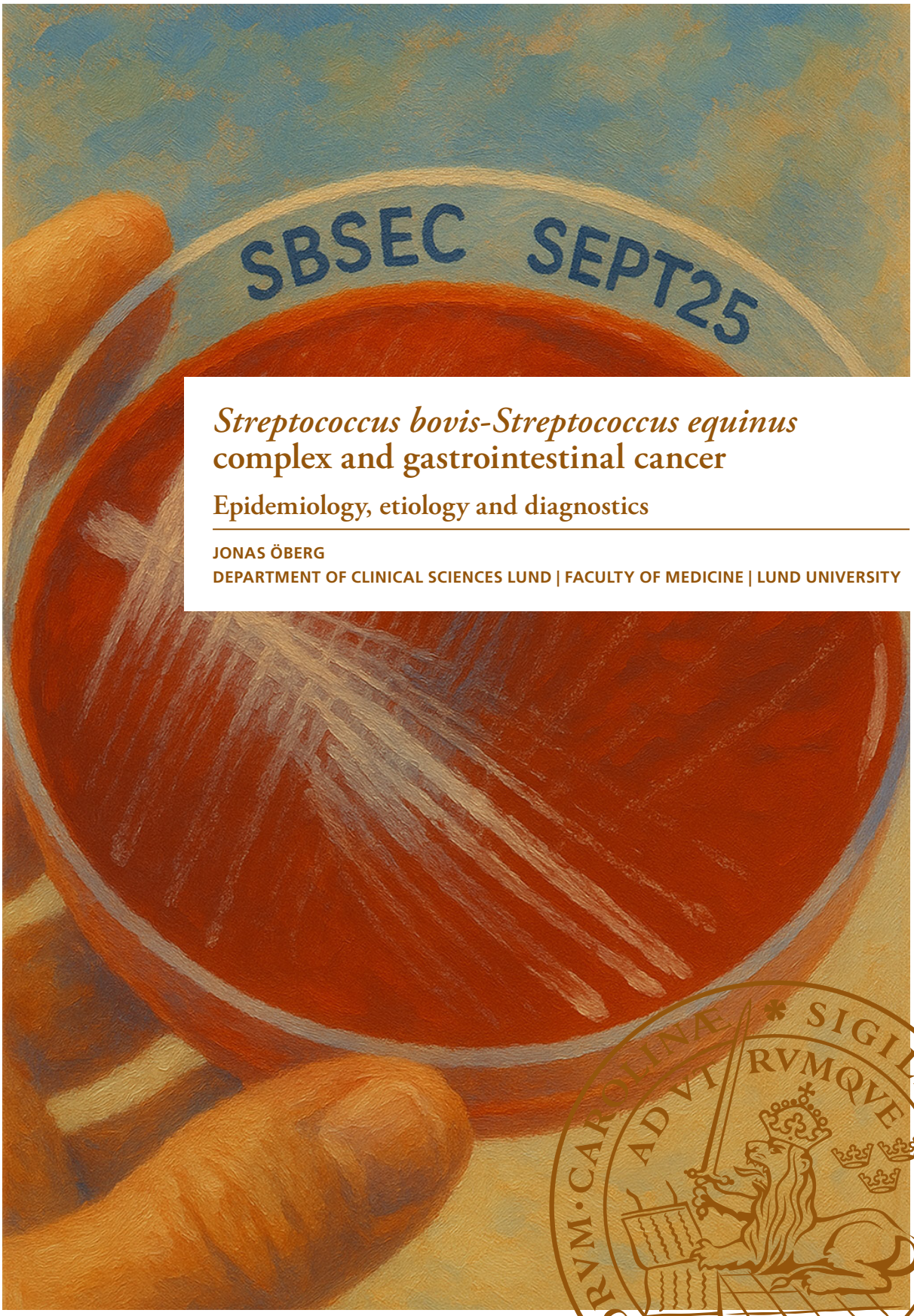
Streptococcus bovis-Streptococcus equinus complex and gastrointestinal cancer: Epidemiology, etiology and diagnostics

Öberg, Jonas

2025

Document Version:
Publisher's PDF, also known as Version of record



Streptococcus bovis-Streptococcus equinus complex and gastrointestinal cancer

Epidemiology, etiology and diagnostics

JONAS ÖBERG
DEPARTMENT OF CLINICAL SCIENCES LUND | FACULTY OF MEDICINE | LUND UNIVERSITY

Printed by Media-Tryck, Lund 2025 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences Lund

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2025:85
ISBN 978-91-8021-738-5
ISSN 1652-8220

Streptococcus bovis-Streptococcus equinus complex and gastrointestinal cancer

Streptococcus bovis-Streptococcus equinus complex and gastrointestinal cancer

Epidemiology, etiology and diagnostics

Jonas Öberg

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on Friday, September 19, 2025 at 09.00 in Belfragesalen, BMC, Lund

Faculty opponent
Oddvar Oppegaard, Associate Professor, Universitetet i Bergen

Organization: LUND UNIVERSITY

Document name: Doctoral dissertation **Date of issue** 2025-09-19

Author(s): Jonas Öberg **Sponsoring organization:**

Title and subtitle: *Streptococcus bovis-Streptococcus equinus* complex and gastrointestinal cancer: Epidemiology, etiology and diagnostics

Abstract:

An association exists between Streptococcus bovis-Streptococcus equinus complex (SBSEC) bacteraemia and colorectal cancer (CRC). SBSEC comprises seven subspecies, but it is unclear if all are linked to CRC, and potential associations with other gastrointestinal cancers have been suggested. The timing of CRC diagnosis relative to bacteraemia is understudied. There is limited knowledge on the proportions of infective endocarditis (IE) diagnosed in bacteraemia for each SBSEC subspecies. The current routine bacterial identification method, Matrix-assisted desorption/ionisation-Time of flight mass spectrometry (MALDI-TOF MS), cannot discriminate between the different (sub-)species using available libraries.

Paper I used 1:10 matched population controls and was conducted nationally in Sweden 2010-2019 using registry data and examined when in time CRC is diagnosed following SBSEC-bacteraemia and the frequency of other gastrointestinal cancers. Most CRCs were diagnosed within the first year, although CRC diagnoses were elevated through the remainder of follow-up. There were no differences in the frequency of other gastrointestinal cancers following bacteraemia.

Paper II was conducted in Skane region 2003-2018 and examined whether the rates of CRC-diagnoses following bacteraemia differed between the SBSEC (sub-)species. SBSEC subspecies was identified by whole genome sequencing. CRC was primarily associated with the subspecies *S. gallolyticus* ssp. *gallolyticus (Sg gallolyticus)*.

Paper III utilised the same cohort as Paper II and examined the proportions of infective endocarditis (IE) diagnosed in bacteraemia for the different SBSEC subspecies and found that *Sg gallolyticus* carried the highest proportions of IE.

Papers IV and **V** developed libraries for SBSEC subspecies identification using MALDI-TOF MS, using the previously identified isolates of Paper II. The novel libraries could identify SBSEC subspecies.

In conclusion, most SBSCE-associated CRCs are diagnosed within the first year, with elevated rates over time. SBSEC is not linked to subsequent gastrointestinal cancers. *Sg gallolyticus* poses the highest CRC and IE risk. MALDI-TOF-MS can identify SBSEC subspecies.

Key words: SBSEC, *S. bovis*, infective endocarditis, colorectal cancer, MALDI-TOF MS

Classification system and/or index terms (if any) Supplementary bibliographical information

Language English **Number of pages**: 116

ISSN: 1652-8220

ISBN: 978-91-8021-738-5

Recipient's notes Price Security classification

Signature Date 2025-07-31

Streptococcus bovis-Streptococcus equinus complex and gastrointestinal cancer

Epidemiology, etiology and diagnostics

Jonas Öberg

LUND
UNIVERSITY

Main supervisor:
Malin Inghammar, MD, PhD, Associate Professor

Co-supervisors:
Bo Nilson, PhD
Pamela Buchwald, MD, PhD, Associate Professor
Anton Nilsson, PhD, Associate Professor

Faculty of Medicine
Department of Clinical Sciences Lund

ISBN 978-91-8021-738-5
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2025

“Streptococcus bovis is the most common streptococcus in cow dung.”

— Orla-Jensen, 1919

Table of Contents

Abstract

An association exists between *Streptococcus bovis-Streptococcus equinus* complex (SBSEC) bacteraemia and colorectal cancer (CRC). SBSEC comprises seven subspecies, but it is unclear if all are linked to CRC, and potential associations with other gastrointestinal cancers have been suggested. The timing of CRC diagnosis relative to bacteraemia is understudied. There is limited knowledge on the proportions of infective endocarditis (IE) diagnosed in bacteraemia for each SBSEC subspecies. The current routine bacterial identification method, Matrix-assisted desorption/ionisation-Time of flight mass spectrometry (MALDI-TOF MS), cannot discriminate between the different (sub-)species using available libraries.

Paper I used 1:10 matched population controls and was conducted nationally in Sweden 2010-2019 using registry data and examined when in time CRC is diagnosed following SBSEC-bacteraemia and the frequency of other gastrointestinal cancers. Most CRCs were diagnosed within the first year, although CRC diagnoses were elevated through the remainder of follow-up. There were no differences in the frequency of other gastrointestinal cancers following bacteraemia.

Paper II was conducted in Skane region 2003-2018 and examined whether the rates of CRC-diagnoses following bacteraemia differed between the SBSEC (sub-)species. SBSEC subspecies was identified by whole genome sequencing. CRC was primarily associated with the subspecies *S. gallolyticus* ssp. *gallolyticus* (*Sg gallolyticus*).

Paper III utilised the same cohort as Paper II and examined the proportions of infective endocarditis (IE) diagnosed in bacteraemia for the different SBSEC subspecies, and found that Sg gallolyticus carried the highest proportions of IE.

Papers IV and **V** developed libraries for SBSEC subspecies identification using MALDI-TOF MS, using the previously identified isolates of Paper II. The novel libraries could identify SBSEC subspecies.

In conclusion, most SBSCE-associated CRCs are diagnosed within the first year, with elevated rates over time. SBSEC is not linked to subsequent gastrointestinal cancers. *Sg gallolyticus* poses the highest CRC and IE risk. MALDI-TOF-MS can identify SBSEC subspecies.

Populärvetenskaplig sammanfattning

Streptococcus bovis-gruppen (internationellt kallat *Streptococcus bovis-Streptococcus equinus* complex) är en del av normalfloran i tarmen. S. bovis-gruppen består egentligen av sju arter och subarter, men i den kliniska verksamheten betraktas de som en bakterie - *S. bovis* - eftersom identifiering endast sker på gruppnivå. Rutinmetoden för bakterieidentifering idag är en typ av masspektrometri som kallas Matrix-assisted desorption/ionisation-Time of flight mass spectrometry (MALDI-TOF MS) vilken inte kan särskilja på *S. bovis* gruppens arter och subarter med befintliga bibliotek.

Det finns sedan tidigare ett välkänt samband mellan fynd av *S. bovis* i blododlingar (bakteriemi) och kolorektalcancer (tjock- och ändtarmscancer). Det är dock bristfälligt studerat när i tid kolorektalcancer upptäcks i förhållande till bakteriemin, huruvida cancern var känd sedan tidigare eller upptäcktes pga bakteriefyndet och när det skedde. Det har också spekulerats i att en liknande association som med kolorektalcancer skulle kunna finnas även för annan cancer i mag- och tarmkanalen.

Vidare så talar tidigare studier för att det framför allt är subarten *S. gallolyticus* subspecies *gallolyticus* (*Sg gallolyticus*) som är associerad med kolorektalcancer, men studierna är små och det är okänt huruvida sambandet finns även för de övriga arterna och subarterna. *S. bovis* kan även orsaka hjärtklaffinfektion (endokardit), och här anses risken också vara högst vid fynd av *Sg gallolyticus* medan risken är oklar för övriga arter och subarter. Kunskapen om de olika arterna och subarterna är begränsad då fynd av *S. bovis* i blododlingar är relativt ovanligt i Sverige och världen i övrigt. I Skåne har ca 30-40 patienter per år drabbats av detta de senaste åren.

Vid fynd av *S. bovis* i blododlingar genomförs ofta undersökning för kolorektalcancer med koloskopi och för endokardit med transesofagalt hjärtultraljud (undersökning genom att ultraljudsprob sväljs ner i matstrupen) – två ofta obehagliga och resurskrävande undersökningar.

I den första studien undersöktes när i tid kolorektalcancer diagnosticerades i relation till fynd av *S. bovis* i blododlingar, samt om även annan cancer i mag- och tarmkanalen diagnosticerades i ökad omfattning efter infektionen. Studien omfattade *S. bovis*-fynd i blododlingar i hela Sverige 2010-2019 och baserades på registerdata. Vi jämförde förekomsten av kolorektalcancer och annan mag- och tarmcancer hos patienter som hade haft en *S. bovis*-infektion med en matchad kontrollgrupp från totalbefolkningen. Jämförelsen justerades även för samsjuklighet. Vi fann ökad frekvens av kolorektalcancer-diagnoser hos individer med *S. bovis* innan och efter infektionen. Totalt hade patienterna med *S. bovis*-infektion 7-10 gånger ökad risk att diagnosticeras med kolorektalcancer, med störst ökning inom det första året (20-23 gånger ökad risk), men ökningen kvarstod även senare under uppföljningsperioden. De flesta andra mag- och tarmcancrar var

vanligare hos individer med *S. bovis* innan infektionstillfället, men individerna med *S. bovis*-infektion diagnosticerades inte med förhöjd andel av andra mag- och tarmcancer än befolkningen i övrigt efter infektionen.

I den andra studien undersöktes hur frekvensen av kolorektalcancer var innan och efter infektion med *S. bovis*-gruppens respektive arter och subarter, samt hur fördelningen såg ut mellan dessa i blododlingsfynd i Skåne 2003-2018. Bakterier från blododlingar hade förvarats frysta sedan tidigare, dessa tinades, odlades och noggrann identifiering av art och subart genomfördes med helgenomsekvensering. Vi fann att 20% av fallen med infektion orsakade av subarten *Sg gallolyticus* diagnosticerades med kolorektalcancer efter infektion, jämfört med 0-4 % för övriga subarter, men de flesta fall av bakteriemi orsakades av subarten *S. gallolyticus* subspecies *pasteurianus (Sg pasteurianus)*.

I den tredje studien användes samma patienter och bakterier som i den andra studien, men hur stor andel som diagnosticerats med endokardit för respektive *S. bovis* art och subart undersöktes. Vi fann att andelen som diagnosticerats med endokardit var högst för *Sg gallolyticus* 33 % och *Streptococcus infantarius* 16% jämfört med 5% för övriga.

I den fjärde och femte studien utgick vi från samma bakteriestammar som i den andra och tredje studien, och utvecklade nya bibliotek för art och subartsidentifiering av *S. bovis* med MALDI-TOF MS. I den fjärde studien användes bakterier odlade i buljong, men eftersom främst blodagarplattor används i klinisk rutin odlar skapades även kompletterande bibliotek baserade på denna odlingsmetod. Samtliga nya bibliotek kunde framgångsrikt identifiera *S. bovis* på art och subartsnivå, och det bäst presterande biblioteket uppnådde 100 % korrekt identifiering av isolaten korrekt.

Våra studier kan vägleda behandlande läkare genom att laboratoriet kan identifiera *S. bovis* art och subart med de nya biblioteken för MALDI-TOF MS. Detta kan ske på minuter, utan extra kostnader eller arbetsinsats. Behandlande läkare kan sedan göra en riskbedömning avseende sannolikheten för odiagnostiserad kolorektalcancer eller endokardit utifrån vilken *S. bovis* subart den enskilde individen är infekterad med. Risken för båda dess tillstånd är överlägset högst vid infektion med subarten *Sg gallolyticus*. Förhoppningen är att färre patienter därmed behöver genomgå resurskrävande och potentiellt obehagliga undersökningar såsom koloskopi och transesofagalt hjärtultraljud i onödan. Dessutom är risken för annan odiagnostiserad mag- och tarmcancer låg, och vidare utredning för detta är sällan motiverad om inte andra kliniska skäl föreligger.

List of scientific papers

Paper I

Öberg J, Buchwald P, Nilsson A, Nilson B, Inghammar M. Risk and prognosis of colorectal cancer following bacteremia with *Streptococcus bovis*-*Streptococcus equinus*-complex: A Swedish nationwide retrospective cohort study. Manuscript undergoing peer review in Epidemiology & Infection.

Paper II

Öberg J, Rasmussen M, Buchwald P, Nilson B, Inghammar M. *Streptococcus bovis*-bacteremia: subspecies distribution and association with colorectal cancer: a retrospective cohort study. Epidemiol Infect. 2021 Nov 26;150:e8.

Paper III

Öberg J, Nilson B, Gilje P, Rasmussen M, Inghammar M. Bacteraemia and infective endocarditis with *Streptococcus bovis*-*Streptococcus equinus*-complex: a retrospective cohort study. Infect Dis. 2022 Oct;54(10):760-765.

Paper IV

Öberg J, Inghammar M, Nilson B. Improved identification of *Streptococcus bovis*-*Streptococcus equinus*-complex species and subspecies by MALDI-TOF MS using a novel library. Diagn Microbiol Infect Dis. 2023 Nov;107(3):116045.

Paper V

Öberg J, Engberg A, Inghammar M, Nilson B. Creation and validation of improved MALDI-TOF MS libraries for *S. bovis*-*S. equinus*-complex subspecies identification adapted to diagnostic culturing and extraction conditions. Diagn Microbiol Infect Dis. 2025 Nov;113(3):117000.

Abbreviations

BHI broth	Brain Heart Infusion broth
cgMLST	Core-genome Multilocus sequence typing
CI95%	Confidence interval 95%
CRC	Colorectal cancer
HR	Hazard ratio
IE	Infective endocarditis
IQR	Interquartile range
MALDI-TOF MS	Matrix-assisted desorption/ionisation-Time of flight mass spectrometry
MLST	Multilocus sequence typing
MSP	Main spectrum profile
m/z	Mass-to-charge
NGS	Next-generation sequencing
NNS	Number needed to screen
OR	Odds ratio
PCR	Polymerase chain reaction
SBSEC	*Streptococcus bovis-Streptococcus equinus* complex
Sg gallolyticus	*Streptococcus gallolyticus ssp. gallolyticus*
Sg macedonicus	*Streptococcus gallolyticus ssp. macedonicus*
Sg pasteurianus	*Streptococcus gallolyticus ssp. pasteurianus*
SNPs	Single nucleotide polymorphisms
TTE	Transthoracic echocardiography
TOE	Transoesophageal echocardiography
WGS	Whole genome sequencing

Glossary

S. infantarius	Refers to *Streptococcus infantarius ssp. infantarius* unless specified otherwise
S. lutetiensis	Refers to *Streptococcus infantarius ssp. coli*
Endocarditis	Refers to infective endocarditis

An introduction to the *Streptococcus bovis-Streptococcus equinus* complex

The human microbiome is estimated to comprise 100 trillion cells—10 times more than the human cells in the body and encoding 100 times more unique genes than the human genome (1, 2). Most of the human microbiome colonises the gastrointestinal tract, mainly the large intestine (1, 3). We are born germ-free, and over a lifetime, we acquire this vast microbiota with which our relationship is primarily commensal or mutualistic, although some components may also be harmful (4). Part of this microbiota is the *Streptococcus bovis-Streptococcus equinus* complex (SBSEC), a group of gram-positive non-β-haemolytic bacteria—known for not only its association with colorectal cancer (CRC) and propensity to cause infective endocarditis (IE) but also its role in food fermentation and its beneficial properties in the gastrointestinal tract of ruminants (5).

> "Streptococcus bovis is the most common streptococcus in cow dung."

In 1919, Orla-Jensen first described *Streptococcus bovis* using these words in the publication "The Lactic Acid Bacteria," whereas Andrewes and Horder discovered *Streptococcus equinus* in horse dung in 1906 (6, 7). The close relation between these two species was noted in 1937, and they are now commonly referred to as the SBSEC (8, 9).

Although it is known primarily to colonise the rumen and intestines of cows and horses, the SBSEC has been found in lambs and goats, sea otters, koalas, and birds (5, 10-12). Some SBSEC species play an important role in the microbial ecology of ruminant animals, in which they influence the digestion of food, possibly by affecting cellulose digestion, amino acid degradation, and starch fermentation, and have even been used as a probiotic to improve the development and function of the rumen (12-14). However, the SBSEC may also harm ruminants by causing acute acidosis through bacterial overgrowth, bloat, and mastitis and by causing IE in birds and sea otters (10, 12, 15, 16). Also, a correlation between cattle density and the incidence of SBSEC bacteraemia has been found, suggesting zoonotic potential (10, 17). Further, the SBSEC species *S. gallolyticus* ssp. *macedonicus,* and *S. infantarius* ssp. *infantarius* (*S. infantarius*) have been found in cheese and fermented milk, plants, and fish products in Europe, Asia, and Africa (10, 18-20).

In humans, the SBSEC is known predominantly for its propensity to cause IE and its association with colorectal neoplasia (CRC and its precursor lesions, colorectal adenomas) (9, 21-24). There is uncertainty regarding the extent to which the SBSEC is part of the normal gut microbiota in humans. Reported faecal carriage rates range between 5% and 60% (25-31). However, whether these values can be applied to the general population remains unknown, because most studies comprise hospital-derived study populations (10). The SBSEC is primarily considered a coloniser of the intestines, but oral colonisation has been observed (32-34).

Although the SBSEC has been studied extensively due to its role in livestock, food production, its ability to induce IE, and its link to CRC, many aspects of the SBSEC remain unknown. This thesis delves into the species and subspecies of the SBSEC; their association with CRC, other gastrointestinal cancers, and IE; and methods for determining SBSEC subspecies using matrix-assisted desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS).

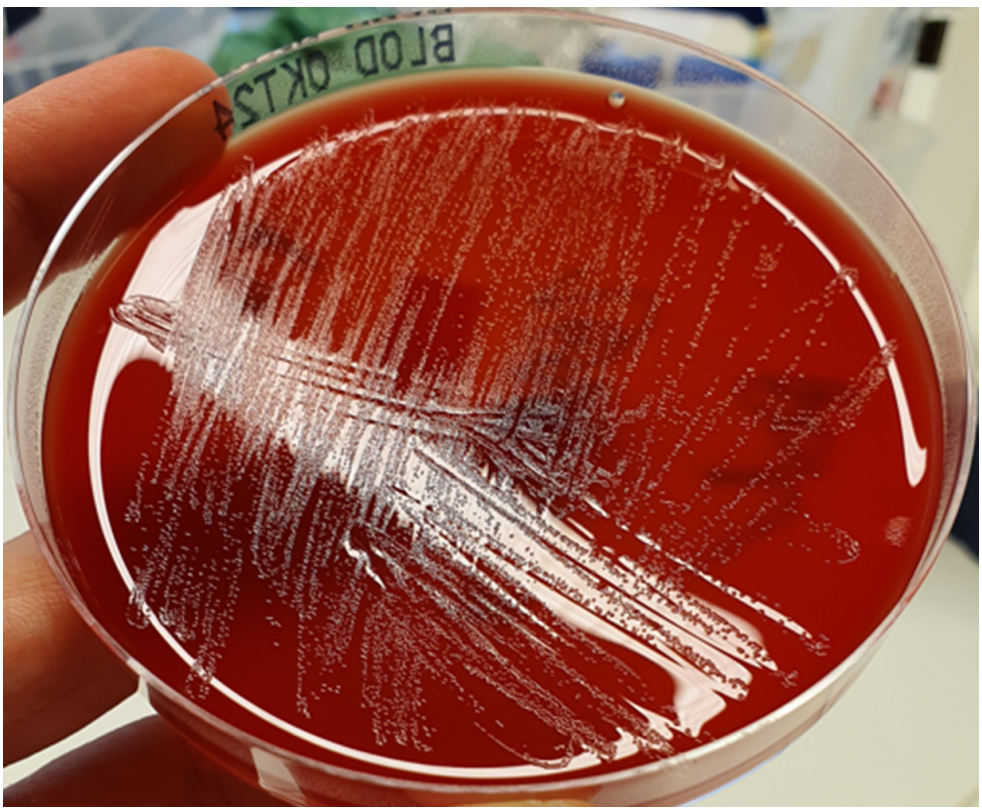


Figure 1. SBSEC, cultured on sheep blood agar plate under aerobic conditions. Photograph by Jonas Öberg.

The taxonomy of the *Streptococcus bovis-Streptococcus equinus* complex

Since their discovery, the taxonomy of the SBSEC has changed frequently, following discoveries of new species and subspecies. The ever-changing taxonomy and reuse of names have confused the interpretation of older scientific studies in the context of the current taxonomy.

The SBSEC carries Lancefield D antigen in its cell wall, a feature that is shared by enterococci, rendering it a member of Group D streptococci (9). The SBSEC is also closely related to viridans streptococci and causes similar disease spectra in humans (9). As discussed, *S. equinus* was first described in 1906, followed by the discovery of *S. bovis* in 1919, and their close relationship was established in 1937 (6-8). In 1972 and 1976, *S. bovis* was categorised by mannitol fermentation ability into biotypes I (positive) and II (negative), and in 1984, biotype II was subdivided into biotypes II/1 and II/2 according to the latter subtype's ability to ferment trehalose and produce b-galactosidase and b-glucuronidase (35-37).

S. alactolyticus was described in 1984, based on its characterisation using DNA-DNA hybridisation techniques, and Osawa et al. proposed that gallate-degrading strains of *S. bovis* be called *Streptococcus gallolyticus* due to their genetic similarities (38, 39). *Streptococcus macedonicus* was first reported in 1998, and *Streptococcus infantarius* was discovered in 1997, of which its subgroups *Streptococcus infantarius ssp. infantarius* and *Streptococcus infantarius ssp. coli* were proposed in 2000 due to genetic and biochemical differences and later reclassified as *Streptococcus infantarius* and *Streptococcus lutetiensis,* respectively (19, 40-42). During this reclassification by Poyart et al. in 2002, *Streptococcus pasteurianus* was introduced, consisting of strains that had been classified as biotype II/2 (42). In 2003, Schlegel et al. noted the genetic similarities between *S. gallolyticus*, *S. macedonicus*, and *S. pasteurianus* and proposed that they be considered *S. gallolyticus* subspecies and renamed *Streptococcus gallolyticus ssp. gallolyticus*, *Streptococcus gallolyticus ssp. macedonicus,* and *Streptococcus*

gallolyticus ssp. pasteurianus, respectively (43). However, this suggestion has been debated, because some groups maintain that *Streptococcus gallolyticus ssp. pasteurianus* should still be recognised as the separate species *S. pasteurianus (44)*. Other species, such as *Streptococcus waius* and *Streptococcus caprinus,* were included briefly until they were found to be synonyms of other SBSEC species (42, 45).

Although controversy remains regarding this issue, the SBSEC taxonomy that was established by Poyart in 2002 and Schlegel in 2003 is the most frequently used system for SBSEC species and subspecies (9, 46). Most classifications consider *S. gallolyticus ssp. gallolyticus* (henceforth abbreviated as *Sg gallolyticus*), *S. gallolyticus ssp. macedonicus* (abbreviated as *Sg macedonicus*), *S. gallolyticus ssp. pasteurianus* (abbreviated as *Sg pasteurianus*), *S. infantarius* (synonym *S. infantarius ssp. infantarius*), *S. lutetiensis* (synonym *S. infantarius ssp. coli*), *S. alactolyticus*, and *S. equinus* to be the seven members that constitute the SBSEC (9, 46). This taxonomy will be used throughout the thesis unless otherwise stated.

However, in 2021 and 2022, two additional SBSEC species, *Streptococcus vicugnae* and *Streptococcus ruminicola*, were proposed (47, 48). *S. ruminicola* has genetic similarities to *S. infantarius,* whereas *S. vicugnae* is more similar to *S. equinus*, *S. lutetiensis,* and *S. infantarius,* based on an analysis of 16s rRNA genes *(47, 48)*. Human infections with these species have not been reported, and their status as new species of the SBSEC has not been fully established. They have not been included in studies of the SBSEC in a clinical or diagnostic microbiological context.

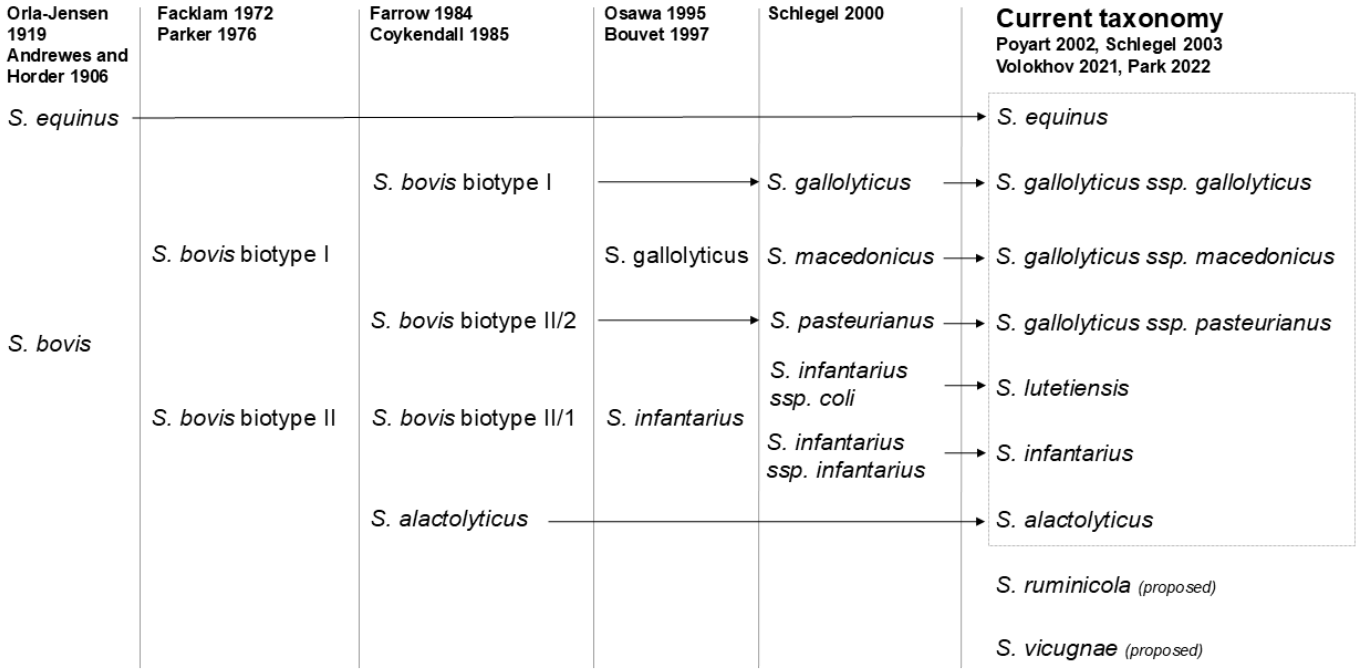


Figure 2. An overview of the taxonomy of the *Streptococcus bovis-Streptococcus equinus* complex over time. Image by Jonas Öberg.

Differentiation of *Streptococcus bovis-Streptococcus equinus* complex species and subspecies

Identification of the species and subspecies of the SBSEC is difficult in clinical microbiology laboratories and research studies due to the current limitations of widely available methods for routine bacterial species identification.

Biochemical/phenotypical methods

The SBSEC is classified as non-enterococcal Group D streptococci, and older classifications into biotypes are based on biochemical properties, such as the ability to ferment mannitol. Neverthelss, the current taxonomy is based primarily on genetic differences (44). Whereas the biochemical properties of SBSEC species and subspecies have been described, the capacity for subspecies-differentiation using biochemical methods is limited due to phenotypic variability within the same species and subspecies (44, 49). Further, identification by manual phenotypic biochemical methods is time-consuming and labour-intensive, whereas automated identification systems that are based on biochemical/phenotypical methods have limited or unknown ability to identify SBSEC members according to the current taxonomy (46, 49, 50). Thus, the SBSEC species and subspecies in the current taxonomy cannot be expected to be reliably identified using biochemical/phenotypical methods, nor is it feasible to do so (46, 49, 50).

Amplicon sequencing

The current taxonomy is based on genetic diversity among SBSEC species and subspecies. The most widely used amplicon sequencing method for identifying bacterial species in clinical routine is polymerase chain reaction (PCR) amplification of 16S rRNA genes. 16S rRNA has generally been successful in identifying *Sg pasteurianus, Sg macedonicus,* and *Sg gallolyticus* but appears less reliable in discriminating between *S. infantarius* and *S. lutetiensis* (49-56). The other SBSEC species have seldom been included in the evaluation of clinical bacterial samples (49, 51-56). However, most studies that have applied 16s rRNA analysis and other amplicon sequencing methods have attempted to identify SBSEC species without a reference standard and using limited datasets, excluding rarer species, such as *Sg macedonicus*, *S. equinus,* and *S. alactolyticus*. As a result, 16s rRNA is considered insufficient for identifying all SBSEC species and subspecies definitively (49).

Sequencing of *sodA* genes has been employed in several studies on the SBSEC and outperforms 16s rRNA sequencing (42, 50, 55-62). It can be used to identify most SBSEC members, although limited discrimination between *Sg gallolyticus* and *Sg macedonicus* has been noted (42, 50, 55-62).

Other methods for sequencing targeted gene regions, such as groES/EL, gyrB, rpoB, and 16S-23S ITS, as well as multilocus sequence typing (MLST), have been used, yielding promising results in a handful of studies, but none has been established as a reference method for identifying SBSEC species and subspecies (46, 49, 55, 57, 61, 63-65).

In conclusion, certain approaches for targeted gene sequencing may be sufficient for identifying all SBSEC species and subspecies in a clinical context. However, they are expensive, time-consuming, and rarely available for bacterial identification in clinical microbiology laboratories (49). The lack of a reference standard against which sequences can be compared and the limited number of isolates of rarer SBSEC members render some results inconclusive (46, 49). Further, reference databases are plagued with incorrect sequences, likely due in part to changes in taxonomy (46).

MALDI-TOF MS

Mass spectrometry (MS), developed at the end of the 19th century, measures the mass-to-charge ratio (m/z) of chemical compounds and calculates their molecular weight (66). Matrix-assisted laser desorption/ionisation (MALDI) is an ionisation technique for analysing chemical structures in biological systems (66). At the end

of 1990s, it was demonstrated that intact bacterial cells could be distinguished using MALDI with a time-of-flight (TOF) analyser, introducing MALDI-TOF MS systems in microbiology as a tool for characterising bacteria (66). MALDI-TOF MS systems can apply various strategies using intact or digested proteins—e.g., peptides from protein digestion (peptide sequencing) (66).

In short, bacterial samples are placed on a target plate intact, through the direct transfer (direct sample spotting) method; via on-target extraction (extended direct transfer) method through the addition of formic acid directly on the bacterial sample on the target plate; or after preparation in a test tube by the ethanol-formic extraction (full extraction) method (66-69). The sample is overlaid with an energy-absorbent, organic solution, called a matrix; the sample and matrix co-crystallise when air-dried; and the crystals are then ionised with a laser beam and converted to individually charged gas-phase ions (66). Using the TOF analyser and ion detection system (detector), the separated ions are measured, and a mass spectrum is created, based on relative abundance and mass-to-charge (m/z) ratio, generating an individual protein profile (66). This profile can then be matched to libraries (databases) of previously identified bacterial isolates, yielding a ranking of the closest matches (66). The analysis can be completed in minutes. However, several factors, such as the preparation of the sample, culture medium, culture conditions, extraction method, calibration of the MALDI-TOF MS instrument, and number of laser shots, may affect the system's performance (66). Most importantly, the identification of bacterial species depends directly on the quality of the library that is used for matching individual protein profiles (66).

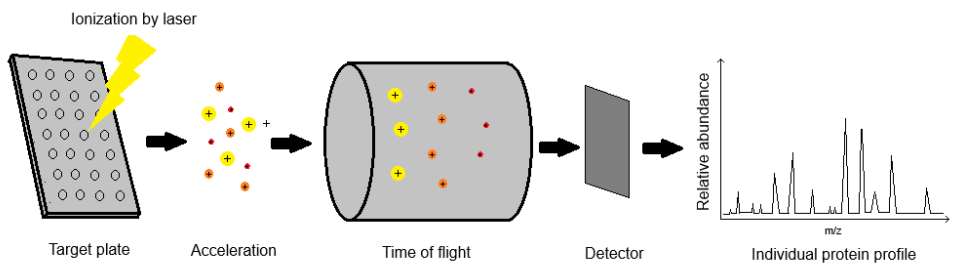


Figure 3. Schematic of MALDI-TOF MS. Image by Jonas Öberg.

The two most widely used MALDI-TOF MS systems in clinical microbiology laboratories in Sweden are the Bruker Biotyper and bioMérieux Vitek MS. Several studies have examined the identification of SBSEC species and subspecies by these systems using the commercial libraries that are supplied by each manufacturer; yet, these studies face the issue of a lack of an established reference standard for identifying SBSEC species. The reports that have been generated have been

hampered by the limited availability of bacterial isolates of rare species in human infections, including *S. equinus*, *S. alactolyticus,* and *Sg macedonicus*. The overall performance of both systems in these studies is influenced heavily by the number of isolates of each SBSEC member that are included.

Romero et al. used the Bruker Biotyper with sequencing of the *sodA* gene as the reference standard, demonstrating that *Sg pasteurianus* and *Sg gallolyticus* can be identified to the *S. gallolyticus* species, but not subspecies, level but that *S. infantarius* and *S. lutetiensis* could not be identified (56). Ben-Chetrit et al. tested both systems on 40 samples of *Sg pasteurianus*, *Sg gallolyticus*, *Sg macedonicus,* and *S. lutetiensis* (*S. infantarius ssp. coli*, per the old taxonomy) compared with 16s rRNA data (52). They noted that no system could identify *Sg macedonicus*, whereas the other subspecies could be identified in 84% and 83% of isolates by the Bruker Biotyper and Vitek MS, respectively; however, identification at the species level was possible in 100% of isolates by both systems (52). Gherardi et al. evaluated both systems on 22 isolates of *Sg pasteurianus* by 16s rRNA sequencing and found that the Bruker Biotyper identified 100% of isolates to the subspecies level, whereas the Vitek MS identified the subspecies in 27% of isolates but the species in 100% (54).

In a study by Agergaard et al. of 66 SBSEC isolates, including all species and subspecies except *S. alactolyticus*, that evaluated both systems using various extraction methods compared with subspecies that were identified by 16s rRNA and 16S-23S ITS, the correct subspecies was obtained in only up to 46% of isolates with the Bruker Biotyper versus 53% with the Vitek MS, but concluded that both systems can still be used for identification to the species level (70). Lopez-Roa et al. examined 24 isolates of *S. equinus*, *Sg gallolyticus*, *Sg pasteurianus,* and *S. lutetiensis* on the Bruker Biotyper versus *sodA* and found a sensitivity of 63% (71). The type strains of eight SBSEC members were analysed on both systems by Putnam et al., of which the Bruker Biotyper identified subspecies in up to 63% of isolates, compared with 88% by the Vitek MS (50).

Hinse et al. demonstrated that species and subspecies discrimination is possible based on dendrogram analysis but noted that *Sg gallolyticus* could not always be separated from *Sg pasteurianus* (62). However, the direct identification of SBSEC subspecies by the MALDI-TOF MS systems was not tested (62).

Bacterial identification by MALDI-TOF MS depends highly on the system and the library that is used for identification and may be affected by errors in the library due to changes in taxonomy. Further, commercial libraries are continuously updated, sometimes invalidating older evaluations. Although SBSEC species can be identified using the Bruker Biotyper and bioMérieux Vitek MS, depending on the species, none can reliably determine all SBSEC species and subspecies as of the time of writing (46, 49, 50, 52, 54, 56, 70, 71).

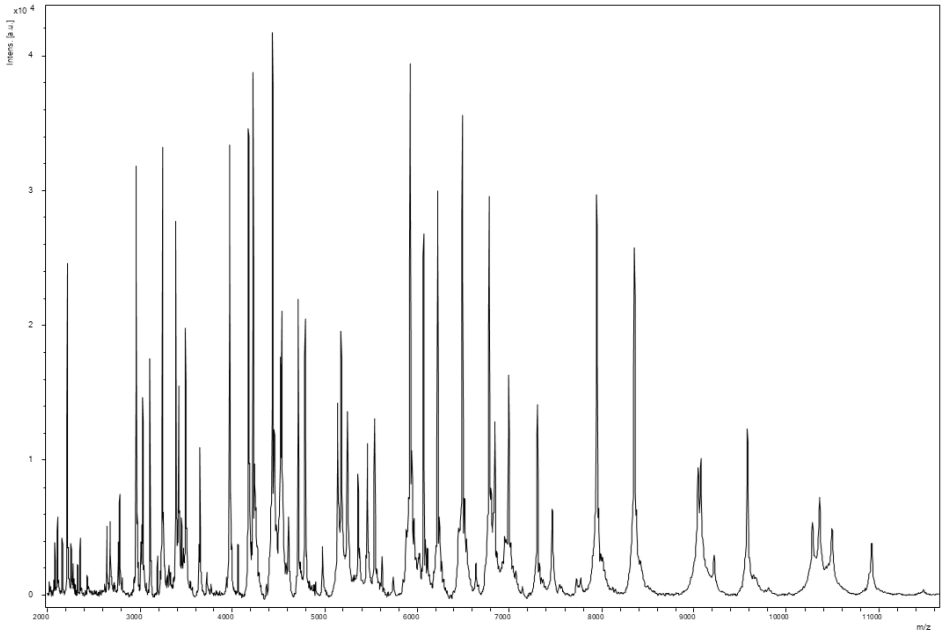


Figure 4. Spectral profile of *S. gallolyticus* ssp. *gallolyticus* reference type strain CCUG 35224T, cultured on a blood agar plate under aerobic conditions. Courtesy of Bo Nilson.

Whole genome sequencing

Sanger sequencing is sometimes referred to as first-generation sequencing (72). Although it laid the foundation for subsequent whole genome sequencing (WGS) techniques, Sanger sequencing was incredibly labour-intensive, time-consuming, and expensive (72). Next-generation sequencing (NGS) was introduced at the beginning of the 21st century and allowed massive parallel sequencing of single DNA strands, resulting in an enormous increase in data sequencing capacity (72). NGS approaches are often categorised into short-read sequencing (second-generation sequencing) and long-read sequencing (third-generation sequencing) (72). In short-read sequencing, the genome is fragmented into shorter segments before sequencing, offering a massive benefit in parallel sequencing, accompanied by the downside of the challenge of reassembling the DNA following sequencing (72). Although long-read sequencing has the advantage of producing longer interrupted sequence reads with less sample preprocessing that is required, its drawbacks include the generation of more errors, higher costs, and the necessity for larger machines (72).

Short-read sequencing is the most widely used WGS technique in clinical microbiology laboratories and will be the focus of the following discussion. The

value of WGS in microbiology in the clinical and research settings is massive, aiding in the identification of microorganisms, typing of viral strains, surveillance of transmissible diseases, prediction of antimicrobial susceptibility, and advances in metagenomics (73). However, the disadvantages of WGS in clinical microbiology are its complexity, the time that it requires, training of laboratory personnel, and the costs and expertise that are needed in bioinformatics, in addition to the challenges of integrating WGS laboratory workflows and the absence of established reference standards and standardised guidelines in clinical microbiology (73).

In principle, short-read sequencing is performed by DNA extraction and library preparation, followed by sequencing, after which bioinformatics, including assembly and mapping, is performed, depending on the query (72, 73). Bacterial genomes contain a core genome, a standard set of genes for each species, and other variable genomic content (74). Variability in bacterial genomes within the same species occurs for various reasons, including point mutations, such as single nucleotide polymorphisms (SNPs)—substitutions of one base for another (74). Bacterial strains may be identified by reference-based mapping of SNPs, often compared with a reference genome—e.g., from the type strain for the species (74). Alternatively, a core genome alignment can be constructed and used as a reference (74). Another option is gene-by-gene comparison, a modification of MLST, called core genome MLST (cgMLST), in which alleles are compared with a set of predefined core genes (74). Public databases for reference genomes, species core genomes, and cgMLST are growing but may be limited for certain species (74).

Whereas WGS has been used to study the classifications and taxonomy of *Streptococcus*, including the SBSEC, it has not been applied to determine SBSEC species and subspecies (75, 76). It has been speculated that the use of WGS will require a re-evaluation of the SBSEC taxonomy, as has been demonstrated by the proposal of the potential new species *S. vicugnae* and *S. ruminicola* (46-48, 75, 76).

Incidence of bacteraemia with *Streptococcus bovis-Streptococcus equinus* complex and geographic distribution of the subspecies

The incidence of SBSEC bacteraemia and the distribution of species and subspecies vary widely across geographic areas.

Incidence

Although few studies have examined the incidence of SBSEC bacteraemia, they have established that such rates, and those of SBSEC IE, likely vary across geographic areas (9, 17, 77, 78). The incidence of SBSEC bacteraemia has been best studied in Spain. Corredoira et al. compared this incidence in Galicia between 2005–2016 and found an overall incidence of 2.0 per 100,000 inhabitants, with varying local rates between provinces, ranging from 0–0.6 to 4.6–10.9 per 100,000 inhabitants, which also correlated with the local cattle density (17). This group also studied the incidence of SBSEC IE in Spain, observing an rate of 0.4–0.7 per 100,000 inhabitants in the urban area of Barcelona to 1.7–2.8 in the more rural area of Lugo (79). The incidence of SBSEC IE has also been measured in France, rising in a more urban area, Ile de France, at 1.1 per 100,000 inhabitants, compared with the rural area of Marne, at 2.4 per 100,000 inhabitants (77).

Studies have also found an increasing incidence of SBSEC bacteraemia (78, 80). Whether this pattern mirrors an actual rise in the incidence of SBSEC bacteraemia or merely reflects microbiological advances in the detection of bacteraemia and the tendency to draw blood cultures remains to be determined.

The incidence of SBSEC bacteraemia in Sweden—specifically in Skåne—is unknown.

Species and subspecies distribution

Few studies have examined the distribution of the species and subspecies in SBSEC bacteraemia, likely due to the rarity of bacteraemia and the challenges of identifying subspecies. The studies are mostly limited to smaller case series (47, 48, 50, 51, 58, 71, 72). Further, depending on their clinical presentation, cases can be selected for tertiary hospital treatment, skewing the distributions of subspecies in case series. Population-based studies are lacking.

Sg gallolyticus has often been reported as the predominant species in studies in the western world, especially northern Europe (4, 47, 57). *Sg pasteurianus* is the predominant species in geographic areas in which *Sg gallolyticus* is less common, most notably in Asian studies (9, 42, 55, 56, 81). *S. lutetiensis* and *S. infantarius* usually account for smaller proportions of SBSEC bacteraemia, of which *S. lutetiensis* is the more common (47, 48, 50, 51, 58, 71, 72). Bacteraemia due to *Sg macedonicus*, *S. alactolyticus*, and *S. equinus* is rare and mostly limited to case reports (47, 48, 50, 51, 58, 71, 72). A summary of the SBSEC species and subspecies distribution in bacteraemia according to current taxonomy is presented in Table 1.

Whereas a small study in Denmark found *Sg gallolyticus* to be the predominant SBSEC subspecies, the distribution of subspecies in Sweden has not been studied (63).

Differences in subspecies distribution are likely multifactorial, perhaps explained in part by such factors as contact with cattle, inhabitation in rural versus urban areas, and consumption of fermented food and certain cheeses (17, 79).

Table 1. Distribution of members in SBSEC bacteraemia by study per the current taxonomy (52, 53, 55, 56, 63, 81, 82).

	Sheng (2014)	**Beck** (2008)	**Marmolin** (2016)	**Romero** (2011)	**Ben-Chetrit** (2016)	**Kaiki** (2021)	**Lazarovitch** (2012)
Country	Taiwan	Germany	Denmark	Spain	Israel	Japan	Israel
Identification method	16S rRNA, *sodA*, groES/groEL	16S rRNA	ITS region	16S rRNA, *sodA*	16S rRNA	*sodA*	groES/groEL
SBSEC, total	172	58	53	45	40	39	24
Sg gallolyticus	31 (18%)	29 (50%)	19 (36%)	14 (31%)	6 (15%)	5 (13%)	6 (25%)
Sg pasteurianus	126 (73%)	12 (21%)	12 (23%)	24 (53%)	26 (65%)	29 (74%)	13 (54%)
Sg macedonicus	0 (0%)	0 (0%)	2 (4%)	0 (0%)	2 (5%)	0 (0%)	0 (0%)
S. lutetiensis	4 (2%)	17 (29%)	12 (23%)	5 (11%)	6 (15%)	4 (10%)	4 (17%)
S. infantarius	11 (6%)	0 (0%)	7 (13%)	2 (4%)	0 (0%)	1 (3%)	1 (4%)
S. alactolyticus	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
S. equinus	0 (0%)	0 (0%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

The *Streptococcus bovis-Streptococcus equinus* complex and colorectal cancer

It has been estimated that 20% of the global cancer burden is linked to infectious agents, such as bacteria, viruses, and parasites.

Since the 1950s, infective endocarditis and bacteraemia with the SBSEC have been associated with colorectal cancer.

A brief summary of colorectal cancer

Incidence and disease burden

CRC is the third most common cancer overall in Sweden and worldwide and is the most frequent gastrointestinal cancer, contributing to the second most number of deaths from cancer, with incidences varying between geographic areas (83-86). Nearly 2 million new cases and 1 million deaths from CRC per year were estimated in 2020–2022 (Table 2); these numbers are projected to increase to 3.2 million new cases and 1.6 million deaths per year by 2040 (83, 85, 87).

Table 2. The number of new cases and deaths for the most common gastrointestinal cancers worldwide in 2022, according to Global Cancer Statistics, ranked for all cancers (85).

	Incidence		Mortality	
	Rank	New cases	Rank	Deaths
Colorectal	3	1,926,000	2	904,000
Gastric	5	968,000	5	660,000
Hepatic	6	865,000	3	758,000
Oesophagus	11	511,000	7	445,000
Pancreatic	12	511,000	6	467,000
Gallbladder	22	122,000	20	89,000

Risk factors

CRC is approximately 1.5 times as common in men (88). The risk of CRC increases with age, wherein those aged over 65 years are roughly three times more likely to be diagnosed than those aged between 50–64 years and about 30 times more likely than 25–49-year-olds (88). However, the incidence of CRC in the latter group is rising, and it has been estimated that 1 of 10 colon cancers and 1 of 4 rectal cancers will be diagnosed in those aged <50 years within the next decade (88-90). Other non-modifiable risk factors are ethnicity, hereditary mutations, and inflammatory bowel disease (88). Risk factors that can be modified are obesity, physical inactivity, diet, smoking, and alcohol intake (88).

Early prevention

Nearly all CRCs are adenocarcinomas, and the adenoma-carcinoma sequence provides a rationale for detecting and preventing early CRC (24, 91). This hypothesis describes the morphological evolution—from benign colorectal tumours, adenomas and, ultimately, to CRC (Figure 5) (24). Because early adenomas and CRC are often asymptomatic, earlier detection and removal of precancerous tumours can decrease morbidity and mortality (83, 88, 92). Biennial CRC screening for those aged 60–74 years by faecal haemoglobin tests was introduced nationally in Sweden in 2021 and is estimated to save at least 300 lives annually (86, 93, 94).

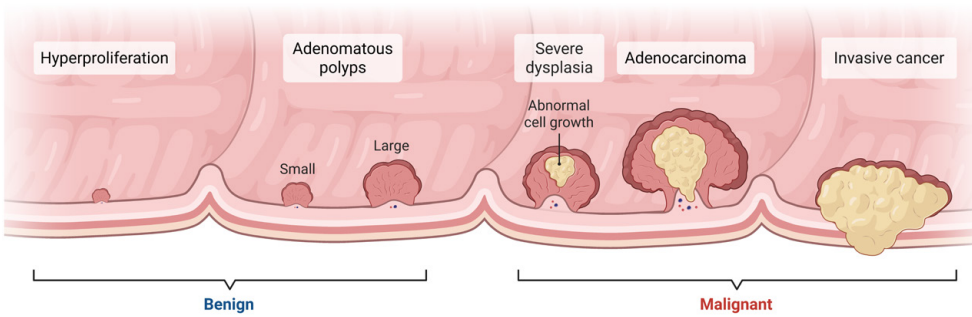


Figure 5. The adenoma-carcinoma sequence of colorectal cancer. Created with BioRender.com.

Classifications and staging

CRC and other malignant tumours are staged per the Tumour, Node, Metastasis (TNM) classification, which describes the anatomical extent of cancer (86, 95, 96). The TNM system is divided into preoperative clinical (c) or pathological (p) assessment: cTNM and pTNM. Based on the TNM system, CRC is further divided into stages I–IV, reflecting the spread and prognosis of the cancer. Adenocarcinomas are classified as low-grade (well-differentiated cancer cells) or

high-grade (poorly differentiated cancer cells) (86). Further, CRC is categorised into left-sided and right-sided CRC, given their disparate molecular and immunological characteristics, disease progression, and overall survival (97). Tables 3 and 4 describe the CRC TNM classification and staging, respectively.

Table 3. Classification of CRCs by TNM classification (86, 95).

Primary tumour (T)	
Tis	Tumour/carcinoma in situ - no invasion through muscularis mucosa.
T1–T2	Tumour invades submucosa (T1) or muscularis propria (T2). It cannot be differentiated by preoperative imaging.
T3	Tumour invades through the muscularis propria into the pericolorectal tissues. Further classified into T3a–T3d, depending on the degree of invasion: <1 mm to >15 mm.
T4	Tumour invades through the visceral peritoneum (T4a) or other organs (T4b).
Regional lymph node metastasis (N)	
N0	No regional lymph node metastasis.
N1	1–3 regional lymph node metastases. Further classified depending on the number or character of metastases: 1 (N1a), 2–3 (N1b), or tumour deposits (N1c).
N2	4 or more regional lymph node metastases. Further classified depending on the number of metastases; 4–6 (N2a) or 7 or more (N2b).
Distant metastasis (M)	
M0	No distant metastasis.
M1	Distant metastasis. Further classified depending on metastases in one (M1a) or several (M1b) organs or the peritoneal surface with or without other organ metastases (M1c).

Table 4. Staging of CRC (86, 95).

	T	**N**	**M**
Stage 0	Tis	N0	M0
Stage I	T1–T2	N0	M0
Stage II	T3–T4	N0	M0
IIA	T3	N0	M0
IIB	T4a	N0	M0
IIC	T4b	N0	M0
Stage III	Any T	N+	M0
IIIA	T1–T2	N1/N1c	M0
	T1	N2a	M0
IIIB	T3–T4a	N1/N1c	M0
	T2–T3	N2a	M0
	T1–T2	N2b	M0
IIIC	T4a	N2a	M0
	T3–T4a	N2b	M0
	T4b	N1–2	M0
Stage IV	Any T	Any N	M+
IVA	Any T	Any N	M1a
IVB	Any T	Any N	M1b

Prognosis and treatment

Surgical treatment has been considered the cornerstone of CRC treatment with a curative aim, often in conjunction with neoadjuvant or adjuvant chemotherapy or radiotherapy, depending on cancer stage and patient factors (86, 92). Newer treatment options include immunotherapy, but also a so-called "watch and wait" approach—comprising only oncological treatment without surgical resection—has been used internationally (98, 99).

CRC survival rates have increased over the past several decades (Figure 6) (100, 101). The overall 5-year survival rate is approximately 70% in Sweden and around or above 90% for those without metastatic disease (Stage I–II) (100-102).

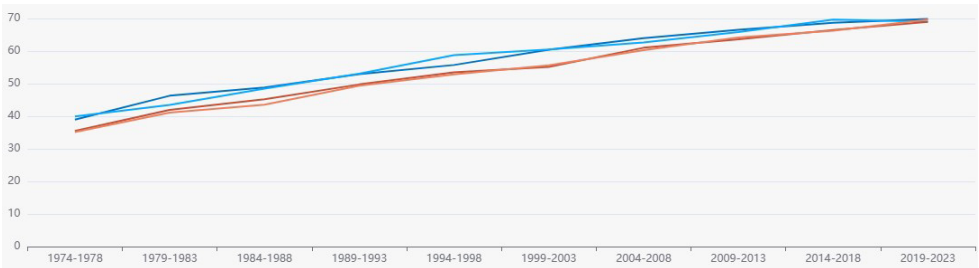


Figure 6. Colorectal cancer 5-year survival rates in Sweden from 1974–2023, 20-89 year olds, age-standardised. Source: Socialstyrelsen (100). Blue: colon cancer. Orange: rectal cancer. Darker colours: male. Lighter colours: female.

Overview of bacterial infections and colorectal cancer

Certain infectious agents are causative factors in carcinogenesis—directly through their oncogenic properties or indirectly through inflammation and immunomodulatory effects—and infectious agents have been linked to 20% of the global cancer burden (103, 104). Most evidence supports the role of viral infections in carcinogenesis, including that of human papillomavirus in cervical cancer, hepatitis B and C in liver cancer, and Epstein-Barr virus in Burkitt's lymphoma (99). The influence of parasitic infections in cancer is not as extensively studied, but examples include liver flukes in bile duct cancer and schistosomiasis in bladder cancer (103).

The role of specific bacterial pathogens and cancer is less clear. Yet, there are notable and well-studied associations, such as infection with *Helicobacter pylori* and the development of gastric cancer, likely through chronic inflammation (103, 105). Whereas the role of other bacterial infections is uncertain, the bacterial microbiota is likely important in the development of certain cancers, especially in the gastrointestinal tract, where the exposure to bacteria is high (3). Further, frequent

antibiotic use has been linked to an increased risk of the development or diagnosis of CRC, strengthening the putative role of intestinal microbiota in cancer development (106-109).

The function of the intestinal microbiota in cancer development is challenging to study due to its vastness and the slow process by which cancer evolves. Findings of specific bacteria in blood cultures have been implicated as a marker for CRC, and registry studies have shown that diagnoses of CRC are overrepresented following bacteraemia by *Clostridium* species (especially *Clostridium septicum*), *Fusobacterium* spp., and the SBSEC (110-112). However, their follow-up times have been limited, and these studies have failed to determine whether the bacteria are carcinogenic or merely a marker for established cancer. The bacterial infections that warrant diagnostic cancer work-up and follow-up remain unknown.

A model for bacterial involvement in the development of CRC has been proposed, termed the driver-passenger model (3). In this mechanism, intestinal bacteria are bacterial drivers that may initiate the development of CRC through their carcinogenic properties, such as the ability to establish an asymptomatic but chronic inflammatory response in the colonic mucosa (3). Suggested bacterial drivers include *Shigella* spp., *Salmonella* spp., *Citrobacter* spp., *Escherichia* coli, and *Bacteroides fragilis* (3). Bacterial passengers are worse at colonising a healthy intestinal tract but have competitive advantages once a tumour microenvironment has been established. In this new environment, they may be able to replace the bacterial drivers, but it is unknown whether they also continue to propel tumour development (3). In an advanced tumour, they can migrate from the intestinal lumen and lead to systemic infections, whereas they may be cultured in blood (113). Proposed bacterial passengers include those that have been found in registry studies, such as *C. septicum, Fusobacterium* spp., and the SBSEC (3). The functions of bacterial drivers and passengers are likely to overlap, and some bacteria probably serve as both (3).

An alternative hypothesis centres on alpha-bugs—pro-oncogenic microbes that also remodel the colonic microbiota to one that promotes and further induces more alpha-bugs, accompanied by modulations in the immune response and colonic epithelial cells, ultimately promoting CRC (114). The first alpha-bug to be proposed was the enterotoxigenic *Bacteroides fragilis* (114).

The SBSEC and colorectal cancer

The SBSEC in general

The association between IE and CRC was initially suggested in 1951, but Klein et al. first systematically studied the association between the SBSEC and CRC in 1977 and found that the SBSEC was more common in faeces among patients with CRC, describing symptom-free patients who were diagnosed with CRC following SBSEC endocarditis (25, 115).

Since then, many studies have further examined the link between the SBSEC and CRC, wherein 25–80% of patients with SBSEC bacteraemia have been found to have colorectal tumours (23). However, most have been conducted on the SBSEC group level, and few have evaluated the association of individual species and subspecies according to the current taxonomy, except for *Sg gallolyticus.*

Three registry-based studies have examined the rates of CRC following SBSEC bacteraemia, all of which used hospital populations as controls. Kwong et al. found a hazard ratio (HR) of 3.79 (confidence interval 95% [CI95%] 2.29–6.27) for 662 episodes of SBSEC bacteraemia compared with individuals with negative blood culture in Hong Kong, with an inflection point for normalisation of the HR observed after 1 year (110). A study of 117 episodes of SBSEC bacteraemia in Denmark, also using controls with negative blood cultures, reported an HR of 4.36 overall (CI95% 1.96–9.74) and 8.46 within 1 year (CI95% 3.51–20.4), with a median time to diagnosis of 97.5 days (111). The third study, conducted in Australia, assessed 397 cases of SBSEC bacteraemia versus patients with other pathogens in blood cultures, noting a relative risk of 4.4 (CI95% 2.72–6.80) (112). Whereas some of these studies also examined the CRC rates of some SBSEC species and subspecies, they did not describe the identification methods, rendering the results unreliable (110-112).

A recent meta-analysis found that the relative risk of underlying CRC was 3.73 (CI95% 2.79–5.01) for individuals with versus without SBSEC bacteraemia (116). The SBSEC is often associated with early colorectal neoplasms, primarily colorectal adenomas (9, 117). However, although the risk ratio of colorectal adenomas was elevated in this meta-analysis, it was not statistically significant (5.00; CI95% 0.83–30.03) (116). Also, individuals with CRC have higher rates of SBSEC colonic colonisation and higher levels of immunoglobulin G against SBSEC antigens (23, 116, 118-120).

Although most specialists recommend that a diagnostic work-up for CRC be considered following SBSEC bacteraemia, it is unknown whether it should be repeated unless warranted by the findings of the primary colonoscopy (9, 121). Corredoira et al. followed up patients with IE due to the SBSEC, observing that many were diagnosed with colorectal neoplasms long after the primary infection—

regardless of previous colorectal neoplasms—with a mean time to diagnosis of colorectal neoplasm of 5 years after infection (122). The authors concluded that colonoscopy should be repeated during the follow-up. However, they did not include a control group, and colorectal neoplasms, such as adenomas, are also common in the general population (123). In contrast, Boltin et al. found no differences in the rates of colorectal neoplasms during the follow-up of patients with faecal carriage with the SBSEC compared with controls without faecal carriage, but the number of individuals with SBSEC carriage was limited (29).

The species and subspecies of the SBSEC

Initially, individuals with IE caused by the SBSEC were considered to be at the highest risk of CRC. But, *Sg gallolyticus* was later discovered to carry the highest risk of IE and to be the SBSEC member that was primarily associated with CRC (9, 23, 52, 53, 55, 56, 63, 81, 82).

Corredoira et al. compared the colonoscopy results of 98 individuals with *Sg gallolyticus* bacteraemia with sex-matched and age-matched controls who had undergone colonoscopy due to colorectal symptoms or CRC family history, reporting that invasive carcinoma (12.5% vs. 5%, p=0.03) and colorectal neoplasia overall (70% vs. 32%, p<0.01) were more common among those with bacteraemia (124). Another study by the same group compared bacteraemia with biotypes I (*Sg gallolyticus*) and II, noting higher rates of colorectal neoplasms in individuals who had been infected with biotype I (50.9% vs. 16.6%, p<0.01) (78).

Associations between the other SBSEC members and CRC are more obscure (9). Few studies have examined the potential correlation between other SBSEC members and CRC, because results from studies that used older taxonomies are not applicable to all species and subspecies that are currently recognised. Further, few have included a control population without SBSEC infection. Colorectal neoplasia is less common in bacteraemia with *S. bovis* biotype II versus I infection, and rates of colorectal neoplasia are similar in bacteraemia with biotype II and *Streptococcus salivarius*, a pathogen that is not traditionally linked to CRC (78, 125-127). Although CRC is common among individuals with non-*Sg gallolyticus* SBSEC bacteraemia, these cancers are more advanced and often diagnosed before bacteraemia (128). There is no established consensus regarding whether individuals with non-*Sg gallolyticus* SBSEC bacteraemia should be evaluated for CRC; some guidelines recommend colonoscopy only if there is no known source of infection (9, 126).

Proposed mechanisms

The degree to which the SBSEC, especially *Sg gallolyticus*, is part of a healthy gastrointestinal microbiota has not been established, because faecal carriage rates differ between studies, with an overrepresentation in individuals with CRC (26, 28, 29, 129). The SBSEC can adhere to the exposed collagen in colorectal neoplasms and is effective at degrading the carbohydrates that abound in the microenvironment of CRC. This property could confer a competitive advantage on the SBSEC in the presence of colorectal neoplasms, whereas other microbiota might hinder them in a healthy colon (9, 129, 130). The SBSEC can grow in bile and potentially access the bloodstream through the hepatic reticulo-endothelial system, a trait that could explain the suggested association with early colorectal neoplasms, given that translocation of bacteria from the gastrointestinal tract to the bloodstream usually correlates with advanced malignancies (23, 131, 132).

In vitro studies have suggested that *Sg gallolyticus* possesses pro-inflammatory properties by inducing the COX2 pathway, which is important in CRC carcinogenesis and the upregulation of inflammatory cytokines (23, 129). Further, in vitro, *Sg gallolyticus* strains stimulate the proliferation of human colon cancer cells, and mice that have been administered *Sg gallolyticus* bear significantly more colon tumours, with more severe dysplasia grades and greater tumour burden than those that are exposed to control bacteria (133, 134). These results could possibly be attributed to a chromosomal locus, termed the *S. gallolyticus ssp. gallolyticus* pathogenicity-associated region (SPAR) (135, 136). Moreover, the *Sgg*T7SS^{T05} locus is important for the type VII secretion systems that are required for adherence to host cells and has been shown to promote CRC in a mouse model (136). *Sg gallolyticus* also up-regulates and down-regulates cancer-associated genes in vitro (137).

The SBSEC, particularly *Sg gallolyticus,* could serve as both a driver and passenger in the driver-passenger model of bacterial involvement in the development of CRC (3).

Gaps in knowledge

Whereas an epidemiological association between the SBSEC, particularly *Sg gallolyticus*, and CRC has been established, many questions remain unanswered. How high are rates of CRC in individuals with SBSEC bacteraemia compared with the general population? Is SBSEC-associated CRC diagnosed regardless of infection or because of infection? Does the risk of developing CRC increase following infection, and should individuals with SBSEC infection be followed up with repeat colonoscopies? Can examination by colonoscopy sometimes be omitted following SBSEC bacteraemia? Are SBSEC members other than *Sg gallolyticus* also associated with CRC? Is SBSEC carcinogenic or merely a marker of CRC?

What about the SBSEC and other cancers?

Gastrointestinal cancers other than CRC, as well as other cancers, are common among patients with SBSEC bacteraemia (55, 63, 125). *S. lutetiensis* was observed to have potential carcinogenic properties in a recent laboratory study, in which this bacteria was enriched in the tumour tissue of gastric cancer patients and had tumour-suppressing properties, promoting gastric cancer in a mouse model (138). An association between *Sg pasteurianus* and hepatocellular, biliary tract, pancreatic, and gastric cancer has been suggested in epidemiological studies (55, 63, 139). However, few studies have systematically examined whether these cancers were detected before or after bacteraemia (55, 63, 139). Whereas some studies have proposed that patients with SBSEC bacteraemia should also be evaluated for non-colorectal gastrointestinal cancer, there is limited evidence that supports such recommendations (55, 139).

Infective endocarditis and other clinical manifestations of *Streptococcus bovis-Streptococcus equinus* complex infection

Infective endocarditis, a disease that is characterised by high morbidity and mortality, is frequently associated with SBSEC bacteraemia. However, the prevalence of infective endocarditis in bacteraemia regarding the respective SBSEC species and subspecies has not been well studied.

Overview of infective endocarditis

IE, first described in 1674, is an infection of the endocardial surface of the heart, most commonly in the heart valves (140). It is commonly categorised as subacute or acute. Subacute IE is generally associated with bacteria of lesser virulence, such as non-β-haemolytic streptococci (e.g., the SBSEC), in which symptoms can persist for several weeks to months before a diagnosis is made (140, 141). Although acute IE is most often associated with more virulent bacteria, such as *Staphylococcus aureus*, its onset is acute and can debut with symptoms of sepsis. Most cases of IE are caused by staphylococci, streptococci, and enterococci; less common microorganisms in IE include fungi, gram-negative bacteria, and low virulence pathogens, such as corynebacteria and cutibacteria (140). The epidemiology of IE has shifted over time, and although the disease occurs at all ages, it primarily affects males and the elderly (140).

IE may lead to uncontrolled infection; cardiac complications, such as heart failure; and embolic complications, including ischemic stroke; overall mortality rates of 30% have been reported (140). Treatment for IE has traditionally entailed intravenous antibiotic therapy for 2–6 weeks in a hospital, but partial oral treatment

and outpatient parenteral antimicrobial therapy have been implemented more recently and are now recommended in management guidelines (140-144). Moreover, open heart surgery is often required to control infection, manage heart failure due to valve dysfunction, and prevent emboli (140-142).

Newer diagnostic modalities for detecting valve vegetations and paravalvular complications include positron emission tomography-computed tomography (PET-CT) and cardiac computed tomography (cardiac CT), but the mainstay for diagnosing IE remains a positive blood culture, accompanied by findings on transthoracic and transoesophageal echocardiography (TTE and TOE) (140-142, 145).

The diagnosis of IE can be challenging, because it is difficult to rule out the infection in patients with high clinical suspicion due to findings of high-risk pathogens in blood cultures and patient risk factors, such as the presence of prosthetic valves or degenerative valve disease (140-142). This hurdle complicates the design and interpretation of studies on IE, because a definitive diagnosis of IE is often impossible to obtain or rule out.

Several diagnostic criteria have been designed to improve the diagnosis of IE (141, 142, 145-148). In 1981, the Von Reyn criteria for diagnosing IE were published, in which cases were classified as definite, probable, possible, or rejected, according to pathological, clinical, and microbiological findings (148). These standards evolved into the Duke Criteria, published in 1994, and the 2000 Modified Duke Criteria (146, 147). These criteria were all based on cases with a clinical or pathological diagnosis of IE (146-148). The major differences in the latter two Duke's criteria compared to the earlier criteria, were that they categorised cases into definite, possible, or rejected based on the fulfilment of major and minor criteria for the diagnosis of IE and introduced an imaging criterion that was based on echocardiography findings (146, 147).

The Modified Duke Criteria have been the predominant system for diagnosing IE in clinical settings and in studies of IE for most of the 21st century and constitute the basis for all of the widely used systems that have followed (140-142, 145). The newer 2015 and 2023 ESC guidelines and the 2023 Duke-ISCVID criteria have revised the criteria introducing additional imaging criteria, based on cardiac CT and PET-CT, and have updated the microbiological criteria with new diagnostic methods and expansion of the number of organisms that are considered to be associated with a high risk for IE (141, 142, 145). These newer criteria were based on previously available evidence and, unlike previous criteria, were not evaluated or validated in the publications. However, subsequent publications have assessed these changes retrospectively, often with greater sensitivity for the diagnosis of IE but sometimes at the cost of decreased specificity and an increase in cases that are defined as possible endocarditis (149-152).

Infective endocarditis and other disease manifestations of SBSEC species and subspecies

SBSEC infection in humans is generally associated with bacteraemia and IE, and unlike viridans streptococci, which cause similar diseases, detection of the SBSEC in blood cultures is typically considered to have clinical significance (9, 77, 78). IE that is induced by the SBSEC presents as symptoms that are similar to those by viridians streptococci, involving subacute presentation with prolonged duration of symptoms (9, 77, 153). The treatment for IE by the SBSEC usually consists of 2–6 weeks of intravenous therapy with penicillin G or ceftriaxone, sometimes in combination with aminoglycosides—as for other non-β-haemolytic streptococci (141, 142). The species and subspecies of SBSEC is generally considered penicillin-susceptible, although decreased penicillin susceptibility has been reported (9, 49, 55).

Other infections by the SBSEC include hepatobiliary and other intra-abdominal infections, osteomyelitis, septic arthritis, spondylodiscitis, and, in rare cases, meningitis (9, 51, 55, 78, 154). Bacteraemia with the SBSEC is often polymicrobial, especially if the source is an intra-abdominal infection (9, 55). Urinary tract infections have been reported, although the frequency of SBSEC-induced urinary tract infections is difficult to establish, because streptococci in urine cultures are often viewed by clinical microbiology laboratories as contaminants (9, 55). SBSEC bacteraemia is most common in adults, especially elderly men, but can also occur in children and neonates (9, 53, 55, 154, 155).

The risk of IE and other clinical manifestations varies between SBSEC species and subspecies (9, 55, 63, 79). However, such risk between subspecies is challenging to calculate due to the disparate taxonomic changes that have occurred over time and the difficulty in determining SBSEC species and subspecies (9, 49). Further, the variable distribution of SBSEC species and subspecies between geographic areas renders it impossible to estimate the risk of IE for the individual SBSEC species and subspecies using risk values for the SBSEC overall (52, 53, 55, 56, 63, 81, 82). Differences in the incidence of SBSEC bacteraemia and in the distribution of subspecies could explain in part the high variance in the frequency of SBSEC endocarditis between countries, wherein 58% of non-β-haemolytic streptococci IE has been found to be caused by the SBSEC in France compared with 9% in Sweden (9, 77, 78, 156, 157).

Streptococcus gallolyticus ssp. *gallolyticus*

Sg gallolyticus is usually considered synonymous with biotype I in clinical studies (4, 47, 57). Because biotype I was established as a separate entity from other SBSEC

members early on, *Sg gallolyticus* has been studied more extensively than the other species and subspecies (9).

Previous demographic studies of all the different SBSEC subspecies have found that IE is diagnosed in 17–53% of bacteraemias by *Sg gallolyticus* (52, 53, 55, 56, 63, 81, 82). However, a study by Corredoira et al. in Spain of 109 episodes of only *Sg gallolyticus* bacteraemia reported considerably higher proportions, in which as many as 79% cases were diagnosed with IE, and a literature review reported IE in 64% of bacteraemias (124, 158). Most studies agree that bacteraemia with *Sg gallolyticus* has a high risk of IE—greater than that with other SBSEC species and subspecies (9, 124, 158). Other infections include osteomyelitis, spondylodiscitis, and biliary tract and other intra-abdominal infections (9, 51, 55).

Streptococcus gallolyticus ssp. *pasteurianus*

Sg pasteurianus is considered synonymous with biotype II/2 in clinical studies (9). It is commonly associated with biliary tract infections, such as cholangitis and cholecystitis; other intra-abdominal infections are also frequent (9, 51, 55, 81). Infections with *Sg pasteurianus* are often polymicrobial (55, 81). Previous studies of all SBSEC subspecies have described IE in 0–33% of bacteraemias, and 18% in the review referred to in the previous section (52, 53, 55, 56, 63, 81, 82, 158). *Sg pasteurianus* is the SBSEC member that is foremost associated with neonatal sepsis and meningitis (9, 154).

Streptococcus lutetiensis and *Streptococcus infantarius*

Because *S. lutetiensis* (previously *S. infantarius ssp. coli*) and *S. infantarius* (previously *S. infantarius ssp. infantarius*) used to be encompassed by the previous species *S. infantarius* (previously biotype II/1), it is often impossible to determine the species to which *S. infantarius* refers unless specified or implied from the context of a study (9, 41, 42). Further, many studies use 16s rRNA to identify species and subspecies, but the capacity of this method to distinguish between these two species has been questioned (49).

Generally, few bacteraemias of each species have been included in studies on the SBSEC, making it difficult to estimate the risks of endocarditis and other infectious manifestations. Endocarditis rates of 0–29% for *S. lutetiensis* and 0–14% for *S. infantarius* have been reported in previous demographic studies of all SBSEC subspecies, and a case series by Corredoira et al. found IE in 7% (7/96) of bacteremia with *S. lutetiensis* and 60% (6/10) of *S. infantarius* (52, 53, 55, 56, 63, 81, 82, 158). Intra-abdominal infections, such as cholecystitis, cholangitis, and intra-abdominal abscesses by *S. lutetiensis* and *S. infantarius* have been described (51-53, 55, 56, 63, 82).

Streptococcus alactolyticus, *Streptococcus equinus,* and *Streptococcus gallolyticus* ssp. *macedonicus*

Bacteraemia with *S. alactolyticus*, *S. equinus,* and *Sg macedonicus* has been reported in humans, albeit rarely. Rare case reports have described IE by all three pathogens (52, 159-163).

Marmolin et al. characterised 53 SBSEC bacteraemias; one was noted as *S. equinus,* and two were attributed to *Sg macedonicus* (63). None had IE, but the infections were not otherwise detailed. Ben-Chetrit reported a series of 40 SBSEC isolates from blood cultures, two of which were found to harbour *Sg macedonicus:* one with definite IE and one with probable IE (52).

The few episodes that have been noted are likely to be due, in part, to the use of an older taxonomy, in which *Sg macedonicus* was not recognised, and the inability of MALDI-TOF MS—the routine method for identifying species today—to properly distinguish these three bacteria from other SBSEC species and subspecies (9, 49). Although human infections with these three SBSEC members are rare, they could also be underreported.

Aims of the thesis

Despite extensive previous research on the SBSEC, much remains unknown. When a patient is found to have bacteraemia with the SBSEC, the subspecies is likely to be unidentified, because routine identification can be expected to be performed by MALDI-TOF MS in most modern clinical microbiology laboratories. But can MALDI-TOF MS identify SBSEC species or subspecies with current libraries, or can this technique be improved?

When the SBSEC is detected in a blood culture, most clinicians are likely to consider a routine colonoscopy, but how great is the risk of CRC for the patient? Is a routine colonoscopy necessary for every SBSEC subspecies? Should colonoscopy be repeated every third year thereafter, as some clinicians recommend (9)? Should the cancer work-up also include other gastrointestinal cancers? What is the risk of IE for each SBSEC subspecies, and when can echocardiography be omitted? What is the incidence of SBSEC bacteraemia and the distribution of its subspecies in Sweden?

This thesis examined the following questions in this regard:

- Can SBSEC species and subspecies be distinguished by MALDI-TOF MS?
- How prevalent is colorectal and other gastrointestinal cancers following SBSEC bacteraemia?
- When is CRC diagnosed following SBSEC bacteraemia? Is there an elevated incidence of CRC after the first year following bacteraemia?
- Is bacteraemia with all SBSEC species and subspecies associated with a diagnosis of CRC following infection?
- Is bacteraemia with certain SBSEC species and subspecies associated with diagnoses of other gastrointestinal cancers than CRC following infection?
- What is the proportion of IE in bacteraemia with each respective SBSEC species and subspecies? Is the risk of IE greater for certain species and subspecies?
- What is the distribution of SBSEC species and subspecies in Sweden?

Present investigations

An overview of epidemiological research in Sweden and in general

Data sources in Sweden

Sweden has a long tradition of maintaining national health registries with comprehensive coverage. All inhabitants in Sweden are assigned a unique personal identification number upon birth or migration to Sweden. Combined with extensive national health registries, this practice provides ample opportunities for large-scale register-based epidemiological research (164, 165).

The Total Population Register

The Total Population Register (in Swedish: Registret för totalbefolkningen/ Befolkningsregistret) contains information on birth dates, death dates, migrations, name changes, and family relationships (164, 166). It is maintained by the government agency Statistics Sweden (in Swedish: SCB) and is often used to select matched controls in register-based research.

The National Patient Register

The National Patient Register (in Swedish: Patientregistret) is maintained by the National Board of Health and Welfare (in Swedish: Socialstyrelsen), a government agency, and contains data on diseases and treatments in Swedish specialised care by public and private healthcare providers, coded according to the Swedish International Classification of Disease (ICD) system (124, 126). It was created in 1964 and provides complete coverage of all completed inpatient stays since 1987, with a reported coverage of over 99% (124). Since 2001, this register has also included data on patients who have been treated in specialised outpatient care. It does not contain data on primary care.

The National Prescribed Drug Register

The National Prescribed Drug Register (in Swedish: Läkemedelsregistret) is managed by the National Board of Health and Welfare (167). It was established in 2005 and provides data on all prescribed drugs that are dispensed in pharmacies.

The National Cause of Death Register

The National Cause of Death Register (in Swedish: Dödsorsaksregistret) is also maintained by the National Board of Health and Welfare (167). It lists the official cause-of-death statistics, and its data are used to describe the health of the Swedish population and provide information for healthcare interventions and research.

The National Cancer Register

The National Cancer Register (in Swedish: Cancerregistret) was established in 1958 and contains the official cancer statistics in Sweden (167). It is used to monitor the incidence of cancer and its changes over time. It has been coded according to the 3rd edition of International Classification of Diseases for Oncology (ICD-O-3) since 2005. This register contains details on the time of diagnosis, clinical and morphological diagnosis; extent of tumour at the time of diagnosis; and patient demographics, such as gender, age, and place of residence. The coverage and accuracy of coding are estimated to be approximately 99%, according to the National Board of Health and Welfare, which maintains this register (168).

The Swedish Colorectal Cancer Register (SCRCR) and Colorectal cancer data Base Sweden (CRCBaSe)

The Swedish Colorectal Cancer Register (SCRCR) (in Swedish: Svenska kolorektalcancerregistret) contains national data on all rectal cancers since 1995 and all colon cancers since 2007 (169, 170). The Colorectal Cancer data Base Sweden (CRCBaSe) is a register-based data research platform that links data in the SCRCR to the registers of Statistics Sweden and the National Board of Health and Welfare. Their purpose is to facilitate epidemiological research, containing detailed data on diagnosed CRCs and their follow-up.

Regional databases

Sweden is divided into 21 administrative regions, each of which is responsible for providing healthcare services for its population. The regions have separate medical records systems—sometimes for the specialised and primary care in a respective region, as is the case for Region Skåne. Private healthcare providers are occasionally included in the regions' medical records systems.

There is no national microbiology database. The regions are responsible for providing and managing microbiological databases for their respective laboratories, sometimes in collaboration with other regions. Local laboratories and databases have changed over time, and public and private healthcare services have managed microbiological laboratories. At the time of data collection for Paper I, 23 microbiological laboratories were listed in The Public Health Agency in Sweden. Region Skåne has only one microbiological laboratory—Clinical Microbiology, Region Skåne—which serves the entire population of Skåne, enabling population-based epidemiological research in infectious disease in this area.

Study designs in epidemiology

Although randomised controlled studies are considered the gold standard for clinical interventional trials, they are unfeasible to implement for observational studies that follow individuals over a long period or for the study of rare exposures or outcomes.

For epidemiological studies that are based on healthcare databases, such as registry data, one must select an alternative study design and strive to mitigate confounding biases—for instance, by including a matched control population and adjusting for potential confounders in the statistical analysis.

An often-used design for observational studies is the case-control study, in which a group of individuals—i.e., cases—with an outcome, such as cancer, are compared with individuals who do not experience the outcome—controls (166). This study design is suitable for investigating the exposures that are associated with the outcome, comparing how common the exposure is among those with the outcome versus those without it—for example, smoking among those with and without lung cancer. A case-control study is particularly useful if the outcome is rare and if the time from exposure to outcome is long, and it has the advantage of allowing multiple exposures to be studied simultaneously while being quick to conduct (166). The main limitations are that such studies are retrospective, prone to multiple biases, and limited primarily to odds ratios (OR) as effect measures (166).

Another common observational study design is the cohort study. In contrast to case-control studies, which begin with the outcome, an observational study is based on an exposure rather than an outcome (166, 167). The exposed group is then matched to an unexposed control population to determine whether one or multiple outcomes correlate with the exposure. A cohort study can be prospective or retrospective. The strength of the cohort study lies in its capacity to examine the effect of an exposure on an outcome, with the potential to assess other effect measures, such as risks and incidences (167).

A third option is the cross-sectional study design, which evaluates data at a specific point in time without retrospective or prospective follow-up (171). This type of study is applied primarily to assess the prevalence of diseases in a select population. It is faster to plan, but its ability to reach conclusions about associations and causal relationships is limited, because exposure and outcome are measured at the same time (171).

A major pitfall of observational studies concerns how they select study and control populations. Although case-control and cohort studies aim to include controls that are as similar to the study group as possible in all aspects other than exposure or outcome, depending on the study design, such a practice is often unfeasible in a retrospective study and is prone to introducing bias (172).

Several strategies can be used to mitigate such known and unknown confounders. One approach is to match controls on pre-selected parameters, such as age and sex,

to balance covariates between study groups (168). Another method is weighting, in which data is adjusted based on weights that are assigned to individuals. A typical example of this technique is inverse probability of treatment weights, based on propensity scores, whereas entropy balancing is a method that has been introduced more recently (173-175). A potential drawback of weighting methods is that extreme weights lead to bias, potentially due to differences in the covariates of the study groups. This issue can be addressed through weight trimming, although there is limited guidance on the optimal methods for doing so (173-176). A third option is stratification, in which the sample is divided into strata of similar individuals—e.g., based on age, sex, or propensity scores (173). Alternatively, a study design, called "Target Trial Emulation," has proposed that observational studies be designed more like randomised controlled studies to decrease selection bias (177, 178).

The STROBE guidelines, comprising a 22-item checklist, have been created for epidemiological research (179). This checklist is considered good practice and often required for publication in medical journals. These guidelines were developed to increase the generalisability of epidemiological research by assisting writers in preparing their manuscripts and editors, reviewers, and readers in critically appraising articles.

Paper I

Risk and prognosis of colorectal cancer following bacteremia with Streptococcus bovis-Streptococcus equinus-complex: A Swedish nationwide retrospective cohort study

The aim of this study was to examine the risk of CRC and its prognosis, and the frequencies of other gastrointestinal malignancies, and determine when the malignancies were diagnosed following SBSEC bacteraemia compared with the general population.

Methods

Study design and setting

A nationwide retrospective cohort study of SBSEC bacteraemia in Sweden from 2010–2019. A personal identity number is issued to all individuals who are registered in Sweden, which allows cross-referencing with national health registries and hospital and laboratory databases. Data on all blood cultures with growth of SBSEC were requested from all 23 microbiological laboratories in Sweden, of which 18 delivered them.

We included individuals aged ≥18 years old with a positive SBSEC blood culture, defining an episode and its index date in the exposed cohort. The exposed episodes were matched 1:10 to the general population for index year, sex, year of birth, and geographic region using the Swedish Total Population Register. End of follow-up was December 31, 2021; death; emigration from Sweden; or cross-over to the exposed cohort. The Swedish Total Population Register supplied data on this information and date of birth, and data on the underlying cause of death were retrieved from the National Cause of Death Register.

Colorectal and other gastrointestinal cancers

Cancer diagnoses from January 1, 2000 until the end of follow-up were retrieved from the National Cancer Register, and details on CRCs that were diagnosed during follow-up were obtained from the Colorectal Cancer data Base Sweden (CRCBaSe).

CRC was defined according to ICD-O/3 topography codes C18–C20 (excluding C18.8) with a histopathological diagnosis of adenocarcinoma, unspecified malignant tumour, or an unavailable histopathological diagnosis, and other gastrointestinal cancers were defined per topography codes C15–C17 and C21–C26 with a histopathological diagnosis of invasive malignancy with a risk of metastasis, according to the coding in the National Cancer Register.

Comorbidities and infective endocarditis

Data were retrieved from the National Patient Register, which contains data on diagnoses in inpatient or specialised care. Diagnoses of IE during the episode with SBSEC bacteraemia and data on comorbidities and colonoscopies from 5 years before the index date until the end of follow-up were collected. The Charlson comorbidity index, adapted to Swedish ICD-10 lists, was used to score comorbidities (180). Data from the National Prescribed Drug Register on prescribed drugs for treating diabetes mellitus and chronic pulmonary disease from 1 year prior until the index date were obtained and added to the Charlson comorbidity index score to improve the coverage of comorbidities in primary care.

Statistical analysis

Pearson chi-square test was used to analyse associations between categorical variables and exposure status, and Mann-Whitney U-test was applied for continuous variables. Individuals with previously diagnosed gastrointestinal cancers were excluded from the calculation of rates for specific cancers. Missing data were only available on details on diagnosed cancer, and the missing data were excluded from the analyses.

A CRC follow-up cohort was created, excluding episodes with previously diagnosed CRC, SBSEC relapses, those who died within 14 days of bacteraemia, and their corresponding controls. HRs for a diagnosis of CRC following SBSEC bacteraemia in this cohort were estimated by Cox regression, stratified by matching cohorts, and via weights based on entropy balancing, using index year, geographic region, sex, age, and the comorbidities of the Charlson comorbidity index as separate covariates.

Results

In total, 930 SBSEC bacteraemia episodes were identified in the databases, of which 908 SBSEC episodes and 9080 controls were included in the complete cohort; ultimately, 731 SBSEC episodes and 7158 controls were included in the CRC follow-up cohort (Figure 7).

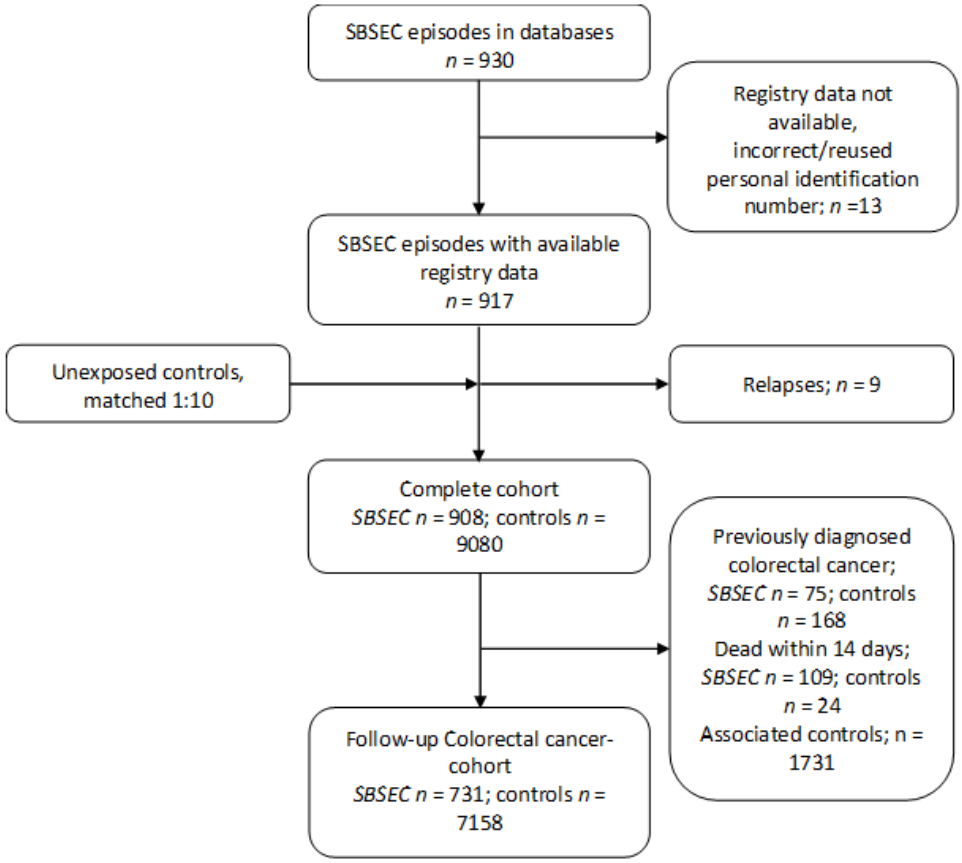


Figure 7. Flowchart of inclusion
The flowchart shows inclusion of the SBSEC episodes in databases, the complete cohort, and the CRC follow-up cohort.

Gastrointestinal cancer overall

Most gastrointestinal cancers were more common in the SBSEC cohort prior to bacteraemia. Only CRC was diagnosed more frequently in the exposed versus unexposed cohort following bacteraemia (p <0.01) (Table 5).

Table 5. Gastrointestinal cancer diagnosed prior to and after bacteraemia compared with controls

All gastrointestinal cancers: Previously diagnosed and diagnosed during follow-up of the complete cohort		**Exposed – SBSEC (n = 908)**	**Unexposed controls (n = 9080)**	***P***
Follow-up total, years (median, IQR)		2.1 (0.3–4.1)	4.1 (2.6–6.3)	<0.01
Colorectal cancer	Previous	75 (8.3)	168 (1.9)	<0.01
	Diagnosed	45 (5.4)	114 (1.3)	<0.01
Oesophagus	Previous	2 (0.2)	4 (0.0)	0.04
	Diagnosed	0 (0)	7 (0.1)	0.40
Gastric	Previous	12 (1.3)	9 (0.1)	<0.01
	Diagnosed	1 (0.1)	16 (0.2)	0.65
Hepatic	Previous	4 (0.4)	10 (0.1)	0.01
	Diagnosed	2 (0.2)	18 (0.2)	0.89
Gallbladder, extrahepatic bile ducts	Previous	12 (1.3)	3 (0.0)	<0.01
	Diagnosed	1 (0.1)	9 (0.1)	0.91
Pancreatic	Previous	27 (3.0)	7 (0.1)	<0.01
	Diagnosed	1 (0.1)	17 (0.2)	0.62
Small bowel	Previous	8 (0.9)	6 (0.1)	<0.01
	Diagnosed	1 (0.1)	3 (0.0)	0.27
Anal	Previous	1 (0.1)	4 (0.0)	0.40
	Diagnosed	0 (0.0)	2 (0.0)	0.66
Other gastrointestinal	Previous	0 (0.0)	0 (0.0)	-
	Diagnosed	1 (0.1)	5 (0.1)	0.52

Colorectal cancer

The incidence of CRC was higher overall in the exposed SBSEC cohort during follow-up (p <0.01), wherein the largest differences were seen during the first year following bacteraemia (p <0.01). There were no differences in this incidence during the second year, but the differences throughout the remainder of the follow-up were significant (p=0.01). The median time to diagnosis was significantly shorter in the exposed cohort, at 0.21 years (interquartile range [IQR] 0.02–0.75) versus 2.34 years (1.22–4.41) (p <0.01). Ten individuals were diagnosed with CRC at >12 months in the SBSEC cohort, but only two of them had been examined by colonoscopy during the first 12 months. CRCs diagnosed and HRs for CRC diagnoses following SBSEC bacteraemia are presented in Table 6 and Figure 8.

Of the CRCs that were diagnosed during follow-up, few significant differences were observed between the exposed and unexposed cohorts. The Charlson comorbidity score was lower in unexposed individuals (p <0.01). The follow-up time was shorter among the exposed cohort—2.6 years (IQR 1.8–5.1)—compared with the unexposed, (5.2; IQR 3.0–7.2) (p <0.01). However, no significant differences appeared with regard to tumour location, TNM stage, or tumour grade at diagnosis.

Table 6. Colorectal cancer diagnosed during follow-up of the colorectal cancer cohort

Colorectal cancer diagnosed during follow-up of the colorectal cancer cohort.			**Exposed – SBSEC (n = 731)**	**Unexposed controls (n = 7158)**	***P***
Colonoscopy		In total	195 (26.7)	437 (6.1)	<0.01
		<12 months	154 (21.1)	108 (1.5)	<0.01
		≥12 months	41 (5.6)	329 (4.6)	0.22
Benign colorectal tumours and polyps[a)]		In total	106 (14.5)	187 (2.6)	<0.01
Colorectal cancer	Total	Colorectal cancer, *n*	45	89	
		Incidence per 1000 person-years	20.4	2.6	<0.01
		Incidence difference (CI95%)	17.7 (11.7–23.7)		
	Year <1	Colorectal cancer, *n*	35	20	
		Incidence per 1000 person-years	59.5	2.9	<0.01
		Incidence difference (CI95%)	56.6 (36.9–76.4)		
	Year ≥1–<2	Colorectal cancer, *n*	2	20	
		Incidence per 1000 person-years	4.3	3.1	0.62
		Incidence difference (CI95%)	1.2 (-4.9–7.2)		
	Year ≥2	Colorectal cancer, *n*	8	49	
		Incidence per 1000 person-years	6.9	2.4	0.01
		Incidence difference (CI95%)	4.5 (-0.3–9.4)		

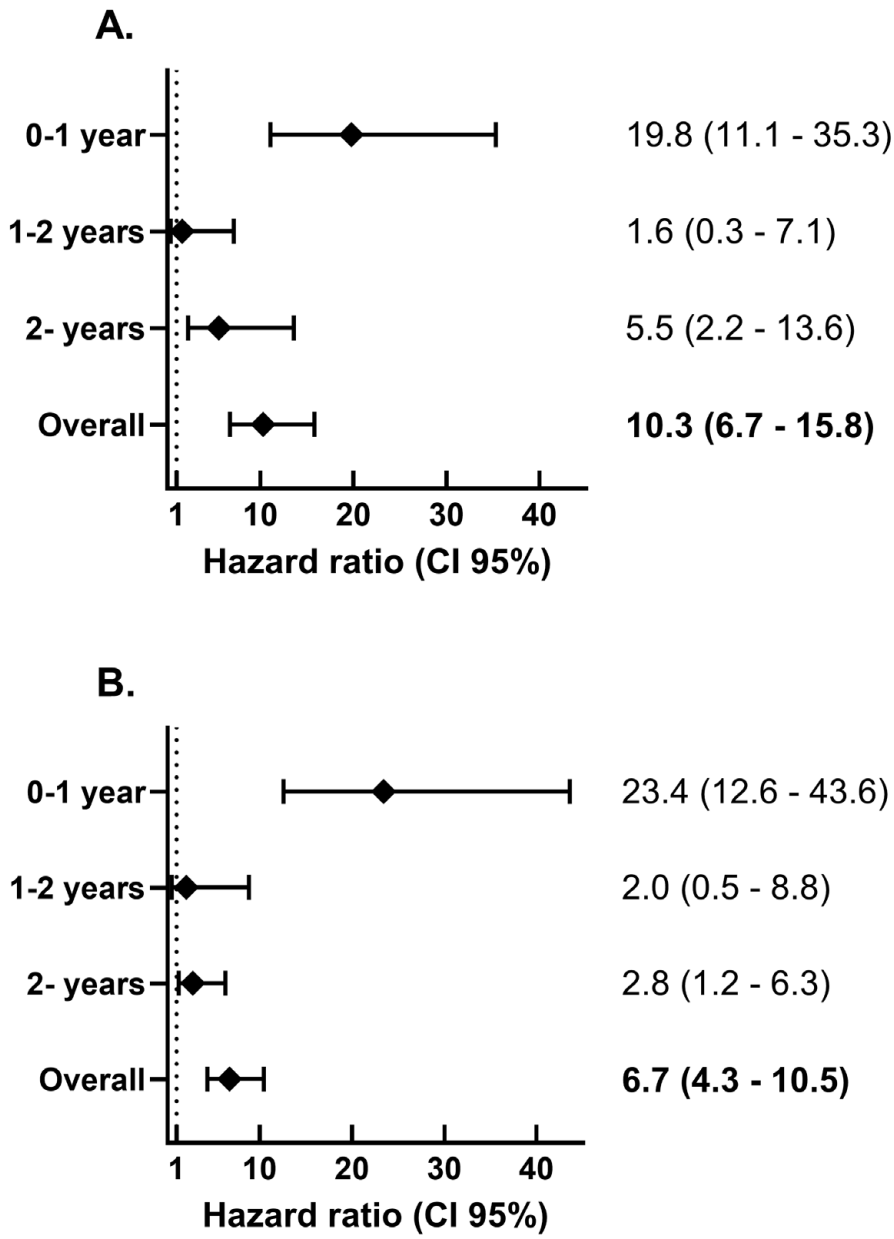


Figure 8. Hazard ratios for colorectal cancer diagnosis following SBSEC bacteraemia compared with matched controls. A. Stratified by matching groups. B. Entropy balancing using index year, geographic region, sex, age, and the comorbidities of the Charlson comorbidity index.

Paper II

Streptococcus bovis-bacteremia: subspecies distribution and association with colorectal cancer: a retrospective cohort study

The aim of this study was to analyse the distribution of SBSEC species and subspecies in bacteraemia in the Skåne region and examine the risk of colorectal and other gastrointestinal cancers following bacteraemia for each species and subspecies.

Methods

Setting

The study was conducted retrospectively in Skåne, where a single laboratory, Clinical Microbiology, Region Skåne, serves all hospitals and stores all bacterial isolates from blood cultures for a minimum of 10 years. All episodes of SBSEC-positive blood cultures from 2003–2018 were identified and retrieved from Clinical Microbiology, Region Skåne. Microbiological and clinical records were reviewed retrospectively. Individuals could be included multiple times as episodes of SBSEC bacteraemia but only once per hospital admission.

Microbiology

SBSEC isolates from blood cultures were stored at -70°C at Clinical Microbiology, and reference type strains for each SBSEC species and subspecies were precultured for 24–120 h on blood agar plates in 5% CO_2 at 37°C to check for bacteria viability and contamination. They were then re-cultured in Brain Heart Infusion (BHI) broth at 37°C and 120 shakes/min for 12 to 48 hours to increase the amount of cultured bacteria.

DNA was extracted and purified using the DNeasy Blood & Tissue Kit according to the manufacturer's instructions. The Center for Translational Genomics, Lund University then performed short-read WGS. SeaView was used to construct a neighbour-joining phylogenetic tree, based on SNPs, using Kimura-2-parameters (181). Species and subspecies were confirmed by clustering around the reference type strains (Figure 9).

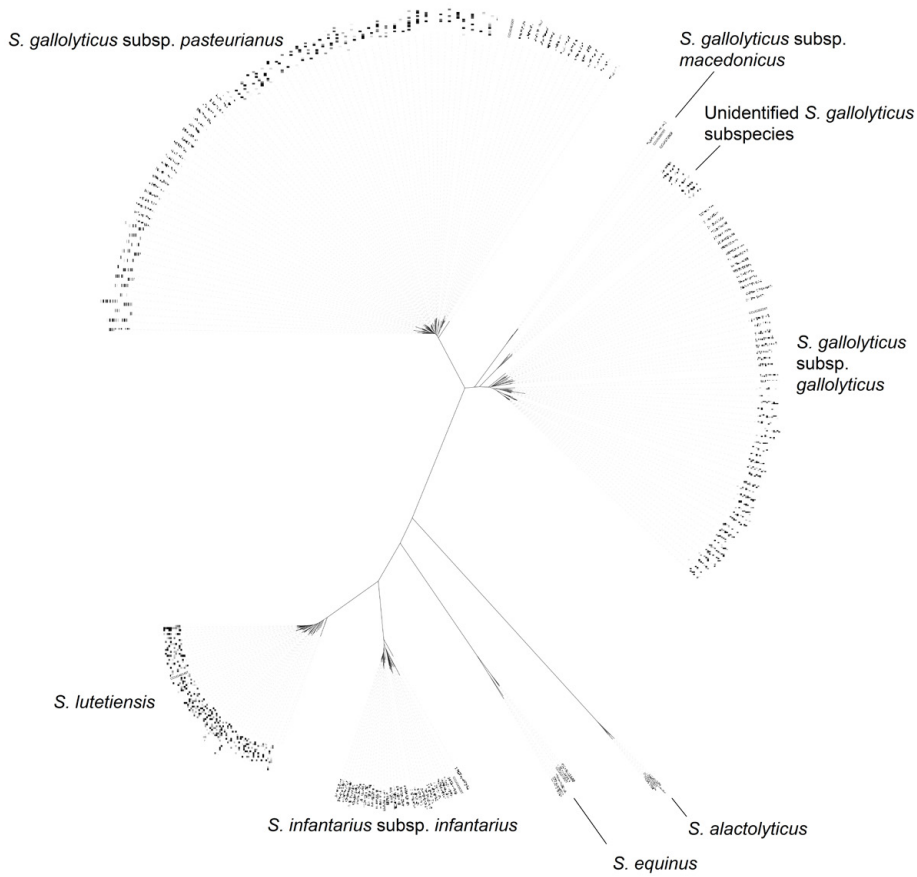


Figure 9. Clustering of SBSEC species and subspecies in a neighbour-joining phylogenetic tree (21).

Data collection

Data on comorbidities and cancers were collected from medical records retrospectively according to a predetermined protocol. Data collected included colonoscopies within 12 months, and findings of cancer and adenomas were registered. Cancers that were diagnosed prior to a positive blood culture (prevalent cancer) and those following a positive blood culture (incident cancer) were gathered from the regional pathological database, covering the entire population of Skåne. Comorbidities were registered and categorised per the updated Charlson comorbidity index (182). Date of bacteraemia was considered the index date, and episodes were followed until the date of death, emigration, or March 15, 2021.

Statistics

An incidence cohort that consisted of episodes without prior CRC or metastasised cancer was included in the follow-up after positive blood cultures to determine the risks of CRC and other gastrointestinal cancers. Logistic regression was used to analyse the association between SBSEC species and subspecies and CRC, adjusted for predetermined potential confounders (90, 183). The covariates were analysed independently and in a multivariable logistic regression model.

The numbers that were needed to screen for a diagnosis of CRC within 12 months were calculated for the incidence cohort, for all SBSEC overall and only *Sg gallolyticus*-bacteraemia, both overall and for episodes without an identified infection focus.

The incidence of CRC in Skåne, stratified by calendar year, sex, and 5-year interval age groups, was used to calculate the standardised incidence ratio of CRC within 12 months following *Sg gallolyticus* and other SBSEC bacteraemias for the complete cohort.

Results

The study cohort and the distribution of subspecies and species

Episodes of SBSEC bacteraemia were included according to the flowchart in Figure 10.

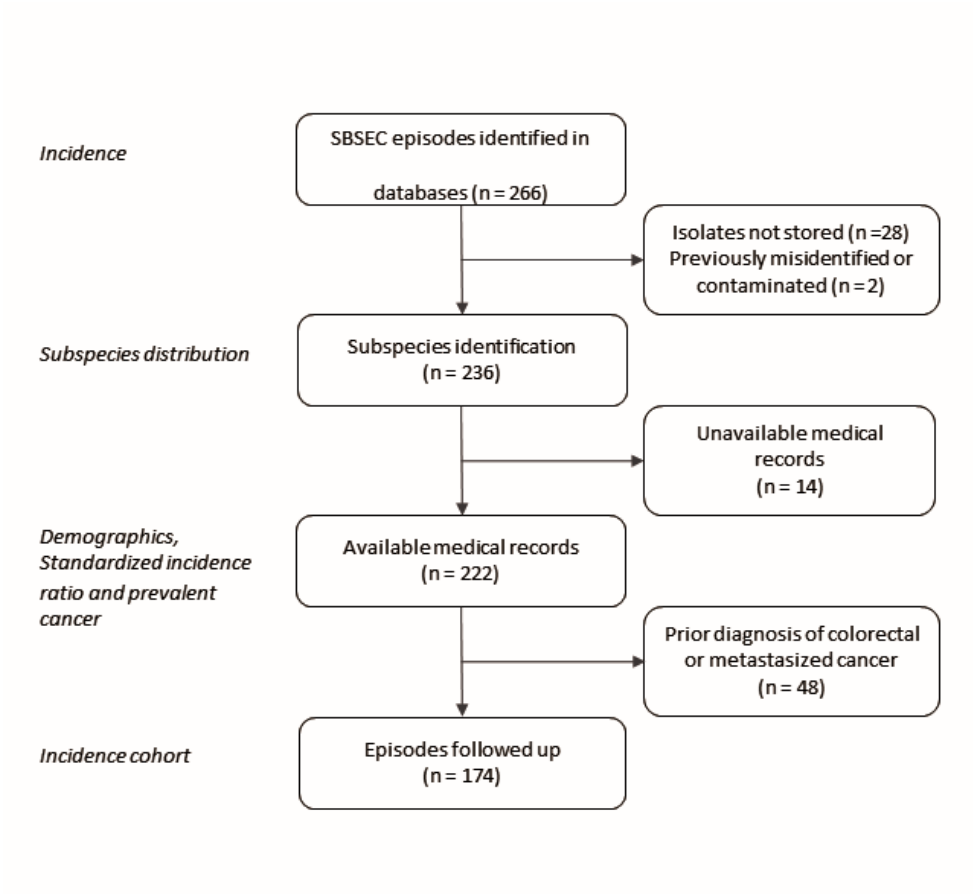


Figure 10. Flowchart of selection of episodes of SBSEC bacteraemia.

In total, 236 isolates were available for the determination of species and subspecies, of which *Sg pasteurianus* was the most common (88; 37%), followed by *Sg gallolyticus* (58; 25%), *S. lutetiensis* (47; 20%), and *S. infantarius* (32; 14%). Nine isolates formed a separate cluster among the *S. gallolyticus* subspecies and could not be assigned to a specific subspecies. The distribution of SBSEC species and subspecies in SBSEC bacteraemia in Skåne from 2003-2018 is presented in Figure 11.

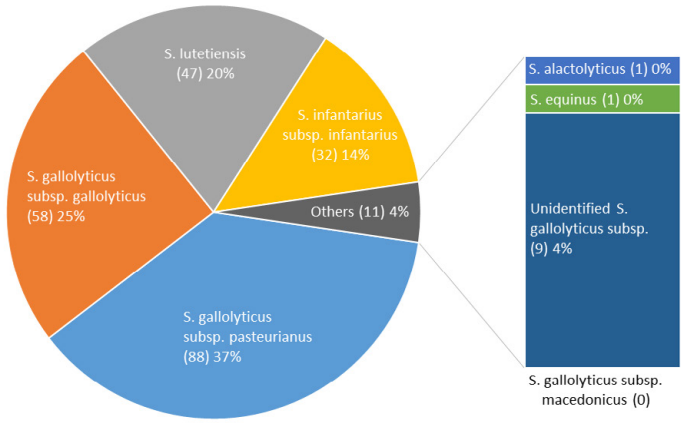


Figure 11. Distribution of SBSEC species and subspecies in Skåne from 2003-2018.

Prevalent colorectal and other gastrointestinal cancers

There were no significant differences in the prevalence of known CRCs or colorectal adenomas/polyps before bacteraemia. Pancreatic cancer was more common among episodes of bacteraemia with *S. lutetiensis* and *S. infantarius*.

Incidence of colorectal and other gastrointestinal cancers

In total, 13 of 174 (7%) bacteraemia episodes were diagnosed with CRC during the follow-up, of which 12 were diagnosed in the first 12 months. CRC was significantly more frequent among bacteraemias with *Sg gallolyticus* than with all other SBSEC species and subspecies—within 12 months and in total during follow-up (Table 7). Only one case of other gastrointestinal cancer was diagnosed during follow-up: pancreatic cancer following bacteraemia with *S. lutetiensis*.

Table 7. Discovered cancer following bacteraemia
Other subspecies (n=9) are omitted from the table, among which no cancers were diagnosed following bacteraemia. Variables are presented as n (% of observations per subspecies).

	Sg pasteurianus **(n = 67)**	***Sg gallolyticus*** **(n = 46)**	***S. lutetiensis*** **(n = 29)**	***S. infantarius*** **(n = 23)**	***P***
Cancer diagnoses within 12 months					
Colorectal cancer	2 (3)	9 (20)	0 (0)	1 (4)	<0.01
Other gastrointestinal cancers	0 (0)	0 (0)	0 (0)	0 (0)	NA
Cancer diagnoses in total during the follow-up period					
Colorectal cancer	3 (4)	9 (20)	0 (0)	1 (4)	<0.01
Other gastrointestinal cancers	0 (0)	0 (0)	1 (3)	0 (0)	0.28

In univariate analysis, the OR for a diagnosis of CRC within 12 months for *Sg gallolyticus* compared with all other SBSEC species was 10.14 (CI95% 2.61–39.38; p<0.01) and 9.61 (CI95% 2.27–40.61; p<0.01) in the multivariable model, adjusted for sex, age, and whether an infection focus was detected.

The standardised incidence ratio of CRC within 12 months for *Sg gallolyticus* was 59.8 (CI95% 27.3–113.3) and 7.2 for all other SBSEC species (CI95% 1.5–20.9).

The numbers that were needed to screen to detect one CRC were 15 for all SBSEC species in the incidence cohort and 9 for those without an identified focus of infection, versus 5 and 3, respectively, for *Sg gallolyticus*.

Paper III

Bacteraemia and infective endocarditis with Streptococcus bovis-Streptococcus equinus complex: a retrospective cohort study

The aim of this study was to examine the proportions of IE and presentation of the disease in bacteraemia with the SBSEC species and subspecies that are most common in infection in humans: *Sg pasteurianus*, *Sg gallolyticus*, *S. lutetiensis*, and *S. infantarius*.

Methods

A retrospective study on SBSEC bacteraemia in Skåne from 2003-2018, using the same bacteria and cohort as in Paper II. Species and subspecies were identified by WGS as described in Paper II, and episodes of bacteraemia by *Sg pasteurianus*, *Sg gallolyticus*, *S. lutetiensis*, and *S. infantarius* were included. Medical records were reviewed retrospectively according to a predefined protocol. An episode of bacteraemia was defined as when a positive blood culture was drawn and, ending at hospital discharge. Relapse was defined as a new episode of bacteraemia with the same subspecies within 6 months.

The infection focus was defined according to the attending physician's diagnosis in the medical charts, and multiple sites of infection were possible per episode.

Definite IE was defined per ESC 2015 (142). In the case of inconclusive imaging results or if a clinical diagnosis of IE did not fulfil the criteria of definite IE, imaging results and/or medical charts were reviewed by another physician who did not have knowledge of SBSEC species and subspecies.

Differences in proportions were assessed by chi-squared or Kruskal-Wallis test, as appropriate.

Results

The proportions of IE in bacteraemia were significantly higher with *Sg gallolyticus* (33%) and *S. infantarius* (16%) versus *Sg pasteurianus* (5%) and *S. lutetiensis* (5%) (p <0.01) (Table 8). Intra-abdominal infections were more common in bacteraemia with *S. lutetiensis* (21/44; 48%) and *Sg pasteurianus* (37/83; 45%) compared with *S. infantarius* (5/31; 16%) and *Sg gallolyticus* (8/52; 15%) (p <0.01). No episode of bacteraemia with intra-abdominal infections was diagnosed with IE, but echocardiography was only performed in 22 of 71 cases (31%). Polymicrobial infections were more frequent in bacteraemia with *S. lutetiensis* and *Sg pasteurianus* (p=0.02).

Table 8. Clinical characteristics
Variables are presented as n (% of observations per subspecies).

	***Sg pasteurianus* (*n* = 83)**	***Sg gallolyticus* (*n* = 52)**	***S. lutetiensis* (*n* = 44)**	***S. infantarius* (*n* = 31)**	***P***
Microbiology and examinations					
Polymicrobial infection	27 (33)	7 (13)	17 (39)	6 (19)	0.02
2 or more positive cultures	46 (55)	43 (83)	18 (41)	21 (68)	<0.01
Echocardiography: TTE or TOE	43 (52)	46 (88)	18 (41)	23 (74)	<0.01
Echocardiography: TOE	28 (34)	33 (63)	10 (23)	16 (52)	<0.01
Site of infection					
Unknown	32 (39)	23 (44)	16 (36)	16 (52)	0.52
Infective endocarditis	4 (5)	17 (33)	2 (5)	5 (16)	<0.01
Cholecystitis/ cholangitis/ pancreatitis/hepatic infection	17 (20)	3 (6)	13 (30)	3 (10)	0.10
Colitis/appendicitis/ diverticulitis/other intra-abdominal	20 (24)	5 (10)	8 (18)	2 (6)	0.06
Spondylodiscitis, osteomyelitis, or arthritis	6 (7)	6 (12)	0 (0)	0 (0)	0.04
Skin/tissue/muscle	1 (1)	1 (2)	2 (5)	1 (3)	0.68
Other[1]	6 (7)	3 (6)	4 (9)	4 (13)	0.69
Outcome					
In-hospital mortality	9 (11)	2 (4)	8 (18)	6 (19)	0.09
Relapse	4 (5)	1 (2)	0 (0)	0 (0)	0.26

[1]Urinary, respiratory, or female genital tract and central venous catheter infection.

Paper IV

Improved identification of Streptococcus bovis-Streptococcus equinus complex species and subspecies by MALDI-TOF MS using a novel library

The aim of this study was to evaluate the current diagnostic accuracy of the Bruker MALDI Biotyper system in identifying SBSEC species and subspecies using the ethanol-formic acid extraction method and to develop a novel library to improve diagnostic accuracy in the identification of SBSEC species and subspecies with MALDI-TOF MS.

Methods

Setting and reference standard test

We included unique SBSEC isolates in Skåne from 2003–2018 that were identified by WGS, as described in Paper II. Type strains and isolates of SBSEC species and subspecies that are rare in human infections were obtained from culture collections.

MALDI-TOF MS

Isolates were precultured for 24–120 h on blood agar plates in 5% CO_2 at 37°C to assess bacteria viability and contamination and then recultured in BHI broth at 37°C and 120 shakes/min for 12–48 h. The broth was then centrifuged, and the supernatant was removed. Standard ethanol-formic acid extraction was performed on the remaining pellet according to the MALDI-TOF MS manufacturer's protocol (Bruker Daltonics, Germany). The resulting supernatant was then used to collect spectral profiles, 1 µL per sample was pipetted onto 8–10 spots of a stainless steel MALDI target plate and left to dry. After the drying step, 1 µL of fresh matrix solution was applied to each spot and left to dry again at room temperature. Spectra were collected using a Microflex LT/SH SMART MALDI-TOF MS with flexControl v. 3.4 (Bruker Daltonics, Germany) after calibration according to the manufacturer's instructions using Bruker Bacterial Tests Standard, with a spectral intensity of 10,000–25,000. The spots were analysed three times, and a total of 24–30 spectra were generated.

Half of the isolates per species and subspecies (including type strains) were selected to create the library, termed "*Streptococcus bovis*/*Streptococcus equinus* complex-Clinical Microbiology Region Skåne" (SBSEC-CMRS), and the remaining isolates were used for library evaluations.

flexAnalysis v. 3.4 (Bruker Daltonics, Germany) was used for further analysis. Smoothing and baseline subtraction were performed, and spectra with outlier peaks (not within 500 ppm per 1,000 m/z interval) were excluded. For every isolate, 20–

24 spectra were used to generate Main spectrum profiles (MSPs) and included in the SBSEC-CMRS library.

The SBSEC-CMRS-library version 1.0 included MSPs for 117 unique SBSEC isolates (18 from culture collections, including type strains). Isolates included per SBSEC species and subspecies are listed in Table 9.

An evaluation set was created from the remaining 119 isolates (13 from culture collections). Two random spectra from different spots of each isolate were analysed further in the evaluation set, using MALDI Biotyper (MBT) Compass v.4.1 (Bruker Daltonics, Germany). The evaluation set was analysed with the complete Bruker reference library—MBT Compass Library Revision K (DB-11897, 2022)—applying a predetermined identification cutoff score of ≥2.0. This set was then reanalysed using the SBSEC-CMRS library, with cutoff scores of ≥2.0 and ≥2.3, as recommended by the manufacturer of the mass spectrometer.

The result with the highest score (best match) is reported for every isolate.

Table 9. Isolates in the SBSEC-CMRS library and evaluation set
Number of isolates in total, with those from culture collections in parentheses.

	SBSEC-CMRS library	Evaluation set
S. gallolyticus ssp gallolyticus	28 (1)	28 (0)
S. gallolyticus ssp pasteurianus	36 (1)	36 (0)
S. gallolyticus ssp macedonicus	3 (3)	4 (4)
S. lutetiensis	24 (2)	24 (0)
S. infantarius	16 (1)	16 (0)
S. equinus	4 (4)	5 (4)
S. alactolyticus	6 (6)	6 (5)
Total	117 (18)	119 (13)

Results

The Bruker MBT library could only identify the correct species and subspecies in 72 of 119 isolates (61%; CI95% 51–69%) with a cutoff score ≥2.0, in none of which *Sg gallolyticus* was successfully identified (Table 10 and Figure 12). However, the species *S. gallolyticus* was correctly identified in 66 of 68 isolates (97%), and no other species was misidentified as *S. gallolyticus*.

The SBSEC-CMRS library identified the correct species and subspecies in 116 of 119 isolates, (97%; CI95% 92–99%) with the same cutoff (≥2.0); the scores for all isolates exceeded this cutoff (Table 10 and Figure 12). At a cutoff of ≥2.3, the same number of isolates was correctly identified, but an *Sg gallolyticus* isolate, which was misidentified as *Sg pasteurianus*, attained a score of 2.21 and was thus considered indeterminate, resulting in a total of 116 correct identifications among 119 isolates, 2 misidentifications, and 1 indetermined subspecies.

Table 10. Comparison of results between the Bruker MBT Compass Library and the SBSEC-CMRS Library

Numbers of correct subspecies identified at a cutoff score ≥2.0 (%).

Evaluation set	Bruker MBT Library	SBSEC-CMRS library
***S. gallolyticus ssp gallolyticus* (n=28)**	0 (0)	27 (96)
***S. gallolyticus ssp pasteurianus* (n=36)**	36 (100)	36 (100)
***S. gallolyticus ssp macedonicus* (n=4)**	4 (100)	4 (100)
***S. lutetiensis* (n=24)**	14 (58)	24 (100)
***S. infantarius* (n=16)**	10 (63)	16 (100)
***S. equinus* (n=5)**	2 (40)	4 (80)
***S. alactolyticus* (n=6)**	6 (100)	5 (83)
Total (n=119)	72 (61)	116 (97)

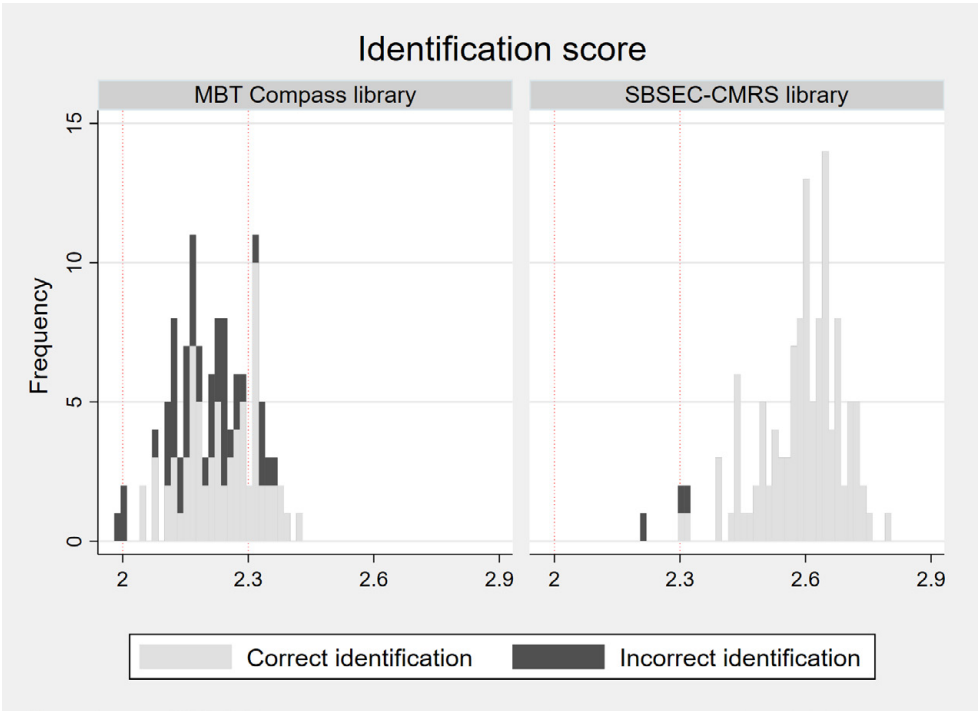


Figure 12. Identification score.
Distribution of identification scores for the two libraries.

Paper V

Creation and validation of improved MALDI-TOF MS libraries for S. bovis-S. equinus complex subspecies identification adapted to diagnostic culturing and extraction conditions

The aim of this study was to evaluate and improve the diagnostic accuracy of the SBSEC-CMRS library under various culture conditions by direct transfer method, on-target extraction with formic acid, and ethanol-formic acid extraction.

Methods

Setting and reference standard test

We included unique SBSEC isolates in Skåne from 2003–2018 that were identified by WGS, as detailed in Paper II, and type strains and additional SBSEC isolates were obtained from culture collections, as described in Paper IV. The SBSEC-CMRS library v1.0 was also included, as described in Paper IV.

Redesigning the SBSEC-CMRS library

The isolates included in the SBSEC-CMRS library v1.0 were included but limited to a maximum of 10 per subspecies (184). To reduce the likelihood of misidentifications of subspecies with over 10 available isolates (*Sg gallolyticus*, *Sg pasteurianus*, *S. lutetiensis*, *S. infantarius*), isolates causing incorrect top hits in the previous evaluation were omitted first, followed by those with the most incorrect hits in the top 10 list. Thus, a maximum of 10 isolates per subspecies, resulting in high-quality subspecies-specific identifications with version 1.0 of the SBSEC-CMRS library, were prioritised. A total of 53 unique isolates remained in the redesigned library (Table 11).

Two new libraries were constructed: one that consisted of isolates that were cultured on horse blood agar plates under aerobic conditions with 5% CO2 and another that was cultured with the same isolates on chocolate blood agar under anaerobic conditions.

Cultures were assessed visually for contamination and confirmed by analysis using the direct transfer method on a Microflex LT/SH SMART MALDI-TOF MS instrument (Bruker Daltonics, Germany) with MBT Compass Library Revision K (DB-11897, 2022). MSPs were created according to the protocol supplied by the manufacturer (Bruker Daltonics, Germany) using the standard ethanol-formic acid extraction method. The resulting supernatant (1 µl) was pipetted onto a minimum of 8 spots on a stainless steel MALDI target plate and left to air-dry, after which 1 µl of fresh matrix solution was pipetted onto each spot and air-dried. Spectra were collected on a Microflex LT/SH SMART MALDI-TOF MS instrument with

flexControl, v.3.4 (Bruker Daltonics, Germany) after calibration according to the manufacturer's protocol using Bruker Bacterial Tests Standard, with a spectra intensity between 10,000 and 25,000. The spots were analysed three times, and a minimum of 24 spectra per sample were collected. flexAnalysis, v.3.4 (Bruker Daltonics, Germany) was used for further analysis. Smoothing and baseline subtraction were performed, spectra with outlier peaks (not within 500 ppm per 1,000 m/z interval) were excluded, and MSPs that included 20–24 spectra were generated.

The "*Streptococcus bovis/Streptococcus equinus*-complex-Clinical Microbiology Region Skåne blood agar aerobic" (SBSEC-CMRS-BAE) library, v.1.0 was created from MSPs using isolates that were cultured on blood agar plates under aerobic conditions, and the "*Streptococcus bovis/Streptococcus equinus*-complex-Clinical Microbiology Region Skåne chocolate agar anaerobic" (SBSEC-CMRS-CAN) library, v.1.0 was constructed from those that were cultured on chocolate agar plates under anaerobic conditions.

Table 11. Isolates in the redesigned SBSEC-CMRS libraries
Number of isolates in total, with those from culture collections in parentheses.

	Redesigned SBSEC-CMRS libraries
S. gallolyticus ssp gallolyticus	10 (1)
S. gallolyticus ssp pasteurianus	10 (1)
S. gallolyticus ssp macedonicus	3 (3)
S. lutetiensis	10 (2)
S. infantarius	10 (1)
S. equinus	4 (4)
S. alactolyticus	6 (6)
Total	53 (18)

Evaluation set

The evaluation set consisted of the same isolates as in the evaluation of the SBSEC-CMRS, v.1.0 in paper IV—a total of 119 unique SBSEC isolates (13 isolates from culture collections), comprising *Sg gallolyticus* 28, *Sg pasteurianus* 36, *Sg macedonicus* 4, *S. lutetiensis* 24, *S. infantarius* 16*, S. alactolyticus* 6, and *S. equinus* 5. The isolates in the evaluation set were cultured under the same conditions as in the creation of the MSPs—under aerobic conditions on blood agar plates and anaerobic conditions on chocolate agar plates.

Extraction was performed using the direct transfer method, the on-target extraction method with 1 µl 70% formic acid, and the ethanol-formic acid extraction method as described, on a Microflex LT/SH SMART MALDI-TOF MS instrument. The samples were analysed in duplicate at different spots. MALDI Biotyper (MBT) Compass, v.4.1 (Bruker Daltonics, Germany) was used for further analysis.

Subspecies were identified using MBT Compass Library Revision x (DB-12438, 2023); the original BHI broth-based SBSEC-CMRS library, v.1.0; the new SBSEC-CMRS-BAE library; and the SBSEC-CMRS-CAN library alone, and the new SBSEC-CMRS-BAE and SBSEC-CMRS-CAN libraries in combination. A predetermined cutoff score of ≥2.0 was used, as established by the MALDI-TOF MS manufacturer. An identification score of ≥2.0 against the MBT Compass Library for the spot with the highest score was used as a surrogate marker of sufficient spectral quality. Culturing, sampling, and extraction were repeated for isolates with lower scores until ≥2.0 was reached. The result with the highest score (best match) for every bacterial isolate is reported.

Results

Bruker MBT Compass library

The Bruker MBT Compass library generally performed poorly in identifications at the subspecies level, correctly determining subspecies in 463 of 714 isolates in total (65%; CI95% 61–68%), all with a score ≥2.0. Results per culture condition and extraction method are presented in Figure 13.

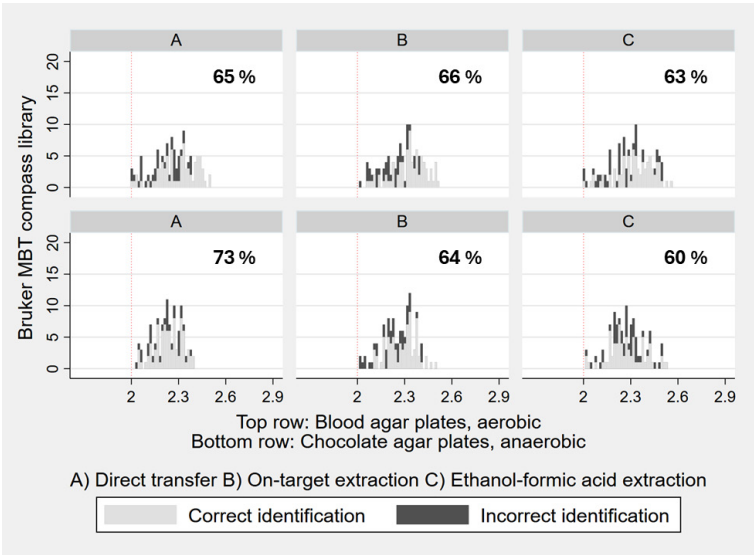


Figure 13. Identification scores and correct identification at the subspecies level by the Bruker MALDI Biotyper Compass library. Comparison of correct identification at the species level and identification scores for the various culture conditions and sample preparation methods. The results of the spot with the highest score are presented. An identification cutoff score of 2.0 is marked (vertical dotted red line), and percentages of correct identifications to the subspecies level per culture condition and sample preparation method are shown.

The results were improved when limiting identification to the SBSEC species level, with the best results achieved with isolates that were cultured under anaerobic conditions on chocolate agar and the direct transfer method—114 of 119 isolates correctly identified (96%) (Table 12).

Table 12. Bruker MBT Compass – correct identification to the SBSEC species level

Bruker MBT Compass	***S. alactolyticus* (n = 6)**	***S. equinus* (n = 5)**	***S. gallolyticus* (n = 68)**	***S. infantarius*[a] (n = 16)**	***S. lutetiensis* (n = 24)**	**Total (n = 119)**
Blood agar plates, aerobic culture conditions						
Direct transfer						
≥2.0, *n* (%)	6/6 (100)	1/5 (20)	68/68 (100)	6/16 (38)	24/24 (100)	105/119 (88)
On-target extraction						
≥2.0, *n* (%)	6/6 (100)	3/5 (60)	68/68 (100)	4/16 (25)	24/24 (100)	105/119 (88)
Ethanol-formic acid extraction						
≥2.0, *n* (%)	6/6 (100)	3/5 (60)	68/68 (100)	0/16 (0)	24/24 (100)	101/119 (85)
Chocolate agar plates, anaerobic culture conditions						
Direct transfer						
≥2.0, *n* (%)	6/6 (100)	3/5 (60)	68/68 (100)	13/16 (81)	24/24 (100)	114/119 (96)
On-target extraction						
≥2.0, *n* (%)	6/6 (100)	2/5 (40)	68/68 (100)	4/16 (25)	24/24 (100)	104/119 (87)
Ethanol-formic acid extraction						
≥2.0, *n* (%)	6/6 (100)	3/5 (60)	68/68 (100)	1/16 (6)	21/24 (88)	99/119 (83)

[a]*S. infantarius* is synonymous with *S. infantarius* ssp. *infantarius*.

The new libraries

Of the new libraries, the SBSEC-CMRS-BAE library performed best overall, correctly identifying subspecies in 99% of isolates (CI95% 98–100) with a cutoff score of ≥2.0; only one isolate fell below the cutoff. The SBSEC-CMRS-BAE library performed best with blood agar plates under aerobic conditions and the on-target method and with chocolate agar plates that were cultured anaerobically with full extraction—both methods yielded correct subspecies identification in 119 of 119 isolates (100%), with none receiving a score of under 2.0. The performance was also good with the SBSEC-CMRS-CAN library, albeit slightly lower than that of the SBSEC-CMRS-BAE library. Combining the libraries did not improve performance. Total identification results for all culture conditions and extraction methods for all libraries are presented in Table 13. Results per culture condition and extraction method are presented in Figure 14 and 15.

Table 13. Comparison of the results with the various libraries: correct subspecies identification of the spot with the highest score.

Library	Correct subspecies identification, identification score ≥2.0[a-c]	Identification score <2.0, n
Bruker MBT Compass	463/714, 65% (61–68)	0
SBSEC-CMRS (BHI broth)	674/698, 97% (95–98)	16
SBSEC-CMRS-BAE[d]	707/713, 99% (98–100)	1
SBSEC-CMRS-CAN[d]	684/706, 97% (95–98)	8
Combined -BAE and -CAN[d]	704/713, 99% (98–99)	1

[a]Total results of all culture and extraction conditions. [b]n, % (CI95%) [c]Results <2.0 excluded. [d]SBSEC-CMRS-BAE library, v.1.0 and SBSEC-CMRS-CAN library, v.1.0, individually and combined.

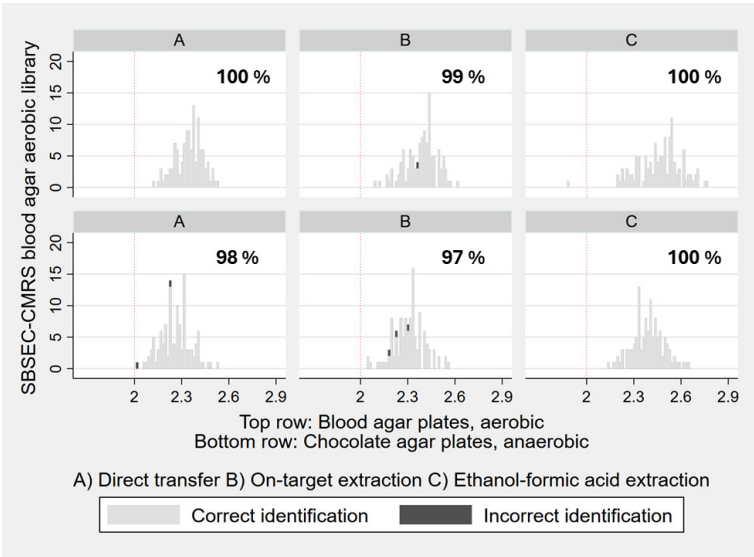


Figure 14. Identification scores and correct identification to the subspecies level by the SBSEC-CMRS-BAE library, v.1.0. Comparison of correct identification at the species level and identification scores for the various culture conditions and sample preparation methods. The results of the spot with the highest score are presented. An identification cutoff score of 2.0 is marked (vertical dotted red line), and percentages of correct identifications to the subspecies level per culture condition and sample preparation method are shown.

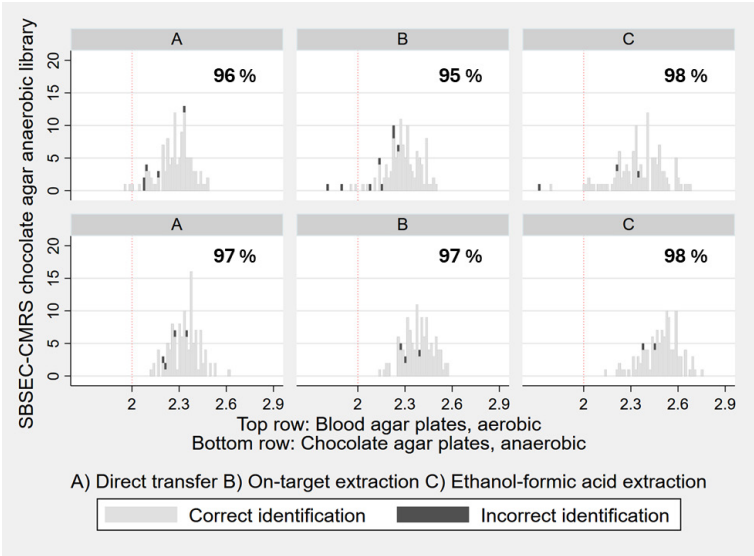


Figure 15. Identification scores and correct identification to the subspecies level by the SBSEC-CMRS-CAN library v. 1.0. Comparison of correct identification at the species level and identification scores for the various culture conditions and sample preparation methods. The results of the spot with the highest score are presented. An identification cutoff score of 2.0 is marked (vertical dotted red line), and percentages of correct identifications to the subspecies level per culture condition and sample preparation method are shown.

Discussion

These studies delve into the associations between bacteraemia with SBSEC and CRC, the proportions of IE in bacteraemia with the various species and subspecies, and the potential use of MALDI-TOF MS in differentiating SBSEC species and subtypes.

In Paper I, we reaffirmed the association between SBSEC bacteraemia and CRC and observed elevated rates of CRC over a long-term follow-up. However, we did not find increased rates of other gastrointestinal cancers after bacteraemia compared with the general population.

In Paper II, we established a WGS method for determining SBSEC species and subspecies and showed that CRC following SBSEC bacteraemia is more frequent with *Sg gallolyticus* versus other subspecies, whereas *Sg pasteurianus* was the most common subspecies.

In Paper III, the proportions of IE were considerably higher in bacteraemia with *Sg gallolyticus* and *S. infantarius* than with other SBSEC species and subspecies.

Papers IV and V developed methods for identifying SBSEC species and subspecies by MALDI-TOF MS using novel libraries.

However, the broader implications of these findings remain to be discussed, as do certain methodological considerations.

Strengths, limitations, and methodological considerations

Paper I

In this retrospective cohort study, we used matched controls from the general population. This may have affected the effect estimates due to surveillance bias, because patients in a hospital might undergo more exams than the general population—particularly when SBSEC is identified in blood cultures. However, recent studies in other geographic areas have used hospital populations as controls, whereas effect estimates are lacking compared with the general population (110-112).

Further, the study's goal was to increase the understanding of possible associations or causal relationships between the SBSEC bacteraemia and CRC over time, for which the general population is more suitable as a control population. In addition, the potential differences in the rates of other gastrointestinal cancers might be underestimated in a hospital-derived control group. Initially, we planned to recruit a matched control group with bacteraemia by *Escherichia coli*. However, this proved unfeasible due to the many local microbiological laboratories from which data were retrieved. In hindsight, it may have been prudent to have added a control group of blood culture-negative individuals and use propensity score matching or a similar method to adjust for unmeasured confounders by better reflecting the population with SBSEC bacteraemia.

Also, due to the well-known association between SBSEC bacteraemia and CRC, individuals with SBSEC bacteraemia are more likely to be examined by colonoscopy than other populations, which will be difficult to adjust for in a statistical analysis.

A major shortcoming of this study is its lack of proper identification of SBSEC species and subspecies. Because *Sg gallolyticus* is primarily associated with CRC following bacteraemia and because the distribution of SBSEC species and subspecies varies between regions, all effect measures in studies of SBSEC without proper microbiological identification are likely to be plagued by this unquantifiable error (9, 21, 55, 63). We chose to estimate effect measures for *S. gallolyticus* species versus other species, because MALDI-TOF MS can likely manage this distinction (70, 184, 185). Nevertheless, these estimates are also affected by the local distribution of *S. gallolyticus* subspecies. A longer follow-up with reliably identified species and subspecies would have been preferable but remains unfeasible with currently available data.

Another limitation concerns the time at which the cancer was diagnosed relative to bacteraemia. Because the study was register-based and centred on the date when cancer was confirmed or definitively diagnosed—rather than when it was suspected—there was bound to have been a delay from the point of suspicion to pathologically confirmed diagnosis. It is possible that although cancer could have been detected before the date of bacteraemia, the formal date of diagnosis could have been after bacteraemia. In this scenario, a diagnosis of CRC would not have been a consequence of a positive SBSEC blood culture in these cases.

Further, there are multiple risk factors for cancers and many potential confounders that affect which patients are examined for cancer. Although matching was done on factors commonly associated with cancer, such as gender and age, other factors, including alcohol intake and smoking, were rare in the National Patient Register, likely due to underreporting, because the register does not include primary care data, which is why we refrained from adjusting from these risk factors (83, 183). Also, we did not adjust for, nor did we have data on, socio-economic factors. Examples

of factors potentially associated with the development or diagnosis of CRC are illustrated in Figure 16.

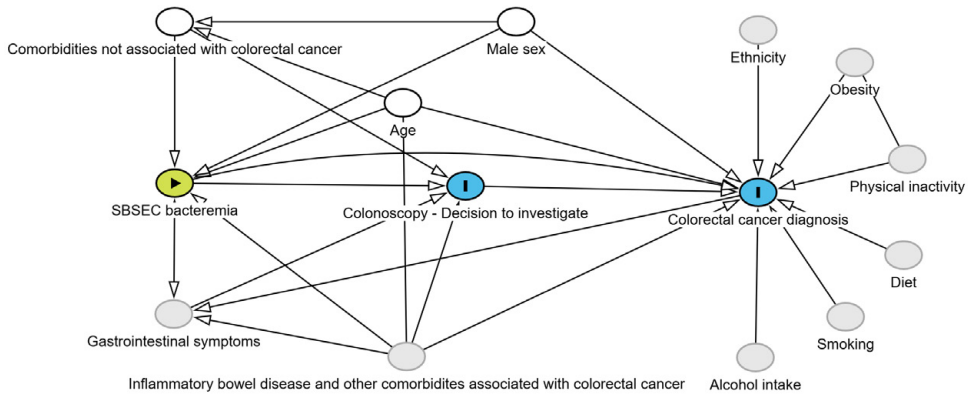


Figure 16. Factors that are potentially associated with the development or diagnosis of colorectal cancer or the decision to investigate by colonoscopy. Picture created using DAGitty (186). White covariates were adjusted for, whereas grey covariates were not.

Initially, we only adjusted the stratified analysis for Charlson comorbidity score. However, one referee raised a valid concern over the unbalanced groups regarding the distribution of comorbidities and suggested using propensity score matching or inverse probability of treatment weighting to adjust for this imbalance. We attempted to apply propensity score matching, but this method is prone to data loss due to the exclusion of individuals without a valid match. Instead, we tried inverse probability of treatment weighting, but we could only incorporate matching variables and Charlson comorbidity score into this model, because including more covariates adversely increased the bias from other covariates. Also, the extreme weights comprised all exposed individuals, and by excluding them, we lost a disproportionate number of exposed individuals who were later diagnosed with CRC.

Ultimately, we chose to use entropy balancing, allowing us to incorporate the comorbidities of the Charlson comorbidity index score as separate covariates, in addition to the matching variables. This step led to a balanced weight distribution without extreme weights and with no extreme weights in the exposed cohort, given the nature of the weighting method.

Detailed data on diagnosed CRC from SCRCR/CRCBaSe were lacking. Although less information on cTNM stage was missing—which was considered the most important data—more substantial portions on tumour grade were absent, as were follow-up data regarding local recurrences and distant metastases. Because the

proportions of missing data were similar between the exposed and control groups, we excluded these data from our statistical analyses. Alternative statistical strategies could have been used instead. One straightforward solution might have involved adding a category for the missing data in the statistical analysis. Another approach could have been mean substitution, replacing the missing data with the mean of the variables (187). A third option might have been multiple imputation, in which the missing data are replaced with predicted values, based on existing data from other variables (187).

A strength of this study was its nationwide population-based design, despite the inability or refusal of some laboratories to deliver data. This design led to one of the largest SBSEC cohorts to be assembled, with longer follow-ups than previous studies. Nevertheless, we were limited by the inclusion of fewer SBSEC episodes than expected, based on preliminary numbers on episodes of bacteraemia in Skåne. We speculate that SBSEC bacteraemia is more common in Skåne due to its large cattle population (17).

Another strength is the study's use of matched controls from the general population, because effect estimates compared with the general population have been lacking. Further, we provided detailed data on diagnosed cancers and follow-up data, dispelling preconceived notions of SBSEC-associated CRC being less advanced than in the general population, and supplied numbers on other gastrointestinal cancer diagnoses following infection.

Paper II

This retrospective study examined the distribution of various SBSEC species and subspecies in bacteraemia and their respective associations with CRC and other gastrointestinal cancers. Whereas Paper I evaluated the associations between SBSEC and CRC over time, Paper II investigated the findings of cancers following known bacteraemia and, consequently, studied the implications of deciding whether to assess patients with the various SBSEC subspecies for CRC and other gastrointestinal cancers.

The retrospective nature of this study limited the scope of information that was gathered. Although we initially collected data on gastrointestinal symptoms prior to bacteraemia, these data were soon deemed to be unreliable and were not examined further. A prospective design would have improved the reliability of the baseline information; however, such a design was impractical due to the low number of cases of SBSEC bacteraemia each year and the extensive follow-up that is required.

Although the study was population-based—limited to Skåne—all hospitals in Skåne were served by a single laboratory during the study period, limiting the number of episodes of bacteraemia that were included. The laboratory also has an established routine of storing bacterial isolates for at least 10 years—a practice that not all laboratories follow in Sweden.

Most isolates that were unavailable for culture were obtained at the start of the study period. However, we observed increasing rates of SBSEC bacteraemia over time, and most of the more recent isolates (stored <10 years) were available; thus, few isolates were unavailable for WGS. It is unknown whether this pattern reflects an actual increase in incidence or is attributed to more blood cultures being drawn, improved sensitivity in blood culturing, or older isolates being identified as streptococci without identification of group or species. Further, when a relapse in SBSEC occurred shortly after primary bacteraemia, the isolate from the relapse was not always stored and was thus presumed to be the same subspecies, because the SBSEC in blood cultures is relatively rare.

The study was also limited in the number of episodes, particularly of less common species and subspecies, which impeded the statistical analyses capabilities to adjust for confounders. We also assessed differences in CRC rates between SBSEC species and subspecies. Although we calculated standardised incidence ratios, the study would have benefited from inclusion of an additional group of individuals with infections by other bacteria less associated with CRC, such as *E. coli* and viridans streptococci.

Whereas Paper I was register-based and included a larger cohort, Paper II generated more detailed data and identified SBSEC species and subspecies by collecting data from medical records and isolates from the local laboratory. This study design also allowed the exact dates on which SBSEC was detected and when cancer was discovered to be determined, rather than the dates for reported pathological diagnosis. As a result, we drew more robust conclusions regarding which CRC was diagnosed due to a positive blood culture.

Further, differences in the proportions of colonoscopies that were performed were observed between species and subspecies. However, the treating clinician was likely blinded to specific SBSEC subspecies, because most bacteraemias occurred after the introduction of Bruker MALDI-TOF MS, which does not separate subspecies within the SBSEC using the supplied library, as noted by the manufacturer.

Although WGS likely reduced misclassifications by providing accurate identifications, it also resulted in the appearance of a subgroup of *S. gallolyticus* without definite subspecies, which likely would have been avoided by using an alternative method, such as MLST, or by combining multiple amplicon sequencing methods, because the taxonomic and clinical relevance of the unidentified isolates is unknown (46, 49, 55, 57, 61, 63-65).

Overall, at the time of writing, this study is one of the largest SBSEC studies, identifying species and subtypes according to the current taxonomy (52, 53, 55, 56, 63, 81, 82).

Paper III

This study determined the proportions of IE in bacteraemia between SBSEC species and subspecies and examined the clinical presentation of bacteraemia.

Like the other papers, this study was limited by its retrospective nature. The proportions that were determined by echocardiography varied between SBSEC members, wherein those that were most frequently diagnosed with definite IE were also examined more often. Nevertheless, relapses in bacteraemia were not observed more frequently among the subspecies that were examined less often by echocardiography. As in Paper II, the clinicians were likely unaware of the SBSEC species or subspecies and thus blinded de facto, and the uneven proportions of echocardiography were likely attributed to symptoms that were associated with IE.

The diagnosis of IE is based on criteria, and there is no clinical gold standard for such a diagnosis (141, 145). Because most SBSEC bacteraemias are likely to fulfil the criteria for possible IE—given that the SBSEC is regarded as a typical IE pathogen—we opted to focus solely on definite IE (136, 137, 140). Patients who are classified as having possible IE constitute a recurring problem in studies of risks and proportions of IE in bacteraemia, because there is no consensus on how to consider and manage this group of patients. To this end, one potential approach is to use another reference standard, such as the consensus of a multidisciplinary endocarditis team or of several physicians, or the physician's decision to treat as IE (149). Nevertheless, a preconceived risk of IE for a group of bacteria, such as the SBSEC, is likely to influence these decisions, prompting us to report numbers only on definite IE.

To further reduce the risk of misclassifications, all patients who were treated as IE but did not meet the criteria for definite IE—accompanied by inconclusive or ambiguous echocardiography results, medical charts, and/or imaging—were reviewed by experienced clinicians who were blinded to SBSEC species and subspecies. Also, the treating physicians were likely blinded to SBSEC species and subspecies, and subspecies misclassifications were mitigated through identification by WGS.

Another limitation of the study was that only relapses with growth of the SBSEC in cultures were identified. Although a few suspected clinical relapses and undiagnosed cases of IE were listed in the manuscript, more clinical relapses could have been missed during review of the medical records, as this was not done systematically in this regard.

The study was further limited by its descriptive nature and the relatively few episodes that were included, despite it being one of the larger cohorts of SBSEC subspecies, particularly for *S. lutetiensis* and *S. infantarius* (52, 53, 55, 56, 63, 81, 82).

Papers IV & V

In Papers IV and V, we evaluated the performance of Bruker's current MALDI-TOF MS libraries in identifying SBSEC species and subspecies and created in-house libraries to improve this identification.

In Paper IV, we cultured the isolates in BHI broth and utilised ethanol-formic acid extraction to create the library and evaluation set. Certain SBSEC isolates tend to grow poorly on blood agar plates, and culturing them in broth improved their bacterial yield, making extraction easier. However, this practice does not reflect MALDI-TOF MS routines in clinical laboratories, in which bacteria are usually cultured on blood agar plates under aerobic conditions and analysed using the direct transfer method or by on-target extraction with formic acid. Previous studies have shown that culture conditions and extraction methods may affect the performance of MALDI-TOF MS in identifying bacteria (70, 188). In hindsight, applying culture and sampling methods that are used in clinical routine would have been advisable.

Thus, we deemed it necessary to create the new libraries in Paper V. Nevertheless, this endeavour was an opportunity to remove bacterial isolates or MSPs that were likely to cause misidentifications. We elected to omit the isolates that caused misidentifications in the lists of top 10 hits in Paper IV, but an alternative approach could have been to select isolates based on clustering in the dendrogram. Other authors have previously suggested this type of refinement of reference library data (189). However, it could be argued that this may be selectively biasing, causing artificially inflated results, despite other culture and sampling conditions being used. Testing the libraries from Paper V on another evaluation set could confirm the findings. Dividing our SBSEC culture collection into three parts from the outset—using the first two to create and evaluate the library and the third for validation—would have been a more sound approach.

Further, we opted to use the manufacturer's established cutoff identification score of ≥2.0, which the manufacturer has deemed to be reliable for identification to the species level when using the supplied commercial libraries. We also used duplicate sampling to report the best match (the sample with the highest score), as is established in clinical routine and research (70). However, alternative criteria have been proposed, such as an additional criterion of a difference in score of >0.2 or 10% to the second species (190, 191).

An important limitation of these studies is the absence of a recognised gold-standard method for identifying SBSEC subspecies. Although identification through short-read WGS, based on reference-based mapping of SNPs, is likely to be more reliable than other methods that have been used to identify SBSEC subspecies, this strategy has not been applied previously to the identification of SBSEC species and subspecies. An alternative approach was also tested, using Kraken for subspecies identification (192). It was apparent that the accuracy of Kraken with regard to identifying SBSEC subspecies was poor, likely due to database errors or limitations,

resulting in numerous misidentifications; thus, it was not used further. At the time that the libraries were developed, public genome references were not available for all subspecies.

Moreover, in Paper II, we identified a group of isolates that formed a separate cluster among the *S. gallolyticus* subspecies. At the time, they could not be assigned to a specific subspecies, and it is unknown whether they constitute a new *S. gallolyticus* subspecies or belong to *Sg gallolyticus*. They were not included in the published paper, but in an internal evaluation, they were identified as *Sg gallolyticus* using the new libraries. Also, the potentially new SBSEC species *S. vicugnae* and *S. ruminicola* were not included in the libraries, but at the time of writing, infections in humans have not been reported (47, 48)

Further, *Sg macedonicus*, *S. equinus*, and *S. alactolyticus* infections are rare in humans. Most of these isolates were obtained from culture collections, limiting the number of isolates that were included in the libraries and evaluations. These bacteria were rarely derived from human infections, and it could be speculated that they may differ from those that infect humans. Thus, the performance of the libraries in identifying these rare pathogens is somewhat uncertain but likely to have limited consequences in blood cultures from human infections. We also noted that several isolates from culture collections were previously misidentified (unpublished results), further limiting the scarce supply of these rare pathogens.

Another potential limitation is that the various species and subspecies were not completely separated in the dendrograms in Papers IV and V, indicating that high-quality spectra are needed to make reliable identifications.

None of the new libraries has been validated externally. To evaluate the external validity of the results, their performance should be assessed on other MALDI-TOF MS instruments in external laboratories and on SBSEC isolates from different geographic regions. Developing similar libraries for MALD-TOF MS systems by other manufacturers would be prudent.

Comparisons with other studies

Incidence and subspecies distribution

There are little data on the overall incidence of SBSEC bacteraemia. Corredoira et al. studied such rates from 2005–2016 in the provinces of Galicia, Spain (17), reporting an overall incidence of 2.0 per 100,000 inhabitants in Galicia, peaking in Lugo at 4.6–10.9 and bottoming in Orense at 0–0.6. They also noted a correlation between incidence and cattle density between provinces. We found an overall incidence of 2.0 (range 1.35–2.94) per 100,000 inhabitants in Paper II (21).

In Paper I, we observed that the Stockholm region had notably fewer episodes of SBSEC bacteraemia than Skåne (151 vs. 246) despite having nearly twice the population of Skåne (193). Skåne is known for agricultural production—both crops and cattle—and its cattle density is higher than in Stockholm (194). Previous studies in Spain and France have published similar results, reporting a higher incidence of SBSEC IE in rural compared with urban areas (77, 79). However, because Stockholm is served by several microbiological laboratories, which changed during the study period, we refrained from publishing incidence numbers from the various regions, because the quality of data was difficult to evaluate.

As discussed, there are a few studies on the distribution of the subspecies in SBSEC bacteraemia using the current taxonomy, although most are limited in size (52, 53, 55, 56, 63, 81, 82). *Sg gallolyticus* is often described as the predominant subspecies in Europe and North America, but exceptions exist (9). A study of 53 episodes by Marmolin et al. that was conducted in Denmark—a setting that presumably has similar demographics and lies near Skåne—showed *Sg gallolyticus* to predominate, contributing to 36% of episodes, followed by *Sg pasteurianus* (23%) and *S. lutetiensis* (23%) (63). In contrast, in our population-based cohort in Paper II, *Sg gallolyticus* constituted 25% of episodes and was the second most common subspecies (21). This rate could have been higher if the unidentified subspecies had been determined to be *Sg gallolyticus*, which is likely, had an amplicon-based method been used instead of WGS.

In contrast, *Sg pasteurianus* is often the dominant subspecies in Asian studies on SBSEC bacteraemia. The largest cohort was recruited by Sheng et al., comprising 172 episodes, 73% of which were Sg pasteurianus, followed by *Sg gallolyticus* (18%) (49). *Sg pasteurianus* also predominated in Paper II, contributing to 37% of episodes.

Previous studies have reported proportions of 2–29% for *S. lutetiensis* and 0–13% for *S. infantarius*, whereas we observed rates of 20% and 14%, respectively (21, 52, 53, 55, 56, 63, 81, 82). However, the results of several of these studies could be affected by the limited ability of 16S rRNA sequencing to discriminate between these two species (49).

The distribution of SBSEC species and subspecies may have been skewed in earlier studies, because several of them were likely not population-based and were possibly from tertiary centres to which cases of IE could have been selected. *Sg gallolyticus* is the most likely to cause IE, perhaps effecting the selection of bacteraemia due to *Sg gallolyticus*, explaining the higher proportions of this subspecies in some studies (9). Moreover, if not all species are systematically identified to the subspecies level, bacteraemia episodes with IE could have their subspecies identified to a higher degree than episodes without IE, resulting in a type of detection bias.

Colorectal cancer

Few studies have compared CRC rates following SBSEC bacteraemia with those of the general population, as we did in Paper I. Kwong et al. contrasted such rates for 662 SBSEC bacteraemia episodes with those of a hospital-derived population with negative blood cultures in Hong Kong, using propensity score matching that incorporated age, sex, and comorbidity index score (110). They found an HR of 3.9 (CI95% 2.3–6.4) for SBSEC overall and 5.7 (CI95% 2.2–15.1) for *Sg gallolyticus*. However, the study was registry-based, and the microbiological methods were not disclosed, raising doubts about the identification of *Sg gallolyticus* and the reliability of these rates. Moreover, they found an inflection point of normalised CRC rates for SBSEC after 374 days and noted that SBSEC was primarily associated with early-stage CRC (stages 1 and 2).

Another registry-based study by Laupland et al. in Australia examined the rates of CRC within 1 year of SBSEC bacteraemia compared with other unspecified bacteraemias, finding a relative risk of 4.4 (CI95% 2.7–6.8) (112). Despite this study design and its omission of detailed microbiological methods, they reported such ratios for individual SBSEC subspecies: 4.5 (CI95% 1.7–9.9) for *Sg gallolyticus*, 4.7 (CI95% 1.3–12.0) for *Sg pasteurianus,* and 10.6 (CI95% 2.9–27.3) for *S. lutetiensis*.

Justesen et al. conducted a registry-based study in Denmark, also using blood-culture-negative episodes as controls, matched for age and sex (111). They included 117 SBSEC episodes and reported HRs of 8.5 (CI95% 3.5–20.4) within 1 year and 4.4 overall (CI95% 2.0–9.7), with a median number of days to a diagnosis of CRC of 97.5.

A meta-analysis, comprising Kwong et al., Laupland et al., and three other studies with various control populations, found an overall HR of CRC of 3.7 (CI95% 2.8–5.0) (116).

In Paper I, we noted higher CRC rates overall (HR 10.3 in the stratified analysis and 6.7 with the adjusted method using entropy balancing) and during the first year (19.8 and 23.4, respectively). We used controls from the general population, whereas the three registry studies above chose hospital-derived control populations. These control populations are likely to have different risk factors for CRC, as well as differences in expected life spans, and may be more or less likely to undergo colonoscopy. When we applied entropy balancing to adjust for comorbidities, the HRs became more similar to those in other studies. The primary goal of Paper I was to study CRC rates in SBSEC bacteraemia compared with the general population, because the few existing cohort studies on this topic have primarily examined rates versus hospital populations. Further, the rates of CRC and the distribution of SBSEC subspecies may differ between geographic regions and populations (21, 55).

We also observed increased rates of CRC during follow-up of over 2 years, which has been reported in other studies (110-112). However, this rise was slight and may not have been detected in a smaller sample size. Corredoira et al. presented a case series of 109 episodes of SBSEC IE (102 with *Sg gallolyticus*), of which 70 were diagnosed with colorectal neoplasia (8 were CRC), in conjunction with the acute episode (122). During follow-up, 43 episodes were diagnosed with colorectal neoplasia (12 had previously had a normal colonoscopy, and 31 had previous colorectal neoplasia), 5 of which were CRC, with a mean duration of 60.6 months following the primary infection. Although no control group was included and although the study group consisted of older individuals, in whom colorectal adenomas are likely to occur, the study authors concluded that colonoscopy should be repeated following SBSEC bacteraemia. Regardless, their findings highlight the need for long follow-up times when studying carcinogenesis.

In Paper I, the CRCs diagnosed after SBSEC bacteraemia were not less advanced than those of the general population, in contrast to the findings in Kwong et al. and as has been reported elsewhere (9, 110). Early colorectal neoplasms correlate primarily with *Sg gallolyticus*, and we did not determine the distribution of SBSEC subspecies in Paper I. The subspecies distribution may have differed in the study by Kwong et al., in which a *Sg gallolyticus* constituted a greater proportion. However, previous Asian studies have reported higher percentages of *Sg pasteurianus* than *Sg gallolyticus* (55, 139). Although we reported higher rates of colorectal adenomas in SBSEC compared with controls in Paper I, this pattern could be attributed to surveillance bias, because more colonoscopies were conducted among SBSEC patients, likely due to its established association with CRC.

Further, we reported rates of CRC for the *S. gallolyticus* and other SBSEC species at the species level in Paper I, but we chose not to examine their various subspecies due to the unreliability of MALDI-TOF MS in distinguishing SBSEC subspecies. The other registry-based studies did not acknowledge this severe methodological limitation, and their reported rates per subspecies could be unreliable (110, 112).

In Paper II, we presented the results of CRC following bacteraemia with the various SBSEC species and subspecies, which has been reported using reliable microbial methods but previously only in limited cohorts. Further, several such studies have omitted whether CRC was diagnosed before or after bacteraemia—instead presenting overall CRC numbers.

The largest study on *Sg gallolyticus* bacteraemia (at the time of writing) was conducted by Corredoira et al., who observed CRC overall in 25 of 257 bacteraemia episodes (9.7%) (195) but did not state whether these cases were diagnosed before or after the infection. In our study, Paper II, the overall rate of CRC among cases of *Sg gallolyticus* bacteraemia was 25% (13/53), of which nine patients were diagnosed after bacteraemia.

Sheng et al. presented the largest study of *Sg pasteurianus,* noting CRC overall in 19 of 126 cases (15%), but they did not distinguish how many CRCs were known before and diagnosed after presenting with bacteraemia (55). In our cohort in Paper II, CRC was diagnosed overall in 15 of 83 cases of *Sg pasteurianus* (18%), three of whom were diagnosed after bacteraemia.

Another study by *Corredoira* et al. presented a cohort of *S. infantarius* episodes according to a previous taxonomy, including *S. infantarius* ssp. *infantarius* and *S. lutetiensis,* in which 6 of 58 episodes (10%) had CRC, but all were diagnosed prior to known infection, although some were diagnosed simultaneously as blood cultures were drawn (128). They did not report any case of CRC that was diagnosed following a confirmed infection. These values match the rates of CRC in our cohort in Paper II—14% for *S. lutetiensis* and 10% for *S. infantarius*—with only one CRC diagnosed during follow-up. CRC rates overall are summarised in the table below. The findings of these studies are summarised in Table 14.

Table 14. CRC associated with the most common SBSEC subspecies – overall proportions (52, 55, 56, 63, 81, 195).
Kaiki et al. presented CRC diagnoses following bacteraemia, not overall proportions.

	Sg gallolyticus	***Sg pasteurianus***	***S. lutetiensis***	***S. infantarius***
Corredoira (2017)	25/257 (10%)	N/A	N/A	N/A
Sheng (2014)	5/31 (16%)	19/126 (15%)	N/A	N/A
Marmolin (2016)	3/19 (16%)	0/12 (0%)	0/12 (0%)	0/7 (0%)
Romero (2011)	1/14 (7%)	0/24 (0%)	0/5 (0%)	0/2 (0%)
Ben-Chetrit (2016)	2/6 (33%)	4/26 (15%)	0/6 (0%)	0/0 (0%)
Kaiki (2021)	3/5 (60%)	4/29 (14%)	1/4 (25%)	0/1
Total	39/372 (10%)	27/205 (13%)	1/27 (4%)	0/10 (0%)

Other gastrointestinal cancers

Few studies have presented rates of non-CRC gastrointestinal cancers following SBSEC bacteraemia. Sheng et al. reported more non-CRC gastrointestinal cancers among bacteraemias with *Sg pasteurianus* versus *Sg gallolyticus* (32 of 126 [25%] vs. 2 of 31 [6%]; p=0.04), but as discussed, they failed to note how many followed bacteraemia (55). There were no differences between the individual non-CRC gastrointestinal cancers. Kale et al. presented a series of *S. gallolyticus* species in blood and ascites cultures from 68 patients, most of which were *Sg pasteurianus* (exact number not disclosed), and observed HCC in 5 of 68 episodes (7%) (139).

They did not differentiate between prevalent and incident cancer but presented overall numbers. Moreover, they used the phenotypic identification method VITEK II, which is not considered sufficient for determining SBSEC subspecies (49). In Paper I, we found that oesophageal, gastric, hepatic, and pancreatic cancers were more prevalent among individuals with SBSEC bacteraemia than the general population before bacteraemia, but no other gastrointestinal cancers other than CRC were diagnosed more often during follow-up. In Paper II, only 1 patient was diagnosed with a non-CRC gastrointestinal cancer during follow-up, but previously diagnosed pancreatic cancer was more prevalent among patients with *S. lutetiensis* bacteraemia compared with the other SBSEC species and subspecies (21).

Infective endocarditis

As discussed, the lack of a clinical gold standard for diagnosing of IE complicates its study. Most studies have applied established criteria, such as Duke's criteria, modified Duke's criteria, and, more recently, ESC2015, ESC2023, and Duke-ISCVID2023 (141, 142, 145-147). Whereas infections that fulfil the definitions of definite IE are subject to less debate, the category of possible IE constitutes a major challenge in research and the clinic.

Several studies have examined the rates of IE in SBSEC bacteraemia, but few have used the current taxonomy. Further, many of these studies are plagued by uncertainties in methodology, such as poor definition of the IE that is assessed and uncertain or unreliable methods for identifying subspecies (55, 63, 124, 158).

In Paper III, the proportion of IE in *Sg gallolyticus* bacteraemia was 33% (CI95% 21–47%), compared with 79% of 109 episodes in Corredoira et al., 52% of 31 episodes in Sheng et al., and 53% of 19 episodes in Marmolin et al. (55, 63, 124). A review that included some of these studies found a 64% rate of IE in *Sg gallolyticus* bacteraemia (158).

The proportion of IE in *Sg pasteurianus* bacteraemia in Paper III was 5% (CI95% 2–12%), versus 13% of 126 episodes in Sheng et al., 8% of 106 episodes in Nasomsong et al., and 33% of 12 episodes in Marmolin et al. (55, 63, 196). In the review above, 18% of cases of *Sg pasteurianus* bacteraemia were IE (158).

Fewer bacteraemia episodes by *S. lutetiensis* and *S. infantarius* have been described previously. Our proportions of IE were 5% by *S. lutetiensis* (CI95% 1–17%) and 16% by *S. infantarius* (95% CI 7–34%), whereas Marmolin et al. reported a 17% rate of *S. lutetiensis* IE in 12 episodes and a 14% rate of *S. infantarius* IE in 7 cases; Corredoira noted rates of 7% and 60% in 96 in 10 episodes, respectively (63, 158).

Although the proportion of IE has been consistently higher in bacteraemia by *Sg gallolyticus* compared with *Sg pasteurianus* and *S. lutetiensis* in most published studies, we found lower proportions of IE for all SBSEC (sub)species (55, 63, 124, 158, 196). Among several explanations for this pattern, our cohort was population-

based, whereas the cohorts in other studies might have been derived from tertiary health care centres, potentially selecting for complicated infections, such as IE (22, 55, 196). Further, unless the SBSEC species and subspecies are systematically identified in all SBSEC bacteremia, it could be speculated that the laboratories might have opted to identify species and subspecies in bacteraemias with IE, increasing the rates of IE in retrospective analyses of such episodes.

Another possibility is misclassification of SBSEC (sub)species due to insufficient accuracy of bacterial identification, as perhaps was the case in Namsong et al., who used the VITEK 2 automated system (196). Many studies have failed to define whether only definite IE was included, because the proportions of IE would have been higher had possible IE been included (55, 63, 124, 196). These issues are likely present in the review above, potentially overestimating rates of IE, whereas our values may be underestimated due to the application of more rigorous definitions of IE (22, 158). Another explanation is differences in demographics with regard to other comorbidities, because the studies were performed in disparate geographic areas.

Identification of SBSEC species and subspecies by MALDI-TOF MS

Few studies have evaluated the performance of MALDI-TOF MS systems in identifying SBSEC species and subspecies. Agergaard et al. assessed the performance of the Bruker Biotyper and bioMérieux VITEK MS using direct transfer sampling, on-target extraction, and ethanol-formic acid extraction on isolates that were grown under aerobic and anaerobic conditions on horse blood and chocolate agar, respectively (70). A total of 66 clinical SBSEC isolates from blood cultures were included, representing every species and subspecies except *S. alactolyticus.* This group sequenced 16S rRNA and the ITS region as its reference method but also noted the limitation of a lack of an established reference method for identifying SBSEC species and subspecies. Generally, direct transfer performed worse than on-target and ethanol-formic acid extraction. However, neither system could reliably identify subspecies, yielding rates of correct identification from 4.5–62.1%. The authors observed that although both systems are useful for identifying SBSEC species, only *S. lutetiensis* was reliably identified without misidentifications. These results are consistent with our evaluation of the Bruker Biotyper in Papers IV and V.

Hinse et al. used the AXIMA Confidence MALDI-TOF MS instrument to compare taxonomic resolution with sequencing of the *SodA* gene, on 88 SBSEC isolates from culture collections (62). When culturing in BHI broth, they could not generate a dendrogram that was consistent with that based on *SodA* sequencing—a similar limitation as we noted in Paper IV. However, a sufficient dendrogram was constructed when the isolates were cultured on sheep blood agar. All SBSEC members could be discriminated by MALDI-TOF MS in the dendrogram, except *Sg pasteurianus* and *Sg gallolyticus,* which did not separate completely but were

grouped into sub-clusters—a tendency that could be seen in the dendrograms of our novel libraries in Papers IV and V. The group also noted that several SBSEC isolates in culture collections were misidentified. They concluded that MALDI-TOF MS could potentially be used to identify SBSEC species and subspecies but did not evaluate the actual performance of any library that was associated with the system.

As discussed, alternative methods for bacterial species identification such as amplicon sequencing can be used to determine SBSEC species and subspecies (49, 62). However, compared with MALDI-TOF MS, these approaches are more time-consuming, expensive, and often unavailable in clinical microbiology laboratories (49, 66).

Conclusions

Although the incidence of SBSEC bacteraemia appears to be rising, the small annual number of cases per subspecies renders high-quality studies challenging to perform. Moreover, SBSEC studies should be conducted at the subspecies level, but the lack of identification of SBSEC members in clinical routine impedes retrospective studies from being conducted. Further, Papers I–III were limited by their retrospective nature. It has become increasingly clear that SBSEC species and subspecies are distinct entities, and objections to the methodology will always be raised when the SBSEC is studied on a group level. Ultimately, due to the low incidence of SBSEC bacteraemia, all contributions to our knowledge of the SBSEC are worthwhile, but proper subspecies identification remains the cornerstone for further studies.

Association with colorectal and other gastrointestinal cancer: Carcinogenic or not?

In Paper I, we studied the association between SBSEC bacteraemia and CRC over time. This endeavour was challenging, because the life expectancy following SBSEC bacteraemia is limited due to the high median age and numerous comorbidities in such patients, whereas carcinogenesis is a slow process, taking several years. It is evident that CRC is diagnosed early following infection, with most cases diagnosed within the first year (110, 111). Our study also found increasing rates of CRC during long-term follow-up, whereas few other studies of larger SBSEC cohorts with long follow-up have been conducted. However, most CRCs that were diagnosed after the first year had not undergone a previous colonoscopy, several of which could have been overlooked CRCs rather than those that developed during follow-up.

Other groups have recommended regular colonoscopies following *Sg gallolyticus* bacteraemia (9, 121, 122). While data on SBSEC members at the group level cannot

be translated directly to individual subspecies, there are little data to support this practice, since the recommendation is based on a limited case series without population controls or systematically conducted colonoscopies (122). One must also consider whether colonoscopies are performed solely to detect CRC or whether it should be used for precancerous lesions, such as colorectal adenomas. Is diagnosis and excision of precancerous lesions necessary in an ageing population that may never develop symptomatic cancer?

In Paper II, we confirmed the association between *Sg gallolyticus* and a diagnosis of CRC following bacteraemia. Whereas the standardised incidence ratio of CRC was elevated for the other SBSEC subspecies compared with the general population in Skåne, few CRCs were diagnosed in a group that is more likely to be examined by colonoscopy than the general population, introducing possible surveillance bias. Few studies have been conducted on all SBSEC subspecies using the current taxonomy, and a recurring problem is that most do not distinguish between CRC that is diagnosed before and after bacteraemia. Our results do not clearly support systematic colonoscopy following bacteraemia with other SBSEC species and subspecies. However, a matched control group of non-SBSEC controls was not included.

There is undoubtedly a link between the SBSEC in blood cultures—specifically *Sg gallolyticus*—and a subsequent diagnosis of CRC. There is also increasing in vitro evidence that *Sg gallolyticus* has carcinogenic properties (125-128, 189). However, no sufficient clinical study has shown definitive causation between the SBSEC or *Sg gallolyticus* and the development of CRC. It remains unknown whether the SBSEC, and specifically *Sg gallolyticus,* is merely a passenger or also a driver in the driver-passenger hypothesis. Further, the association between the SBSEC and CRC has been based primarily on blood culture results, which are sufficient to determine whether these patients should undergo colonoscopy based on blood culture findings. However, a more ideal examination of the causal relationship between the SBSEC and CRC would be based on faecal carriage. Yet, studying SBSEC faecal carriage presents challenges, and prospective studies on the development of CRC would need to encompass a large number of individuals who are followed over extended times.

It has been suggested that workups for other gastrointestinal cancers should be conducted following SBSEC bacteraemia. Although all gastrointestinal cancers were common among the SBSEC bacteraemia cases in Papers I and II, they were known before the bacteraemia. There were no differences in the diagnoses of other gastrointestinal cancers after bacteraemia. If other gastrointestinal cancers are more frequent after SBSEC bacteraemia, at the group or subspecies level, they are likely to have limited clinical relevance.

Infective endocarditis?

There is ample evidence that the risk of IE in SBSEC bacteraemia is greatest for *Sg gallolyticus* (9, 22, 63, 158). Although most groups would likely agree that the risk is lower for other species and subspecies, there is no consensus on how much lower that risk is.

All SBSEC studies that are conducted on the subspecies level are limited by the low number of bacteraemia cases. Further, many are not population-based. If performed in tertiary centres, there is likely to be referral bias of patients with IE, which could yield inflated proportions of patients who are diagnosed with IE.

In addition, all research on IE is limited by the lack of a gold standard, and clinical criteria are used instead (141, 145-147). Whereas cases that fulfil the criteria for "definite IE" are undisputed, the category of "possible IE" represent a major challenge in research studies and clinical settings. Because the SBSEC is considered a typical IE pathogen regardless of species and subspecies, most patients will be classified into "possible IE" by fulfilling the microbiological criteria, in addition to the minor criteria regarding fever. Thus, in comparing the risks of IE in SBSEC bacteraemia, only comparison of the rates of "definite IE" has clinical relevance.

Identification of SBSEC species and subspecies

There are several methods for identifying SBSEC species and subspecies, each with advantages and drawbacks. Whereas WGS is likely the most reliable approach, it is currently not feasible for clinical use due to a lack of availability, high costs, and long processing times. Several amplicon methods can identify SBSEC subspecies sufficiently, of which 16S rRNA is the most widely available and capable of distinguishing subspecies to a level that is acceptable for clinical use.

MALDI-TOF MS is now the most widely used method for routine bacterial identification in clinical microbiology laboratories using modern methods, given its speed, cost-effectiveness, and ease of use. Our new libraries demonstrate that MALDI-TOF MS can be used to identify SBSEC species and subspecies. It would be preferable to replicate these results by internal and external validation, preferably on isolates from a different geographic region. The new libraries would also have to be adapted to other manufacturers' MALDI-TOF MS systems.

The question remains whether the unidentified *S. gallolyticus* group is a novel subspecies or whether it belongs to the other *S. gallolyticus* subspecies. Although constituting few episodes, of which several were relapses, this group had clinically similar properties as *Sg gallolyticus* with high rates of IE and was also identified as *Sg gallolyticus* using the novel libraries in Papers IV and V. A potential misidentification as *Sg gallolyticus* is unlikely to cause patients harm by a missed diagnosis of IE or CRC, because bacteraemia by *Sg gallolyticus* is likely to be examined for both. Other potential new SBSEC species have been suggested

recently; they were not included in the novel libraries, but such infections in humans have not yet been reported (47, 48).

Clinical implications

When should colorectal and other gastrointestinal cancers be screened for?

The number needed to screen (NNS) to diagnose 1 CRC in Paper II was 15 for SBSEC bacteraemia overall and 5 for *Sg gallolyticus* bacteraemia. The NNS for such a diagnosis in patients with gastrointestinal symptoms varies across studies: two meta-analyses reported values that correspond to an NNS of 12 and 17 (197, 198). A large study in which colonoscopies were conducted on asymptomatic patients aged over 50 years, almost only men, detected CRCs at a rate that corresponded to an NNS of 100 (199).

Although the exact percentages vary by studies, the proportions of patients who are diagnosed with CRC following bacteraemia by *Sg gallolyticus* warrant CRC screening following the infection, and it should be mandatory if the patient is eligible for CRC treatment. One should also consider that the most pronounced association for bacteraemia by *Sg gallolyticus* is with early CRC and precancerous lesions, rendering invasive examinations, such as colonoscopies, questionable in the most elderly population, a group with many comorbidities. Further, although the correlation between *Sg gallolyticus* bacteraemia and a subsequent diagnosis of CRC increases with age, placing older patients at greater risk of CRC, a missed diagnosis of CRC could have greater consequences for younger patients.

It remains unknown whether colonoscopies should be repeated regularly, and little data support this practice following *Sg gallolyticus* bacteraemia unless warranted by findings at the first colonoscopy. Nevertheless, repeated colonoscopy is recommended in some international guidelines (200).

Whether colonoscopy should be performed following bacteraemia that is caused by SBSEC species and subspecies other than *Sg gallolyticus* is unknown. Although the frequency of CRC appears to be elevated, this trend may indicate surveillance bias due to a greater likelihood of undergoing colonoscopy. Our findings in Paper II do not support mandatory colonoscopies. However, it would be prudent to consider a colonoscopy if no source of infection is identified. Colonoscopy should also be strongly considered if the SBSEC subspecies is not identified.

With regard to other gastrointestinal cancers, there is no evidence that supports their screening unless symptoms present otherwise.

When should echocardiography be conducted?

The claim that examination for IE by TTE or TOE should be mandatory in bacteraemia in *Sg gallolyticus* without an identified source of infection is

undisputed, given that most clinicians agree that the risk of IE is high. However, there is no clear answer for bacteraemia with other subspecies, and there are no established criteria or risk levels that dictate when routine investigation for IE can be omitted.

Our findings and those of other groups indicate that the risk of IE is likely to be relatively high in *S. infantarius* bacteraemia, particularly when the patient has a prosthetic valve (22, 63, 158); thus, routine investigation for IE should be considered in such a case.

The risk of IE is lower in bacteraemia with *S. lutetiensis* and *Sg pasteurianus,* but the proportion of IE vary between studies. If there is low clinical suspicion of IE in bacteraemia with these pathogens, TTE may be sufficient to rule out IE and could possibly be omitted entirely in the setting of an acute-onset intra-abdominal infection. The remaining SBSEC species and subspecies are so rare that it is impossible to determine the risks of IE in bacteraemia with such members.

In addition to the subspecies, a long duration of symptoms (>1 week), monomicrobial bacteraemia, patient risk factors for IE, and the lack of an identified source of infection should prompt an investigation for IE, as for other non-β-haemolytic streptococci (201).

How should SBSEC species and subspecies be identified?

The novel MALDI-TOF MS libraries that were constructed in Papers IV and V provided sufficient accuracy in identifying SBSEC species and subspecies. The blood agar aerobic library SBSEC-CMRS-BAE can be incorporated directly into clinical routine using the direct transfer method on bacterial isolates that are cultured on blood agar plates under aerobic conditions, requiring minimal additional effort and cost. It should be noted that the novel libraries are not CE IVD-certified in the EU nor FDA-cleared in the USA, and therefor considered in-house developed tests and subjected to local regulations. Also, none of the libraries has been evaluated directly on positive blood culture bottles with regard to performance (202).

Current MALDI-TOF MS libraries are insufficient for identifying SBSEC subspecies. Nevertheless, the Bruker MBT Compass library can distinguish *S. gallolyticus* on a species level from other SBSEC species. This property may be beneficial, because non-*S. gallolyticus* SBSEC bacteraemia has a lower indication for colonoscopy, and the risk of IE could be lower than in bacteraemia with the *S. gallolyticus* species.

Other alternatives for identifying SBSEC species and subspecies are WGS and amplicon sequencing-based methods, whereas 16s rRNA is the most widely used technique. Although the latter has certain limitations in identifying SBSEC subspecies, it remains a viable option for clinical use.

Ethical considerations

Definitely answering the questions of when to conduct a work-up for CRC by colonoscopy or for IE by TOE would require randomised controlled trials or the examination of all patients with SBSEC bacteraemia and controls. It is simply not feasible or ethical to submit patients to unnecessary invasive examinations or to omit tests that would typically be performed. In addition, although CRC is common in the general population, large control groups would still be required to achieve sufficient statistical power.

Moreover, whereas screening for CRC is less controversial or perhaps even uncontroversial, which patients should be screened for potential precancerous tumours, such as colorectal adenomas? At what prevalence and for how many remaining years of life is screening meaningful? Perhaps more importantly, at what age should screening begin?

In addition, at what risk and proportion of IE in bacteraemia should an investigation for endocarditis be conducted? What is the acceptable NNS? These levels have not been established.

Another consideration is the performance of a microbiological method. In the clinical setting, we accept certain failures and missed diagnoses when treating and examining patients, but laboratory methods are often held to much higher standards.

Future research

Future studies must be larger per SBSEC species and subspecies and require longer follow-ups in the context of gastrointestinal cancer, necessitating population-based studies, likely conducted on national level, for long periods.

- It has not been fully established whether colonoscopy can be omitted in bacteraemia with a non-*Sg gallolyticus* SBSEC species or subspecies. Further studies must be conducted at the SBSEC subspecies level, and suitable control groups must be selected carefully.
- The need for performing colonoscopy at regular intervals after bacteraemia with *Sg gallolyticus* has not been established or studied systematically.
- Whether *Sg gallolyticus* is carcinogenic or merely a marker for established CRC remains undetermined in epidemiological studies on bacteraemia and faecal carriage. Research on *Sg gallolyticus* faecal carriage with long-term follow-up is needed.
- The risk of IE for various SBSEC species and subspecies must be assessed in larger studies.

- Because bacteraemia due to individual SBSEC species and subspecies is rare, meta-analyses of population-based studies of IE and CRC in various regions are likely needed, and microbiological methods for subspecies identification must meet acceptable standards.
- External validation of our novel MALDI-TOF MS libraries would be beneficial.
- Can our novel MALDI-TOF MS libraries be applied directly to positive blood culture bottles?
- Are there subsets of *Sg gallolyticus* that possess oncogenic properties, and can these subsets of bacteria be identified by WGS or MALDI-TOF MS?
- Are there more SBSEC species and subspecies waiting to be discovered?

Acknowledgements

Min huvudhandledare **Malin Inghammar** och min bihandledare **Bo Nilson.** Tack för att ni alltid tagit er tid för att lära ut både laborativ och statistik metodik som ofta känts övermäktig, men också för att ni lika ofta har fått mig att inse att det ofta inte endast finns ett rätt svar. Tack för all tid ni har ägnat åt mig och våra alster – jag har ofta undrat hur ni hunnit med era ”riktiga” jobb. Jag har känt mig privilegierad över att ha er som handledare.

Mina bihandledare **Pamela Buchwald** och **Anton Nilsson**. Tack för vägledning och expertis inom era respektive ämnesområden. Tack för rak feedback och orädsla för att tycka till – även om det som varit mindre bra. Med facit i hand borde ni ha varit med redan från början, då hade det blivit ännu bättre!

Alla underbara **kollegor på Infektionsenheten** på Helsingborgs lasarett, och särskilt min nuvarande chef **Ulrika Wåhlin** och före detta chef **Cecilia Löfgren**. Tack för att ni värdesätter och möjliggör forskningen på kliniken även om det ibland medför ökad belastning i det kliniska arbetet för er andra. Tack för att ni alla alltid ställer upp då det kniper – både i forskningen och kliniken, men framför allt i livet i övrigt. Jag uppskattar er enormt mycket.

Mina vänner och kollegor **Jonas Tverring** och **Oskar Ljungquist**. Tack för att ni svarat på otaliga frågor om allt från vägledning i praktiska aspekter av doktorandutbildningen och forskningsanslag till statistikprogramvara (dags för R?), och för möjligheten att ventilera både mindre och större frustrationer.

Tack till **Magnus Rasmussen** som gav upphov till idén för arbetet och som introducerade mig för Malin. Tack för att du alltid är ödmjuk, inkluderande och uppmuntrande.

Tack till **Agnes Engberg**, **Patrik Gilje** och **Jonas Björkman**. Era insatser på de olika delarbetena har varit ovärderliga.

Tack till alla på klinisk mikrobiologi i Lund för alltid bra bemötande och trevliga diskussioner, och för att jag fått möjlighet att laborera hos er. Ett särskilt tack till **Lena Hyllbusk** och biomedicinsk analytiker **Niklas**. Tack till alla mikrobiologiska lab i Sverige som varit behjälpliga med att fram data.

Tack till **Simon Heissler**, **Ulla-Britt Karlsson** och **Anita Berglund** för hjälp med all forskningsadministration.

Tack till alla **finansiärer** som möjliggjort min forskarutbildning: Stig och Ragna Gorthons stiftelse, Anna-Lisa och Sven Eric Lundgrens stiftelse, Stiftelsen Hedda och John Forssmans fond genom Kungliga Fysiografiska Sällskapet i Lund, ALF, Region Skåne/Södra sjukvårdsregionen och Helsingborgs lasarett.

Allra mest tack till min **Sara**. För att du härdat ut. För ditt tålamod och kärlek. Och för **Ted**.

References

1. Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell. 2006;124(4):837-48.
2. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59-65.
3. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. Nat Rev Microbiol. 2012;10(8):575-82.
4. Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science. 2005;307(5717):1915-20.
5. Jans C, Meile L, Lacroix C, Stevens MJ. Genomics, evolution, and molecular epidemiology of the Streptococcus bovis/Streptococcus equinus complex (SBSEC). Infect Genet Evol. 2015;33:419-36.
6. Orla-Jensen S. The Lactic Acid Bacteria. Mem R Acad Sci Denmark Sect Sci Ser 8. 1919;VIII 5:81-197.
7. F Andrewes TH. A study of the streptococci pathogenic for man. The Lancet. 1906;171:708-13.
8. Sherman J. The streptococci. Bacteriol Rev. 1937;1:3-97.
9. Corredoira J, Rabuñal-Rey R, Alonso M. Streptococcus bovis: 100 Years of an Intriguing Pathogen. Clin Microbiol Newsl. 2017;39:1-9.
10. Jans C, Boleij A. The Road to Infection: Host-Microbe Interactions Defining the Pathogenicity of Streptococcus bovis/Streptococcus equinus Complex Members. Front Microbiol. 2018;9:603.
11. Herrera P, Kwon YM, Ricke SC. Ecology and pathogenicity of gastrointestinal Streptococcus bovis. Anaerobe. 2009;15(1-2):44-54.
12. Russell JB, Rychlik JL. Factors that alter rumen microbial ecology. Science. 2001;292(5519):1119-22.
13. Kumar K, Chaudhary LC, Agarwal N, Kamra DN. Effect of feeding tannin-degrading bacteria Streptococcus gallolyticus strain TDGB 406 on meat quality of goats fed with Quercus semicarpifolia leaves. Trop Anim Health Prod. 2016;48(7):1513-6.
14. Kmet V, Flint HJ, Wallace RJ. Probiotics and manipulation of rumen development and function. Arch Tierernahr. 1993;44(1):1-10.
15. Cheng KJ, McAllister TA, Popp JD, Hristov AN, Mir Z, Shin HT. A review of bloat in feedlot cattle. J Anim Sci. 1998;76(1):299-308.

16. Garvie EI, Bramley AJ. Streptococcus bovis—an Approach to its Classification and its Importance as a Cause of Bovine Mastitis. Journal of Applied Bacteriology. 1979;46(3):557-66.
17. Corredoira J, Miguez E, Mateo LM, Fernandez-Rodriguez R, Garcia-Rodriguez JF, Perez-Gonzalez A, et al. Correlation between Streptococcus bovis bacteremia and density of cows in Galicia, northwest of Spain. Infection. 2019;47(3):399-407.
18. Jans C, Kaindi DW, Bock D, Njage PM, Kouame-Sina SM, Bonfoh B, et al. Prevalence and comparison of Streptococcus infantarius subsp. infantarius and Streptococcus gallolyticus subsp. macedonicus in raw and fermented dairy products from East and West Africa. Int J Food Microbiol. 2013;167(2):186-95.
19. Tsakalidou E, Zoidou E, Pot B, Wassill L, Ludwig W, Devriese LA, et al. Identification of streptococci from Greek Kasseri cheese and description of Streptococcus macedonicus sp. nov. Int J Syst Bacteriol. 1998;48 Pt 2:519-27.
20. Isono Y, Shingu I, Shimizu S. Identification and Characteristics of Lactic Acid Bacteria Isolated from Masai Fermented Milk in Northern Tanzania. Bioscience, Biotechnology, and Biochemistry. 1994;58(4):660-4.
21. Öberg J, Rasmussen M, Buchwald P, Nilson B, Inghammar M. Streptococcus bovis - bacteremia: subspecies distribution and association with colorectal cancer – a retrospective cohort study. Epidemiol Infect. 2022;150(E8):1-7.
22. Öberg J, Nilson B, Gilje P, Rasmussen M, Inghammar M. Bacteraemia and infective endocarditis with Streptococcus bovis-Streptococcus equinus-complex: a retrospective cohort study. Infect Dis. 2022;54(10):760-5.
23. Abdulamir AS, Hafidh RR, Abu Bakar F. The association of Streptococcus bovis/gallolyticus with colorectal tumors: the nature and the underlying mechanisms of its etiological role. J Exp Clin Cancer Res. 2011;30:11.
24. O'Brien MJ, Gibbons D. The adenoma-carcinoma sequence in colorectal neoplasia. Surg Oncol Clin N Am. 1996;5(3):513-30.
25. Klein RS, Recco RA, Catalano MT, Edberg SC, Casey JI, Steigbigel NH. Association of Streptococcus bovis with carcinoma of the colon. N Engl J Med. 1977;297(15):800-2.
26. Noble CJ. Carriage of group D streptococci in the human bowel. J Clin Pathol. 1978;31(12):1182-6.
27. Al-Jashamy K, Murad A, Zeehaida M, Rohaini M, Hasnan J. Prevalence of colorectal cancer associated with Streptococcus bovis among inflammatory bowel and chronic gastrointestinal tract disease patients. Asian Pac J Cancer Prev. 2010;11(6):1765-8.
28. Chirouze C, Patry I, Duval X, Baty V, Tattevin P, Aparicio T, et al. Streptococcus bovis/Streptococcus equinus complex fecal carriage, colorectal carcinoma, and infective endocarditis: a new appraisal of a complex connection. Eur J Clin Microbiol Infect Dis. 2013;32(9):1171-6.
29. Boltin D, Goldberg E, Bugaevsky O, Kelner E, Birkenfeld S, Gingold-Belfer R, et al. Colonic carriage of Streptococcus bovis and colorectal neoplasia: a prospective 17-year longitudinal case-control study. Eur J Gastroenterol Hepatol. 2015;27(12):1449-53.

30. Dumke J, Vollmer T, Akkermann O, Knabbe C, Dreier J. Case-control study: Determination of potential risk factors for the colonization of healthy volunteers with Streptococcus gallolyticus subsp. gallolyticus. PLoS One. 2017;12(5):e0176515.
31. Kaindi DWM, Kogi-Makau W, Lule GN, Kreikemeyer B, Renault P, Bonfoh B, et al. Investigating the association between African spontaneously fermented dairy products, faecal carriage of Streptococcus infantarius subsp. infantarius and colorectal adenocarcinoma in Kenya. Acta Trop. 2018;178:10-8.
32. Sissons CH, Hancock EM, Perinpanayagam HE, Cutress TW. The bacteria responsible for ureolysis in artificial dental plaque. Arch Oral Biol. 1988;33(10):727-33.
33. Shen S, Samaranayake LP, Yip HK. Coaggregation profiles of the microflora from root surface caries lesions. Arch Oral Biol. 2005;50(1):23-32.
34. Arul AS, Palanivelu P. Biofilm forming ability of a new bacterial isolate from dental caries: An atomic force microscopic study. J Nat Sci Biol Med. 2014;5(2):278-83.
35. Facklam RR. Recognition of group D streptococcal species of human origin by biochemical and physiological tests. Appl Microbiol. 1972;23(6):1131-9.
36. Parker MT, Ball LC. Streptococci and aerococci associated with systemic infection in man. J Med Microbiol. 1976;9(3):275-302.
37. Coykdendall AL, Gustafson KB. Deoxyribonucleic Acid Hybridizations Among Strains of Streptococcus salivarius and Streptococcus bovis. International Journal of Systematic and Evolutionary Microbiology. 1985;35(3):274-80.
38. Farrow JAE, Kruze J, Phillips BA, Bramley AJ, Collins MD. Taxonomic Studies on Streptococcus-Bovis and Streptococcus-Equinus - Description of Streptococcus-Alactolyticus Sp-Nov and Streptococcus Saccharolyticus Sp-Nov. Syst Appl Microbiol. 1984;5(4):467-82.
39. Osawa R, Fujisawa T, Sly LI. Streptococcus Gallolyticus Sp-Nov - Gallate Degrading Organisms Formerly Assigned to Streptococcus-Bovis. Syst Appl Microbiol. 1995;18(1):74-8.
40. Bouvet A, Grimont F, Collins MD, Benaoudia F, Devine C, Regnault B, et al. Streptococcus infantarius sp. nov. related to Streptococcus bovis and Streptococcus equinus. Adv Exp Med Biol. 1997;418:393-5.
41. Schlegel L, Grimont F, Collins MD, Regnault B, Grimont PA, Bouvet A. Streptococcus infantarius sp. nov., Streptococcus infantarius subsp. infantarius subsp. nov. and Streptococcus infantarius subsp. coli subsp. nov., isolated from humans and food. Int J Syst Evol Microbiol. 2000;50 Pt 4:1425-34.
42. Poyart C, Quesne G, Trieu-Cuot P. Taxonomic dissection of the Streptococcus bovis group by analysis of manganese-dependent superoxide dismutase gene (sodA) sequences: reclassification of 'Streptococcus infantarius subsp. coli' as Streptococcus lutetiensis sp. nov. and of Streptococcus bovis biotype 11.2 as Streptococcus pasteurianus sp. nov. Int J Syst Evol Microbiol. 2002;52(Pt 4):1247-55.

43. Schlegel L, Grimont F, Ageron E, Grimont PAD, Bouvet A. Reappraisal of the taxonomy of the Streptococcus bovis/Streptococcus equinus complex and related species: description of Streptococcus gallolyticus subsp. gallolyticus subsp. nov., S. gallolyticus subsp. macedonicus subsp. nov. and S. gallolyticus subsp. pasteurianus subsp. nov. Int J Syst Evol Microbiol. 2003;53(Pt 3):631-45.
44. Vos P, Garrity G, Jones D, Krieg N, Ludwig W, Rainey F, et al. Bergey's Manual of Systematic Bacteriology, Volume 3: The Firmicutes. 2 ed: Springer New York, NY; 2009.
45. Sly LI, Cahill MM, Osawa R, Fujisawa T. The tannin-degrading species Streptococcus gallolyticus and Streptococcus caprinus are subjective synonyms. Int J Syst Bacteriol. 1997;47(3):893-4.
46. Dekker JP, Lau AF. An Update on the Streptococcus bovis Group: Classification, Identification, and Disease Associations. Journal of Clinical Microbiology. 2016;54(7):1694-9.
47. Park SY, Muhammad N, Kim SG, Kwon H, Seo S, Kim JH. Streptococcus ruminicola sp. nov., new species of the Streptococcus bovis/Streptococcus equinus complex (SBSEC) isolated from the rumen of Korean domestic ruminants. Arch Microbiol. 2022;204(10):636.
48. Volokhov DV, Zagorodnyaya TA, Shen Z, Blom J, Furtak VA, Eisenberg T, et al. Streptococcus vicugnae sp. nov., isolated from faeces of alpacas (Vicugna pacos) and cattle (Bos taurus), Streptococcus zalophi sp. nov., and Streptococcus pacificus sp. nov., isolated from respiratory tract of California sea lions (Zalophus californianus). Int J Syst Evol Microbiol. 2021;71(5).
49. Pompilio A, Di Bonaventura G, Gherardi G. An Overview on Streptococcus bovis/Streptococcus equinus Complex Isolates: Identification to the Species/Subspecies Level and Antibiotic Resistance. Int J Mol Sci. 2019;20(3).
50. Putnam NE, Youn JH, Wallace MA, Luethy PM, Burnham CD, Butler-Wu S, et al. Comparative Evaluation of Current Biochemical-, Sequencing-, and Proteomic-Based Identification Methods for the Streptococcus bovis Group. J Clin Microbiol. 2023;61(4):e0171222.
51. Corredoira J, Alonso MP, Garcia-Garrote F, Garcia-Pais MJ, Coira A, Rabunal R, et al. Streptococcus bovis group and biliary tract infections: an analysis of 51 cases. Clin Microbiol Infect. 2014;20(5):405-9.
52. Ben-Chetrit E, Wiener-Well Y, Kashat L, Yinnon AM, Assous MV. Streptococcus bovis new taxonomy: does subspecies distinction matter? Eur J Clin Microbiol Infect Dis. 2017;36(2):387-93.
53. Beck M, Frodl R, Funke G. Comprehensive study of strains previously designated Streptococcus bovis consecutively isolated from human blood cultures and emended description of Streptococcus gallolyticus and Streptococcus infantarius subsp. coli. J Clin Microbiol. 2008;46(9):2966-72.
54. Gherardi G, Palmieri C, Marini E, Pompilio A, Crocetta V, Di Bonaventura G, et al. Identification, antimicrobial resistance and molecular characterization of the human emerging pathogen Streptococcus gallolyticus subsp. pasteurianus. Diagn Microbiol Infect Dis. 2016;86(4):329-35.

55. Sheng WH, Chuang YC, Teng LJ, Hsueh PR. Bacteraemia due to Streptococcus gallolyticus subspecies pasteurianus is associated with digestive tract malignancies and resistance to macrolides and clindamycin. J Infect. 2014;69(2):145-53.
56. Romero B, Morosini MI, Loza E, Rodriguez-Banos M, Navas E, Canton R, et al. Reidentification of Streptococcus bovis isolates causing bacteremia according to the new taxonomy criteria: still an issue? J Clin Microbiol. 2011;49(9):3228-33.
57. Hoshino T, Fujiwara T, Kilian M. Use of phylogenetic and phenotypic analyses to identify nonhemolytic streptococci isolated from bacteremic patients. J Clin Microbiol. 2005;43(12):6073-85.
58. Teles C, Smith A, Ramage G, Lang S. Identification of clinically relevant viridans group streptococci by phenotypic and genotypic analysis. Eur J Clin Microbiol Infect Dis. 2011;30(2):243-50.
59. Sasaki E, Osawa R, Nishitani Y, Whiley RA. Development of a diagnostic PCR assay targeting the Mn-dependent superoxide dismutase gene (sodA) for identification of Streptococcus gallolyticus. J Clin Microbiol. 2004;42(3):1360-2.
60. Garcia-Pais MJ, Rabunal R, Armesto V, Lopez-Reboiro M, Garcia-Garrote F, Coira A, et al. Streptococcus bovis septic arthritis and osteomyelitis: A report of 21 cases and a literature review. Semin Arthritis Rheum. 2016;45(6):738-46.
61. Glazunova OO, Raoult D, Roux V. Partial sequence comparison of the rpoB, sodA, groEL and gyrB genes within the genus Streptococcus. Int J Syst Evol Microbiol. 2009;59(Pt 9):2317-22.
62. Hinse D, Vollmer T, Erhard M, Welker M, Moore ER, Kleesiek K, et al. Differentiation of species of the Streptococcus bovis/equinus-complex by MALDI-TOF Mass Spectrometry in comparison to sodA sequence analyses. Syst Appl Microbiol. 2011;34(1):52-7.
63. Marmolin ES, Hartmeyer GN, Christensen JJ, Nielsen XC, Dargis R, Skov MN, et al. Bacteremia with the bovis group streptococci: species identification and association with infective endocarditis and with gastrointestinal disease. Diagn Microbiol Infect Dis. 2016;85(2):239-42.
64. Lopes PG, Cantarelli VV, Agnes G, Costabeber AM, d'Azevedo PA. Novel real-time PCR assays using TaqMan minor groove binder probes for identification of fecal carriage of Streptococcus bovis/Streptococcus equinus complex from rectal swab specimens. J Clin Microbiol. 2014;52(3):974-6.
65. Shibata Y, Tien LHT, Nomoto R, Osawa R. Development of a multilocus sequence typing scheme for Streptococcus gallolyticus. Microbiology (Reading). 2014;160(Pt 1):113-22.
66. Torres-Sangiao E, Leal Rodriguez C, Garcia-Riestra C. Application and Perspectives of MALDI-TOF Mass Spectrometry in Clinical Microbiology Laboratories. Microorganisms. 2021;9(7).
67. Freiwald A, Sauer S. Phylogenetic classification and identification of bacteria by mass spectrometry. Nat Protoc. 2009;4(5):732-42.

68. Haigh J, Degun A, Eydmann M, Millar M, Wilks M. Improved Performance of Bacterium and Yeast Identification by a Commercial Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry System in the Clinical Microbiology Laboratory. Journal of Clinical Microbiology. 2011;49(9):3441-.
69. Calderaro A, Chezzi C. MALDI-TOF MS: A Reliable Tool in the Real Life of the Clinical Microbiology Laboratory. Microorganisms. 2024;12(2).
70. Agergaard CN, Knudsen E, Dargis R, Nielsen XC, Christensen JJ, Justesen US. Species identification of Streptococcus bovis group isolates causing bacteremia: a comparison of two MALDI-TOF MS systems. Diagn Microbiol Infect Dis. 2017;88(1):23-5.
71. Lopez Roa P, Sanchez Carrillo C, Marin M, Romero F, Cercenado E, Bouza E. Value of matrix-assisted laser desorption ionization-time of flight for routine identification of viridans group streptococci causing bloodstream infections. Clin Microbiol Infect. 2013;19(5):438-44.
72. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. Hum Immunol. 2021;82(11):801-11.
73. Price TK, Realegeno S, Mirasol R, Tsan A, Chandrasekaran S, Garner OB, et al. Validation, Implementation, and Clinical Utility of Whole Genome Sequence-Based Bacterial Identification in the Clinical Microbiology Laboratory. J Mol Diagn. 2021;23(11):1468-77.
74. Schurch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. Clin Microbiol Infect. 2018;24(4):350-4.
75. Hinse D, Vollmer T, Ruckert C, Blom J, Kalinowski J, Knabbe C, et al. Complete genome and comparative analysis of Streptococcus gallolyticus subsp. gallolyticus, an emerging pathogen of infective endocarditis. BMC Genomics. 2011;12:400.
76. Kumar S, Bansal K, Sethi SK. Comparative genomics reconciliations of genus Streptococcus resolves its taxonomy and elucidates biotechnological importance of their constituent species. Ecological Genetics and Genomics. 2023;29:100205.
77. Giannitsioti E, Chirouze C, Bouvet A, Beguinot I, Delahaye F, Mainardi JL, et al. Characteristics and regional variations of group D streptococcal endocarditis in France. Clin Microbiol Infect. 2007;13(8):770-6.
78. Corredoira J, Grau I, Garcia-Rodriguez JF, Alonso-Garcia P, Garcia-Pais MJ, Rabunal R, et al. The clinical epidemiology and malignancies associated with Streptococcus bovis biotypes in 506 cases of bloodstream infections. J Infect. 2015;71(3):317-25.
79. Corredoira J, Grau I, Garcia-Rodriguez JF, Romay E, Cuervo G, Berbel D, et al. Species and biotypes of Streptococcus bovis causing infective endocarditis. Enfermedades Infecciosas y Microbiología Clínica. 2021.
80. Coffey S, Nadarasa K, Pan A, van der Linden A, Chu J, Schultz M. The increasing incidence of Streptococcus bovis endocarditis and bacteraemia: a case series from 1997 to 2010. Int J Cardiol. 2012;161(2):111-3.

81. Kaiki Y, Kitagawa H, Tadera K, Taogoshi H, Ikeda M, Kano M, et al. Laboratory identification and clinical characteristics of Streptococcus bovis/Streptococcus equinus complex bacteremia: a retrospective, multicenter study in Hiroshima, Japan. BMC Infect Dis. 2021;21(1):1192.
82. Lazarovitch T, Shango M, Levine M, Brusovansky R, Akins R, Hayakawa K, et al. The relationship between the new taxonomy of Streptococcus bovis and its clonality to colon cancer, endocarditis, and biliary disease. Infection. 2013;41(2):329-37.
83. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. Gut. 2023;72(2):338-44.
84. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209-49.
85. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-63.
86. Tjock- och ändtarmscancer, Nationellt vårdprogram, Version: 4.0. 2025.
87. Collaborators GBDCC. Global, regional, and national burden of colorectal cancer and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Gastroenterol Hepatol. 2022;7(7):627-47.
88. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol. 2019;14(2):89-103.
89. Collaborative R, Zaborowski AM, Abdile A, Adamina M, Aigner F, d'Allens L, et al. Characteristics of Early-Onset vs Late-Onset Colorectal Cancer: A Review. JAMA Surg. 2021;156(9):865-74.
90. Gutlic I, Schyman T, Lydrup ML, Buchwald P. Increasing colorectal cancer incidence in individuals aged < 50 years-a population-based study. Int J Colorectal Dis. 2019;34(7):1221-6.
91. Stewart SL, Wike JM, Kato I, Lewis DR, Michaud F. A population-based study of colorectal cancer histology in the United States, 1998-2001. Cancer. 2006;107(5 Suppl):1128-41.
92. Duan B, Zhao Y, Bai J, Wang J, Duan X, Luo X, et al. Colorectal Cancer: An Overview. In: Morgado-Diaz JA, editor. Gastrointestinal Cancers. Brisbane (AU)2022.
93. Pålsson B. Den nationella screeningen för kolorektalcancer – en översikt. 2023.
94. Tjock- och ändtarmscancer screening, Nationellt vårdprogram, Version: 1.0. 2022.
95. Brierley J GM WC. TNM Classification of Malignant Tumours. 8th ed: Wiley; 2016.
96. O'Sullivan B, Brierley J, Byrd D, Bosman F, Kehoe S, Kossary C, et al. The TNM classification of malignant tumours-towards common understanding and reasonable expectations. Lancet Oncol. 2017;18(7):849-51.

97. Baran B, Mert Ozupek N, Yerli Tetik N, Acar E, Bekcioglu O, Baskin Y. Difference Between Left-Sided and Right-Sided Colorectal Cancer: A Focused Review of Literature. Gastroenterology Res. 2018;11(4):264-73.
98. Cerdan-Santacruz C, Sao Juliao GP, Vailati BB, Corbi L, Habr-Gama A, Perez RO. Watch and Wait Approach for Rectal Cancer. J Clin Med. 2023;12(8).
99. Cornista AM, Giolito MV, Baker K, Hazime H, Dufait I, Datta J, et al. Colorectal Cancer Immunotherapy: State of the Art and Future Directions. Gastro Hep Adv. 2023;2(8):1103-19.
100. The National Board of Health and Welfare, Cancer in Sweden [Available from: https://dataanalys.socialstyrelsen.se/superset/dashboard/CiS_relativ_overlevnad/].
101. Rektalcancer 2023 - Nationell kvalitetsrapport för diagnosår 2023 från Svenska Kolorektalcancerregistret. 2024.
102. Koloncancer 2023 - Nationell kvalitetsrapport för diagnosår 2023 från Svenska Kolorektalcancerregistret. 2024.
103. Jacqueline C, Tasiemski A, Sorci G, Ujvari B, Maachi F, Misse D, et al. Infections and cancer: the "fifty shades of immunity" hypothesis. BMC Cancer. 2017;17(1):257.
104. zur Hausen H. The search for infectious causes of human cancers: where and why (Nobel lecture). Angew Chem Int Ed Engl. 2009;48(32):5798-808.
105. Reyes VE. Helicobacter pylori and Its Role in Gastric Cancer. Microorganisms. 2023;11(5).
106. Lu SSM, Mohammed Z, Haggstrom C, Myte R, Lindquist E, Gylfe A, et al. Antibiotics Use and Subsequent Risk of Colorectal Cancer: A Swedish Nationwide Population-Based Study. J Natl Cancer Inst. 2022;114(1):38-46.
107. Armstrong D, Dregan A, Ashworth M, White P, McGee C, de Lusignan S. The association between colorectal cancer and prior antibiotic prescriptions: case control study. Br J Cancer. 2020;122(6):912-7.
108. Dik VK, van Oijen MG, Smeets HM, Siersema PD. Frequent Use of Antibiotics Is Associated with Colorectal Cancer Risk: Results of a Nested Case-Control Study. Dig Dis Sci. 2016;61(1):255-64.
109. Boursi B, Haynes K, Mamtani R, Yang YX. Impact of antibiotic exposure on the risk of colorectal cancer. Pharmacoepidemiol Drug Saf. 2015;24(5):534-42.
110. Kwong TNY, Wang X, Nakatsu G, Chow TC, Tipoe T, Dai RZW, et al. Association Between Bacteremia From Specific Microbes and Subsequent Diagnosis of Colorectal Cancer. Gastroenterology. 2018;155(2):383-90 e8.
111. Justesen US, Nielsen SL, Jensen TG, Dessau RB, Moller JK, Coia JE, et al. Bacteremia With Anaerobic Bacteria and Association With Colorectal Cancer: A Population-based Cohort Study. Clinical Infectious Diseases. 2022;75(10):1747-53.
112. Laupland KB, Edwards F, Furuya-Kanamori L, Paterson DL, Harris PNA. Bloodstream infection and colorectal cancer risk in Queensland Australia, 2000-2019. The American Journal of Medicine. 2023;136(9):896-901.
113. van Santvoort HC, Besselink MG, Timmerman HM, van Minnen LP, Akkermans LM, Gooszen HG. Probiotics in surgery. Surgery. 2008;143(1):1-7.

114. Sears CL, Pardoll DM. Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. J Infect Dis. 2011;203(3):306-11.
115. McCoy W, Mason JM, 3rd. Enterococcal endocarditis associated with carcinoma of the sigmoid; report of a case. J Med Assoc State Ala. 1951;21(6):162-6.
116. Ouranos K, Gardikioti A, Bakaloudi DR, Mylona EK, Shehadeh F, Mylonakis E. Association of the Streptococcus bovis/Streptococcus equinus Complex With Colorectal Neoplasia: A Systematic Review and Meta-analysis. Open Forum Infect Dis. 2023;10(11):ofad547.
117. Boleij A, van Gelder MM, Swinkels DW, Tjalsma H. Clinical Importance of Streptococcus gallolyticus infection among colorectal cancer patients: systematic review and meta-analysis. Clin Infect Dis. 2011;53(9):870-8.
118. Butt J, Werner S, Willhauck-Fleckenstein M, Michel A, Waterboer T, Zornig I, et al. Serology of Streptococcus gallolyticus subspecies gallolyticus and its association with colorectal cancer and precursors. Int J Cancer. 2017;141(5):897-904.
119. Butt J, Romero-Hernandez B, Perez-Gomez B, Willhauck-Fleckenstein M, Holzinger D, Martin V, et al. Association of Streptococcus gallolyticus subspecies gallolyticus with colorectal cancer: Serological evidence. Int J Cancer. 2016;138(7):1670-9.
120. Boleij A, Roelofs R, Danne C, Bellais S, Dramsi S, Kato I, et al. Selective antibody response to Streptococcus gallolyticus pilus proteins in colorectal cancer patients. Cancer Prev Res (Phila). 2012;5(2):260-5.
121. Corredoira J, Romay E, Pericas JM, Miro JM. Associating enterococcal endocarditis and colorectal neoplasia: is colonoscopy mandatory? Eur J Intern Med. 2021;85:112-3.
122. Corredoira J, Garcia-Pais MJ, Coira A, Rabunal R, Garcia-Garrote F, Pita J, et al. Differences between endocarditis caused by Streptococcus bovis and Enterococcus spp. and their association with colorectal cancer. European Journal of Clinical Microbiology & Infectious Diseases. 2015;34(8):1657-65.
123. Lieberman DA, Weiss DG, Harford WV, Ahnen DJ, Provenzale D, Sontag SJ, et al. Five-year colon surveillance after screening colonoscopy. Gastroenterology. 2007;133(4):1077-85.
124. Corredoira-Sanchez J, Garcia-Garrote F, Rabunal R, Lopez-Roses L, Garcia-Pais MJ, Castro E, et al. Association between bacteremia due to Streptococcus gallolyticus subsp. gallolyticus (Streptococcus bovis I) and colorectal neoplasia: a case-control study. Clin Infect Dis. 2012;55(4):491-6.
125. Corredoira JC, Alonso MP, Garcia JF, Casariego E, Coira A, Rodriguez A, et al. Clinical characteristics and significance of Streptococcus salivarius bacteremia and Streptococcus bovis bacteremia: a prospective 16-year study. Eur J Clin Microbiol Infect Dis. 2005;24(4):250-5.
126. Corredoira JC, Alonso MP, Garcia-Pais MJ, Rabunal R, Garcia-Garrote F, Lopez-Roses L, et al. Is colonoscopy necessary in cases of infection by Streptococcus bovis biotype II? Eur J Clin Microbiol Infect Dis. 2014;33(2):171-7.
127. Ruoff KL, Miller SI, Garner CV, Ferraro MJ, Calderwood SB. Bacteremia with Streptococcus bovis and Streptococcus salivarius: clinical correlates of more accurate identification of isolates. J Clin Microbiol. 1989;27(2):305-8.

128. Corredoira J, Coira A, Iniguez I, Pita J, Varela J, Alonso MP. Advanced intestinal cancer associated with Streptococcus infantarius (former S. bovis II/1) sepsis. Int J Clin Pract. 2013;67(12):1358-9.
129. Boleij A, Tjalsma H. The itinerary of Streptococcus gallolyticus infection in patients with colonic malignant disease. Lancet Infect Dis. 2013;13(8):719-24.
130. Boleij A, Muytjens CM, Bukhari SI, Cayet N, Glaser P, Hermans PW, et al. Novel clues on the specific association of Streptococcus gallolyticus subsp gallolyticus with colorectal cancer. J Infect Dis. 2011;203(8):1101-9.
131. Berg RD. Bacterial translocation from the gastrointestinal tract. Adv Exp Med Biol. 1999;473:11-30.
132. Kouzu K, Tsujimoto H, Kishi Y, Ueno H, Shinomiya N. Bacterial Translocation in Gastrointestinal Cancers and Cancer Treatment. Biomedicines. 2022;10(2).
133. Kumar R, Herold JL, Schady D, Davis J, Kopetz S, Martinez-Moczygemba M, et al. Streptococcus gallolyticus subsp. gallolyticus promotes colorectal tumor development. PLOS Pathogens. 2017;13(7):1006440.
134. Kumar R, Herold JL, Taylor J, Xu J, Xu Y. Variations among Streptococcus gallolyticus subsp. gallolyticus strains in connection with colorectal cancer. Scientific Reports. 2018;8(1):1514.
135. Taylor JC, Kumar R, Xu J, Xu Y. A pathogenicity locus of Streptococcus gallolyticus subspecies gallolyticus. Scientific Reports. 2023;13(1):6291.
136. Taylor JC, Gao X, Xu J, Holder M, Petrosino J, Kumar R, et al. A type VII secretion system of Streptococcus gallolyticus subsp. gallolyticus contributes to gut colonization and the development of colon tumors. PLOS PATHOGENS. 2021;17(1):e1009182.
137. Pasquereau-Kotula E, du Merle L, Sismeiro O, Pietrosemoli N, Varet H, Legendre R, et al. Transcriptome profiling of human col\onic cells exposed to the gut pathobiont Streptococcus gallolyticus subsp. gallolyticus. PLoS One. 2023;18(11):e0294868.
138. Wang H, You W, Zhu Z, Zhang Y, Hu C, Lu J, et al. Streptococcus lutetiensis inhibits CD8(+) IL17A(+) TRM cells and leads to gastric cancer progression and poor prognosis. NPJ Precis Oncol. 2025;9(1):43.
139. Kale P, Khillan V, Sarin SK. Novel association of Streptococcus gallolyticus subspecies pasteurianus and hepatocelluar carcinoma: opening new frontiers. Scandinavian Journal of Gastroenterology. 2018;53(10-11):1354-7.
140. Li M, Kim JB, Sastry BKS, Chen M. Infective endocarditis. Lancet. 2024;404(10450):377-92.
141. Delgado V, Ajmone Marsan N, de Waha S, Bonaros N, Brida M, Burri H, et al. 2023 ESC Guidelines for the management of endocarditis. Eur Heart J. 2023;44(39):3948-4042.
142. Habib G, Lancellotti P, Antunes MJ, Bongiorni MG, Casalta JP, Del Zotti F, et al. 2015 ESC Guidelines for the management of infective endocarditis: The Task Force for the Management of Infective Endocarditis of the European Society of Cardiology (ESC). Endorsed by: European Association for Cardio-Thoracic Surgery (EACTS), the European Association of Nuclear Medicine (EANM). Eur Heart J. 2015;36(44):3075-128.

143. Pries-Heje MM, Wiingaard C, Ihlemann N, Gill SU, Bruun NE, Elming H, et al. Five-Year Outcomes of the Partial Oral Treatment of Endocarditis (POET) Trial. N Engl J Med. 2022;386(6):601-2.
144. Iversen K, Ihlemann N, Gill SU, Madsen T, Elming H, Jensen KT, et al. Partial Oral versus Intravenous Antibiotic Treatment of Endocarditis. N Engl J Med. 2019;380(5):415-24.
145. Fowler VG, Durack DT, Selton-Suty C, Athan E, Bayer AS, Chamis AL, et al. The 2023 Duke-International Society for Cardiovascular Infectious Diseases Criteria for Infective Endocarditis: Updating the Modified Duke Criteria. Clin Infect Dis. 2023;77(4):518-26.
146. Li JS, Sexton DJ, Mick N, Nettles R, Fowler VG, Jr., Ryan T, et al. Proposed modifications to the Duke criteria for the diagnosis of infective endocarditis. Clin Infect Dis. 2000;30(4):633-8.
147. Durack DT, Lukes AS, Bright DK. New criteria for diagnosis of infective endocarditis: utilization of specific echocardiographic findings. Duke Endocarditis Service. Am J Med. 1994;96(3):200-9.
148. Von Reyn CF, Levy BS, Arbeit RD, Friedland G, Crumpacker CS. Infective endocarditis: an analysis based on strict case definitions. Ann Intern Med. 1981;94(4 pt 1):505-18.
149. Lindberg H, Berge A, Jovanovic-Stjernqvist M, Hagstrand Aldman M, Krus D, Oberg J, et al. Performance of the 2023 Duke-International Society of Cardiovascular Infectious Diseases Diagnostic Criteria for Infective Endocarditis in Relation to the Modified Duke Criteria and to Clinical Management-Reanalysis of Retrospective Bacteremia Cohorts. Clin Infect Dis. 2024;78(4):956-63.
150. Goehringer F, Lalloue B, Selton-Suty C, Alla F, Botelho-Nevers E, Chirouze C, et al. Compared Performance of the 2023 Duke-International Society for Cardiovascular Infectious Diseases, 2000 Modified Duke, and 2015 European Society of Cardiology Criteria for the Diagnosis of Infective Endocarditis in a French Multicenter Prospective Cohort. Clin Infect Dis. 2024;78(4):937-48.
151. Avogadri N, Ivarsson A, Burup Kristensen C, Ragnarsson S, Rasmussen M. Evaluation of the 2023 Duke-ISCVID criteria for infective endocarditis in patients referred to the Lund infective endocarditis team. Infect Dis (Lond). 2024:1-9.
152. Papadimitriou-Olivgeris M, Monney P, Frank M, Tzimas G, Tozzi P, Kirsch M, et al. Evaluation of the 2023 Duke-International Society of Cardiovascular Infectious Diseases Criteria in a Multicenter Cohort of Patients With Suspected Infective Endocarditis. Clin Infect Dis. 2024;78(4):949-55.
153. Corredoira J, Alonso MP, Coira A, Casariego E, Arias C, Alonso D, et al. Characteristics of Streptococcus bovis endocarditis and its differences with Streptococcus viridans endocarditis. Eur J Clin Microbiol Infect Dis. 2008;27(4):285-91.
154. Hede SV. Diagnosis and Treatment of Childhood Meningitis Caused by Streptococcus bovis Group. Curr Infect Dis Rep. 2016;18(4):11.

155. Jean SS, Teng LJ, Hsueh PR, Ho SW, Luh KT. Bacteremic Streptococcus bovis infections at a university hospital, 1992-2001. J Formos Med Assoc. 2004;103(2):118-23.
156. Hoen B, Chirouze C, Cabell CH, Selton-Suty C, Duchene F, Olaison L, et al. Emergence of endocarditis due to group D streptococci: findings derived from the merged database of the International Collaboration on Endocarditis. Eur J Clin Microbiol Infect Dis. 2005;24(1):12-6.
157. Nilson B, Olaison L, Rasmussen M. Clinical presentation of infective endocarditis caused by different groups of non-beta haemolytic streptococci. Eur J Clin Microbiol Infect Dis. 2016;35(2):215-8.
158. Corredoira J, Grau I, Garcia-Rodriguez JF, Romay E, Cuervo G, Berbel D, et al. Species and biotypes of Streptococcus bovis causing infective endocarditis. Enferm Infecc Microbiol Clin (Engl Ed). 2023;41(4):215-20.
159. Sechi LA, Ciani R. Streptococcus equinus endocarditis in a woman with pulmonary histiocytosis. JAMA. 2000;283(8):1005.
160. Keerty D, Yacoub AT, Nguyen TC, Haynes E, Greene J. First Case of Infective Endocarditis With Streptococcus equinus in an Immunocompetent Patient in North America: A Case Report and Review of Literature. Cureus. 2021;13(11):e19473.
161. Malkin J, Kimmitt PT, Ou HY, Bhasker PS, Khare M, Deng Z, et al. Identification of Streptococcus gallolyticus subsp. macedonicus as the etiological agent in a case of culture-negative multivalve infective endocarditis by 16S rDNA PCR analysis of resected valvular tissue. J Heart Valve Dis. 2008;17(5):589-92.
162. Mylonas CC, Gomatou G, Poulakou G, Moraitou E, Syrigos K. Human disease caused by Streptococcus alactolyticus: a case report of native valve infective endocarditis and review of the literature. Monaldi Arch Chest Dis. 2020;90(4).
163. Vinciguerra M, Santamaria V, Romiti S, D'Abramo M, Toto G, De Bellis A, et al. Case Report: Streptococcus alactolyticus as a Rare Pathogen of Mitral Endocarditis. Front Cardiovasc Med. 2021;8:648213.
164. Ludvigsson JF, Almqvist C, Bonamy AK, Ljung R, Michaelsson K, Neovius M, et al. Registers of the Swedish total population and their use in medical research. Eur J Epidemiol. 2016;31(2):125-36.
165. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011;11:450.
166. Statistics Sweden [Available from: https://www.scb.se/].
167. The National Board of Health and Welfare [Available from: https://www.socialstyrelsen.se/].
168. Socialstyrelsen. Statistical register's production and quality: National Cancer Register. 2023.
169. Colorectal Cancer data Base (CRCBaSe) [Available from: https://cancercentrum.se/samverkan/cancerdiagnoser/tjocktarm-andtarm-och-anal/tjock--och-andtarm/kvalitetsregister/forskning/forskningsdatabas/].

170. Weibull CE, Boman SE, Glimelius B, Syk I, Matthiessen P, Smedby KE, et al. CRCBaSe: a Swedish register-based resource for colorectal adenocarcinoma research. Acta Oncol. 2023;62(4):342-9.
171. Wang X, Cheng Z. Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations. Chest. 2020;158(1S):S65-S71.
172. Dey T, Mukherjee A, Chakraborty S. A Practical Overview of Case-Control Studies in Clinical Practice. Chest. 2020;158(1S):S57-S64.
173. Wong RKW. Handbook of Matching and Weighting Adjustments for Causal Inference. Journal of the American Statistical Association. 2024;119(545):791-.
174. Hainmueller J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Political Analysis. 2012;20(1):25-46.
175. Chesnaye NC, Stel VS, Tripepi G, Dekker FW, Fu EL, Zoccali C, et al. An introduction to inverse probability of treatment weighting in observational research. Clin Kidney J. 2022;15(1):14-20.
176. McMullin J, Schonberger B. When Good Balance Goes Bad: A Discussion of Common Pitfalls When Using Entropy Balancing. Journal of Financial Reporting. 2022;7.
177. Matthews AA, Danaei G, Islam N, Kurth T. Target trial emulation: applying principles of randomised trials to observational studies. BMJ. 2022;378:e071108.
178. Hubbard RA, Gatsonis CA, Hogan JW, Hunter DJ, Normand ST, Troxel AB. "Target Trial Emulation" for Observational Studies - Potential and Pitfalls. N Engl J Med. 2024;391(21):1975-7.
179. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453-7.
180. Ludvigsson JF, Appelros P, Askling J, Byberg L, Carrero JJ, Ekstrom AM, et al. Adaptation of the Charlson Comorbidity Index for Register-Based Research in Sweden. Clinical Epidemiology. 2021;13:21-41.
181. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27(2):221-4.
182. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. 2011;173(6):676-82.
183. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359-86.
184. Öberg J, Inghammar M, Nilson B. Improved identification of Streptococcus bovis-Streptococcus equinus-complex species and subspecies by MALDI-TOF MS using a novel library. Diagn Microbiol Infect Dis. 2023;107(3):116045.

185. Öberg J, Engberg A, Inghammar M, Nilson B. Creation and validation of improved MALDI-TOF MS libraries for S. bovis-S. equinus-complex subspecies identification adapted to diagnostic culturing and extraction conditions. Diagn Microbiol Infect Dis. 2025;113(3):117000.
186. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol. 2016;45(6):1887-94.
187. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol. 2013;64(5):402-6.
188. Anderson NW, Buchan BW, Riebe KM, Parsons LN, Gnacinski S, Ledeboer NA. Effects of solid-medium type on routine identification of bacterial isolates by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2012;50(3):1008-13.
189. Nybakken EJ, Oppegaard O, Gilhuus M, Jensen CS, Mylvaganam H. Identification of Streptococcus dysgalactiae using matrix-assisted laser desorption/ionization-time of flight mass spectrometry; refining the database for improved identification. Diagn Microbiol Infect Dis. 2021;99(1):115207.
190. De Bel A, Wybo I, Vandoorslaer K, Rosseel P, Lauwers S, Pierard D. Acceptance criteria for identification results of Gram-negative rods by mass spectrometry. J Med Microbiol. 2011;60(Pt 5):684-6.
191. Papalia M, Figueroa-Espinosa R, Steffanowski C, Barberis C, Almuzara M, Barrios R, et al. Expansion and improvement of MALDI-TOF MS databases for accurate identification of Achromobacter species. J Microbiol Methods. 2020;172:105889.
192. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(1):257.
193. Statistics Sweden, Population statistics [Available from: http://www.scb.se/be0101-en].
194. The Swedish Board of Agriculture [Available from: https://jordbruksverket.se/statistik].
195. Corredoira J, Grau I, Garcia-Rodriguez JF, Garcia-Pais MJ, Rabunal R, Ardanuy C, et al. Colorectal neoplasm in cases of Clostridium septicum and Streptococcus gallolyticus subsp. gallolyticus bacteraemia. Eur J Intern Med. 2017;41:68-73.
196. Nasomsong W, Vasikasin V, Traipattanakul J, Changpradub D. Is Echocardiography Mandatory for All Streptococcus gallolyticus Subsp. pasteurianus Bacteremia? Infect Drug Resist. 2020;13:2425-32.
197. Ford AC, Veldhuyzen van Zanten SJ, Rodgers CC, Talley NJ, Vakil NB, Moayyedi P. Diagnostic utility of alarm features for colorectal cancer: systematic review and meta-analysis. Gut. 2008;57(11):1545-53.
198. Astin M, Griffin T, Neal RD, Rose P, Hamilton W. The diagnostic value of symptoms for colorectal cancer in primary care: a systematic review. Br J Gen Pract. 2011;61(586):e231-43.
199. Lieberman DA, Weiss DG, Bond JH, Ahnen DJ, Garewal H, Harford WV, et al. Use of Colonoscopy to Screen Asymptomatic Adults for Colorectal Cancer. New England Journal of Medicine. 2000;343(3):162-8.

200. Hoen B. Infections due to Streptococcus bovis/Streptococcus equinus complex (SBSEC; formerly group D streptococci). In: RF C, editor. UpToDate: Wolters Kluwer; 2025.
201. Sunnerhagen T, Tornell A, Vikbrant M, Nilson B, Rasmussen M. HANDOC: A Handy Score to Determine the Need for Echocardiography in Non-beta-Hemolytic Streptococcal Bacteremia. Clin Infect Dis. 2018;66(5):693-8.
202. La Scola B, Raoult D. Direct identification of bacteria in positive blood culture bottles by matrix-assisted laser desorption ionisation time-of-flight mass spectrometry. PLoS One. 2009;4(11):e8041.