

Kartläggning av företagspopulationen inom life science

- Metodbeskrivning

For information on obtaining additional copies, permission to reprint or translate this work, and all other correspondence. Please contact ADC.

Innehållsförteckning

1	Inledning	4
2	Arbetsprocessen	5
2.1	Fjolårets arbetsprocess 2022/2023	5
2.2	Arbetsprocess för vidareutvecklingen 2023/2024	6
3	Datakällor	7
3.1	Identifiering, selektering, klassificering av företagspopulationen	7
3.2	Beräkning av företagsdata	10
4	Metodbeskrivning	12
4.1	Förändringar och kvalitetssäkrande arbete under 2023/2024	12
4.2	Identifikation av företag	14
4.3	Selektering av företag	15
4.4	Klassificering av branschtillhörighet	17
4.5	Klassificering av marknadssegment	19
5	Känslighetsanalys	21
5.1	Om statistik för life science-sektorn	21
5.2	Avväganden i utvecklingsarbetet	21
5.3	Datakällor och -variabler	23
6	Rutiner vid uppdatering av statistiken	25
6.1	Insamling av datakällor	25
6.2	Framtagning av företagslistor via öppna data	26
6.3	Kvalitetskontroll	26
6.4	Manuella steg	27
6.5	Beräkning av företagsdata från Proff	28
7	Rekommendation för underhåll och vidare utveckling av metoden	29
7.1	Löpande kvalitetssäkring	29
7.2	Klassificering av marknadsaktiviteter	29
8	Appendix	32
8.1	Överblick över inkluderings- och exkluderingskriterium	32
8.2	Exkluderande SNI-koder	33
8.3	Nyckelord för identifiering	35
8.4	Nyckelord för klassificering	39
8.5	Exkluderade nyckelord	42
8.6	Variabellista från Proff	43
8.7	Sök-queries för patent och publikationer	44

1 Inledning

ADC (Amsterdam Data Collective) har på uppdrag av och tillsammans med Vinnova vidareutvecklat den modell för att identifiera företagspopulationen inom life science-sektorn i Sverige som utvecklades under 2022–2023 inom ramen för regeringsuppdraget N2021/02243. Utöver fjolårets arbete tar arbetet även denna gång utgångspunkt i tidigare utförda kartläggningar av företagspopulationen genomförda av Tillväxtanalys och Vinnova. Modellutvecklingen har vidare tagit hänsyn både till de rekommendationer som Tillväxtanalys lämnade inom ramen för sitt regeringsuppdrag att utreda metoden för att följa life science-företag (Regeringsuppdrag N2020/02707) och ADC:s rekommendationer om vidareutveckling av metoden från fjolårets metodbeskrivning.¹

Målsättningen i arbetet har varit att vidareutveckla fjolårets modell som med en hög grad av automatisering kan identifiera företagspopulationen inom life science-sektorn. Detta för att säkerställa långsiktighet i projektet och i förlängningen möjliggöra för återkommande kartläggningar och statistik för sektorn.

Arbetet har genomförts i nära samarbete mellan ADC och Vinnova. Vinnova har varit direkt involverad i samtliga delar av projektet och olika inriktningsbeslut har tagits gemensamt och löpande. Detta dokument syftar till att vara en transparent och tillgänglig dokumentation av metoden för kartläggning och berikning av företagspopulationen inom life science. Metoden beskrivs även för att understödja överlämning av arbetet till Vinnova samt möjliggöra förbättringar eller vidareutveckling genom andra parter i framtiden. Utöver detta dokument omfattar leveransen av arbetet även bilagor i form av företagslistor i Excel-format, en utvecklingskatalog i Vinnovas AzureDevOps innehållandes versionshantering och de R-script som används till att ta fram samt klassificera företagspopulationen inom life science. Slutligen omfattar leveransen en word-fil som kan användas som handbok till R-scripten.

¹ <https://www.vinnova.se/globalassets/mikrosajter/hallbar-precisionshalsa/life-science-a4-0326-bilaga-1-230414.pdf>

2 Arbetsprocessen

Årets arbete tar avstamp i arbetet från 2022–2023 med att identifiera företagspopulationen inom life science-sektorn samt kategorisera dessa företag i olika branscher. Nedan följer en summering av 2022–2023 års arbete, som låg till grund för Vinnovas första rapport av statistik inom life science-sektorn², samt hur arbetsprocessen i vidareutvecklingen av metoden har gått till.

2.1 Fjolårets arbetsprocess 2022/2023

Arbetet har genomförts i tre övergripande faser. I den första utforskande fasen sammanställdes existerande kunskap på området och olika möjliga datakällor identifierades. Genom workshops och dialogmöten med en av Vinnova sammansatt referensgrupp samlades olika perspektiv in gällande metodval. Den utforskande fasen resulterade i en genomförandeplan som utgjorde grunden för det fortsatta utvecklingsarbetet.

Ett viktigt inriktningsbeslut som togs under den utforskande fasen och på begäran av Vinnova var att arbetet med att kartlägga företagspopulationen, så väl som klassificering i branscher, helt skulle baseras på öppna datakällor³. Huvudanledning till varför detta beslut togs var för att möjliggöra identifiering och således kvalitetssäkring av företagen, vilka var viktiga delar av projektets andra fas. Vidare innebär valet att arbeta med öppna datakällor till ökad transparens och möjligheten för andra parter att nyttja statistiken utifrån sina behov.

I projektets andra fas utvecklades själva metoden i nära samverkan mellan Vinnova och ADC. Arbetet strukturerades i sprintar som genomfördes gemensamt av Vinnova och ADC. Mellan sprintarna genomfördes utvecklings- och kvalitetsarbete av både ADC och Vinnova. Arbetet genomfördes iterativt och inriktningsbeslut togs löpande och gemensamt.

Arbetet i utvecklingsfasen genomfördes i tre övergripande steg vilka syftade till att 1) identifiera, 2) selektera och avslutningsvis 3) klassificera företag. Respektive steg beskrivs kortfattat nedan:

1. Det första steget syftade till att identifiera relevanta företag: Detta inbegrep att identifiera och utvärdera möjliga datakällor som kunde bidra till att identifiera företag i life science-sektorn. Utifrån dessa datakällor genererades en bruttolista av företag.
2. Det andra steget syftade till att utifrån den framtagna bruttolistan selektera de företag som bedömdes tillhöra life science-sektorn utifrån de inkluderings- och exkluderingskriterier som sattes upp. Arbetet genomfördes iterativt och kvalitetssäkrades löpande. Resultatet från detta steg var en nettolista med företag som bedömdes tillhöra life science-sektorn.
3. I det avslutande steget klassificerades företagspopulationen som selekterats i steg ovan. Arbetet utgick från en klassificering som utvecklats tillsammans med Vinnova där branschföreträdare fått möjlighet att komma med inspel. Arbetet resulterade i ett första försök till klassificering som skulle fortsätta utvecklas i årets vidareutvecklingsarbete.

² [Statistik över svenska life science-företag \(vinnova.se\)](https://www.vinnova.se/statistik/over-svenska-life-science-foretag)

³ Öppna data är data, eller information, som vem som helst kan använda utan avgift.

Efter att företagspopulationen selekterats berikades populationen med officiella företags- och individdata från SCB i linje med den data som tidigare sammanställts av Tillväxtanalys.

2.2 Arbetsprocess för vidareutvecklingen 2023/2024

Det vidareutvecklande arbete som skett under 2023/2024 kan även delas upp i tre delar: vidareutveckling av metoden för identifiering av företagspopulationen inom life science, utforskande av nya frågeställningar och berikning av data kopplat till företagspopulationen.

Målet har varit att utveckla en helautomatisk metod baserad på öppna datakällor. Beslutet att arbeta med öppna datakällor har tagits för att öka metodtransparens samt möjliggöra kvalitetssäkring av populationen utifrån de metodval som görs. För att kunna kvalitetssäkra implementering av nya regler samt kunna medge viss flexibilitet kopplat till en föränderlig sektor krävs dock vissa manuella steg. Ambitionen är att dessa ska vara så få som möjligt.

Vidareutvecklingen av metoden har bestått av att genomföra förbättringar av fjolårets klassificeringar av företag i olika branscher samt att klassificera vilka marknadssegment företagen är verksamma inom. Arbetet har strukturerats i sprintar som genomförts gemensamt av Vinnova och ADC, där utvecklings- och kvalitetsarbetet har genomförts mellan sprintarna av både ADC och Vinnova.

I den utforskande delen har frågeställningar kopplade till företag med noll anställda, jämställdhetsfrågor, regionala skärningar, koncerntillhörighet, utländskt ägande samt investeringar i och finansiering av life science-företag undersökts. Utforskande av frågeställningarna har genomförts genom workshops tillsammans med experter på Vinnova, involvering av regionala aktörer och den av Vinnova sammansatta referensgruppen och SCB. Därefter har ADC undersökt möjligheten att inkludera variabler för att kunna besvara frågeställningarna.

Baserat på resultaten från vidareutvecklingen och utforskandet av nya frågeställningar har slutligen en marginellt uppdaterad företagspopulation tagits fram som även har berikats med företagsdata från Proff.

Till skillnad från fjolårets arbete har ADC inte berikat populationen med officiella företags- och individdata från SCB. I årets arbete har ADC tagit fram företagspopulationen inom life science enligt hur processen beskrivs i denna metodrapport. ADC har sedan delat med sig av företagspopulationen samt det regelverk som krävs för att ta fram företagspopulationen till SCB. Därefter har SCB tagit fram en egen företagspopulation baserat på det regelverk ADC utvecklat tillsammans med Vinnova och berikat den med officiella företags- och individdata från SCB. Inom ramen för detta uppdrag berikas även företagspopulationen med data från andra källor för analys, exempelvis Proff och Svenska Riskkapitalföreningen.

3 Datakällor

I detta kapitel beskrivs de datakällor som används i nuvarande arbete för att identifiera, selektera, klassificera, och förbättra företagspopulationen. Slutligen beskrivs företagsdata från Proff som används till att berika företagspopulationen.

3.1 Identifiering, selektering, klassificering av företagspopulationen

Ett stort antal öppna datakällor utvärderades under arbetet 2022/2023 för att kartlägga företagspopulationen inom life science. Arbetet utgick från att datakällorna var öppna för att genomgående under arbetet kunna kvalitetssäkra företagen som identifieras samt jämföra utfallet av olika inkluderings- och exkluderingskriterier. I detta vidareutvecklande arbete har två datakällor från det tidigare arbetet exkluderats samtidigt som fyra ytterligare datakällor har undersökts.

Datakällorna som ej längre används är Dealroom och ansökningsdata från Vinnova och EU Horisont. Dealrooms databas bidrog inte till att ytterligare företag identifierades varför datakällan bedömdes som överflödigt. Efter en manuell kvalitetskontroll av ansökningsdata från Vinnova och EU horisont bedömdes denna datakälla vara för ojämn i vilka företag som inkluderas – en stor majoritet av företagen identifierades som tillhörande andra sektorer än life science varför datakällan ej längre används i detta arbete.

Av de fyra datakällor som undersökts har tre lagts till. En (Naturvårdsverket) används till att selektera företag medan två (PATSTAT och OpenAlex) används till att klassificera företagen utifrån vilket marknadssegment de är verksamma inom. Utöver dessa tre nya datakällor undersöktes även data över kliniska studier.⁴ Data över kliniska studier förbättrade inte arbetet varför beslut togs om att ej addera dessa datakällor. Tabell 3.1 sammanfattar de datakällor som använts i det arbetet. Respektive datakälla beskrivs därefter mer detaljerat i löptext.

TABELL 3.1

Datakälla	Beskrivning	Ursprung
Branschföreningar	Listor över medlemmar i branschföreningarna ASCRO, LIF, Swedish Labtech, Swedish Medtech och SwedenBIO.	Inrapporterad information från branschföreningar.
Proff	Utdrag ur näringslivsregistret omfattande SNI-koder, verksamhetsbeskrivningar, omsättning och antal anställda.	Bokslutsdata hos Bolagsverket via tjänsten Proff.

⁴ Följande datakällor över kliniska studier har undersökts: <https://euclinicaltrials.eu/>, <https://clinicaltrials.gov/>, <https://www.who.int/clinical-trials-registry-platform/the-ictrp-search-portal> och https://www.isrctn.com/?gclid=EAlaIqobChMI9_6Xk-u9-AIV2BkGAB3BXgEHEAAYASAAEgJC5vD_BwE.

EUDAMED	Utdrag av svenska aktörer registrerade i EU-databasen EUDAMED över registrerade medicintekniska produkter.	Hämtad från EUDAMED:s webbplats genom sökning efter ekonomiska operatörer med säte i Sverige. (länk)
Nationellt produktregister för läkemedel (NPL)	Utdrag av aktörer registrerade i Nationellt produktregister för läkemedel (NPL).	Data hämtas från Läkemedelverkets webbsida där senaste versionen av NPL kan hämtas vilken innehåller en sammanställning av organisationer. (länk)
Tillväxtanalys	Företag som identifierades som life science-företag av Tillväxtanalys 2018.	Från Tillväxtanalys rapport från 2018. (Länk)
Naturvårdsverket*	Anläggningar för industriell framställning av läkemedel genom kemiska eller biologiska processer	Hämtad från Naturvårdsverkets hemsida. (Länk)
PATSTAT*	Patentansökningar inom life science-kategorierdata	Data hämtas från Europeiska patentorganisationens databas.**
OpenAlex*	Vetenskapliga publikationer	Data hämtas via den öppna publikationsdatabasen OpenAlexs API.**

* Indikerar att dessa källor är nya för arbetet under 2023/2024.

** De sök-queries som används för att identifiera patent och publikationer inom life science presenteras i Appendix 8.7.

3.1.1 Medlemslistor branschföreningar

Branschföreningarna inom life science (ASCRO, LIF, Swedish Labtech, Swedish Medtech, SwedenBIO) är en viktig informationsmängd för att kartlägga företagspopulationen. Listorna samlas in direkt från branschföreningarna och innehåller medlemskapsinformation för år 2022 (Swedish Medtech) och 2023 (ASCRO, Lif, SwedenBIO och Swedish Labtech). Listorna innehåller företag som är medlemmar i respektive branschförening. Listorna har hög träffsäkerhet då en stor andel av medlemmarna kan definieras som tillhörande life science-sektorn. Underlaget utgör även en viktig datamängd i klassificeringen av företag tillhörande olika branscher. Samtidigt som medlemslistorna har hög träffsäkerhet är täckningsgraden begränsad då endast en mindre del av den samlade företagspopulationen inom life science är medlemmar i en branschförening. Att medlemskapsinformation endast representerar nuvarande medlemmar innebär en osäkerhet i historisk statistik. Detta innebär exempelvis att företag som tidigare varit medlemmar men gått i konkurs eller blivit uppköpta inte kommer att identifieras bak i tiden.

3.1.2 Företagsinformation

Företagsinformation för senast tillgängliga bokslutsår (2022 eller 2023) hämtas från årsredovisningar som registrerats hos Bolagsverkets via tjänsten Proff. Årsredovisningar sammanställer information om samtliga svenska organisationers räkenskaper och förvaltning för ett räkenskapsår. Variabler som används i arbetsprocessen omfattar bland annat SNI-koder, verksamhetsbeskrivningar, antal anställda och omsättning. Från dessa data identifieras företag genom sökningar efter företag med utvalda SNI-koder. Vidare används SNI-koder i kombination i inkluderings- och exkluderingskriterier. Se kapitel 4 för beskrivning av urval av SNI-koder.

3.1.3 EUDAMED

EU-databasen EUDAMED innehåller information över registrerade medicintekniska produkter. Produkter som omfattas av såväl MDR (EU-förordningen om medicintekniska produkter, trädde i kraft 2021) som IVDR (EU-förordningen om medicintekniska produkter för in vitro-diagnostik, trädde i kraft i maj 2022) ska registreras i denna databas. Registret har uppgifter om de aktörer som har registrerade produkter i databasen och det är denna information som används i arbetet. Information från EUDAMED används i samtliga steg (identifiera, selektera och klassificera) av processen. Databasen har hög träffsäkerhet gällande företag som tar fram medicintekniska produkter samtidigt som den är begränsad genom att den endast innehåller information om medicintekniska produkter som registrerats sedan databasen upprättades från och med 2022.

3.1.4 Nationellt produktregister för läkemedel (NPL)

Nationellt produktregister för läkemedel, NPL, är ett register med information om samtliga läkemedel som är godkända i Sverige. Registret upprätthålls av Läkemedelsverket. Registret har uppgifter om de aktörer som har registrerade produkter i databasen och det är denna information som används i arbetet. Information från NPL används i samtliga steg (identifiera, selektera och klassificera) av processen. NPL används för att identifiera, selektera och klassificera läkemedelsföretag.

3.1.5 Naturvårdsverket

Naturvårdsverket har det svenska utsläppsregistret Utsläpp i siffror på uppdrag av regeringen. Utsläpp i siffror är det svenska svaret på Kievprotokollet. Det ska underlätta för medborgare att ta del av information om utsläpp av farliga ämnen till miljön. Registret är också ett redskap för att öka företagens ansvar, samt att minska utsläppen och nedsmutsningen av miljön. Registret har bland annat uppgifter om utsläpp från anläggningar för industriell framställning av läkemedel genom kemiska eller biologiska processer. Denna datakälla används i kombination med SNI-koder för att identifiera life science-företag.

3.1.6 PATSTAT

PATSTAT underhålls av Europeiska patentorganisationen (EPO) och tillhandahåller omfattande information om patent världen över. Den täcker patent från mer än 100 länder och inkluderar data om patent, bibliografiska detaljer, rättslig status, patentansökande och uppfinnare. Patentansökningar

registrerade i PATSTAT används till att klassificera life science-företag verksamma inom marknadssegmentet forskning och utveckling.

3.1.7 OpenAlex

OpenAlex är en öppen publikationsdatabas skapad av den ideella verksamheten <https://ourresearch.org/>. OpenAlex aggregerar och standardiserar data från en mängd olika publikations- och citeringskällor, däribland [ORCID](#), [PubMed](#), [Microsoft Academic](#) och [Crossref](#). Databasen innehåller data över vetenskapliga publikationer och används till att klassificera life science-företag verksamma inom marknadssegmentet forskning och utveckling.

3.1.8 Tillväxtanalys

Ytterligare en datakälla som har används är listan över företag som identifierades som life science-företag av Tillväxtanalys 2018.⁵ Totalt återfinns 3356 företag på denna lista som togs fram under 2017. I detta projekt används företagslistan från Tillväxtanalys i den första identifieringen av företag till en bruttolista samt i kvalitetssäkrande syfte.

3.2 Berikning av företagsdata

I tabell 3.2 presenteras och beskrivs datakällorna från Proff och Riskkapitalföreningen som används till att berika företagsdata. Variablerna har valts baserat på den statistik som efterfrågats inom ramen för förfrågningsunderlaget. I appendix 8.6 finns variabellistan från Proff över samtliga variabler som används för att berika data.

TABELL 3.2

Källa	Beskrivning	Ursprung
Proff	Statistiken innehåller nyckeltal på företagsnivå, arbetsställe, koncern tillhörighet, ordförande och styrelseledamöters kön.	Bokslutsdata hos Bolagsverket via tjänsten Proff.
Riskkapitalföreningen	Investeringar i privatägda företag som oftast inte finns på börsen. Innefattar i den här rapporten både venturekapital i företag i tidiga faser, och investeringar som görs i mer mogna företag.	Dessa data är från den svenska riskkapitalföreningen Swedish Private Equity & Venture Capital Association (SVCA) som i sin tur samlar in data tillsammans med den

5

Europeiska
riskkapitalföreningen
Invest Europe och sexton
europiska
riskkapitalföreningar.
Invest Europe samlar in
och bearbetar
riskkapitaldata för en rad
europeiska länder.

4 Metodbeskrivning

Årets arbete har bland annat haft som mål att förbättra fjolårets klassificeringar av företagens branschtillhörighet inom företagspopulationen samt vilka marknadssegment företagen är verksamma inom. Årets metod för att identifiera, selektera och klassificera företagspopulationen följer således i stora drag den logik och de regler som utvecklades under fjolårets arbete:

- Identifiera – fokuserar på att säkra analysens täckningsgrad genom att säkra att alla företag inom life science sektorn identifieras genom datakällorna.
- Selektera – fokuserar på att säkra exakthet i analysen genom att endast företag som är verksamma inom life science inkluderas.
- Klassificera – fokuserar på att klassificera selekterade life science-företag utifrån följande branscher: läkemedel, medicinteknik, bioteknik, labtech och diagnostik.

Utöver dessa tre steg har årets modell byggts ut genom att även klassificera vilka marknadssegment företagen är verksamma inom. I årets arbete har följande två marknadssegment varit möjliga att klassificera bland företagen: 1) forskning och utveckling samt 2) konsult.

Detta kapitel syftar till att beskriva den uppdaterade metoden i sin helhet. För att underlätta för läsaren inleds kapitlet med en sammanfattning av genomförda förändringar. Därefter följer en beskrivning av metoden i sin helhet, inklusive genomförda förändringar.

4.1 Förändringar och kvalitetssäkrande arbete under 2023/2024

Nedan sammanfattas genomförda förändringar och uppdaterat kvalitetssäkrande arbete jämfört med fjolårets rapport. Avsnittet är uppdelat utifrån metodens olika steg.

4.1.1 Identifiering

- Sammanställningen av företag görs mot organisationsnummer och för datakällor där organisationsnummer saknas matchas dessa på namn mot Näringslivsregistret. Efter att matchningen på namn har skett genomförs en manuell kvalitetskontroll av de matchade företag för att säkerställa att namnmatchningen har genomförts korrekt. Därefter berikas företagen med organisationsnummer genom tjänsten Proff. Nytt för årets modell är att samtliga företag som matchas på namn markeras för att underlätta den manuella kvalitetskontrollen.
- Dealroom och ansökningsdata från Vinnova (område: Hållbar precisionshälsa) och EU horisont-ansökningar används ej eftersom de inte bidrog med nya företag utöver de som redan identifieras via de andra datakällorna.

4.1.2 Selektion

- Naturvårdsverkets lista över företag med anläggningar för industriell framställning av läkemedel genom kemiska eller biologiska processer läggs till som datakälla för selektion till

nettolistan. Om ett företag från bruttolistan finns med på denna lista samt har en av de SNI-koder som presenteras i Tabell 4.1 selekteras företaget till att inkluderas på nettolistan.

- Inkluderingskriteriet från fjolårets modell kopplad till Dealroom används ej i år.
- Ett exkluderingskriterium som skulle varit med i fjolårets selekteringsprocess men av misstag inte användes har nu lagts till. Kriteriet innebär att ett företag som har SNI-kod 32.502 och inte använder ordet "utveckling" i sin verksamhetsbeskrivning exkluderas givet att någon av företagets SNI-koder inte tillhör en SNI kategori 1 (se Tabell 4.2). Att kriteriet inte användes i fjolårets arbete resulterade i att 271 företag som ej skulle inkluderats inkluderades. Dessa är i år exkluderade.
- 15 ytterligare exkluderingsnyckelord har lagts till.

4.1.3 Klassificering

- Klassificeringen av konsulter flyttas till den nya marknadssegmenteringsdelen.
- Om ett företag har SNI-kod 32.502 blir företaget endast klassificerat som medicinteknik och ingen annan bransch.
- Tre nyckelord har uppdaterats utifrån vilka branscher de klassificerar företag till. ADC och Vinnova gick tillsammans igenom samtliga nyckelord som används i branschklassificeringen i syfte att minimera att företag klassificeras i fler än en bransch. Följande tre ord var de enda som bedömdes kunde justeras enligt detta syfte:
 - "Cellteknik" har gått från att klassificera företag till branscherna läkemedel och biotech till endast läkemedel
 - "Kemiteknik" har gått från att klassificera företag till branscherna läkemedel och biotech till endast läkemedel
 - "Biosensorer" har gått från att klassificera företag till branscherna labtech och medicinteknik till läkemedel och biotech.⁶

4.1.4 Marknadssegmentering

Detta är ett nytt steg som genomförts i årets arbete jämfört med fjolårets. Vidareutvecklingen av metoden har bestått av att klassificera vilka marknadssegment företagen är verksamma inom. Företagen har klassificerats utifrån huruvida de bedriver forskning och utveckling samt om de bedriver konsultverksamhet. I den utvecklade metoden är det möjligt att bedriva forskning och utveckling och samtidigt bedriva konsultverksamhet. Detta steg beskrivs i detalj i avsnitt 4.5.

⁶ Uppdateringen av hur ordet "biosensorer" klassificerar företag är ett misstag och kommer ändras i kommande uppdateringar av metoden att klassificeras till branscherna labtech och medicinteknik. Totalt rör detta två företag.

4.1.5 Berikning av företagsdata

Ett ytterligare nytt steg för årets arbete är att företagsdata berikas med koncerntillhörighet, placering av företagens arbetsställen och könsfördelning bland företagens VD och styrelsemedlemmar.

4.2 Identifikation av företag

En bruttolista över företag som bedöms täcka majoriteten av den svenska life science-sektorn år 2023 identifieras genom sökningar i tidigare beskrivna databaser och resultaten sammanställs. Sammanställningen görs mot organisationsnummer. För datakällor där organisationsnummer saknas matchas dessa på namn mot Näringslivsregistret. Efter att matchningen på namn har skett genomförs en manuell kvalitetskontroll av de matchade företagen för att säkerställa att namnmatchningen har genomförts korrekt. Därefter berikas företagen med organisationsnummer med data från Proff. Företagen berikas även med data över placering av huvudkontor, SNI-koder, verksamhetsbeskrivning, antal anställda och omsättning i detta steg.

Följande kriterier används för sammanställning av bruttolistan:

- Samtliga företag som återfinns i datamängderna EUDAMED, NPL, medlemslistor från branschföreningar samt listan över life science-företag som identifierades av Tillväxtanalys rapport 2018 inkluderas i bruttolistan.
- Företag som har någon av nedanstående SNI-koder inkluderas i bruttolistan.

TABELL 4.1

SNI-kod	Branschbeteckning
26.600	Industri för strålningsutrustning samt elektromedicinsk och elektroterapeutisk utrustning
21.100	Industri för farmaceutiska basprodukter
21.200	Läkemedelsindustri
32.501	Industri för medicinska och dentala instrument och tillbehör
32.502	Tandteknikerlaboratorier
46.460	Partihandel med medicinsk utrustning och apoteksvaror
72.110	Biotekniska FoU-institutioner
72.190	Andra naturvetenskapliga och tekniska FoU-institutioner
86.901	Medicinska laboratorier o.d.

Urvalsprocessen i fjolårets arbete med att ta fram dessa SNI-koder skedde genom att ADC tog ut fullständiga företagslistor via näringslivsregistret Proff för ett stort urval av SNI-koder. Därefter matchades företagen i respektive SNI-kod-lista i ett första steg mot företagslistan från Tillväxtanalys rapport i syfte att bedöma SNI-kodens relevans ur ett life science-perspektiv. Om företagen inom en SNI-kod hade hög träffsäkerhet gentemot Tillväxtanalys företagslista identifierades denna SNI-kod som relevant. Om företagen inom en SNI-kod hade låg träffsäkerhet gentemot Tillväxtanalys företagslista men ändå bedömdes som relevant för life science-sektorn, exempelvis 62.010 (dataprogrammering)

och 71.122 (teknisk konsultverksamhet inom industriteknik), granskades dessa företag manuellt av ADC. Om det under denna granskning framkom att företagen inom dessa SNI-koder i relativt stor utsträckning var relevanta för life science-sektorn inkluderades dessa SNI-koder. Den manuella granskningen ledde dock inte till att ytterligare SNI-koder lades till.

Sammantaget identifieras i årets arbete 10 614 företag i detta steg vilket utgör resultatet från steg 1.

4.2.1 Manuella, kvalitetssäkrande steg

För datakällor (exempelvis EUDAMED) där organisationsnummer saknats har dessa företag matchats på namn mot Näringslivsregistret. Efter att matchningen på namn har skett har Vinnova och ADC genomfört en manuell kvalitetskontroll av de matchade företag för att säkerställa att namnmatchningen har genomförts korrekt. Därefter har företagen berikats med organisationsnummer genom tjänsten Proff.

4.3 Selektion av företag

Metodens andra steg syftar till att selektera de företag som har verksamhet som faller inom den definition av life science som har tagits fram för projektet. I metoden görs definitionen om till ett format som kan arbetas med för att hitta företagen i datamängderna och representeras därigenom av ett antal inkluderings- och exkluderingskriterier. Detta arbete baseras på de datakällor som presenterades i Tabell 3.1. Dessutom berikas företagen med följande data: koncern tillhörighet, placering av företagets arbetsställen och könsfördelning bland företagets VD och styrelsemedlemmar. Sammantaget används fem inkluderingskriterier och fyra exkluderingskriterier.⁷

4.3.1 Inkluderingskriterier

Totalt ställs fem inkluderingskriterier upp. Företag skulle potentiellt kunna fångas upp av fler än ett kriterium men inkluderas utifrån första träffen. Följande inkluderingskriterier ställs upp:

- Inkluderingskriterium 1 - Företaget har första eller andra SNI-kod som återfinns i kategori 1⁸.
- Inkluderingskriterium 2 - Företaget återfinns på EUDAMED- eller NPL-listan och har första eller andra SNI-kod som återfinns i kategori 1 eller 2.
- Inkluderingskriterium 3 - Företaget återfinns på någon av medlemslistorna år 2022/2023 från branschorganisationerna ASCRO, LIF, Swedish Labtech, Swedish Medtech eller SwedenBIO och har första eller andra SNI-kod som återfinns i kategori 1 eller 2.
- Inkluderingskriterium 4 - Företaget återfinns på Naturvårdsverkets lista över företag med anläggningar för industriell framställning av läkemedel genom kemiska eller biologiska processer och har första eller andra SNI-kod som återfinns i kategori 1 eller 2.
- Inkluderingskriterium 5 - Minst ett ord i företagets verksamhetsbeskrivning matchar mot nyckelordlista.

⁷ Se Appendix 8.1 för överblick över hur respektive inkluderings- och exkluderingskriterium påverkar företagspopulationen.

⁸ Se hur SNI-koderna fördelas mellan kategorierna under rubriken "Användning av SNI-koder" längre ned på sidan.

Användning av SNI-koder:

SNI-koder grupperas i två kategorier: kategori 1 för primära life science-SNI-koder och kategori 2 för sekundära life science-SNI-koder:

TABELL 4.2

Kategori	SNI-koder
1 (primära)	21.100, 21.200, 26.600
2 (sekundära)	32.501, 32.502, 46.460, 72.110, 72.190, 86.901

Företag kan ha flera SNI-koder kopplade till sig. Ordningen av SNI-koderna bestäms av företagen själva men ska i regel spegla verksamhetens inriktning. I selektionskriterierna i denna arbetsprocess används kriterier kopplade till SNI-kodernas ordning där enbart företag med relevanta (för life science-sektorn) första eller andra SNI-kod inkluderas.

Användning av nyckelord:

ADC och Vinnova har för uppdraget utvecklat en nyckelordlista som innehåller nyckelord med bäring mot life science. Nyckelordlistan tar utgångspunkt i ett tidigare arbete som är genomfört av SCB där nyckelord har identifierats baserat på ord som återfinns i företags verksamhetsbeskrivningar. Inom projektet har den ursprungliga listan bearbetats vidare för att matcha den definition av life science som används inom projektet. Nyckelordlistan används i inkluderingskriterium 5.

Nyckelordlistan som används för selektion av life science-företag innehåller 522 nyckelord och återfinns som appendix till denna rapport.

4.3.2 Exkluderingskriterier

Efter att företag selekterats genom inkluderingskriterium kan företag selekteras bort baserat på fyra exkluderingskriterier. Dessa syftar till att ta bort företag som bedöms falla utanför den definition som används för life science inom detta arbete. Företag skulle potentiellt kunna fångas upp av fler än ett kriterium men exkluderas utifrån första träffen. De fyra exkluderingskriterierna genomförs i fallande ordning enligt nedan:

- Exkluderingskriterium 1 - Företag exkluderas baserat på en lista av SNI-koder som har bedömts som irrelevanta. Kriteriet gäller för företagets samtliga SNI-koder. Lista över dessa SNI-koder återfinns i appendix.
- Exkluderingskriterium 2 - Företag exkluderas baserat på en lista av nyckelord som har bedömts som irrelevanta. Dessutom har företag som bedöms vara verksamma som tex veterinärer och således inte anses tillhöra målpopulationen, exkluderas på basen av en kombination av ord: företag som nämner "veterinär" i sin verksamhetsbeskrivning och samtidigt inte nämner "human" eller "mänsklig" utesluts. Lista över exkluderingsnyckelord återfinns i appendix.
- Exkluderingskriterium 3 - Företag som inte är ett aktiebolag exkluderas.

- Exkluderingskriterium 4 - Företag som har SNI-kod 32.502 och inte använder ordet "utveckling" i sin verksamhetsbeskrivning exkluderas givet att någon av företagets SNI-koder inte tillhör en SNI kategori 1 (se Tabell 4.2).

Utöver dessa exkluderingskriterier har ADC och Vinnova löpande under projektet samt i den avslutande delen av kvalitetssäkringsprocessen manuellt gått igenom majoriteten av den identifierade företagspopulationen och tagit beslut om att exkludera företag. Trots att majoriteten av de identifierade företagen har granskats är det således möjligt att företag som borde inkluderas har exkluderats samt att inkluderade företag borde exkluderats. Granskningen har genomförts genom att läsa den identifierade företagspopulationens verksamhetsbeskrivningar samt undersöka dessa företags hemsidor och därefter göra en bedömning huruvida företagen tillhör life science-sektorn eller ej.

Användning av SNI-koder:

En lista över irrelevanta SNI-koder har upprättats baserat på analys av företag som inkluderats genom inkluderingskriterier. Listan omfattar ett stort antal SNI-koder som bedöms indikera att ett företag faller utanför den definition av life science-sektorn som används inom ramen för analysen. Ett exempel på företag som exkluderas är praktiserande läkare. SNI-kodslistan som används för exkludering återfinns i Appendix 8.2.

Användning av nyckelord:

En lista som innehåller nyckelord som används för att exkludera företag från att tillhöra företagspopulationen inom life science har under projektet utvecklats tillsammans av ADC och Vinnova. Nyckelordlistan för exkludering har utvecklats dynamiskt under projektet med huvudsyftet att säkerställa att företag som inte arbetar inom life science men som använder SNI-koder som i detta projekt klassificeras som life science-SNI-koder exkluderas. Exempel på nyckelord som exkluderats är *skidspårsutrustning* och *ölbryggeri*.

Nyckelordlistan som används för exkludering av företag innehåller 51 nyckelord och återfinns i Appendix 8.5. Jämfört med förra året har 15 ytterligare exkluderingsnyckelord lagts till.

4.4 Klassificering av branschtillhörighet

Metodens tredje steg fokuserar på klassificering av de identifierade och selekterade företagen i olika branscher. Precis som de två föregående stegen baseras data i detta steg på senast tillgängliga data, det vill säga 2022 eller 2023. Företagen kategoriseras i sex grupper:

- Läkemedel
- Bioteknik
- Medicinteknik
- Labtech
- Diagnostik
- Övrigt

En viktig aspekt att notera med klassificeringen är att företag kan klassificeras som tillhörande flera branscher. Exempelvis kan ett företag både klassificeras som tillhörande läkemedel och bioteknik om verksamhetsbeskrivningen innehåller olika nyckelord som klassificerats som nyckelord för läkemedel och bioteknik.

Klassificeringen är baserad på fyra olika faktorer som presenteras nedan:

- Medlemskap i branschföreningar
- Återfinns på NPL och EUDAMED
- Nyckelord i verksamhetsbeskrivningar eller företagsnamn⁹
- SNI-koder

Notera att samtliga datakällor som presenteras ovan förutom NPL-data och SNI-koder enbart är baserad på data från innevarande år. Detta innebär större osäkerhet i statistiken som sammanställs från längre bak i tiden eftersom medlemskap och verksamhetsbeskrivningar kan ha sett annorlunda ut än vad de gör idag. Om ett företag exempelvis var medlem i någon av branschorganisationerna under perioden 2009-2020 men sedan upphört att vara medlem kommer det således inte kunna klassificeras baserat på branschorganisationsmedlemskap eftersom historiska medlemslistor inte har använts i detta projekt. Dessa företag kan dock klassificeras baserat på någon av de andra ovan nämnda faktorerna.

Metoden för att klassificera företag har skett iterativt där ett stort antal klassificeringskriterier har testats varefter resultaten har verifierats av både ADC och Vinnova. Branschaktörer har även fått möjlighet att komma med inspel kopplat till klassificering av företag genom ett referensgruppmöte samt återkoppling till Vinnova.

Nyckelordlistan för klassificering av företagen innehåller 197 nyckelord och återfinns i Appendix 8.3.

4.4.1 Klassificering av läkemedelsföretag

Företag klassificeras som läkemedel om det återfinns på NPL-listan, om företaget är medlem i LIF, om företaget har ett nyckelord i sin verksamhetsbeskrivning som kategoriserats som läkemedel eller om nyckelordet "pharma" finns med i företagsnamnet. Listan över nyckelord som används för att klassificera företag som läkemedel återfinns i Appendix 8.4.

4.4.2 Klassificering av bioteknikföretag

Företag klassificeras som bioteknik om företaget har ett nyckelord i sin verksamhetsbeskrivning som kategoriserats som bioteknik eller om nyckelordet "bio" finns med i företagsnamnet. Listan över nyckelord som används för att klassificera företag som bioteknik återfinns i Appendix 8.4.

⁹ Tre nyckelord har uppdaterats utifrån vilka delsektorer de klassificerar företag till: "Cellteknik" har gått från att klassificera företag till delsektorerna läkemedel och biotech till endast läkemedel. "Kemiteknik" har gått från att klassificera företag till delsektorerna läkemedel och bioteknik till endast läkemedel. "Biosensorer" har gått från att klassificera företag till delsektorerna labtech och medicinteknik till läkemedel och biotech.

4.4.3 Klassificering av medicinteknikföretag

Företag klassificeras som medicinteknik om det återfinns på EUDAMED-listan, om företaget är medlem i Swedish Medtech, om företaget har ett nyckelord som kategoriserats som medicinteknik eller om företaget har SNI-kod 32.502. Om företaget har SNI-kod 32.502 blir företaget endast klassificerat som medicinteknik och inget annat. Listan över nyckelord som används för att klassificera företag som medicinteknik återfinns i Appendix 8.4.

4.4.4 Klassificering av labtech

Företag klassificeras som labtech om företaget har ett nyckelord som kategoriserats som labtech. Listan över nyckelord som används för att klassificera företag som labtech återfinns i Appendix 8.4.

4.4.5 Klassificering av diagnostik

Företag klassificeras som diagnostik om det har ett nyckelord i sin verksamhetsbeskrivning som kategoriserats som diagnostik. Listan över nyckelord som används för att klassificera företag som diagnostik återfinns i Appendix 8.4.

4.4.6 Övrigt-kategori

Under fjolårets arbete gjordes ett manuellt klassificeringsarbete för företagen som ej kunde klassificeras automatiskt utifrån ovan nämnda kriterier. Det rörde sig om cirka 600 företag. Företagen som klassificerades manuellt förra året har behållit den klassificering de manuellt tilldelades till detta års arbete. Företag som har varit svåra att klassificera tillhörande en specifik bransch har tilldelats en övrigt-kategori. Dessa 135 företag återfinns i en ny övrigt-kategori.

4.5 Klassificering av marknadssegment

Detta är metodens nyaste tillskott gällande att beskriva företagspopulationen inom life science. Steget syftar till att klassificera vilka marknadssegment företagen är verksamma inom. Detta steg identifierades i fjolårets rapport som en rekommenderad vidareutveckling av metoden och har varit ett prioriterat steg från Vinnovas sida i årets arbete. I fjolårets rekommendationer identifierades fyra möjliga marknadssegmenteringar: *forskning och utveckling, tillverkning, försäljning och marknadsföring* samt *konsulter*. I årets utforskande fas togs beslut om att undersöka möjligheten att inkludera ytterligare ett marknadssegment: *marknadsbolag*.

Under årets arbete har det utifrån identifierade datakällor endast varit möjligt att automatiskt identifiera företag tillhörande marknadssegmenten *forskning och utveckling* samt *konsulter*. Misslyckade försök har gjorts att automatiskt identifiera företag som *tillverkning, försäljning och marknadsföring* samt *marknadsbolag* men det har varit svårt att identifiera datakällor som på ett tillförlitligt sätt fångar in dessa företag. Definitionerna som har skapats har blivit för breda varför kvaliteten inte uppnått önskad nivå.

Kvalitetskontrollen av kriterier och företagen som blivit identifierade som tillhörande *forskning och utveckling* samt *konsult* har genomförts iterativt av ADC och Vinnova enligt samma metod som används

för att kvalitetssäkra branschklassificeringen. Notera att företagen i detta steg i linje med branschklassificeringen kan klassificeras både som *forskning och utveckling* samt *konsult*. Företag som inte identifierats som tillhörande någon av dessa kategorier har samlats i en övrigt-grupp. Nedan följer en detaljerad beskrivning av kriterier för att bli klassificerad som *forskning och utveckling* samt *konsult*.

4.5.1 Klassificering av forskning och utveckling

Företag klassificeras som forskning och utveckling om något av följande kriterier uppfylls:

- Företag har gjort minst 50 patentansökningar sedan 2009.
- Företaget har någon av de två SNI-koderna 72 110 eller 72 190 i kombination med någon av följande två kriterier:
 - Verksamhetsbeskrivningen inkluderar något av orden "forskning" eller "biovetenskap"
 - Företaget har publicerat minst en vetenskaplig artikel sen 2009.

Till kriteriet för patent finns ett undantag:

- Om företaget har gjort minst 50 patentansökningar men har noll anställda och tillhör en koncern klassificeras företaget inte som forskning och utveckling. En manuell kvalitetssäkring visade att dessa företag i hög utsträckning var holdingbolag för patent än ett företag som är aktivt verksamt inom forskning och utveckling.

4.5.2 Klassificering av konsult

Företag klassificeras som konsulter om det har ett nyckelord i sin verksamhetsbeskrivning som kategoriserats som konsulter eller om de har en SNI-kod som är klassificerat som "konsult" (70 220, 71.129, 71.122 och 62.020). Nyckelorden innehåller ord som förknippas med kontraktstillverkare, kontraktsforskning och regulatoriska konsulter. Listan över nyckelord som används för att klassificera företag som konsult återfinns i Appendix 8.4.

5 Känslighetsanalys

5.1 Om statistik för life science-sektorn

Utveckling av statistik för en sektor handlar om att definiera vilka företag som tillhör utvald sektor. Detta är ett komplext arbete och innebär ett stort antal avväganden. En genomgång av hur olika regioner i Sverige arbetar med statistik för life science-sektorn visar att många olika definitioner används och ingen allmänt accepterad definition existerar. Detta kan till stor del förklaras med att gränserna för en sektor sällan är skarp och företag verksamma inom life science -sektorn kan även vara verksamma inom andra sektorer. Ytterligare utmaningar kan härledas till att företag kan vara verksamma inom flera olika branscher inom samma sektor, till exempel medicinteknik och läkemedelsbranschen inom life science-sektorn.

ADC och Vinnova har i detta arbete vidareutvecklat metoden som togs fram i fjol vilken utgick från definitionen som används i den svenska life science-strategin och därefter gemensamt beslutat om hur den kan operationaliseras på bästa sätt.¹⁰

5.2 Avväganden i utvecklingsarbetet

5.2.1 Användning av öppna datakällor

I arbetet med att identifiera företagspopulationen inom life science-sektorn har flera olika datamängder utvärderats. Efter önskemål från Vinnova togs ett inriktningsbeslut i fjolårets arbete om att i så hög grad som möjligt använda öppna data om företag.¹¹ Det primära syftet med detta inriktningsbeslut var att kunna generera en lista över företag i sektorn där enskilda företag kan identifieras och således kvalitetssäkras. Samtidigt innebar inriktningsbeslutet att ett antal möjliga datakällor valdes bort. Primärt rörde sig detta om information om företagen som återfinns i SCB:s registerdata, bland annat kopplat till företagens exportverksamhet. Detta inriktningsbeslut står fast i årets vidareutveckling av metoden.

Det är vår övergripande uppfattning att inriktningsbeslutet inte har haft någon negativ påverkan på täckningsgrad och exakthet i den slutgiltiga företagslistan. Arbetet har inneburit en viss begränsning av de datakällor som kunnat användas. Samtidigt har inriktningen medfört en stor fördel genom att det varit möjligt att löpande granska vilka företag som inkluderats på listan.

5.2.2 Förutsättningar för automation

En uttalad målsättning även med årets arbete har varit att vidareutveckla metoden med syfte att i statistiken i möjligaste mån kan uppdateras automatiskt. Detta innebär att element som omfattar manuella genomgångar av företag i möjligaste mån begränsas. Detta har stora fördelar då det innebär

¹⁰ Regeringskansliet, (2019). En nationell strategi för life science, N2019.06

¹¹ De två icke-öppna datakällorna som används i arbetet är medlemslistorna från branschorganisationerna samt patentansökningsdatabasen PATSTAT.

att uppdateringen av statistiken genomförs mer effektivt och utan den stora arbetsinsats som tidigare krävts vid framtagning av statistiken hos både Tillväxtanalys och Vinnova, vilket identifierats som ett hinder för att upprätthålla en uppdaterad statistik.

Samtidigt som det finns stora fördelar med en automatiserad metod för att identifiera och klassificera företagspopulationen medför metoden risker. Framför allt föreligger det risker att företag som inte tillhör branschen inkluderas i populationen, att företag som tillhör branschen inte inkluderas, samt att stora företag men med en mycket begränsad life science-verksamhet inkluderas. Dessa utmaningar har stått i centrum i vidareutvecklingsarbetet.

För att minimera risken att företag inkluderas felaktigt har företagspopulationen kontinuerligt under arbetets gång granskats och kvalitetssäkrats av ADC och Vinnova, samt även i viss mån med hjälp av branschföreningarna i referensgruppen. Genom en iterativ process har regelverk utvecklats för att minimera antalet företag som identifieras och klassificeras felaktigt. Riskerna föreligger dock fortsatt och det är inte möjligt att garantera att samtliga företag är korrekt identifierade och klassificerade, nu och i framtiden. Vår bedömning är dock, baserat på det omfattande arbete som genomförts med att manuellt granska listorna som metoden genererat att antalet felklassificerade företag är få och att deras påverkan på den övergripande statistiken är begränsad.

I år har extra fokus lagts på att kvalitetssäkra de större företagen: de med fler än 50 anställda. Dessa företag påverkar i hög grad statistiken varför detta prioriterats. För att minimera risken att företag som tillhör sektorn felaktigt inte identifieras har framför allt två åtgärder genomförts: involvering av referensgruppen och jämförelser med den tidigare identifierade företagspopulationen.¹² Sammantaget är bilden att de företag som omfattas av definitionen som används i arbetet i hög grad identifieras genom metoden.

Ovan nämnda risker att felaktigt inkludera och exkludera företag ökar dessutom per definition vid en automatisk uppdatering av statistiken om sektorn eller de datakällor som används för identifiering och klassificering förändras. Exempel på förändringar som skulle påverka statistiken är uppdateringar av SNI-koder och verksamhetsbeskrivningar som görs att företag felaktigt exkluderas eller inkluderas. Vidare föreligger en risk att resurser avsatta för granskning av den identifierade populationen är begränsad. Detta bidrar sammantaget till en ökad risk att automatiseringen leder till en över tid försämrad statistik. Rutiner för hur detta kvalitetsarbete bör upprätthållas över tid beskrivs i avsnitt 6.3.

5.2.3 Förutsättningar för klassificering av branscher och marknadssegmentering

Branschklassificeringen och marknadssegmenteringen av företagen är det arbetsmoment som varit mest utmanande. Detta beror bland annat på att det finns överlapp mellan olika branscher, att företag kan vara verksamma i flera branscher samt avsaknad av relevanta och heltäckande datakällor. I jämförelse med referensarbetet som genomförts i Danmark (se avsnitt 5.2.2) kan exportverksamhet

12

https://www.tillvaxtanalys.se/download/18.62dd45451715a00666f194a2/1586366154910/pm_2018_06_Den%20svenska%20life%20science-industrins%20utveckling.pdf

inte användas för klassificering av verksamheter på grund av den initiala avvägningen att använda öppna datakällor för att identifiera, selektera och klassificera företagen.

Utifrån fjolårets rekommendationer om att bygga ut metoden och inkludera marknadssegmentering har resurser i år lagts på detta. Initialt fanns en önskan om att kunna marknadssegmentera företagen baserat på deras verksamhet inom *forskning och utveckling, tillverkning, försäljning* samt huruvida de är *marknadsbolag* eller *konsulter*. Avsaknad av relevanta eller heltäckande datakällor för att kunna klassificera företag som är verksamma inom tillverkning och försäljning eller identifiera de som marknadsbolag begränsar möjligheterna för att ta fram en automatisk lösning på detta. Däremot har datakällorna varit bättre sett utifrån att automatiskt identifiera företag som verksamma inom forskning och utveckling samt konsultverksamhet.

5.3 Datakällor och -variabler

Nedan presenteras en känslighetsanalys av datakällor och -variabler. De utvalda datakällorna- och variablerna är de vi bedömer är förknippade med störst risk i förhållande till att göra Vinnova beroende av externa parter eller förändras över tid.

5.3.1 SNI-koder

SNI-koder används för att identifiera, selektera och klassificera företag. Vi har ingen kontroll över vilka SNI-koder företagen väljer att förknippa sig med vilket innebär en risk sett till att dessa kan vara både irrelevanta och inaktuella eftersom de inte bestäms centralt av exempelvis Bolagsverket. Detta kan påverka statistiken över tid om en viss typ av företag är bättre eller sämre på att klassificera sig enligt SNI-koderna. Ytterligare faktor som måste tas hänsyn till är att företagens SNI-koder kan uppdateras mellan åren varför det årliga arbetet med att uppdatera företagspopulationen måste inkludera ett steg där företag som försvinner och läggs till i listan manuellt kontrolleras.

5.3.2 Verksamhetsbeskrivningar

Vi använder oss i huvudsak av verksamhetsbeskrivningar för klassificering av företag. Detta möjliggör en mer granulär segmentering än vad enbart SNI-koder tillåter, även om detaljnivån i verksamhetsbeskrivningarna varierar. Samtidigt är syftet med företagens verksamhetsbeskrivningar inte att de ska användas för att klassificera företag och företagen har således inte utformat dessa med hänsyn till att de ska användas för klassificering. Ytterligare faktor som måste tas hänsyn till är att företagens verksamhetsbeskrivningar kan uppdateras mellan åren varför det årliga arbetet med att uppdatera företagspopulationen måste inkludera ett steg där företag som försvinner och läggs till i listan manuellt kontrolleras.

5.3.3 EUDAMED och Nationellt produktregister för läkemedel (NPL)

Genom att använda EUDAMED och NPL får vi en insikt i vilken typ av marknad som företagen verkar på och vilka produkter de erbjuder. Samtidigt har vi ingen kontroll över hur registren används i praktiken. Vi gör oss också beroende av register som kan förändras över tid. Detta kan påverka statistiken på sikt.

5.3.4 Medlemslistor från branschföreningar

Branschlistorna ger oss insikt i företagens perspektiv utifrån vilken bransch de anser sig tillhöra. Vi har dock ingen kontroll över kriterierna företagen baserar sina val av branschorganisationsmedlemskap på, huruvida branschorganisationerna är villiga att dela med sig av medlemslistorna, eller vilken kvalitet medlemslistorna håller. Således gör vi oss även här beroende av register från extern part. Ytterligare riskfaktor med att använda branschlistor är att Vinnova gör sig beroende av att branschföreningarna vill dela med sig av uppdaterade medlemslistor.

5.3.5 Naturvårdsverket

Data från det svenska utsläppsregistret Utsläpp i siffror används i kombination med SNI-koder för att selektera life science-företag. Registret är det svenska svaret på Kievprotokollet och innehåller bland annat uppgifter om utsläpp från anläggningar för industriell framställning av läkemedel genom kemiska eller biologiska processer. Det saknas vid författandet av denna metodrapport en API-tillgång till registret varför det årliga arbetet med att uppdatera företagspopulationen måste inkludera ett manuellt steg där företag från detta register matchas med företag från bruttopopulationen vilket ökar resursförbrukning för statistikframställningen.

5.3.6 OpenAlex

Databasen innehåller data över vetenskapliga publikationer och används till att klassificera life science-företag verksamma inom marknadssegmentet forskning och utveckling. Data hämtas via publikationsdatabasen OpenAlex API. Företagen behöver sedan matchas mot nettolistan av företagspopulationen. Efter matchningen bör användaren kvalitetssäkra resultatet för att se att matchningen är gjord korrekt. Användandet av denna datakälla innebär således ytterligare ett manuellt kvalitetssäkringssteg.

5.3.7 PATSTAT

Patentansökningar registrerade i PATSTAT används till att klassificera life science-företag verksamma inom marknadssegmentet forskning och utveckling. Data hämtas och behöver sedan matchas mot nettolistan av företagspopulationen. Efter matchningen bör användaren kvalitetssäkra resultatet för att se att matchningen är gjord korrekt. Användandet av denna datakälla innebär således ytterligare ett manuellt kvalitetssäkringssteg. Utöver detta manuella steg bör noteras att data från PATSTAT vid skrivandet av denna metodrapport inte är tillgänglig utan kostnad, däremot har Vinnova prenumeration på tjänsten. En risk med användandet av denna datakälla är således att Vinnova gör sig beroende av en betaltjänst.

6 Rutiner vid uppdatering av statistiken

6.1 Insamling av datakällor

Vid uppdatering av statistiken ska information från externa datakällor inhämtas. Dessa beskrivs nedan.

TABELL 6.1

Datakälla/information	Beskrivning	Ursprung
Branschföreningarna ASCRO, LIF, Swedish Medtech, Labtech, Sweden BIO	Listor över medlemmar.	Inrapporterad information från branschföreningar i Excel-format.
Företagsinformation	Utdrag ur näringslivsregistret omfattande SNI-koder, verksamhetsbeskrivningar, omsättning och antal anställda.	Årsredovisningar hos Bolagsverket som hämtas via tjänsten Proffs API.
EUDAMED	Utdrag av svenska aktörer registrerade i EU-databasen EUDAMED över registrerade medicintekniska produkter.	Hämtad från EUDAMED:s webbplats genom sökning efter ekonomiska operatörer med säte i Sverige. (länk)
Nationellt produktregister för läkemedel (NPL)	Utdrag av aktörer registrerade i Nationellt produktregister för läkemedel (NPL).	Data hämtas från Läkemedelverkets webbsida där senaste versionen av NPL kan hämtas vilken innehåller en sammanställning av organisationer. (länk)
Naturvårdsverket	Anläggningar för industriell framställning av läkemedel genom kemiska eller biologiska processer	Hämtad från Naturvårdsverkets hemsida. (Länk)
PATSTAT	Patentansökningar inom life science-kategorierdata	Data hämtas från Europeiska patentorganisationens databas.
OpenAlex	Vetenskapliga publikationer	Data hämtas via den öppna publikationsdatabasen OpenAlexs API.

Identifiering, selektering och klassificering av företagen har genomförts baserat på data från 2022 eller 2023. Företagsinformationen från Bolagsverket utgår i huvudsak från data från år 2023 eftersom dessa baseras på företagets årsredovisningar. Resterande data uppdateras när den hämtas hem vilket för detta projekt har gjorts mellan oktober 2023-februari 2024.

För att få fram en ny nettolista över företag inom den svenska life science-sektorn för år 2023 behöver således data från dessa datakällor hämtas för senast möjliga år. Detta innebär att Vinnova måste begära ut nya medlemslistor från branschföreningarna, hämta nya data från EUDAMED, NPL och Naturvårdsverket. Dessutom behöver Vinnova göra nya utdrag från näringslivsregistret, PATSTAT och OpenAlex över företagsinformation och utifrån dessa uppdaterade datasammanställningar göra nya körningar av de R-script som utvecklats inom ramen för detta projekt för framtagning av uppdaterade företagslistor.¹³

6.2 Framtagning av företagslistor via öppna data

Avsnitt 6.1 beskriver de datakällor som används för att samla in underlag till att ta fram en bruttolista över företag som enligt de definitioner som tagits fram inom ramen för projektet kan tänkas identifieras som life science-företag. För att ta fram bruttolistan behöver dessa olika datakällor slås samman i en och samma bearbetningsbar fil. I detta projekt har vi valt att sammanställa data i en Excel-fil. Excel-filen är uppbyggd så att raderna innehåller unika företag medan företagsinformationen samt från vilka datakällor företaget inkluderats via visas i kolumnerna.

Det är på denna sammanställning av företagsdata (bruttolistan) Vinnova sedan behöver köra det R-script som utvecklats inom ramen för projektet för att identifiera, selektera och klassificera företagen. Detta R-script återfinns i en utvecklingskatalog på Vinnovas AzureDevOps. R-scriptet genomför samtliga kriterier kopplade till identifiering, selektering och klassificering som arbetats fram inom ramen för detta projekt. Slutprodukten av detta R-script är en nettolista över den svenska företagspopulationen inom life science för senast tillgängliga år (givet att datakällorna har uppdaterats med senast tillgängliga års data enligt vad som beskrivs i avsnitt 6.1).

6.3 Kvalitetskontroll

De listor som genereras från metoden bör kontrolleras inför varje uppdatering av statistiken för att säkerställa att företagen klassificerats enligt gällande definition. Utgångspunkten i utformningen av metoden är att företagen ska kunna klassificerats automatisk, men en årlig manuell genomgång kommer bidra med att säkerställa kvaliteten.

De nytillkomna företag som ej kan klassificeras automatiskt enligt framtagna metod och R-script behöver årligen granskas och klassificeras manuellt av Vinnova på motsvarande vis som gjorts inom ramen för detta projekt (se avsnitt 4.4.6 för beskrivning av hur residualen hanterats). Även företag som försvinner från företagspopulationen på grund av förändringar i SNI-koder eller verksamhetsbeskrivningar måste granskas. Denna typ av kontrollsteg är inbyggda i R-scriptet i form av att denna typ av förändringar markeras.

¹³ Se avsnitt 6.2 för beskrivning av processen för framtagning av företagslistor.

6.4 Manuella steg

Detta avsnitt beskriver samtliga manuella steg som används för att sammanställa resultatet från den semi-automatiska identifieringen av den svenska företagspopulationen inom life science.

Identifikation av företag

- Listorna från Naturvårdsverket och Eudamed går inte att exportera direkt till en fil och måste därför samlas in manuellt. Detsamma gäller för eventuella branschlistor om branschföreningen inte har möjlighet att skicka medlemslistor till oss.
- Följande listor ger inte organisationsnummer, varför organisationsnumret måste sökas på basen av företagsnamn: ASCRO, Swedish Labtech, Swedish Medtech, Eudamed, Naturvårdsverket, och NPL. Användaren måste manuellt granska om det har kommit nya företag på dessa listor och manuellt hämta dessa organisationsnummer från Proff.

Selektering

- Företag som under åren kvalitetssäkras och noteras som ej tillhörande life science-sektorn bör dokumenteras med organisationsnummer och namn samt läggas in i Excel-filerna "Identifierade icke-LS företag" och "Manuellt inkluderade företag". Därefter bör användaren köra om scriptet för att säkerställa att de manuella ändringarna har genomförts.
- Till nästa år måste företag som manuellt klassificerat som "Ej LS" tas bort från företagspopulationen. Dessa 15 företag noterades som icke tillhörande life science-sektorn men har av misstag inkluderats i företagspopulationen inom life science.
- När scriptet är kört bör användaren jämföra företagen som utgör nettolistan med föregående års nettolista. Detta för att upptäcka om de företag som exkluderats från den nya nettolistan är korrekt exkluderade, exempelvis om de har uppdaterat sina SNI-koder eller verksamhetsbeskrivningar.

Klassificering

- Efter att användaren har kört funktionen för branschindelning bör användaren gå igenom företagen som fått uppdaterad branschindelning i kvalitetsgranskningssyfte.
- Användaren bör sedan sortera den slutliga datafilen efter alla företag som har värdet 1 i variabeln "residual". Dessa företag bör manuellt klassificeras till olika branscher där möjligt. Om det inte är möjligt att manuellt klassificera företagen ska de klassificeras så att de hamnar i övrigt-kategorin. Användaren samlar in organisationsnummer och namn för dessa företag och skapar en ny kolumn där användaren anger den eller de branscher som hen anser att företaget tillhör. Sedan bör användaren föra in dessa data i filen "vinnova residual" och fliken "Residual_segmentering_22&24" i Excel. Användaren måste använda exakt samma namn som i Excel-filen för den branscher som användaren vill kategorisera företaget inom (t.ex. används "Medtech" för att fördela residualföretagen inom medicinteknik). Användaren måste använda samma struktur som finns i filen.

- Om användaren önskar ändra branschklassificering för något företag måste användaren först dokumentera organisationsnummer och namn för företaget. Därefter, enligt instruktionerna ovan, öppna Excel-filen "vinnova residual". I stället för att gå till fliken "Residual_segmentering" bör användaren gå till fliken "manuell_bransch_fördelning" och följa samma instruktioner som tidigare för att ange den eller de branscher som företaget ska tillhöra. Det är viktigt att användaren använder samma namn och struktur som i Excel-filen för att säkerställa korrekt hantering av branschfördelningen.
- Klassificering av branscher från ASCRO: Användaren bör årligen gå igenom medlemslistan från ASCRO och bedöma huruvida nyttillkomna företag är konsulter eller ej.

Marknadssegment

- Data över patent samt vetenskapliga publikationer måste hämtas via patentdatabasen PATSTAT och publikationsdatabasen OpenAlex. Företagen behöver sedan matchas mot nettolistan av företagspopulationen. För att matcha mot nettolistan måste en Fuzzy matchings-metod användas. Fuzzy innebär att metoden matchar två observationer som är lika varandra men inte exakta matchningar (i detta fall görs det på företagsnamn). Efter matchningen bör användaren kvalitetssäkra resultatet för att se att matchningen är gjord korrekt. Detta sker genom att jämföra namnet som användes för fuzzy-matchningen med det verkliga namnet.
- Kategorisering av företag som *forskning och utveckling* samt *konsult* sker helt automatiskt. Dock måste denna kategorisering kvalitetssäkras av användaren manuellt vid uppdateringar för att säkerställa att reglerna fortsatt är relevanta.

6.5 Berikning av företagsdata från Proff

De R-script som används för framtagande av statistik från Proff har överlämnats till Vinnova. Utöver nya djupdykningar Vinnova väljer att göra behöver scriptet endast uppdateras för att inkludera senast tillgängliga år i data.

7 Rekommendation för underhåll och vidare utveckling av metoden

7.1 Löpande kvalitetssäkring

Metoden som tagits fram för att kartlägga företagspopulationen inom life science är utformad för att med hög grad av automation identifiera och klassificera företag i sektorn. Som en del i utvecklingen av metoden har resultaten som genereras i respektive steg granskats ingående. Samtidigt kan förändringar i sektorn eller i de underliggande datakällorna påverka statistikens kvalitet. Av denna anledning är det viktigt att resultaten från metoden löpande kvalitetssäkras så att eventuella felaktigheter upptäcks och kan korrigeras. Vid uppdatering av metoden bör därför listorna över identifierade och klassificerade företag granskas på liknande sett som beskrivs i avsnitt 6.3.

Den löpande kvalitetssäkringen bidrar till att identifiera underhållsbehov. Eventuella fel kan delvis korrigeras manuellt, men kan även hanteras genom framtida utvecklingsarbete. Ett löpande kvalitetsarbete bör därför genomföras och omfatta översiktlig analys av den deskriptiva statistik som metoden levererar. Vidare bör de ingående listorna översiktligt granskas och feedback från aktörer i branschen systematiskt och löpande tillvaratas.

7.2 Klassificering av marknadsaktiviteter

Vinnova uttryckte inledningsvis i årets arbete en önskan om att utveckla en modell för att klassificera företagens marknadssegment, fördelat på forskning, tillverkning, produktion, försäljning och stödfunktioner (konsulter). Mot denna bakgrund har det i år lagts resurser på att få till denna klassificering. Efter interna arbetsmöten samt involvering av referensgruppen uppdaterades målbilden för klassificeringen till följande fyra marknadsaktiviteter: Forskning & utveckling, Tillverkning, Försäljning och marknadsföring och Marknadsbolag. Utöver dessa identifierades konsultföretagen som en intressant företagsgrupp. Konsultföretagens verksamhet identifierades dock sträcka sig vertikalt över de fyra andra marknadsaktiviteterna, snarare än att utgöra en isolerad marknadsaktivitet. Målbilden för klassificering av marknadsaktiviteter illustreras i Figur 7.1.

FIGUR 7.1

	Marknadsaktiviteter			
	Forskning & utveckling	Tillverkning	Försäljning och marknadsföring	Marknadsbolag
	Konsulter			
Läkemedel				
Medicinteknik				
Bioteknik				
Labtech				
Diagnostik				
Övrigt				

Utifrån denna målbild lyckades projektet endast med att hitta metoder för att klassificera marknadsaktiviteterna Forskning & Utveckling samt Konsulter. Företag som ej identifierades placerades i en övrigt-kategori. Se Figur 7.2 för illustration av hur företagen klassificeras som forskning & utveckling, konsulter och övrigt.

FIGUR 7.2

		Marknadsaktiviteter			
		Forskning & utveckling		Konsulter (Ej F&U)	Övrigt
		Ej konsulter	Konsulter		
Bransch	Läkemedel				
	Medicinteknik				
	Bioteknik				
	Labtech				
	Diagnostik				
	Övrigt				

En vidareutveckling av metoden för att klassificera övriga marknadsaktiviteter är således fortfarande önskvärd. Följande avsnitt beskriver hur ADC inom ramen för detta projekt har försökt klassificera företag utifrån marknadsaktivitet. Detta för att underlätta för framtida uppdateringar av metoden.

7.2.1 Forskning & utveckling

Se avsnitt 4.5.1 för hur företag verksamma inom forskning & utveckling har klassificerats. Ytterligare försök att klassificera företag verksamma inom forskning & utveckling som har testats men inte lyckats är:

- Kliniska studier: Data från fyra register över kliniska studier sammanställdes och matchades med selekterade life science-företag.¹⁴ Efter manuell kvalitetsgranskning konstaterades att denna metod inte hjälpte oss att klassificera företag verksamma inom forskning & utveckling – inte ens i kombination med andra kriterier (patentansökningar eller vetenskapliga publikationer).
- Forskarutbildade anställda: detta har inte testats i praktiken eftersom det skulle kräva användande av registerdata.
- Jobbannonser: tanken bakom att använda jobbannonser var att söka upp jobbannonser från selekterade life science-företag och se vilka som eftersökte forskande personal. Test gjordes på data från Platsbanken men dömdes bort av flera anledningar, bland annat: många företag

¹⁴ De fyra registren: <https://euclinicaltrials.eu/>, <https://clinicaltrials.gov/>, <https://www.who.int/clinical-trials-registry-platform/the-ictrp-search-portal> och https://www.isrctn.com/?gclid=EAlalQobChMI9_6Xk-u9-AIV2BkGAB3BXgEHEAAYASAAEgJC5vD_BwE.

annonserar inte jobbmöjligheter på Platsbanken och ej möjligt att koppla jobbannonser från rekryteringsbolag till det anställande företaget.

7.2.2 Tillverkning

Två försök gjordes för att identifiera tillverkande företag:

- SNI-kod: Vi testade att klassificera de selekterade life science-företagen som använde någon av SNI-koderna inom kategori C (Tillverkning) som verksamma inom tillverkning. Denna grupp blev alldeles för bred.
- Tillstånd för tillverkning av läkemedel: Tanken var att använda öppna listor över olika typer av tillstånd för att producera kemikalier kopplade till läkemedel. Den enda öppna listan vi hittade var listan över anläggningar för industriell framställning av läkemedel genom kemiska eller biologiska processer. Denna lista innehåller endast 19 företag varför den bedömdes som otillräcklig till att användas för att klassificera alla tillverkande företag. Listan används istället till att selektera företag till att vara life science-företag.

7.2.3 Försäljning och marknadsföring

Två försök gjordes för att identifiera företag verksamma inom försäljning och marknadsföring:

- SNI-kod: Vi testade att klassificera de selekterade life science-företagen som använde någon av SNI-koderna inom kategori 46 (partihandel) som verksamma inom försäljning och marknadsföring. Denna grupp blev alldeles för bred.
- Försäljningstillståndslicenser: ADC, Vinnova och referensgruppens idé var att identifiera life science-företag verksamma inom försäljning och marknadsföring genom att hitta öppna listor över försäljningstillståndslicenser för läkemedel eller medicintekniska produkter. Några sådana öppna listor kunde inte identifieras.

7.2.4 Marknadsbolag

Denna kategori av marknadsaktivitet tillkom i slutet av projektet och utifrån nuvarande data var det inte möjligt att automatisera klassificering av dessa företag.

7.2.5 Konsulter

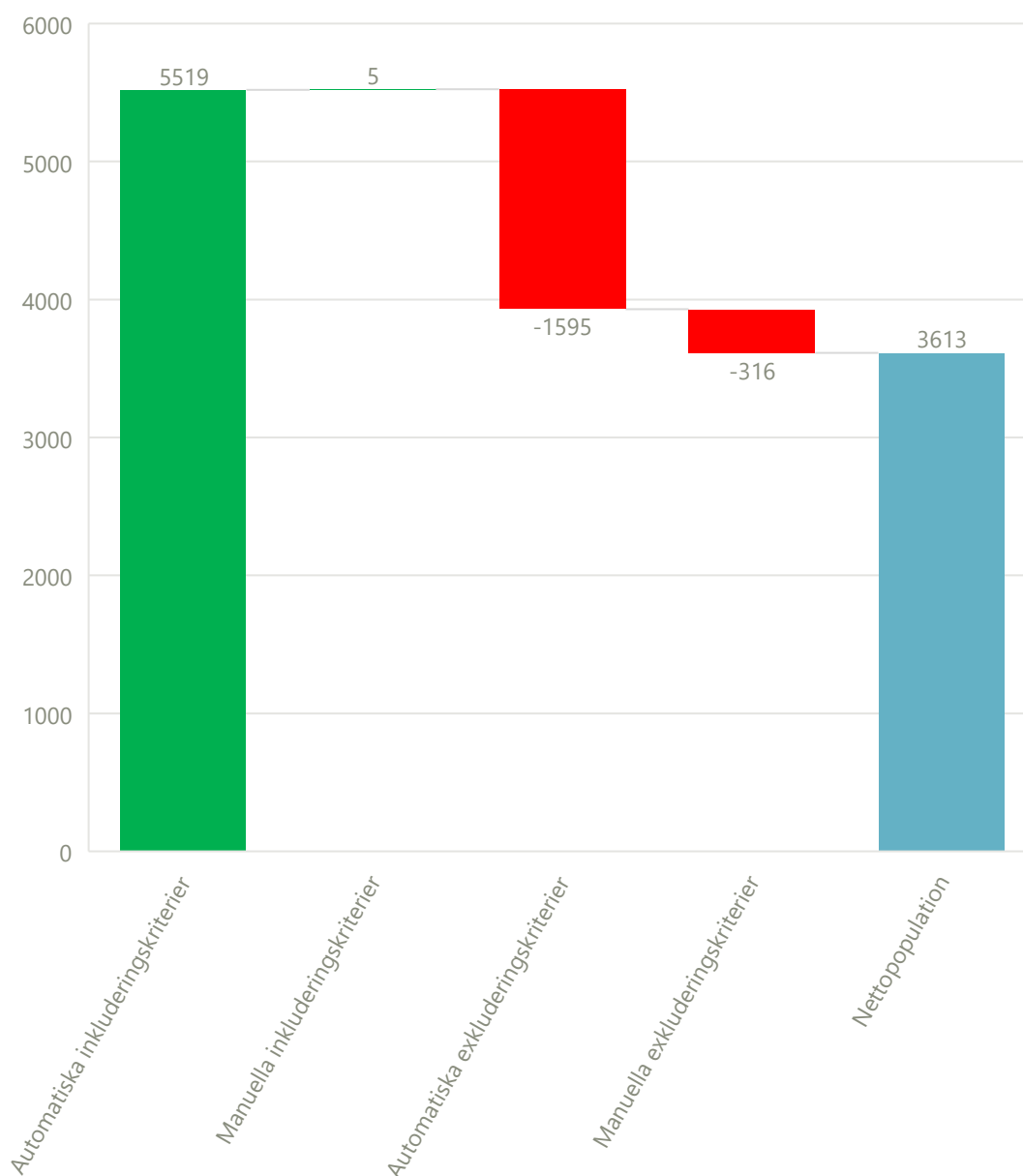
Se avsnitt 4.5.2 för hur konsulter har klassificerats. Inga ytterligare försök har gjorts.

8 Appendix

8.1 Överblick över inkluderings- och exkluderingskriterium

Figur 8.1 presenterar ett vattenfallsdiagram över antal företag som inkluderas respektive exkluderas per kriterium från bruttolistan av företag för att slutligen visa totalt antal företag som ingår i nettopopulationen av life science-företag i Sverige.¹⁵ Bruttolistan som detta diagram utgår ifrån består av 10 614 företag.

FIGUR 8.1



¹⁵ Se avsnitt 4.3 för detaljerad information om respektive inkluderings- och exkluderingskriterier.

8.2 Exkluderande SNI-koder

Företag som har någon av de SNI-koder som presenteras nedan som utesluts från företagspopulationen för life science.

SNI-kod	Benämning
00.009	Huvudnäring okänd
01.280	Odling av kryddväxter, drog- och medicinalväxter
01.430	Uppfödning av hästar och andra hästdjur
01.500	Blandat jordbruk
01.620	Service till husdjursskötsel
01.700	Jakt och service i anslutning härtill
02.101	Skogsförvaltning
02.102	Skogsskötsel
03.210	Fiskodling i saltvatten
03.220	Fiskodling i sötvatten
10.200	Beredning och hållbarhetsbehandling av fisk samt skal- och blötdjur
10.519	Annan mejerivarutillverkning
10.612	Tillverkning av frukostflingor, mixer och andra livsmedelsberedningar av kvarnprodukter
10.830	Framställning av te och kaffe
10.910	Framställning av beredda fodermedel
13.921	Sömnad av gardiner, sängkläder och linnevaror
13.922	Tillverkning av presenningar, tält, segel o.d.
16.102	Hyuling av trä
17.220	Tillverkning av hushålls- och hygienartiklar av papper
19.200	Petroleumraffinering
20.200	Tillverkning av bekämpningsmedel och andra lantbrukskemiska produkter
20.300	Tillverkning av färg, lack, tryckfärg m.m.
20.420	Tillverkning av parfymer och toalettartiklar
20.510	Sprängämnestillverkning
32.400	Tillverkning av spel och leksaker
32.910	Tillverkning av borstbinderiarbeten
43.120	Mark- och grundarbeten
43.222	Ventilationsarbeten
43.320	Byggnadssnickeriarbeten
46.160	Provisionshandel med textilier, kläder, skodon och lädervaror
46.170	Provisionshandel med livsmedel, drycker och tobak
46.340	Partihandel med drycker
46.380	Partihandel med andra livsmedel, bl.a. fisk samt skal- och blötdjur
46.390	Icke specialiserad partihandel med livsmedel, drycker och tobak
46.410	Partihandel med textilier
46.420	Partihandel med kläder och skodon
46.431	Partihandel med elektriska hushållsmaskiner och -apparater

46.450	Partihandel med parfym och kosmetika
46.731	Partihandel med virke och andra byggmaterial
47.241	Specialiserad butikshandel med bröd och konditorivaror
47.291	Specialiserad butikshandel med hälsokost
47.711	Specialiserad butikshandel med herr-, dam- och barnkläder, blandat
47.713	Specialiserad butikshandel med damkläder
47.721	Specialiserad butikshandel med skodon
47.730	Apotekshandel
47.750	Specialiserad butikshandel med kosmetika och hygienartiklar
47.919	Postorderhandel och detaljhandel på Internet med övriga varor
49.320	Taxitrafik
50.102	Icke reguljär sjötrafik över hav och kust av passagerare
64.993	Förvaltning av och handel med värdepapper för en begränsad och slutna krets av ägare
68.310	Fastighetsförmedling
69.201	Redovisning och bokföring
73.111	Reklambyråverksamhet
75.000	Veterinärverksamhet
81.210	Lokalvård
81.300	Skötsel och underhåll av grönytor
85.510	Sport- och fritidsutbildning
85.594	Personalutbildning
86.102	Specialiserad slutna somatisk hälso- och sjukvård på sjukhus
86.103	Specialiserad slutna psykiatrisk hälso- och sjukvård på sjukhus
86.211	Primärvårdsmottagningar med läkare m.m.
86.212	Annan allmän öppen hälso- och sjukvård, ej primärvård
86.221	Specialistläkarverksamhet inom öppenvård, på sjukhus
86.222	Specialistläkarverksamhet inom öppenvård, ej på sjukhus
86.230	Tandläkarverksamhet
86.905	Fysioterapeutisk verksamhet o.d.
86.909	Annan öppen hälso- och sjukvård, utan läkare
87.301	Vård och omsorg i särskilda boendeformer för äldre personer
88.102	Öppna sociala insatser för personer med funktionshinder
90.010	Artistisk verksamhet
90.030	Litterärt och konstnärligt skapande
91.040	Drift av botaniska trädgårdar, djurparker och naturreservat
96.022	Skönhetsvård
96.040	Kroppsvård

8.3 Nyckelord för identifiering

Följande 522 nyckelord används för identifiering av företag som tillhör life science-sektorn.

absorbenter	diagnostiska-produkter	kromatografi	Ortoped
acceleratorer	dialys	kromatografisk	ortopedi
adhesiv	digital uppkopplad vård	kroniska	ortopedisk
adjuvant	disk-implantat	kroppspoteser	ortopedteknisk
aerosolprodukter	DMPK	kroppstemperatur	ortoser
affinering	dosdispensering	kroppsvätskebaserade	osmometrar
agonister	dosering	kroppsvätskor	osteoporosdiagnostik
AIDS	dosimetri	kylväst	osterila
allgener	dragering	kärlintervention	OTC-läkemedel
allergi	drug	kärlsjukdomar	otoskop
allergiska	eHäls	labora- torier	oxalatteknologi
anaerob	e-häls	laborativa	oximeter
analys av biomolekyler	elektrokirurgiska	laboratorie	pacemaker
analysinstrument	endopoteser	labratoriearbete	patient
analysreagensprodukter	endoskop	laparoskopiska	patologi
analysutrustning	enzym	larynx	patologiska
analytisk kemi	enzymatisk katalys	laserutrustning	peptider
anatomisk	ergonomi	life science	pharma
andnings	ergonomiska	lifescience	plasma
ansiktspoteser	exosome	life-science	plåster
ansiktsprotetik	Exosomer	lifesciences	prebiotika
antibakteriell	extraktion	lipidkoncept	precisionsinstrument
antibio	farma	livsvetenskaper	preklinisk
antidrog	farmace	ljusbehandling	pre-klinisk
antigena	farmaceutisk	ljuskänsliga adhesiv	prekliniska
antikropp	farmaceutiska	luftvägs	pre-kliniska
antikroppar	farmaci	lungsjuka	prepa
anti-oxidanter	farmako	lustgas	preventivmedel
antitoxiner	farmakologi	läke	probiotika
antivirala	fermentorer	läkemedel	profylaxutrustning
artros	fertilitet	läkemedels	protaminsulfat
aseptiskt	fertilitetsmonitorer	läkemedelsbolag	proteiner
astma	fetma	läkemedelsbranschen	proteinkemiområdet
audiologi	fibros	läkemedelsforskning	proteinproduktion
audiologisk	fingerpoteser	läkemedelsföretag	proteomics
autoimmuna	flerpatiens	läkemedelsindustri	proteomik
bakterie	fluoriderade	läkemedelsindustrin	poteser
bakteriekultur	flödesrelaterade	läkemedelsområdet	psykiatriska
balansrubbnigar	forskning inom området viktminskning	läkemedelsprodukter	Psykolog
bandage	fysikaliska	läkemedelsprövning	psykosociala
basläkemedel	fysiologi	läkemedelsprövningar	pulsoximeter
behandlingsmetoder	fysiologiska	läkemedelsutveckling	radiologi

bendensiometri	födoämnesöverkänslighet	makromolekyler	radiologiska
benproteser	förbättring av hjärnans funktioner	maligna	radionuklider
benskador	förlossning	masspektrometrar	radiopharmaka
beräkningskemi	förstahjälpnprodukter	med.tekniska	radioterapi
bildanalys- instrument	galektininhibitorer	medcintekniska	reagen
bindarmolekyler	gaskromatografer	medfödda	reagens
bio- och nanoteknik	gastroenterologi	medi- cinsktekniska	recept
bioadditiv	geléartad	medical	regenerativ
bioaktiva	gemmologi	medici	regulatorisk
biobank	geneti	medicin	rehabiliteringsprodukter
biobaserade	genomik	medicin teknik	repositionering
biocid	genprodukter	medicinleveranssystem	reumatism
bioelektronik	gensekvensering	medicinsk	RNA
bioenergiforskning	genspåring	medicinsk forskning	rygggradsimplantat
biofeedback	genteknologi	medicinsk forskning	röntgen
biofysikalisk	genterapi	medicinsk teknik	scintillatorer
biofysikaliska	glasapparater	medicinska	screeningteknologier
bioimpedansbaserade	glaukom	medicinska produkter	serum
bioinformatik	glykosaminoglykaner	medicinska provningar	sjukdomstillstånd
bioinformatisk	GMP	medicinskt	Sjukvård
biokatalysatorer	gynekologi	medicinskteknisk	sjukvårdsartiklar
biokatalytiska	habilitering	medicinsktekniska	sjukvårdsprodukter
biokemi	handikapp	medicinstekniska	skleros
biokemikalier	handinstrument	medicinteknik	skruvfixation
biokemisk	handledsfrakturer	medicinteknisk	smärt
biokemiska	haptiska	medicintekniska	snark
biologi	hematologiområdet	medicintekniskt	spinala
biologiskt	heparinnatrium	medinsk	sprutor
biologiskt material	herpesvirus	medtech	spårmolekyler
biomarkörer	histologisk	membranvesiklar	stamceller
biomaterial	histopatologi	mentala	steril
biomedicin	histoteknik	metabol	stetoskop
biomedicinsk	HIV	microarray	stroboskopi
biomedicinska	hjälpmedel för funktionshindrade	microdialys	strukturbologi
biomimetisk	hjälpmedel för rörelsehindrade	microdialysis	strålbehandling
Biomimetiskt	hjälpmedel inom hälsa	mikrobiell	strålterapi
biomolekyler	hjälpmedelsprodukter	mikrobio	stödbandage
biooptiska	Hjärtstartare	mikrobiologi	substrat
biopharmaceutisk	HLR	mikrodia	svampsjukdomar
biopolymerer	hormon	mikroorganismer	synförbättrande
biopolymerernas	hornhinne	mikroskop	synhjälpmedel
bioprinting	hudparametrar	mikrotom	synkorrektion
bioreaktor	hudplåster	mikrotomer	synskadade
bioreaktorer	hudvävnadsprover	minispirometer	sårläggning
biosciences	human	molekyl analys	sårområdet
biosensorbaserad	huvudvikt	molekylär	sömnapne

biosensorer	hydrofoba	molekylärbiologi	sömnrelaterade
biosensorsystem	Hygien	muskel	tablettdispensering
biosyntetisk	hyperhidros	mykologi	tandimplantat
biotech	hypotermi	nanofiber	tandläkarutrustning
bioteknik	Hälsovård	nanostrukturer	tandmaterial
bioteknikbranschen	högpolymera	nanotrådstrukturer	tandprotetik
bioteknikområdet	hörhjälpmedel	naturpreparat	Tandteknisk
bioteknisk forskning	hörselhjälpmedel	nerv	tandtekniskt
bioteknisk industri	immobilisering	neurala	tarmfloras
biotekniska	immun	neuro	tarmsjukdomar
bioteknologi	immun terapi	neurodegenerativa	terapeutisk
bioteknologisk	immunologi	neurofysiologi	terapeutiska
bioteknologiska	immunreceptorer	neurologi	terapi
biotransformation	immuntoxilogi	neurologiska	tobaksavvänjning
biovetenskap	implantat	neurovetenskap	toxikologi
biovetenskaplig	infektionssjukdomar	nikotin	toxikologiska
biverkningar	inflammat	nukleinsy	tranplantationsområdet
blindhet	inhiber	nålar	transplantationer
blodserum	injektions	näringsprodukter	transplantation
blodsockertestare	inkontinens	näringsstillförsel	tumör
blodtryck	inplantat	obstetrik	ultraljudsteknik
blodtrycksmätning	intensivvård	obstret	uppdragsforskning
bröstproteser	interdentala	obstruktiv	urin
cancer	invitrodiagnostiska	odontologi	urologi
cellbiologi	kanyler	odontologisk	urologiska
celler	karbonater	oftalmologiutrustning	vaccin
cellprodukter	kardio	okularistverksamhet	vaskulära
cellproduktion	katetrar	oligomerer	virus
cellprover	kemiteknik	oligonukleotid	virusproduktion
cellterapi	kirurgi	oligonukleotider	vitro
cro	kirurgiska	oligosackarider	vivo
cytopatologisk	klirik	onkologi	vävnad
datortomografi	klinisk	operation	öroninsatser
defibrillatorer	klinisk forskning	oral	
demenssjuka	klinisk prövning	orphan	
dental	kliniska forskningar	ortoped	
dentalinstrument	kliniskt		
dentalutrustning	klinsk		
dermato	koagulationssjukdomar		
desinfektionsutrustning	koenzym		
detektering	kofaktorer		
detektion	kognitiv		
diabetes	konstgjorda ögon		
diabetsshatteringen	kontaktlinser		
diagnos	kontrastmedel		
diagnositk	kraniofa		

diagnosticering kraniofaciala
diagnostik
diagnostisera
diagnostisk
diagnostiska
diagnostiska tester

8.4 Nyckelord för klassificering

Följande 193 nyckelord används för bransch- eller marknadsklassificering av företagen i life science-sektorn. Nyttillkomna ord markeras med *.

Bioteknik	Labtech	Läkemedel	Medicinteknik
antibakteriell	analysutrustning	agonister	ansiktsproteser
antikroppar	laboratio	antitoxiner	ansiktsprotetik
bindarmolekyler	laboratorie	basläkemedel	audiologi
biofysik	labratorie	biopharmaceutisk	audiologisk
bioinformatik	vitro	biosensorer	behandlingsmetoder
bioinformatisk		cellterapi	benproteser
biokemi		drug	bildanalys- instrument
biokemiska		farma	biooptiska
biologi		farmace	bröstproteser
biologisk		farmaceutisk	datortomografi
biologiska		farmaceutiska	defibrillatorer
biomarkörer		farmako	dental
biomedicin		farmakologi	disk-implantat
biopharma		galektininhibitorer	elektromedicin
biosyntetisk		kemisk-tekniska	endoproteser
biotech		kemiteknik	endoskop
biotekn		kromatografi	ergonomi
bioteknikbranschen		läke	fertilitetsmonitorer
bioteknikområdet		läkemedels	fingerproteser
bioteknik-produkter		läkemedelsbolag	förstahjälpenprodukter
bioteknisk forskning		läkemedelsbranschen	gaskromatografer
biotekniska		läkemedelsforskning	handinstrument
bioteknologi		läkemedelsindustri	hjälpmedel
bioteknologisk		läkemedelsindustrin	hjälpmedel för funktionshindrade
bioteknologiska		läkemedelsområdet	hjälpmedel för rörelsehindrade
biovetenskap		läkemedelsprodukter	hjälpmedel inom hälsa
cellbiologi		orphan	hjälpmedelsprodukter
cellterapi		otc-läkemedel	hudplåster
exosome		pharma	hörselhjälpmedel
exosomer		toxikologi	immobilisering
genmodifierade		vaccin	implantat
genomik			inplantat
gensekvensering			instrument
genspårning			interdentala
hydrofoba			kanyler
högpolymera			katetrar
kemiteknik			kirurgi
koenzym			kirurgiska
kofaktorer			konstgjorda ögon
mikrobiell			kroppsproteser
mikrobio			

Bioteknik

mikrobiologi
oligomerer
oligosackarider
proteinkemiområdet
spårmolekyler
stamceller

Medicinteknik

ljuskänsliga adhesiv
medicintekniska
medical Device
medicin tekniska
medicinsk teknik
medicinsk teknisk
medicinsk utrustning
medicinskteknisk
medicinsk-teknisk
medicinsktekniska
medicinteknik
medicin-teknisk
medicintekniska
medicintekniskt
medtech
odontologisk
oftalmologitrustning
okularistverksamhet
ortoped
ortopedi
ortopedisk
ortopedteknisk
ortoser
otoskop
oximeter
pacemaker
profylaxutrustning
proteser
pulsoximeter
radiologi
radiologiska
radioterapi
rehabiliteringsprodukter
ryggradsimplantat
röntgen
sensor
sjukvårdsartiklar
sjukvårdsprodukter
snark
sprutor
stetoskop
stroboskopi
strålbehandling
strålterapi

Medicinteknik

stödbandage
synhjälpmedel
synkorrektion
tandtekniskt
utrustning
öroninsatser

Konsulter	Diagnostik	Forskning & Utveckling
CMO	diagnostik	Forskning*
CRO	diagnostisering	Biovetenskap*
Kontrakt	diagnostik	
Kontraktsanalys	diagnostisera	
Kontraktsforskning	diagnostisk	
Kontraktstillverkning	diagnostiska	
legotillverkning	diagnostiska tester	
regulatorisk	diagnostiska-produkter	
Regulatoriska		
Regulatoriska konsulter		
uppdragsforskning		

8.5 Exkluderade nyckelord

Företag som nämner något av följande 51 ord i sin verksamhetsbeskrivning utesluts från företagspopulationen för life science. Ord markerade med * är nya för i år:

akupressur
biodrivmedel
brandskydd
dataspel
fordonsindustrin
energilagring
helikopterdelar
honung
hyrläkar
hårdmetall
hästavel
jordförbättring
kennelverksamhet
medicinsk praktik
miljöteknik
miljötillstånd
ortoped¹⁶
naturpreparat
parfym
reseintyg
sexologimottagning
skidspårsutrustning
skogsbruk
skogsresurser
smycken
solceller
tandvårdskliniker
tandvårdsklinik
tatueringsborttagning
travhäst
turism
veterinär¹⁷
viktninskning
vitaminpreparat
ölbryggeri
tandläkarverksamhet
skogs*
frimärken*
sjötransport*
jockey*

¹⁶ Ortoped används som exkluderingsord förutsatt att ordet "skor" nämns i verksamhetsbeskrivningen.

¹⁷ Företag som nämner "veterinär" i sin verksamhetsbeskrivning och samtidigt inte nämner "human" eller "mänsklig" utesluts.

SOFC*
 stuguthyrning*
 målarfärg*
 polartrakterna*
 bordstabletter*
 taxi*
 live-musik*
 hästsportmarknaden*
 sågverk*
 internetdomäner*
 byggnadsvård*

8.6 Variabellista från Proff

Följande tabell visar samtliga variabler som används från Proff för att berika företagsdata.

Variabelnamn
Företagsnamn
organisationsnummer
Grundat
Likviditeringsår
Status(aktiv/inaktiv)
Beskrivning gällande status
antal anställda
Omsättning
verksamhetsbeskrivning
SNI1-5
Län
kommun
Moderbolag
Land (Moderbolag)
Anställda (Historisk data)
Nettoomsättning (Historisk data)
Omsättning (Historisk data)
År
Arbetsställe nr (Cfar)
Arbetsplats namn
Arbetsställe-Län
Arbetsställe-kommun
Postplats
Postnummer
antal anställda(arbetsplats)
Födelsedatum
Titel(person)

8.7 Sök-queries för patent och publikationer

I detta avsnitt presenteras de sökningar i PATSTAT och OpenAlex som har gjorts för att identifiera patentansökningar och publikationer från identifierade life science-företag. I patentdatabasen utgår vi ifrån de identifierade life science-företagen i nettopopulationen och söker huruvida dessa företag har ansökt om patent mellan åren 2009-2023. I publikationsdatabasen undersöker vi huruvida de identifierade företagen har publicerat vetenskapliga artiklar.

8.7.1 Sök-queries för patent

```
SELECT *
```

```
FROM tls207_pers_appln AS appln
```

```
INNER JOIN tls206_person AS person ON appln.person_id = person.person_id
```

```
INNER JOIN tls201_appln ON appln.appln_id = tls201_appln.appln_id
```

```
INNER JOIN tls209_appln_ipc ON appln.appln_id = tls209_appln_ipc.appln_id
```

```
WHERE person.person_ctype = 'SE'
```

```
AND tls201_appln.earliest_filing_year BETWEEN 2009 AND 2023
```

```
AND (tls209_appln_ipc.ipc_class_symbol LIKE 'A61%' OR
```

```
    tls209_appln_ipc.ipc_class_symbol LIKE 'B04%' OR
```

```
    tls209_appln_ipc.ipc_class_symbol LIKE 'C40B%' OR
```

```
    tls209_appln_ipc.ipc_class_symbol LIKE 'G01N%'
```

```
)
```

```
AND person.psn_sector = 'COMPANY' -- Add this line to filter by psn_sector
```

```
ORDER BY appln.appln_id;
```

8.7.2 Sök-queries för publikationer

```
institutions <- oa_fetch(
```

```
  entity = "institutions",
```

```
  country_code = "SE", # Endast svenska institutioner
```

```
  type = c("Company", "Facility", "Other"),
```

```
  verbose = TRUE)
```

ADC Sweden AB
Götgatan 22A
118 46 Stockholm

*Copyright 2024,
ADC*

