REVIEW

Editorial Process: Submission:05/10/2023 Acceptance:12/17/2023

Clinical Epidemiological Studies of Colorectal Cancer by Record Linkage of Cancer Registries and Biospecimen Data: A Systematic Review

Aoi Kataoka[1,2], Masato Ota[3], Kohei Taniguchi[4], Kazumasa Komura[4,5], Yuri Ito[1]*

Abstract

Background: The incidence and prognosis of colorectal cancer are associated with lifestyle, family history, and genetic predisposition. Record linkage between cancer registries and biospecimen data would enable us to conduct clinical epidemiological studies on incidence or prognosis including genome information. In this study, we conducted a systematic review of clinical epidemiological studies of colorectal cancer using record linkage between cancer registries and biospecimen data and examined the possibilities for future use of this linkage. **Methods:** We searched PubMed and Google Scholar for articles regarding cancer registries and biospecimen data use published before December 2021. Selected articles were summarized by cancer registry use, biospecimen use, exposure, outcome, informed consent, and participant numbers by study design and type of cancer registry. **Results:** Of the 2,793 identified articles, 81 studies were included in this review. The most frequently used cancer registries and study design were site specific cancer registries and cohort studies. Most use of cancer registries was for patient selection in cohort studies and case selection in case-control studies. Most use of biospecimen data was for prognostic factors in cohort studies and risk factors in case-control studies. In site specific cancer registries for the examination of familial colorectal cancer, most use of biospecimen data is to examine genome mutation, expression, or deficiency. **Conclusion:** We suggest that record linkage between cancer registries and biospecimen data would enable the accurate capture of outcomes and detailed genome-environmental factors, and to conduct clinical epidemiological studies according to specific research questions and tailored study designs.

Keywords: Biological specimen banks- colorectal neoplasms- medical record linkage- registries- systematic review

Asian Pac J Cancer Prev, **24 (12)**, 4017-4023

Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the second most common cause of cancer death worldwide (Ferlay et al., 2020). The Asia-Pacific region has the highest number of cases of CRC and CRC mortality in the world (Rebeneck et al., 2020), especially in Japan, where colorectal cancer is the most commonly diagnosed cancer and the second most common cause of cancer death (Cancer Information Service NCC, Japan, 2018). Approximately 70% of CRC cases are attributed to environmental factors, which are modifiable risk factors such as lifestyle (e.g., diet, physical activity, alcohol consumption, and smoking) (Wong et al., 2019; Cho and Shin, 2021). Other CRC cases may be caused by non-modifiable risk factors such as sex, age, race, family history, and genetic predisposition (Jasperson, et al., 2010; Rosa et al., 2015). To prevent CRC, we need to deal with both modifiable and non-modifiable risk factors by investigating the role of non-modifiable factors concerning modifiable factors such as changing lifestyle (Boffetta et al., 2012; Rawla et al., 2019; Yang et al., 2019).

Biomarkers play an essential role in the clinical outcomes of CRC. Identifying the existence of biomarkers aids early detection or good vs. poor CRC prognosis. In addition, measuring increases or decreases in biomarkers aids assessment of the effectiveness of adjuvant chemotherapy (Huang et al., 2010; Thirunavukarasu et al., 2011; Duffy et al., 2014). Thus, in clinical and epidemiological studies, examining the association between biomarkers and clinical outcomes can help prevent CRC, manage or treat patients, and predict

*[1]Department of Medical Statistics, Research & Development Center, Osaka Medical and Pharmaceutical University, Takatsuki City, Osaka, Japan. [2]Division of Molecular Epidemiology, Future Medical Science, Kobe University, Kobe City, Hyogo, Japan. [3]Department of General and Gastroenterological Surgery, Osaka Medical and Pharmaceutical University, Takatsuki City, Osaka, Japan. [4]Translational Research Program, Research & Development Center, Osaka Medical and Pharmaceutical University, Takatsuki City, Osaka, Japan. [5]Department of Urology, Osaka Medical and Pharmaceutical University, Takatsuki City, Osaka, Japan. *For Correspondence: yuri.ito@ompu.ac.jp*

prognosis. To investigate the roles and influence of biomarkers in clinical outcomes, registry-based biospecimen studies that link biospecimen data and electronic databases have become more practicable, and biobanks have been created at medical and research institutions around the world (Sudlow et al., 2015; Nagai et al., 2017). A biobank is based on the systematic collection of biospecimens and also health information such as medical, treatment, and lifestyle data collected at the time of biospecimen collection and is further enhanced by the availability of clinical follow-up data to conduct outcome studies (Olson, et al., 2014). Cancer registries are considered influential among the existing electronic databases linked with biospecimen data (Edwards and Bell., 2000; MacCallum, et al., 2018; Tucker, et al., 2019). Population-based cancer registries and hospital-based cancer registries collect information on newly diagnosed cancer patients and followed-up patients, and the criteria for registration and classification of tumor type are internationally standardized. In addition, site-specific cancer registries, which focus on specific cancer and are operated mainly by academic cancer research associations, can collect more detailed clinical information about patients and treatment than that available in population-based or hospital-based cancer registries. Record linkage between cancer registries and biospecimen data enables us to collect both accurate information on cancer diagnosis and prognosis and detailed gene-environment information about biospecimen providers, which can be used to construct valuable databases for molecular epidemiology containing longitudinal data from baseline to follow-up (Langseth, et al., 2010; Dillner, 2015).

In this systematic review, we aim to investigate the possibility of using cancer registry and biospecimen data in the clinical and molecular epidemiology of CRC by identifying and evaluating clinical and molecular epidemiological articles that used record linkage between cancer registries and biospecimen data.

Materials and Methods

Search strategy

Studies were obtained by searching the following two electronic databases: PubMed and Google scholar. We performed the final search in January 2022 and included studies published up to December 2021. The search terms used to identify cancer registries and biospecimen of interest are shown in Table 1.

Selection strategy

We determined inclusion and exclusion criteria before undertaking the review. We included studies meeting the criteria: (1) focused on CRC, including Lynch syndrome (often called hereditary nonpolyposis colorectal cancer) and Familial adenomatous polyposis, (2) used both cancer registries and biospecimen data, (3) covered clinical and molecular epidemiology, (4) were published in peer-reviewed journals, and (5) were written in English or Japanese. Studies were excluded if they met the following criteria: (1) focused on other cancer types, (2) did not use cancer registries and biospecimen data, (3) did not clearly describe the process of obtaining information about participants and biospecimen data, (4) did not clearly describe the details of the type of biospecimen, (5) did not cover clinical and molecular epidemiology, (6) have not been peer-reviewed or were not original articles.

Data extraction

Data extraction was performed by two independent researchers (AK and MO), and differences in judgment between the two researchers were discussed until they reached a consensus. The extracted data included the following information: title, journal name, article type, publication year, country of a survey, type of cancer registry, purpose, study design, acquisition of informed consent (IC), number of participants, research period, exposure or intervention factors, outcome indices, use of cancer registries, use of biospecimen data, and statistical methods.

We classified study design into "cohort study," "case-control study," "cross-sectional study," and "intervention study or randomized control trials" and classified type of cancer registry into "hospital-based cancer registries (HBCRs)," "population-based cancer registries (PBCRs)," "national cancer registries (NCRs)," "site-specific cancer registries (SSCRs), " and "multiple cancer registries (Multiple CRs)" that use more than two cancer registries. Then, we organized the findings of the articles into the following categories according to the classified cancer registries for each study design: use of cancer registries, use of biospecimen data, exposure or intervention factors, outcome indices, acquisition of informed consent, and number of participants (number of cases for case-control studies). Results were reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher, et al., 2009). The review protocol was not registered.

Results

Search and selection

A flow diagram of the identification, screening, eligibility, and inclusion of studies is shown in Figure 1. A total of 2,793 articles were identified through electronic database searching; 1,713 from PubMed and 1,080 from Google scholar. After removing duplicate reports, 2,788 articles remained. We conducted title and abstract screening for these articles, and 95 papers were assessed for eligibility. After reading the complete text, 81 articles were considered to meet the inclusion criteria and included in this review.

Study characteristics

The detailed characteristics of the included articles are described in Table S1. The types of cancer registry by study designs are shown in Table 2. The highest use of a cancer registry was 44.2% (23/52) for SSCRs in cohort studies, 31.8% (7/22) for Multiple CRs in case-control studies, and 42.9% (3/7) for PBCRs in cross-sectional studies. Countries, where the studies were conducted by type of cancer registry are shown in Table S2. NCRs

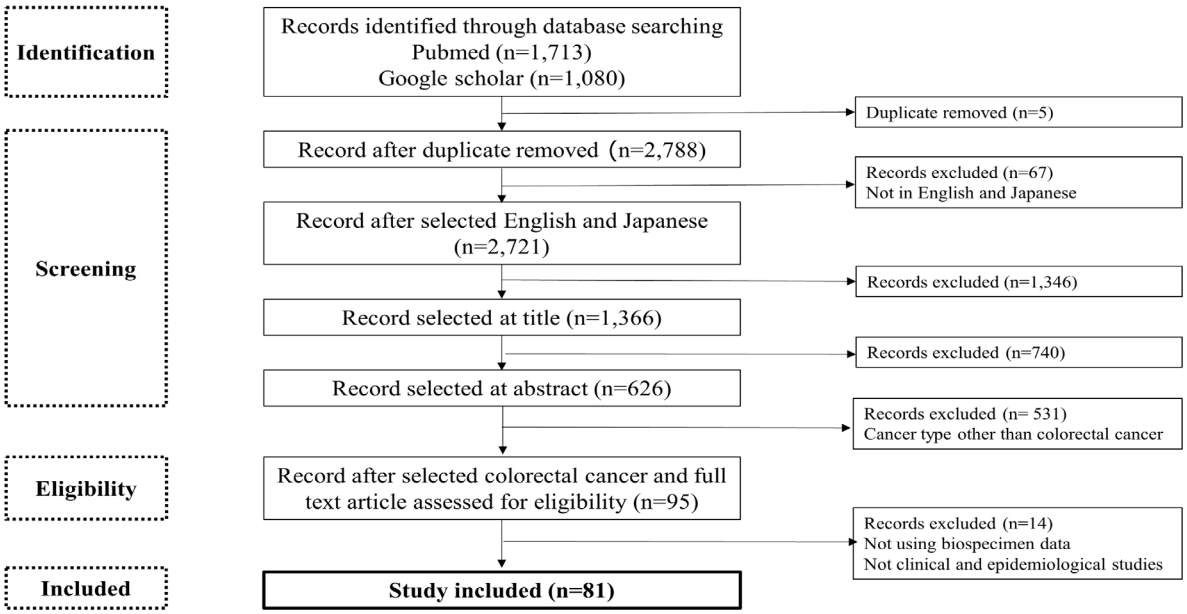

Figure 1. Flow Chart Diagram of the Search Strategy and the Inclusion of Articles along Various Steps

Table 1. Search Terms: Keywords and MeSH terms for systematic review

Concept	Keywords	MeSH terms
Cancer registry	"""cancer registr*“ OR ""tumor registr*"" OR ""hospital cancer registr*"" OR ""hospital tumor registr*"" OR ""hospital based cancer registr*"" OR ""hospital based tumor registr*"""	Neoplasms, Registries, Records
Biospecimen	urine OR blood OR “biological samples” OR specimen OR genome	"Urine Specimen Collection, Biological Specimen Banks, Blood Specimen Collection, Semen Analysis, Hemorheology, Genetic Techniques, Genome"

were the most frequently used registries in Finland (3/11, 27.3%), PBCRs and HBCRs in the United States (PBCRs: 11/17, 64.7%, HBCRs: 9/13, 69.2%), and SSCRs and Multiple CRs were used in collaborations between several countries (SSCRs: 9/29, 31.0%, Multiple CRs: 8/11, 72.7%).

In studies that conducted record linkage between cancer registries and biospecimen data, the main exposure was genome information (68/81,84.0%) and the main outcomes were cancer incidence (36/81, 44.4%) and prognosis (35/81, 43.2%) (Figure 2). Thus, record linkage between cacner registries and biospecimen data in CRC would enable researchers to mainly consider the impact of detailed genome information such as gene murations,

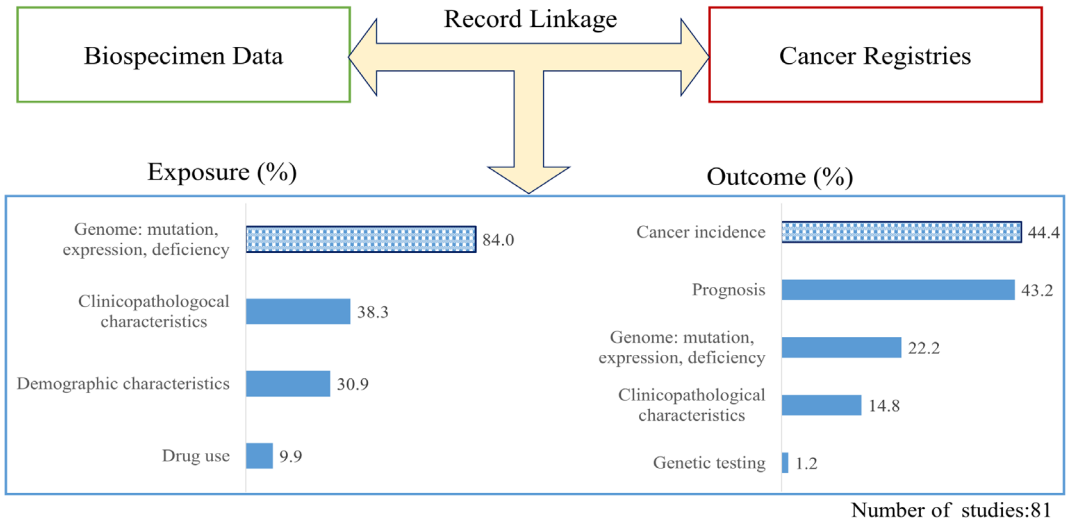

Figure 2. Studies that can be Conducted by Record Linkage between Cancer Registries and Biospecimen Data in Colorectal Cancer

Table 2. Type of Cancer Registry by Study Design

	Cohort study		Case-control study		Cross sectional study		Total	
	n	%	n	%	n	%	n	%
NCR	6	11.5	5	22.7	0	0.0	11	13.6
PBCR	11	21.2	3	13.6	3	42.9	17	21.0
HBCR	9	17.3	3	13.6	1	14.3	13	16.0
SSCR	23	44.2	4	18.2	2	28.6	29	35.8
Multiple CRs	3	5.8	7	31.8	1	14.3	11	13.6
Total	52	100.0	22	100.0	7	100.0	81	100.0

NCR, National cancer registry; PBCR, Population based cancer registry; HBCR, Hospital based cancer registry; SSCR, Site specific cancer registry; Multiple CRs, Using more than two types of cancer registry

expression, deletions or phenotype on the incidence or prognosis of CRC.

Cohort studies

The results of the cohort studies are summarized in Table S3. The highest use of cancer registry data was "Patients selection" in SSCRs (23/23, 100%), HBCRs (6/9, 66.7%), Multiple CRs (2/3, 66.7%), and PBCRs (4/11, 36.4%), and the highest use of biospecimen data was for "Prognostic factors" in all types of cancer registry: HBCRs (9/9, 100%), Multiple CRs (3/3, 100%), PBCRs (8/11, 72.7%), NCRs (4/6, 66.7%),, and SSCRs (9/23, 39.1%). Some studies using SSCRs data also used biospecimen data as "Surrogate endpoint" to investigate genetic mutation, expression, or deficiency (8/23, 34.8%). Most studies using HBCRs, Multiple CRs, and SSCRs data did not describe the acquisition of IC: HBCRs (9/9, 100%), Multiple CRs (2/3, 66.7%), and SSCRs (14/23, 60.9%); however, most studies using NCRs and PBCRs data obtained IC: NCRs (5/6, 83.3%) and PBCRs (6/11, 54.5%). The number of participants was: under 99 people and 100-499 people in HBCRs (under 99 people: 3/9, 30.0%, 100-499 people: 6/9, 66.7%), 500-999 people in PBCRs (7/11, 63.6%), and over 1000 people in NCRs (2/3, 66.7%).

Case-control studies

Results for case-control studies are summarized in Table S4. The highest use of cancer registry data was for "Case selection" in all cancer registries; PBCRs (3/3, 100%), HBCRs (3/3, 100.0%), Multiple CRs (6/7, 85.7%), NCR (4/5, 80.0%),, and SSCRs (2/4, 50.0%),. The highest use of biospecimen data was for "Risk factors" in PBCRs (3/3, 100%), Multiple CRs (7/7, 100%), NCRs (4/5, 80.1%), and SSCRs (2/4, 50.0%). In studies using SSCRs data, biospecimen data was also used as "Surrogate endpoint" (2/4, 50.0%). Most studies in PBCRs, SSCRs, Multiple CRs, and NCRs obtained IC: PBCRs (3/3, 100%), SSCRs (3/4, 75.0%), Multiple CRs (5/7, 71.4%), and NCRs (3/5, 60%). The number of participants was under 99 people in HBCRs (2/3, 66.7%), 100-499 people in NCRs (2/5, 40.0%), 500-999 people in PBCRs (2/3, 66.7%), and over 1000 people in Multiple CRs (5/7, 71.4%).

Cross-sectional studies

Results for cross-sectional studies are shown in Table S5. Use of cancer registry data was for "Comparison with clinicopathological characteristics" in HBCRs (100%, 1/1), "Patient identification" in Multiple CRs (100%, 1/1), and "Patient selection" in PBCRs (100%, 3/3) and SSCRs (100%, 2/2). Use of biospecimen data was for "Factors affecting clinicopathological characteristics" in HBCR (100%, 1/1) and SSCR studies (100%, 1/1), "Risk factors" in Multiple CRs (100%, 1/1), and "Surrogate endpoint" in PBCRs (66.7%, 2/3) and SSCRs (50.0%., 1/2). The acquisition of IC was 100% (1/1) in HBCR, Multiple CRs, and SSCR studies and 66.7% (2/3) in PBCR studies. The number of participants was under 99 in SSCR studies (50.0%, 1/2), 100-499 in Multiple CRs (100%, 1/1), PBCR (66.7%, 2/3), and SSCR studies (50.0%, 1/2), and over 1000 in HBCR studies (100%, 1/1).

Discussion

This is the first systematic review to investigate clinical and molecular epidemiological studies focused on CRC and use record linkage between cancer registries and biospecimen data. The present study showed that the primary use of cancer registry and biospecimen data in cohort studies was for "Patient selection" and "Prognostic factors," and in case-control studies was "Case selection" and "Risk factors." In addition, epidemiological studies investigating genome polymorphism, mutation, or deficiency have been conducted using SSCR data, especially in familial cancer registries.

The results identifying the primary use of cancer registries and biospecimen data in cohort studies and case-control studies were consistent regardless of the type of cancer registry. Study designs were selected based on the outcomes of interest for our investigation (Langseth, et al., 2010). Cohort studies will be chosen when we investigate the risk factors for prognosis or clinical outcomes, and case-control studies will be selected to analyze risk for cancer incidence (Langseth, et al., 2010; Olson, et al., 2014; Dillner., 2015). It will be possible to conduct clinical and epidemiological studies based on specific research questions or study designs because record linkage between cancer registries and biospecimen data can capture details of risk factors including genome mutation, polymorphisms, deficiency, tumor markers, serum, or heredity, and accurate health outcomes such as

diagnosis and prognosis.

Different results for the primary use of biospecimen data were observed in SSCRs based studies, and the biospecimen data were used to investigate genome polymorphism, mutation, or deficiency. This result may be associated with the different characteristics of SSCRs from those of PBCRs or HBCRs. PBCRs and HBCRs collect information on cancer incidence, management, treatment, and outcomes in specific areas or hospitals with uniform data standards and definitions (Edwards and Bell.,2000; Tucker, et al., 2019). In this study, clinical colorectal cancer registries (CCCRs) and familial cancer registries focused on CRC (CFCRs) were regarded as SSCRs. CCCRs contain more detailed clinical information and epidemiological data than PBCRs and HBCRs. Previous studies using CCCRs data investigated surgical and oncological outcomes or prognosis as the primary outcome and proved the association of the care or treatment for the patient with prognosis (Andersson, et al., 2010; MacCallum, et al., 2018). CFCRs were developed to identify high-risk families with CRC, collect detailed personal and family histories, including epidemiological risk factors and biospecimen data, and conduct follow-up of the participants for CRC and other outcomes throughout their lifetime (Newcomb, et al., 2007; Vasen, et al., 2016). Linking these registries with biospecimen data would make it possible to identify the mutation or deficiency as the risk factors for the incidence and mortality of CRC and the risk for CRC in mutation carriers.

We observed a difference in the number of participants according to the type of cancer registry. Registry size varied considerably according to the extent of the defined patient pool and time the registry had been operational (MacCallum, et al., 2018). Studies based on NCR, PBCR, and HBCR data collected information on all newly-diagnosed cancer patients in specific areas or hospitals. Studies that used SSCR data had clearly defined selection criteria to obtain a sufficient sample size and collected more detailed clinical information on the patients and families selected based on the requirements. In the future, we may be able to use the data of a larger sample to investigate rare genome or cancer cases and to conduct sub-group analysis by using NCRs, PBCRs, and HBCRs or SSCRs in multi-center collaborations.

In cohort studies, 56.9% of the studies did not describe the acquisition of IC. This may be because they were retrospective cohort studies using data gathered before the IC rules were introduced. In addition, it would be impossible to obtain retrospective consent from all participants if the studies had a large number of participants or if many of the participants were likely to have died.

Strengths and limitations of record linkage for cancer registries and biospecimen data in CRC

The incidence and mortality of CRC are high worldwide, and many CRC patients are registered in cancer registries. If we linked biospecimen data with these cancer registries, we would be able to investigate the association of already-known gene polymorphisms, mutations, and deficiencies with clinical outcomes and the effect of the unknown genome on clinical outcomes. In addition, CRC caused by both environmental factors such as lifestyle and genetic predisposition, and genetic-environmental factors can be thoroughly investigated by linking cancer registries with biospecimen data, which may help to elucidate the pathology of CRC.

However, record linkage of cancer registries with biospecimen data would not be easy, especially for NCRs and PBCRs, due to the need for standardized data collection between regions and the amount of labor, time, and financial input required for the creation of a record linkage system. In addition, there may be differences between Japan and other countries regarding the use of biobanks or record linkages. In countries that have pioneered record linkage and secondary use of data such as Finland or other Scandinavian countries, comprehensive consent for research using biospecimen data, including the provision of biospecimen and personal information, and record linkage with various registries, has been adopted in accordance with the Biobank Act 2013 (Ministry of Social Affairs and Health, Finland., 2012). This allows the linkage of biospecimen data with medical, social, and economic databases maintained by various hospitals and governments with only an opt-out setting. The United Kingdom has successfully established the UK Biobank, one of the largest biobanks in the world, without a law such as Finland's Biobank Act. The UK Biobank recruited ordinary people registered with a National Health Service (NHS) general practitioner and obtained IC not only to provide biospecimen data but also to record linkage with several health-related databases. The success of this can be attributed to a healthcare system like the UK NHS that facilitated record linkage and public awareness that participation in this research would benefit many people in the future.

In contrast, the obstacles to record linkage for biospecimen data with databases maintained in other hospitals or governments are very challenging in Japan without a law like the Biobank Act (Kimura, et al., 2020). This is because consent must be obtained separately from participants when providing data to an outside organization for record linkage (Ministry of Health, Labour and Welfare., 2022). In particular, genomic data with information that could lead to personal identification is considered highly sensitive, and therefore, the opt-out only setting would not be permitted in principle (Personal Information Protection Commission., 2022). Such strict regulation would make using data through record linkage between databases across hospitals or governments difficult. In addition, while the success of biobanks in Finland or the UK has, in part, been due to public understanding of the research and trust in the reliability of governments or scientific communities (Soini., 2016), in Japan, knowledge of the use of banked biospecimens and databases for research is probably lower than other countries. Therefore, it will be necessary for the future to inform the public more widely about the value of using biobanks or record linkages with databases not only from hospitals or universities but also from national or local governments for research purposes.

Limitation

In this study, search limitation should be considered when interpreting our present findings. This study focused on the use of cancer registries and conducted an extensive literature search using several similar concepts of "cancer" and "registries" as keywords and Mesh Terms, but several articles were not extracted. Although these articles used cancer registries as their research methodology, they may not have been caught by this search strategy because they did not specifically focus on the use of cancer registries or did not mention the use of cancer registries in their manuscripts. To authors' knowledge, the famous cohort studies or hospital-based epidemiologic research in Japan, such as those examining the association between colorectal cancer risk and leptin, colorectal cancer incidence and inflammation markers, or the association between the risk of colorectal cancer and polymorphism (Tamakoshi et al., 2005; Matsuo et al., 2009; Song et al., 2019) were not extracted in this present study.

In conclusion, this review highlighted previous studies that used record linkage between cancer registries and biospecimen data relating to CRC and investigated the possibility of using this linkage in clinical and molecular epidemiological studies of CRC. We found that the primary use of cancer registry and biospecimen data is "Patient selection" and "Prognostic factors" in cohort studies and "Case selection" and "Risk factors" in case-control studies. In addition, we were particularly interested in the genetic predisposition of CRC and found previous studies had focused on genome polymorphism, mutation, and deficiency using Familial cancer registries. Using record linkage between cancer registries and biospecimen data in CRC makes it possible to accurately capture health outcomes and detailed gene-environmental factors and conduct future clinical and epidemiological studies based on specific research questions and tailored study design.

Author Contribution Statement

AK, MO, and YI conceived the project. AK and MO searched, collected, and screened all articles. YI supervised the project. AK wrote this manuscript supported by all authors, and all authors read and approved the final manuscript.

Acknowledgements

The authors acknowledge Tomoharu Sato for assisting to collect articles and Dr Julia Mortimer for her English language editing. The authors acknowledge the funders of this study: Grant-in-Aid for Scientific Research (B) [grant number 21H03174 to Y.I, K.T, and K.K].

Funding statement

This study was supported by The Ministry of Education, Science and Culture of Japan (a Grant-in-Aid for Scientific Research [B] No.21H03174 to Y.I, K.T, and K.K])

Ethics statement

This article is systematic review and does not involved any human or animal study.

Availability of data

All reviewed papers can be obtained from journals.

Study register

This study is not registered in any registering dataset.

Conflict of interest

The authors have no conflict of interest.

References

Andersson K, Bray F, Arbyn M, et al (2010). The interface of population-based cancer registries and biobanks in etiological and clinical research – current and future perspectives. *Acta Oncol*, **49**, 1227-34.

Boffetta P, Winn DM, Ioannidis JP, et al (2012). Recommendations and proposed guidelines for assessing the cumulative evidence on joint effects of genes and environments on cancer occurrence in humans. *Int J Epidemiol*, **41**, 686-704.

Cancer Information Service NCC, Japan (2018). Cancer Registry and Statistics 2018. Accessed January 31, 2022. https://ganjoho.jp/reg_stat/statistics/stat/cancer/67_colorectal.html#anchor1.

Cho S, Shin A (2021). Population Attributable Fraction of Established Modifiable Risk Factors on Colorectal Cancer in Korea. *Cancer Res Treat*, **53**, 480-6.

Dillner J (2015). A basis for translational cancer research on aetiology, pathogenesis and prognosis: Guideline for standardised and population-based linkages of biobanks to cancer registries. *Eur J Cancer*, **51**, 1018-27.

Duffy MJ, Lamerz R, Haglund C, et al (2014). Tumor markers in colorectal cancer, gastric cancer and gastrointestinal stromal cancers: European group on tumor markers 2014 guidelines update. *Int J Cancer*, **134**, 2513-22.

Edwards D, Bell J (2000). Cancer registries – future development and uses in Britain. *J Public Health Med*, **22**, 216-9.

Ferlay J, Ervik M, Lam F, et al (2020). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. https://gco.iarc.fr/today,

Huang Z, Huang D, Ni S, et al (2010). Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. *Int J Cancer*, **127**, 118-26.

Jasperson KW, Tuohy TM, Neklason DW, et al (2010). Hereditary and familial colon cancer. *Gastroenterology*, **138**, 2044-58.

Kimura E, Ohtera S, Sasaki K, et al (2020). Status of health sector registers and data provision in Finland. *Jpn J Stat Data Sci*, **50**, 47-80. (in Japanese)

Langseth H, Luostarinen T, Bray F, et al (2010). Ensuring quality in studies linking cancer registries and biobanks. *Acta Oncol*, **49**, 368-77.

MacCallum C, Skandarajah A, Gibbs P, et al (2018). The Value of Clinical Colorectal Cancer Registries in Colorectal Cancer Research: A Systematic Review. *JAMA Surg*, **153**, 841-9.

Matsuo K, Suzuki T, Ito H, et al (2009). Association between an 8q24 locus and the risk of colorectal cancer in Japanese. *BMC Cancer*, **9**, 379.

Ministry of Health, Labour and Welfare (2022). Guidelines for Research. Accessed January 31, 2022. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hokabunya/kenkyujigyou/i-kenkyu/index.html.

Ministry of Social Affairs and Health, Finland (2012). Biobank

Act. Accessed January 31, 2022. https://www.finlex.fi/en/laki/kaannokset/2012/en20120688.pdf.
Moher D, Liberati A, Tetzlaff J, et al (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*, **62**, 1006–12.
Nagai A, Hirata M, Kamatani Y, et al (2017). Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol*, **27**, 2-8.
Newcomb PS, Baron J, Cotterchio M, et al (2007). Colon Cancer Family Registry: An International Resource for Studies of the Genetic Epidemiology of Colon Cancer. *Cancer Epidemiol Biomarkers Prev*, **16**, 2331-43.
Olson JE, Bielinski SJ, Ryu E, et al (2014). Biobanks and personalized medicine. *Clin Genet*, **86**, 50-5.
Personal Information Protection Commission (2022). Amended Act on the Protection of Personal Information. Accessed January 31, 2022. https://www.ppc.go.jp/en/legal/.
Rawla P, Sunkara T, Barsouk A (2019). Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Prz Gastroenterol*, **14**, 89-103.
Rebeneck L, Chiu HM, Senere C (2020). International Perspectivr on the Burden of Colorectal Cancer and Public Health Effects. *Gastroenterology*, **158**, 447-52.
Rosa MD, Pace U, Rega D, et al (2015). Genetics, diagnosis and management of colorectal cancer (Review). *Oncol Rep*, **34**, 1087-96.
Soini S (2016). Biobanks as a Central Part of the Finnish Growth and Genomic Strategies: How to Balance Privacy in an Innovation Ecosystem?. *J LawMed Ethics*, **44**, 24–34.
Song M, Sasazuki S, Camargo MC, et al (2018). Circulating Inflammatory Markers and Colorectal Cancer Risk: A Prospective Case-cohort Study in Japan. *Int J Cancer*, **143**, 2767-76.
Sudlow C, Gallacher J, Allen N, et al (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med*, **12**, e1001779.
Tamakoshi K, Toyoshima H, Wakai K, et al (2005). Leptin Is Associated with an Increased Female Colorectal Cancer Risk: A Nested Case-Control Study in Japan. *Oncology*, **68**, 454-61.
Thirunavukarasu P, Sukumar S, Sathaiah M, et al (2011). C-stage in colon cancer: implications of carcinoembryonic antigen biomarker in staging, prognosis, and management. *J Natl Cancer Inst*, **103**, 689-97.
Tucker TC, Durbin EB, McDowell JK, et al (2019). Unlocking the Potential of Population-Based Cancer Registries. *Cancer*, **125**, 3729-37.
Vasen H, Velthuizen M, Kleibeuker J, et al (2016). Hereditary cancer registries improve the care of patients with a genetic predisposition to cancer: contributions from the Dutch Lynch syndrome registry. *Fan Cancer*, **15**, 429-35.
Wong M, Ding H, Wang J, et al (2019). Prevalence and risk factors of colorectal cancer in Asia. *Intest Res*, **17**, 317-29.
Yang T, Li X, Montazeri Z, et al (2019). Gene–environment interactions and colorectal cancer risk: An umbrella review of systematic reviews and meta-analyses of observational studies. *Int J Cancer*, **154**, 2315-29.