk Assessment”, was entered for a cited reference search. No additional filters were placed. The search retrieved 350 records. A second search was performed by searching the following: (Cited author = Crump) AND (Cited work = Cancer Research) AND (Cited year = 1976). This search returned two records, one being [Crump et al. \(1976\)](#) and the other one being [Lin et al. \(1976\)](#), which is irrelevant. In other words, the search did not identify any additional records.

For more information, the search included all 6 WOSCC sub-datasets subscribed by the author’s institution, including Science Citation Index Expanded (1970–2026), Social Sciences Citation Index (1956–2026), Arts & Humanities Citation Index (1975–2026), Emerging Sources Citation Index (2005–2026), Conference Proceedings Citation Index–Science (2005–2026), and Conference Proceedings Citation Index–Social Science & Humanities (2005–2026). The disclosure of this information promotes better reproducibility ([Liu, 2019](#)).

2.2. Citation Screening and Coding of Citing Papers

Each citing paper was treated as the unit of analysis and was coded according to how it used [Crump et al. \(1976\)](#) based on the citing passages. Full texts were screened to extract the citing lines. Preceding and following lines were extracted when they were contextually related to the citing lines. Records were excluded if the author had no access to its full text ($n = 22$, listed in Supplementary Table S2; they were distributed as follows: 1977–1985, $n = 14$; 1986–1995, $n = 6$; 1996–2005, $n = 1$; 2006–2015, $n = 0$; 2016–2026, $n = 1$), as well as if it was a duplicate paper ($n = 2$) or incorrect bibliographic record ($n = 3$), written in non-English ($n = 1$), or contained [Crump et al. \(1976\)](#) only in the reference list without any substantive in-text use ($n = 4$). Full texts were sought through journal websites, the institutional library, ResearchGate, and other online sources before exclusion. As a result, 318 citing papers entered the analysis.

For each included paper, four dimensions were coded: citation context, stance, accuracy, and, where relevant, type of accuracy issue. When [Crump et al. \(1976\)](#) was cited more than once in a paper, the most substantive use was coded. If the paper used the citation in several ways, the dominant use in the paper was coded.

The coding scheme was developed using a hybrid deductive–inductive approach. The deductive basis was a careful reading of [Crump et al. \(1976\)](#), especially its argument that additive effects with ongoing background carcinogenic processes imply approximate linearity at sufficiently low doses. The inductive component came from pilot coding of citing papers, which revealed recurring citation contexts and recurring forms of simplification, overgeneralization, and miscitation. The coding scheme was devised and revised iteratively during the pilot stage to sharpen category definitions, reduce overlap, and develop conservative decision rules for distinguishing acceptable shorthand from substantive misrepresentation. Unless the citing text was clearly unrelated or materially distorted the cited paper, the citation was coded as acceptable shorthand rather than as inaccurate. Pilot coding was conducted on 100 papers selected to cover all publication periods and a range of apparent citation uses. During the pilot phase, preliminary context categories were refined—especially the boundary between C6 (policy/regulatory implication) and

C7 (peripheral or unrelated dose–response use), and the accuracy distinction between A2 (acceptable simplification) and A3 (overgeneralization). The codebook was frozen after pilot coding. All pilot records were then recoded using the final definitions before inclusion in the full analysis. The coding scheme is listed in Table 1.

Table 1. Coding scheme of the citing passages to [Crump et al. \(1976\)](#).

Scheme	Explanatory Notes	Example Citing Sentence (the Citation to Crump et al. (1976) Is Denoted by [Citation])
Citation context		
C0 = Minimal or nonspecific citation	Crump et al. (1976) was mentioned only briefly or as part of a general list, with no clear substantive claim beyond bibliographic association.	Models that have received the most attention [citation] are listed in Table 1. (Paper_ID #129)
C1 = Background/ additivity and low-dose linearity	The citing passage used Crump et al. (1976) for its central argument that, when an external carcinogen acts additively with ongoing background carcinogenic processes, the low-dose response is expected to be approximately linear.	There is also a well-based mathematical argument due to [citation], which says that if carcinogens in the environment and a newly introduced one are additive, the effect of low doses of the newly introduced one must be linear. But the additivity assumption is a major one that lacks experimental support. (Paper_ID #1)
C2 = Multistage, multi-hit, or stochastic carcinogenesis modeling	The citation was used in relation to multistage or multi-hit models, single-cell origin of cancer, age-specific incidence, or mathematical models of carcinogenesis.	Our dose–response relation is taken from a general multistage carcinogenesis model developed by Armitage and Doll (1961), extended by Peto (1974) and by [citation], and investigated by Brown (1976). (Paper_ID #12)
C3 = Regulatory or quantitative risk-assessment methodology	The citation was used in connection with linear extrapolation, linearized multistage reasoning, cancer potency estimation, upper-bound risk assessment, or regulatory quantitative risk assessment.	Data from occupational studies has been used to quantify the effect of the exposure to carcinogens in ambient air by extrapolating back with the assumption that cancer incidence is linearly proportional to the exposure (concentration of carcinogens) [citation]. (Paper_ID #29)
C4 = General low-dose extrapolation or cancer risk assessment	Crump et al. (1976) was cited as a general source for low-dose extrapolation, dose–response modeling, or cancer risk assessment, without detailed discussion of the specific biological or mathematical argument.	Whittemore and Keller (1978) have pointed out that theoretical models play an important role since they provide a bridge between animal experiments and epidemiologic studies, and also because they provide a basis for extrapolating dose–response relations downward [citation]. (Paper_ID #67)
C5 = Threshold, non-threshold, safe-dose, or probit-extrapolation discussion	The citation was used in discussions of whether carcinogens have thresholds, whether “safe” doses can be defined, or in comparison with approaches such as Mantel–Bryan/probit extrapolation.	A safe threshold dose consequently does not exist [citation]. (Paper_ID #68)

Table 1. Cont.

Scheme	Explanatory Notes	Example Citing Sentence (the Citation to Crump et al. (1976) Is Denoted by [Citation])
C6 = Policy, public-health, or regulatory implications	The citation was used mainly for implications for environmental regulation, food additives, acceptable risk, public-health decision-making, or conservative risk-management policy.	The fundamental scientific rationale of the default LNT approach used by EPA to make such estimates for presumed genotoxic chemical carcinogens continues to be the 50-year-old multistage somatic mutation/clonal expansion theory of cancer [citation], formulated just a few years after the discovery of DNA, the genetic code, and basic aspects of gene expression. (Paper_ID #250)
C7 = Peripheral, generic, or unrelated dose-response use	The citation was used in a generic or peripheral way, or for a dose-response topic not clearly connected to the specific claims of Crump et al. (1976). This category included cases where the citation appeared to be used only as a broad or possibly misplaced reference to dose-response analysis.	The Crump et al., 1976 approach was used to generate a mouse dose-response curve [citation]. (Paper_ID #298)
Stance towards Crump et al. (1976)		
S = Supportive	The citing passage invoked Crump et al. (1976) in support of its own argument or treated the cited claim favorably.	As the dose decreases, it becomes more difficult to detect the effect, and other factors such as dietary fat become relatively more important, but we must assume that for large populations every stimulus will have an effect, and that separate stimuli will add. The same argument has been powerfully argued for carcinogenic effects [citation]. (Paper_ID #7)
N = Neutral/descriptive	The citing passage described Crump et al. (1976) or cited it as background, without clear endorsement or disagreement.	Mantel et al. (1975) have proposed one model. Crump et al. (1976) [citation] have considered alternative models, and Hartley and Sielken (1977) have proposed a general class of models. (Paper_ID #5)
M = Mixed	The citing passage partly accepted the relevance of Crump et al. but added qualifications, limitations, or alternative interpretations.	While debate over whether toxicity processes that augment existing background processes behave mathematically as linear at low doses [citation], there appears to be little biological analysis that addresses this question. [Paper_ID #183]
C = Critical	The citing passage challenged, criticized, or contrasted its position with Crump et al. (1976).	This led risk assessors to the absurd assumption (from the biological point of view) that time is not important at low doses [citation], when in fact time is less important at high doses but it dominates the low-dose end of the $c \times t$ relationship. (Paper_ID #199)

Table 1. Cont.

Scheme	Explanatory Notes	Example Citing Sentence (the Citation to Crump et al. (1976) Is Denoted by [Citation])
Accuracy towards the context of Crump et al. (1976)		
A0 = No substantive accuracy judgment possible	The citation was too brief, nonspecific, or bibliographic to assess accuracy.	Models that have received the most attention [citation] are listed in Table 1. (Paper_ID #129)
A1 = Accurate and specific	The citing passage represented Crump et al. (1976) accurately and with adequate specificity—for example, by referring to low-dose linearity under assumptions about background carcinogenesis, additivity, single-cell origin, or multistage carcinogenesis.	There is also a well-based mathematical argument due to [citation], which says that if carcinogens in the environment and a newly introduced one are additive, the effect of low doses of the newly introduced one must be linear. But the additivity assumption is a major one that lacks experimental support. (Paper_ID #1)
A2 = Acceptable simplification	The citing passage gave a simplified but broadly fair representation. For example, a statement that Crump et al. (1976) provided theoretical support for low-dose linear extrapolation in carcinogenic risk assessment was coded as A2 even if it did not restate all conditions of the original argument.	Mathematical approaches alone are inadequate to cope with the problem, but they suggest that any amount of carcinogen, however small, will contribute to the total carcinogenic effect in the population by raising, however slightly, the level of response of the total population [citation]. (Paper_ID #14)
A3 = Overgeneralized or potentially misleading	The citing passage remained topically related to Crump et al. (1976) but overstated the scope of the argument. Examples include presenting Crump et al. (1976) as proving that all carcinogens are linear at low dose, that thresholds are impossible, or that the argument applies to all toxic endpoints without qualification.	The conclusion that synergism with arsenic does not change the threshold dose contradicts a biostatistical model [citation] that predicts a zero threshold after the additive interaction of carcinogens with individual threshold dose responses. (Paper_ID #133)
A4 = Clearly inaccurate or unrelated attribution	The citing passage attributed to Crump et al. (1976) a method, result, model, empirical finding, or procedure not contained in the original paper, or used the citation for an unrelated purpose such as generic EC50 estimation, Hill-curve fitting, or another dose–response method not associated with Crump et al. (1976).	Classification of masses into hyperplasia, benign tumor, or malignancy is based upon an opinion of a pathologist [citation]. (Paper_ID #105)
Accuracy issue (applicable to papers coded with A3/A4). Overgeneralization, OG. Inaccurate attribution, IA.		
OG1 = Conditional low-dose linearity stated as unconditional	The conditional argument of Crump et al. (1976) for approximate low-dose linearity was presented as a universal or unconditional claim.	The “additivity to background” concept is a mathematical approach to dose extrapolation that is based on the hypothesis that, like radiation carcinogenesis, chemical carcinogenesis is also caused by direct damage to deoxyribonucleic acid (DNA) and, thus, has a linear dose–response relationship and lacks a threshold [citation]. (Paper_ID #232)

Table 1. Cont.

Scheme	Explanatory Notes	Example Citing Sentence (the Citation to Crump et al. (1976) Is Denoted by [Citation])
OG2 = Threshold or safe-dose implications overstated	The citing passage overstated Crump et al. (1976) as proving the absence of thresholds or safe doses, rather than presenting linear extrapolation as generally appropriate under specified carcinogenic-process assumptions.	A safe threshold dose consequently does not exist [citation]. (Paper_ID #68)
OG3 = Regulatory conclusion overstated	The citing passage treated Crump et al. (1976) as establishing a specific regulatory rule or mandatory policy position, rather than as providing theoretical support relevant to risk assessment.	Methods used for performing risk assessments have also been a major source of uncertainty and controversy. USEPA and other agencies have used the linear multistage model [citation] for cancer risk assessment. (Paper_ID #212)
OG4 = Extension beyond carcinogenesis without qualification	The citing passage generalized the carcinogenesis argument from Crump et al. (1976) to all toxicological endpoints, all chemicals, or all biological dose–response relationships without sufficient caveats.	If an endogenous hormone agonist causes an additive background effect to the external agonist dose, then the dose–response curve will be linear in the low-dose region [citation]. (Paper_ID #221)
OG5 = Critical caricature or mischaracterization	The citing passage criticized Crump et al. (1976) but did so by misstating the argument—for example, by portraying it as an unsupported assertion that all chemicals must have linear non-threshold dose responses.	The conclusion that synergism with arsenic does not change the threshold dose contradicts a biostatistical model [citation] that predicts a zero threshold after the additive interaction of carcinogens with individual threshold dose responses. (Paper_ID #133)
IA1 = Inaccurate empirical attribution	The citing passage attributed specific empirical data, chemical-specific results, or experimental findings to Crump et al. (1976) that were not reported in the 1976 paper.	Also, certain epidemiological data for human exposures indicate, within limits of error, that a one-hit mechanism may play a part in the induction of cancer [citation]. (Paper_ID #20)
IA2 = Inaccurate experimental or bioassay attribution	The citing passage attributed an experimental design, animal bioassay protocol, or laboratory procedure to Crump et al. (1976) that was not part of the paper.	When the experiment was completed, the viability of EAC cells was revealed by the trypan blue exclusion technique [citation]. (Paper_ID #311)
IA3 = Inaccurate risk-level or acceptable-risk attribution	The citing passage attributed a specific acceptable-risk cutoff, safety standard, or numerical regulatory threshold to Crump et al. (1976) in a way not supported by the paper.	N/A
IA4 = Inaccurate NOAEL/LOAEL (No/Lowest Observed Adverse Effect Level) or threshold-method attribution	The citing passage used Crump et al. (1976) as a source for NOAEL, LOAEL, threshold, or uncertainty-factor procedures not developed in the paper.	N/A
IA5 = Inaccurate software or computational-procedure attribution	The citing passage attributed a specific software package, computational implementation, or technical procedure not attributable to Crump et al. (1976).	Statistical comparisons between groups were conducted using GraphPad Prism, with $p < 0.05$ considered significant [citation]. (Paper_ID #306)

Table 1. Cont.

Scheme	Explanatory Notes	Example Citing Sentence (the Citation to Crump et al. (1976) Is Denoted by [Citation])
IA6 = Inaccurate statistical-test attribution	The citing passage attributed a statistical test or general statistical method to Crump et al. (1976) that was not presented in the paper.	Indeed, more often than not, the evidence is summarized by a probability statement in the form of a ‘p’ value [citation]. (Paper_ID #106)
IA7 = Inaccurate non-cancer or non-carcinogenic endpoint attribution	The citing passage used Crump et al. (1976) as if it were a source for a non-cancer endpoint or unrelated toxicological model without explaining the conceptual extension from carcinogenesis.	N/A
IA8 = Inaccurate attribution of a later regulatory model or algorithm	The citing passage treated Crump et al. (1976) as having introduced a later formal regulatory algorithm, such as a specific EPA linearized multistage procedure, benchmark-dose method, or implementation not actually contained in the 1976 paper.	The form of the linearized multistage preferred by the U S Environmental Protection Agency (EPA) uses a truncated polynomial fit model with an upper limit in some statistical sense on the coefficient of the linear slope (dose term ql), with the other coefficients held constant [citation]. (Paper_ID #160)
IA9 = Generic unrelated dose–response miscitation	The citation was used for a generic or unrelated dose–response procedure, such as EC50 calculation, Hill-model fitting, curve plotting, or another method not connected to the argument of Crump et al. (1976).	Dose–response relationships based on the Hill equation are widely used in biological sciences [citation]. (Paper_ID #297)

Please refer to Supplementary Table S1 to identify the citing papers according to the Paper_ID number. In Supplementary Table S1, a blank cell in the Issue column means that no accuracy issue was applicable, because the citation was coded as A0, A1, or A2. Issue codes were assigned only to citations coded as A3 or A4.

Because independent double-coding was not feasible, a validation exercise was conducted. The author recoded a random 20% subset ($n = 64$) after a two-week interval while blinded to the original codes. Intra-coder agreement was calculated with Kappa values. The Kappa values for citation context, citation stance, and citation accuracy were 0.75, 0.90, and 0.81, respectively.

Chi-squared tests were conducted across the survey periods to evaluate whether there were significant differences in the proportion of core citation contexts (C1–C5) against peripheral contexts (C6–C7), as well as in citation accuracy (A1–A2 against A3–A4). C6 citations, while topically adjacent, typically do not engage with the specific theoretical argument (additivity, background processes, single-cell origin, etc.) from Crump et al. (1976) and instead invoke it as a regulatory landmark. Readers should note that C6 citations being grouped as peripheral use is not itself evidence of inaccuracy. Binary logistic regressions were conducted to determine whether a temporal trend existed in citation contexts (C1–C5 against C6–C7) as well as citation accuracy (A1–A2 against A3–A4) across publication years. For citation context, the event was C6–C7. For citation accuracy, the event was A3–A4. Two citing papers coded with C0 (minimal or nonspecific citation) or A0 (no substantive accuracy judgment possible) were excluded from the statistical tests. Statistical tests were conducted with SPSS 28.0 (IBM, New York, NY, USA). Results were significant if $p < 0.05$.

3. Results

Citation context remained strongly concentrated in C1–C5 from 1977 through 2015 (Figure 1). The first period (1977–1985) was shorter, as it began with the first year with a citation to Crump et al. (1976). In other words, there were no citing papers in 1976. Subsequent periods were grouped approximately by decade to maintain interpretable cell sizes while preserving temporal resolution. In those periods, between 89.4% (1986–1995: 59/66) and 97.0% (2006–2015: 32/33) of citations fell within broad original-scope contexts (C1–C5). However, strict C1 (background/additivity and low-dose linearity) use alone was never the majority category, ranging from 37.6% (1977–1985: 35/93) to 47.5% (1996–2005: 29/61) before 2016. The major shift occurred in 2016–2026, when C1–C5 contexts fell to 38.5% (25/65), while C6 (policy, public-health, or regulatory implications)/C7 (Peripheral, generic, or unrelated dose–response use) contexts rose to 60.0% (39/65), driven especially by C7, which accounted for 52.3% (34/65) of all citations in the most recent period. Chi-squared testing revealed a significant difference in the proportion of C1–C5 against C6–C7 across periods ($\chi^2(4) = 113.045, p < 0.001$). Binary logistic regression revealed that publication year was a significant predictor of citation context (Wald (1) = 52.23, $p < 0.001$). The odds of C6–C7 citations increased across years (OR = 1.104, 95% CI: [1.074, 1.133]).

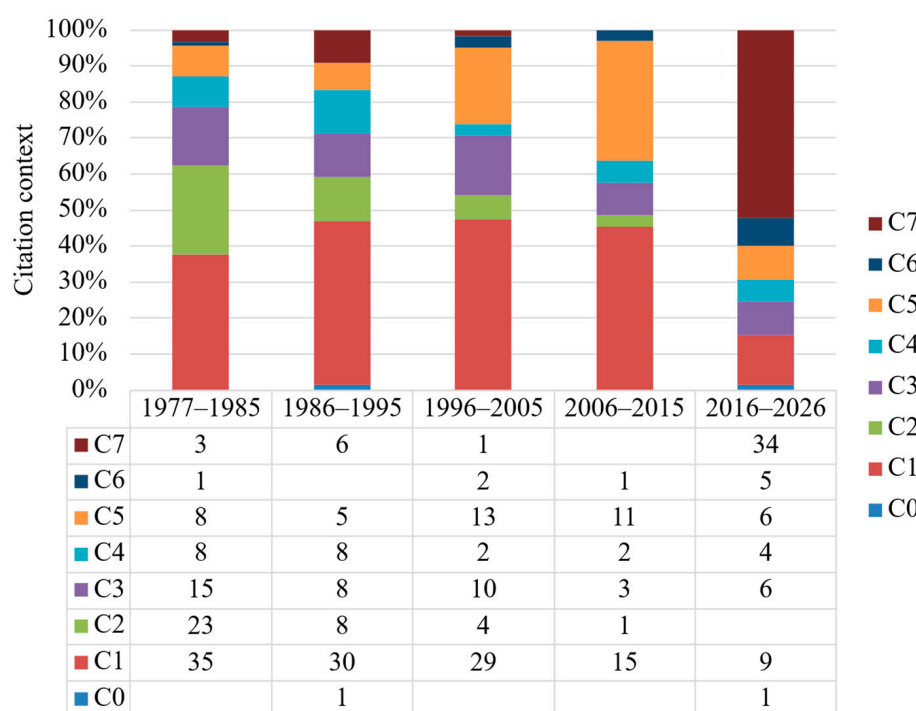


Figure 1. Citation contexts to Crump et al. (1976) over time. Bars show the proportion of citing passages in each time period assigned to each context code. There was a concentration of C1–C5 in earlier periods, a gradual decline in these original-scope contexts, and a sharp rise in C7 (peripheral/generic/unrelated) in 2016–2026. See Table 1 for the explanation of the codes. Data for 2026 was up to 1 June 2026.

Stance remained largely neutral or supportive across the citation history, with critical citations never exceeding 6.2% (2016–2026: 4/65) in any period (Figure 2). The more substantial temporal shift occurred in citation accuracy. From 1977 through 2015, accurate (A1: accurate and specific) or mostly accurate (A2: acceptable simplification) citations constituted more than 70% of citations in every period (Figure 3). In 2016–2026, however, this proportion fell to 33.8% (22/65), while A4 citations (clearly inaccurate or unrelated attribution) rose sharply to 53.8% (35/65). Chi-squared testing revealed a significant difference in the proportion of A1–A2 against A3–A4 across periods ($\chi^2(4) = 55.648, p < 0.001$). Binary

logistic regression revealed that publication year was a significant predictor of citation accuracy (Wald (1) = 43.69, $p < 0.001$). The odds of A3–A4 citations increased across years (OR = 1.062, 95% CI: [1.044, 1.082]).

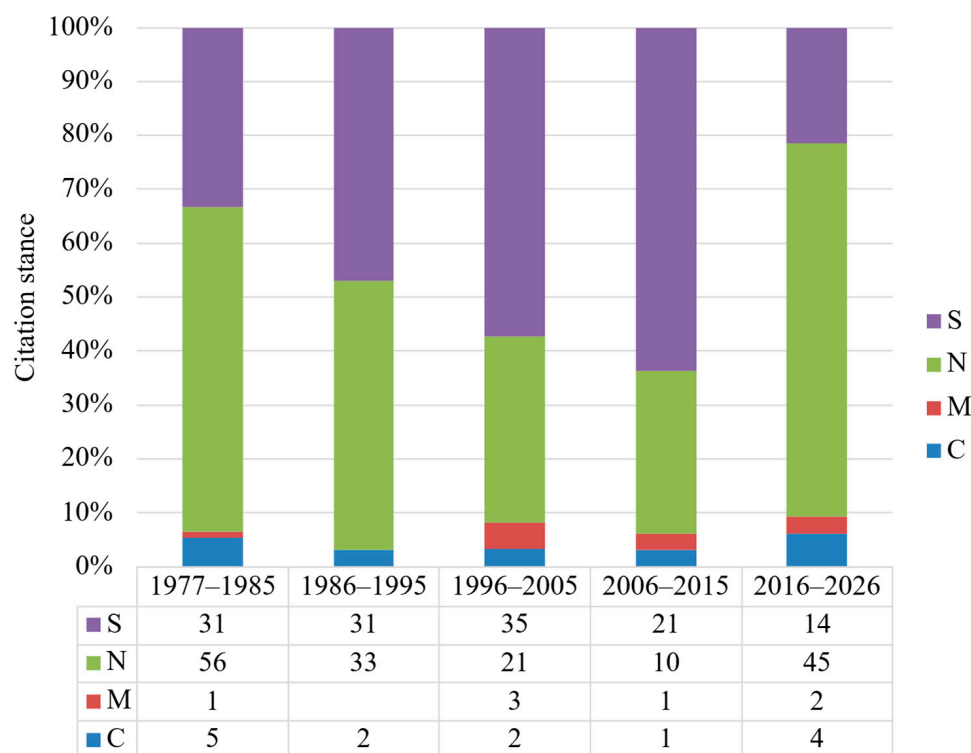


Figure 2. Citation stance to Crump et al. (1976) over time: S, supportive; N, neutral; M, mixed; C, critical. Data for 2026 was up to 1 June 2026.

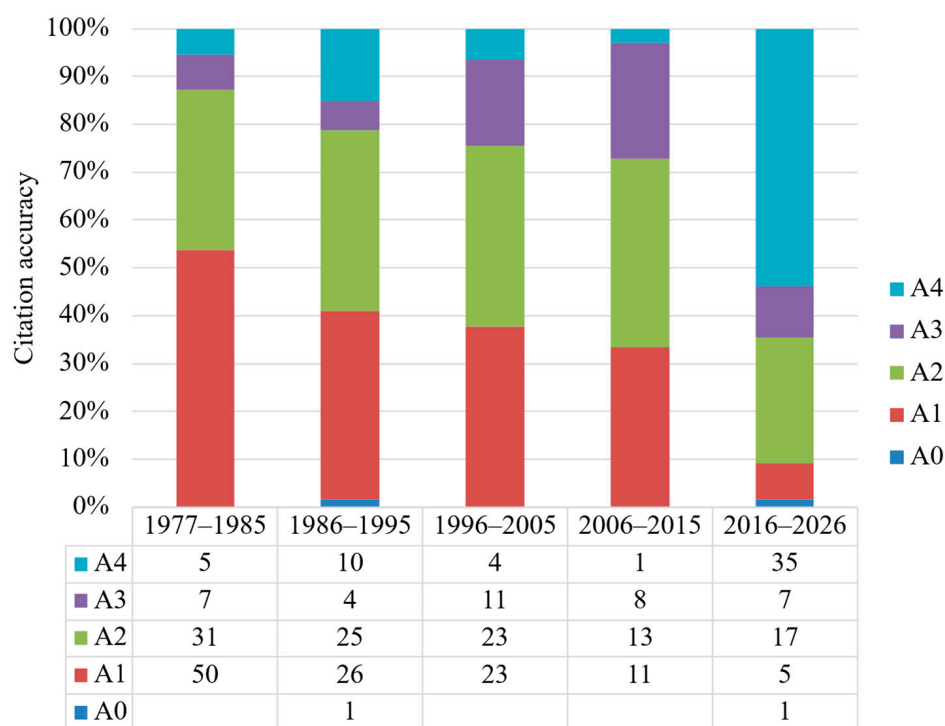


Figure 3. Citation accuracy to Crump et al. (1976) over time: A1, accurate and specific; A2, acceptable simplification; A3, overgeneralized or potentially misleading; A4, clearly inaccurate or unrelated attribution. See Table 1 for the explanation of the codes. Data for 2026 was up to 1 June 2026.

Issue coding indicated that this recent increase in A4 citations was concentrated primarily in IA9 (generic unrelated dose–response miscitation), which accounted for 30 of 35 A4 cases in 2016–2026 (Figure 4). Because recent A4 citations were concentrated in IA9, an additional descriptive clustering check was performed. The 30 IA9 cases in 2016–2026 appeared in 30 different journals and involved 27 distinct author groups. The largest single author group accounted for three cases. Thus, the IA9 increase was not dominated by a single citation cluster. OG4 (extension beyond carcinogenesis without qualification) was the most prevalent OG mode during 2006–2015, whereas OG3 (regulatory conclusion overstated) was the most prevalent during 2016–2026 (Figure 5). Meanwhile, IA3, IA4, and IA7 were logically included in the final codebook, but after coding the entire dataset, no citing papers belonged to these categories.

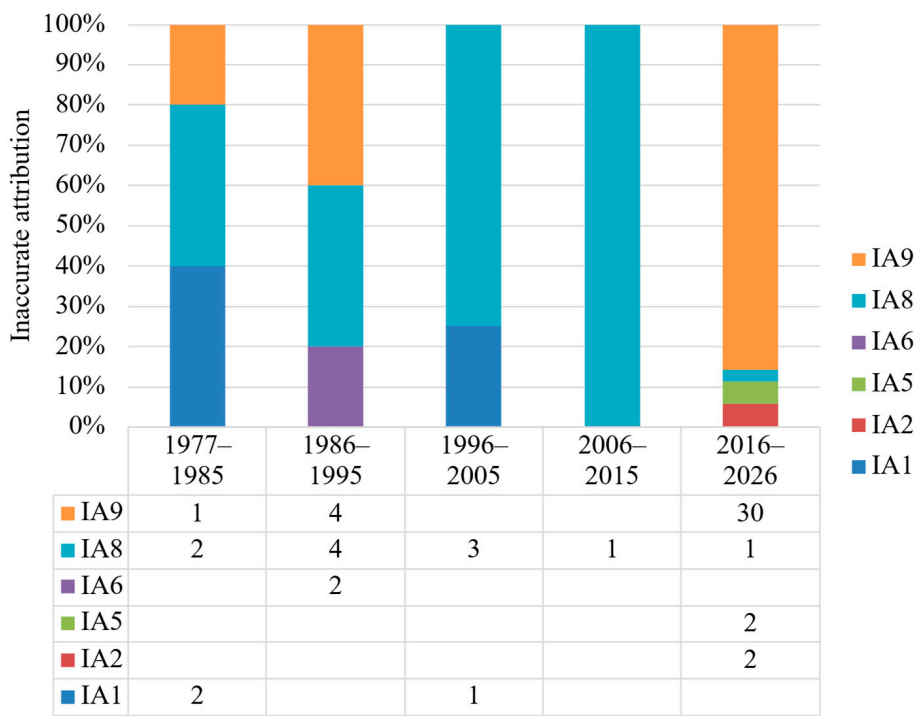


Figure 4. Inaccurate attribution (IA) to Crump et al. (1976) over time. Overall, there were few IAs, except that there were many generic unrelated dose–response miscitations (IA9) during 2016–2026. See Table 1 for the explanation of the codes. Data for 2026 was up to 1 June 2026.

As an additional descriptive analysis, the citing papers were counted according to the journals in which they were published (Table 2). Four journals had over 10 citing papers each, namely, *Environmental Health Perspectives*, *Regulatory Toxicology and Pharmacology*, *Risk Analysis*, and *Critical Reviews in Toxicology*. The most frequent source titles were concentrated in environmental health, regulatory toxicology, risk analysis, and toxicology journals, consistent with the paper’s original role in carcinogenic risk assessment. In contrast, many recent IA9 citations appeared in more diverse biomedical, pharmacological, engineering, and data-science journals, suggesting diffusion of the citation beyond its original disciplinary setting. The coded datasheet of the 318 citing papers is provided as Supplementary Table S1.

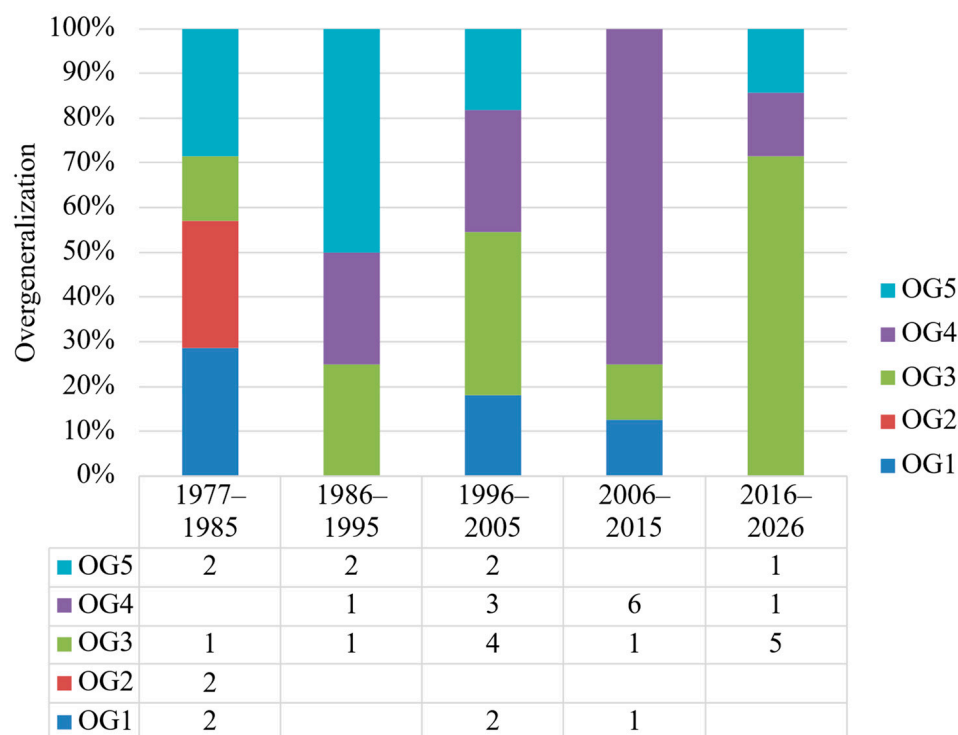


Figure 5. Overgeneralization (OG) to Crump et al. (1976) over time. See Table 1 for the explanation of the codes. Overall, there were few OGs. Data for 2026 was up to 1 June 2026.

Table 2. Frequency count of citing papers by journal.

Journal	No. of Citing Papers Published	% of 318
<i>Environmental Health Perspectives</i>	34	10.7%
<i>Regulatory Toxicology and Pharmacology</i>	26	8.2%
<i>Risk Analysis</i>	15	4.7%
<i>Critical Reviews in Toxicology</i>	12	3.8%
<i>Human and Ecological Risk Assessment</i>	9	2.8%
<i>Chemico-Biological Interactions</i>	8	2.5%
<i>Biometrics</i>	7	2.2%
<i>Cancer Research</i>	7	2.2%

There were another 131 journals that each published less than 2% of the citing papers.

4. Discussion

4.1. Key Take-Home Messages

Citations to Crump et al. (1976) were not predominantly canonical in the narrow sense of directly invoking the low-dose linearity argument made by the paper. However, for nearly four decades, the citations remained overwhelmingly within the paper's broader original domain: carcinogenic risk assessment, multistage modeling, low-dose extrapolation, threshold debate, and related quantitative risk-assessment contexts. This changed sharply in 2016–2026 (up to 1 June 2026), when citations shifted toward generic dose-response or irrelevant contexts, became much more likely to be overgeneralized or potentially misleading (A3) or clearly misleading (A4), and detached from the specific low-dose carcinogenic-risk argument that made the paper important in the first place.

This pattern suggests a transition from disciplinary shorthand to citation distortion. In the former case, a foundational paper is cited in abbreviated form but still within its

conceptual boundary. In the latter case, the citation becomes detached from the actual content of the cited paper and may function as a symbolic or erroneous reference to “dose response” in general. There are many terms proposed by researchers to generally categorize citation irregularities. It is beyond the scope of this study, and readers are referred to [Zhang et al. \(2026\)](#) for further information.

Regardless, this change in citation pattern is important for radiation protection because the LNT debate is already vulnerable to oversimplification. [Crump et al. \(1976\)](#) offered a conditional theoretical argument, not an unrestricted proof of universal low-dose linearity. When later authors cite it as if it established all non-threshold models, all regulatory uses of LNT, or unrelated dose–response procedures, they risk obscuring the assumptions that made the original argument scientifically meaningful. Such miscitation may appear minor at the level of individual papers, but cumulatively it can distort how a field remembers its own conceptual foundations.

It is notable that citations coded IA5 (inaccurate software or computational-procedure attribution) emerged during 2016–2026. Although there were only two such citing papers (Paper_ID #306 and #317)—both from the same author group—that cited [Crump et al. \(1976\)](#) for using GraphPad Prism software, there is a risk that subsequent researchers may follow this citation practice for their own articles if they do not attempt to verify the original source of reference ([Gavras, 2002](#); [Yeung, 2025](#)). A similar situation happened during 1986–1995, when two citing papers (Paper_ID #106 and #130) cited [Crump et al. \(1976\)](#) for general statistical concepts. Those occasions could be considered isolated incidents that did not proliferate.

Excluding such generic or irrelevant citations, one could observe a decline in the number of citations made to [Crump et al. \(1976\)](#) from previous periods to the last 10 years. One potential explanation could be “obliteration by incorporation”, meaning that the research community has become so familiar with the concepts elaborated by a reference that researchers no longer make a citation to the original source but, rather, treat the concepts as common knowledge ([Merton, 1965](#); [Yeung, 2021](#)). Alternative explanations include changes in disciplinary composition, newer canonical sources replacing older ones, database coverage, or changes in terminology. In fact, there have been increasing numbers of publications related to the linear no-threshold model (LNT) and low-dose radiation ([Park et al., 2019](#); [Yeung et al., 2026](#)), implying that researchers in radiation physics or radioprotection have many recent and relevant potential references to cite, instead of a reference published in the distant past.

4.2. Relevance to the LNT and Radiation Protection Scholarship

The findings also help distinguish two related but separate issues: the validity of LNT as a radiological protection model, and the accuracy with which historical sources are cited. This study did not test whether LNT is biologically correct at low doses, nor did it adjudicate between linear, threshold, supralinear, or hormetic models. Instead, it evaluated how one influential low-dose linearity paper has been represented in the later literature. This distinction is useful because debates about LNT often mix empirical evidence, biological plausibility, ethical precaution, regulatory practicality, and historical attribution ([Calabrese, 2022](#)). A citation context approach can help separate what a source actually argues from how it is later mobilized in policy or scientific debate.

Radiation protection relies on models that are simultaneously scientific, practical, and normative. The LNT model is often defended as a prudent basis for protection under uncertainty, even when low-dose risks cannot be estimated with high precision ([Boice, 2017](#)). However, prudential use of a model does not require imprecise citation of the historical literature. On the contrary, careful citation is especially important when scientific models

are used in regulatory and public-health contexts, where misstatements can influence perceptions of certainty, controversy, or authority (August et al., 2026; Rhodes & Lancaster, 2022; Socol et al., 2019).

4.3. Comparison with the Existing Literature

In this study, the overall rate of A3/A4 (A3 = overgeneralized or potentially misleading; A4 = clearly inaccurate or unrelated attribution) citations reached 28.9% (92/318). This rate is slightly lower than the miscitation rate of 34.8% for an influential meta-analysis on active learning (Martella et al., 2021), but higher than the 14.5% reported for the medical literature (Mogull, 2017) and 19% reported for the psychology literature (Cobb et al., 2024). The distinctive contribution of the present study is that it tracks how citation accuracy changed over time for a single foundational paper, revealing a temporal pattern of citation drift rather than a static snapshot of error rates.

4.4. Methodological Limitations

Citation context analysis provides information that cannot be obtained from citation counts alone. Whereas conventional bibliometric indicators measure how often a paper is cited, citation context analysis examines why and how it is cited. This is particularly useful for foundational papers such as the work of Crump et al. (1976), whose influence may include accurate use, shorthand reference, overgeneralization, criticism, and miscitation. However, several limitations should nevertheless be acknowledged. First, the analysis relied on the Web of Science Core Collection (WOSCC). Although the WOSCC is selective and widely used, it may incompletely index older publications, and some records may lack complete metadata or reference information. Second, full-text access was unavailable for some citing papers, which may have led to the exclusion of relevant citation contexts. Third, coding was performed by a single author, so inter-rater reliability could not be assessed. To partially address this limitation, the author recoded a random 20% subset ($n = 64$) after a two-week interval without consulting the original codes; intra-coder agreement was $\kappa = 0.75$ for citation context, $\kappa = 0.81$ for citation accuracy, and $\kappa = 0.90$ for citation stance. Nevertheless, intra-coder reliability is not a substitute for independent double-coding. Fourth, each citing paper was assigned dominant codes even when Crump et al. (1976) was cited multiple times. This approach avoided over-weighting individual papers but may have underrepresented within-paper variation in citation use. Moreover, the coded stance reflects the local citing passage rather than the overall position of the citing article toward LNT, low-dose risk assessment, or radiation protection policy. Finally, this study concerned one foundational paper, Crump et al. (1976), and should not be interpreted as evidence that all foundational papers undergo the same form of citation drift.

4.5. Future Perspectives

Future research could select foundational papers from multiple disciplines (e.g., Kaplan–Meier in biostatistics (Kaplan & Meier, 1958), Shannon in information theory (Shannon, 1948), Coase in economics (Coase, 2013)), apply a common citation context coding framework, and compare temporal patterns of citation drift. Such comparative work could identify whether the pattern observed here is characteristic of foundational papers in general. Such comparative work could also clarify whether the time course of citation drift follows predictable phases—for instance, an early period of accurate disciplinary use, followed by broader diffusion once a paper has accumulated enough citations to make it visible to researchers outside the original field. Identifying discipline-level or publication-level predictors of drift (e.g., audience breadth of the original journal, availability of newer authoritative publications, or interdisciplinary reach of the topic) would help determine whether targeted interventions, such as enhanced reference verification

norms in specific journal communities, could slow the accumulation of inaccurate citations to foundational works.

5. Conclusions

This study shows that the citation history of Crump et al. (1976) reflects both enduring influence and substantial recent citation drift. For several decades, most citations remained broadly connected to the original conceptual domain of the paper, including carcinogenic risk assessment, low-dose extrapolation, multistage modeling, and threshold debate. Although many citing papers used the article as shorthand rather than engaging with its full argument, these uses generally remained within the boundaries of the original work. In contrast, the most recent decade showed a marked increase in peripheral, generic, or unrelated dose–response citations, together with a substantial rise in inaccurate or misleading attributions.

The broader implication is that citation accuracy cannot be inferred from citation counts alone. Citation context analysis can reveal whether influential papers are being cited accurately, used as acceptable disciplinary shorthand, overgeneralized, or misrepresented. Careful source verification is therefore essential, particularly when older foundational papers become familiar references in contemporary literature.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/publications14040072/s1>: Table S1: Coded datasheet of the analyzed papers. Table S2: List of excluded citing papers without access to their full text.

Funding: This research received no external funding.

Data Availability Statement: All data are available in the manuscript and Supplementary Table S1.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LNT	Linear no-threshold model
WOSCC	Web of Science Core Collection

References

- August, E., Mahieu, A., & Brouwer, A. F. (2026). Building better citation practices in public health research: Practical considerations for student researchers, from source discovery to crafting and positioning claims. *PLoS Global Public Health*, 6(9), e0007257. [CrossRef] [PubMed]
- Baldwin, J., & Grantham, V. (2015). Radiation hormesis: Historical and current perspectives. *Journal of Nuclear Medicine Technology*, 43(4), 242–246. [CrossRef] [PubMed]
- Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1), 363–376. [CrossRef]
- Boice, J. D., Jr. (2017). The linear nonthreshold (LNT) model as used in radiation protection: An NCRP update. *International Journal of Radiation Biology*, 93(10), 1079–1092. [CrossRef] [PubMed]
- Bosch de Basea, M., Thierry-Chef, I., Harbron, R., Hauptmann, M., Byrnes, G., Bernier, M.-O., Le Cornet, L., Dabin, J., Ferro, G., & Istad, T. S. (2023). Risk of hematological malignancies from CT radiation exposure in children, adolescents and young adults. *Nature Medicine*, 29(12), 3111–3119. [CrossRef] [PubMed]
- Calabrese, E. J. (2022). Key historical study findings questioned in debate over threshold versus linear non-threshold for cancer risk assessment. *Chemico-Biological Interactions*, 359, 109917. [CrossRef] [PubMed]
- Coase, R. H. (2013). The problem of social cost. *The Journal of Law and Economics*, 56(4), 837–877. [CrossRef]
- Cobb, C. L., Crumly, B., Montero-Zamora, P., Schwartz, S. J., & Martínez, C. R., Jr. (2024). The problem of miscitation in psychological science: Righting the ship. *American Psychologist*, 79(2), 299–311. [CrossRef] [PubMed]

- Crump, K., Hoel, D., Langley, C., & Peto, R. (1976). Fundamental carcinogenic processes and their implications for low dose risk assessment. *Cancer Research*, 36, 2973–2979. [PubMed]
- Daivids, J. R., Weigl, D. M., Edmonds, J. P., & Blackhurst, D. W. (2010). Reference accuracy in peer-reviewed pediatric orthopaedic literature. *Journal of Bone and Joint Surgery*, 92(5), 1155–1161. [CrossRef] [PubMed]
- Gavras, H. (2002). Inappropriate attribution: The “lazy author syndrome”. *American Journal of Hypertension*, 15(9), 831. [CrossRef] [PubMed]
- Hausmann, N. S., McIntyre, T., Bumby, A. J., & Loubser, M. J. (2013). Referencing practices in physical geography: How well do we cite what we write? *Progress in Physical Geography*, 37(4), 543–549.
- ICRP. (2005). Low-dose extrapolation of radiation-related cancer risk. ICRP Publication 99. *Annals of the ICRP*, 35(4), 1–140. [PubMed]
- Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481. [CrossRef]
- Lin, F. K., Banerjee, M. R., & Crump, L. R. (1976). Cell cycle-related hormone carcinogen interaction during chemical carcinogen induction of nodule-like mammary lesions in organ culture. *Cancer Research*, 36(5), 1607–1614. [PubMed]
- Liu, W. (2019). The data source of this study is web of science core collection? Not enough. *Scientometrics*, 121(3), 1815–1824. [CrossRef]
- Luckey, T. D. (2006). Radiation hormesis: The good, the bad, and the ugly. *Dose-Response*, 4(3), 169–190. [CrossRef] [PubMed]
- Martella, A. M., Yacilla, J. K., Martella, R. C., Marchand-Martella, N. E., Ozen, Z., Karatas, T., Park, H. H., Simpson, A., & Karpicke, J. D. (2021). Quotation accuracy matters: An examination of how an influential meta-analysis on active learning has been cited. *Review of Educational Research*, 91(2), 272–308. [CrossRef]
- Merton, R. K. (1965). *On the shoulders of giants: A Shandean postscript*. Free Press.
- Mogull, S. A. (2017). Accuracy of cited “facts” in medical research articles: A review of study methodology and recalculation of quotation error rate. *PLoS ONE*, 12(9), e0184727. [CrossRef] [PubMed]
- Muller, H. J. (1927). Artificial transmutation of the gene. *Science*, 66(1699), 84–87. [CrossRef] [PubMed]
- National Research Council. (1956). *Biological effects of atomic radiation: A summary report to the public from the national academy of sciences*. The National Academies Press.
- Park, J., Kwon, T., Lee, S.-S., Jin, Y. W., & Seong, K. M. (2019). Mapping the research trends on the biological effects of radiation less than 100 mSv: A bibliometric analysis for 30 years publication. *International Journal of Radiation Biology*, 95(5), 527–536. [CrossRef] [PubMed]
- Rhodes, T., & Lancaster, K. (2022). Uncomfortable science: How mathematical models, and consensus, come to be in public policy. *Sociology of Health and Illness*, 44(9), 1461–1480. [CrossRef] [PubMed]
- Rühm, W., Laurier, D., & Wakeford, R. (2022). Cancer risk following low doses of ionising radiation—Current epidemiological evidence and implications for radiological protection. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 873, 503436. [CrossRef] [PubMed]
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. [CrossRef]
- Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126(6), 5113–5142. [CrossRef]
- Socol, Y., Shaki, Y. Y., & Yanovskiy, M. (2019). Interests, bias, and consensus in science and regulation. *Dose-Response*, 17(2), 1559325819853669. [CrossRef] [PubMed]
- UNSCEAR. (2024). *Sources, effects and risks of ionizing radiation*. United Nations. [CrossRef]
- Wojcik, A., & Zölzer, F. (2024). The scientific nature of the linear no-threshold (LNT) model used in the system of radiological protection. *Radiation and Environmental Biophysics*, 63(4), 483–489. [CrossRef] [PubMed]
- Yeung, A. W. K. (2021). Is the influence of Freud declining in psychology and psychiatry? A bibliometric analysis. *Frontiers in Psychology*, 12, 631516. [CrossRef] [PubMed]
- Yeung, A. W. K. (2025). Evaluation of content quality of online health information by global quality score: A case study of researchers misnaming it and citing secondary sources. *Publications*, 13(2), 23. [CrossRef]
- Yeung, A. W. K., Nalley, A., Hung, K. F., & Tanaka, R. (2026). *The evolution of the linear no-threshold model in radiation protection: Insights from a bibliometric perspective*. Health Physics. [Epub ahead of print]. [CrossRef] [PubMed]
- Zhang, Q., Li, Y., & Bouter, L. (2026). How to categorize citation irregularities: A proposal based on an exploration of the literature. *Accountability in Research*, 33, 2654564. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.