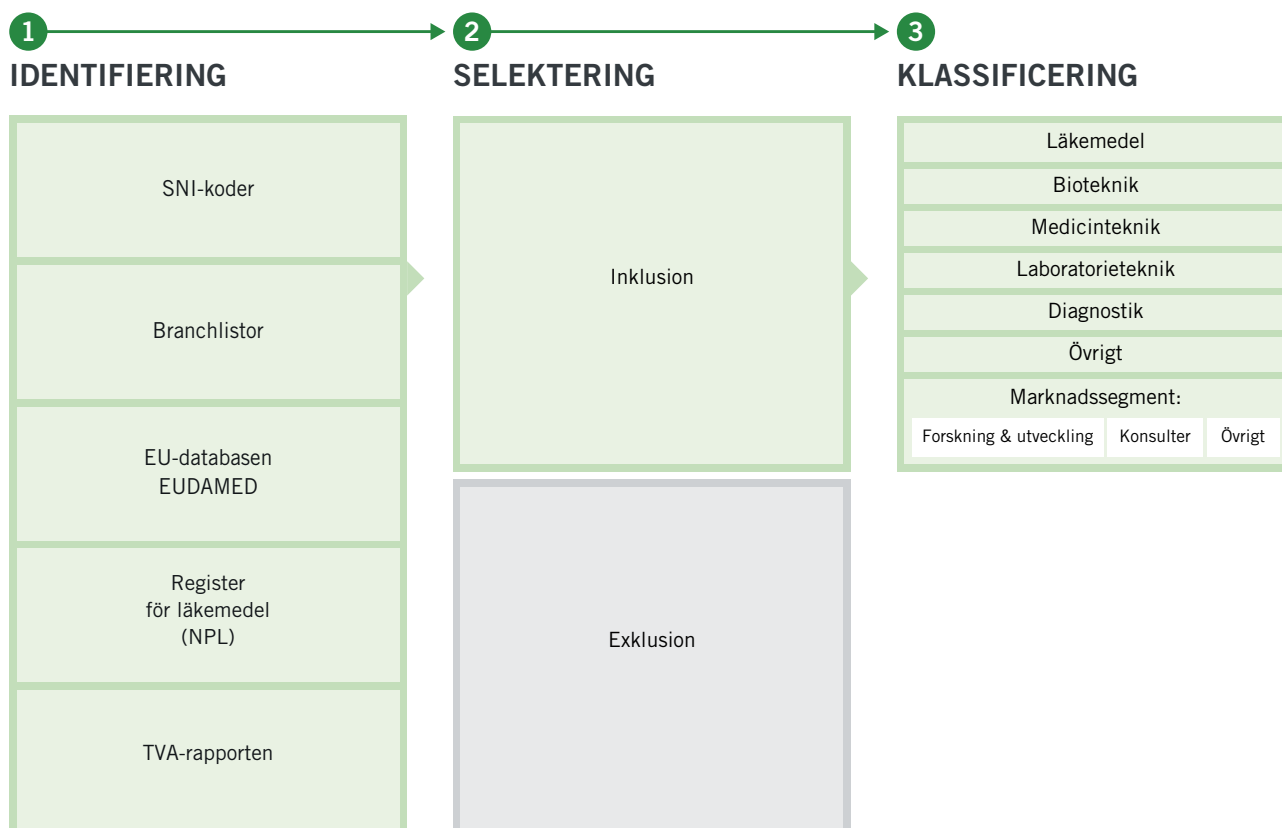


## Bilaga 3. Kort metodbeskrivning för företagspopulationen inom life science

I den här bilagan beskrivs metoden som Vinnova utvecklat för att identifiera life science-företagen i Sverige. Vi beskriver också hur vi arbetat med att sammanställa den statistiken för företagspopulationen. En mer utförlig beskrivning av metoden för att identifiera life science-företagen finns i den metodrapport som skrivits av Amsterdam Data Collective (ADC), ADC är också den konsultfirma som utvecklat metoden tillsammans med Vinnova. Den återfinns i bilaga 4.

### 1 Övergripande metod

Metoden för att identifiera företag inom life science består av tre steg, vilka beskrivs schematiskt i figur 1.



Figur 1. Översikt över metoden för att identifiera, selektera och klassificera företag inom life science-sektorn.

Metodens tre steg är:

- 1. Identifiering** – fokuserar på att säkerställa metodens täckningsgrad genom att identifiera och inkludera samtliga företag i sektorn utifrån valda datakällor.
- 2. Selektion** – fokuserar på att med inklusions- och exklusionskriterier säkra exakthet i metoden genom att endast företag som är verksamma

inom life science-sektorn finns kvar i populationen, och de resterande sorteras bort.

### 3. **Klassificering** – fokuserar på att klassificera life science-företagen i branscher och marknadssegment.

Efter att företagspopulationen identifierats så berikats den med officiella registerdata från SCB, med företagsinformation från Proff, samt data från Svenska riskkapitalföreningen, SVCA.

I kommande avsnitt beskrivs metodens olika steg, samt arbetet med berikning av företagsdata mer i detalj.

#### 1.1 Identifikation av möjligen relevanta företag

Arbetet med identifikation av möjligen relevanta företag börjar med att en bruttolista sammanställs. Bruttolistan innehåller företag som bedöms kunna tillhöra den svenska life science-sektorn för 2023. Genomgående har vi lagt stor vikt vid att använda öppna datakällor<sup>1</sup>, vilket bland annat möjliggör att följa vilka företag som inkluderas i populationen. Att följa vilka företag som inkluderas i populationen är ett viktigt kriterium för att kvalitetssäkra och utveckla populationen och jämföra utfallet av olika identifierings-, selekterings-, och klassificeringskriterier. Det är även betydelsefullt för att kunna tillgängliggöra data om företagen på mikronivå, inklusive företagsnamn och information om anställda, länstillhörighet, med mera.

Medlemslistor från branschorganisationer, medicintekniska- samt läkemedelsregister har bidragit till träffsäkerhet för att skapa bruttolistan, medan företag som inkluderas från registret näringslivsregister och standarden för svensk näringsgrensindelning (SNI-koder) bidragit med hög täckningsgrad. Datakällorna kombinerades sedan och används i en iterativ kvalitetssäkringsprocess, vilket resulterade i en lista på 10 614 företag som bedömdes vara relevanta för vidare analys.

Följande källor användes för att sammanställa den första bruttolistan.

- Samtliga företag som finns i databaserna EUDAMED, nationellt produktregister för läkemedel (NPL), medlemslistor från branschorganisationer, samt listan över life science-företag som identifierades av Tillväxtanalys rapport 2018.
- Företag med ett avgränsat antal utvalda SNI-koder, vilka bedömes som relevanta för life science.

Företagslistan från Tillväxtanalys statistik- och analysrapport från 2018<sup>2</sup> har även använts i kvalitetssäkrande syfte. På listan återfinns 3 356 företag som togs fram under 2017.

<sup>1</sup> Öppna data är data, eller information, som vem som helst har tillgång till och kan använda.

<sup>2</sup> Tillväxtanalys, (2018). Den svenska life science-industrins utveckling – statistik och analys, PM 2018:06

## 1.2 Selektion av företagspopulation inom life science

Med utgångspunkt i bruttolistan från identifieringssteget tillämpas i det här steget ett antal inkluderings- och exkluderingskriterier för att selektera fram life science-populationen. Ett inkluderingskriterium innebär att företaget blir en del av företagspopulationen för life science. Ett exkluderingskriterium innebär att företaget inte blir en del av företagspopulationen för life science.

Följande inkluderingskriterier används för selektion av företag:

- Företaget har relevant första eller andra SNI-kod.
- Företaget återfinns på EUDAMED- eller NPL-listan med första eller andra relevant SNI-kod.
- Företaget återfinns i medlemslista från någon av branschorganisationerna ASCRO, Lif, Swedish Medtech, Swedish Labtech eller SwedenBIO, och har en första eller andra relevant SNI-kod. Medlemslistorna är från senast tillgängliga år (2022 eller 2023).
- Företaget återfinns på Naturvårdsverkets lista över företag med anläggningar för industriell framställning av läkemedel genom kemiska eller biologiska processer och har en relevant första eller andra SNI-kod.
- Företaget har minst ett ord i verksamhetsbeskrivningen som matchar de 522 nyckelord som identifierats som relevanta för life science-företag.

Följande exkluderingskriterier används för exkludering av företag:

- Företaget har SNI-kod som bedömts som irrelevant för inkludering i life science-sektorns företagslista.
- Företaget har något ord i verksamhetsbeskrivningen som matchar de 51 nyckelord som bedöms falla utanför definitionen av life science företag. Dessutom har företag som bedöms vara verksamma som exempelvis veterinärer exkluderats på basen av en kombination av ord.
- Företaget är inte ett aktiebolag<sup>3</sup>.
- 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 av de relevanta SNI-koderna.

I kvalitetssäkringssyfte genomfördes även en manuell granskning av företagspopulationen. I detta moment har ytterligare ett antal företag inkluderats eller exkluderats beroende på om de bedömts tillhöra life science-populationen eller inte, oavsett kriterierna i selektionen. Bedömningen av företagets relevans har genomförts genom att studera företagets verksamhetsbeskrivningar och webbplatser. Trots denna

<sup>3</sup> Innebär att till exempel enskilda företag, kommanditbolag och handelsbolag har exkluderats.

kvalitetssäkring är det möjligt att enskilda företag felaktigt inkluderats eller exkluderats i selekteringsprocessen.

Resultatet av selekteringen blev en "nettolista" med drygt 3 600 företag som bedömdes tillhöra life science-sektorn 2023.

### 1.3 Klassificering av företag i olika branscher och marknadssegment

I det tredje steget klassificerades företagen på nettolistan utifrån branschtillhörighet och marknadssegment. Företagen klassificerades i branscherna; bioteknik, diagnostik, laborieteknik, läkemedel, medicinteknik, eller en övrigt-kategori. Företagen klassificerades också i marknadssegmenten; forskning och utveckling (FoU), konsultbolag, eller en övrigt-kategori.

Följande källor har använts för klassificering av branschtillhörighet:

- Branschorganisationers medlemslistor
- Om företaget finns med på NPL och EUDAMED
- Matchning mot nyckelord i företagets verksamhetsbeskrivningar
- SNI-koder

Klassificeringen av företag i marknadssegment utgick från:

- SNI-koder
- Nyckelord
- Vetenskapliga publikationer
- Patentansökningar

Där klassificeringen av företag som FoU utgick från samtliga dessa källor, medan klassificeringen som konsultföretag enbart utgick från SNI-koder och nyckelord.

Under utvecklingsarbetet genomfördes en rad tester på indelning av företag i de olika branscherna, vilket har påverkat valet av kategorier. Till exempel kunde metoden inte skilja på företag inom in vitro-diagnostik och in vivo-diagnostik, varför diagnostik blir en gemensam kategori.

Under arbetet med metoden 2022 gjordes ett manuellt klassificeringsarbete för företagen som inte tillhörde någon bransch enligt det regelverk som utvecklades. Det rörde sig om cirka 600 företag. Företagen som klassificerades manuellt 2022 har behållit den klassificering tilldelades då. Årets manuella hantering av företag som ej klassificerats automatiskt har varit betydligt mindre omfattande eftersom endast nytillkomna företag som ej kunnat klassificeras automatiskt enligt regelverket har berörts. Företag som varit svåra att klassificera i någon specifik bransch har tilldelats en övrigt-kategori.

### Ett företag kan tillhöra flera branscher och marknadssegment

Ett företag kan tillhöra flera branscher och marknadssegment. I det här avsnittet beskriver vi mer om konsekvenserna av detta på den analys som kan göras över hela sektorn, med branschtillhörigheterna som exempel.

Ett företag kan vara verksamt inom flera branscher och i metoden kan företag därför klassificeras i flera branscher. Exempelvis kan ett företag klassificeras som tillhörande både branschen läkemedel och bioteknik om verksamhetsbeskrivningen innehåller nyckelord för de respektive kategorierna. Därmed är det inte möjligt att addera företagen i kategorierna med varandra och få en total population.

Tabell 1 visar den andel företag inom en bransch som också blivit tilldelad minst en annan bransch. Tabell 2 visar andel företag inom en bransch som överlappar med de andra branscherna, uppdelat på bransch. Tabell 3 visar antal företag inom en bransch som överlappar med de andra branscherna, uppdelat på bransch.

**Tabell 1.** Andelen företag som tillhör flera branscher. Tabellen visar andel företag inom en bransch som samtidigt tillhör minst en annan bransch.

Källa: Vinnova.

Bioteknik	Diagnostik	Laboratorieteknik	Läkemedel	Medicinteknik
56%	99%	80%	52%	36%

**Tabell 2.** Tabellen visar andel företag i procent som tillhör flera branscher. Data är för 2022.

Källa: Vinnova.

	Bioteknik	Diagnostik	Laboratorieteknik	Läkemedel	Medicinteknik	Övrig bransch
Bioteknik	100%	28%	17%	25%	11%	0%
Diagnostik	10%	100%	22%	6%	13%	0%
Laboratorieteknik	3%	12%	100%	2%	5%	0%
Läkemedel	38%	28%	14%	100%	20%	0%
Medicinteknik	29%	96%	73%	34%	100%	0%
Övrig bransch	0%	0%	0%	0%	0%	100%

**Tabell 3.** Tabellen visar överlapp i antal företag vid klassificering i branscher. Data är för 2022.

Källa: Vinnova.

	Bioteknik	Diagnostik	Laboratorieteknik	Läkemedel	Medicinteknik	Övrig bransch
Bioteknik	776	77	25	297	228	0
Diagnostik	77	271	32	75	261	0
Laboratorieteknik	25	32	147	21	108	0
Läkemedel	297	75	21	1192	409	0
Medicinteknik	228	261	108	409	2049	0
Övrig bransch	0	0	0	0	0	114

## 1.4 Berikning av data med hjälp av tre källor

Den identifierade företagspopulationen från selekteringssteget och den indelade populationen från klassificeringssteget berikas med data för att kunna beskriva företagen och dess utveckling över tid. I huvudsak kommer den data som finns presenterad i rapporten från SCB. Men vi presenterar också data från Proff och Svenska riskkapitalföreningen, SVCA.

I fjolårets rapport bearbetade Vinnova själva mikrodata från SCB i MONA<sup>4</sup>. I arbetet med denna rapport har Vinnova övergått till att beställa aggregerade data i form av tabeller från SCB. Beställningen till SCB avser hela företagspopulationen, och inte indelningen i branscher- och marknadssegment. För data kopplat till branscher- och marknadssegment har Vinnova använt dataleverantören Proff som levererar ett mindre urval av öppna data kopplat till företagens utveckling under år 2022. Detta val innebär att företagspopulationerna som berikas skiljer sig något åt, även om de båda har sin grund i den identifierade och selekterade nettolistan. Skillnaden mellan populationerna bedöms som liten, men innebär att Vinnovas data för branscher och marknadssegment ska användas som indikationer på storheter och andelar, snarare än som indelning av statistiken som redovisas för hela life science-företagspopulationen från SCB.

## 2 Risk- och konsekvensanalys

I detta avsnitt redogör vi för några risker och konsekvenser utifrån vald metod.

I utvecklingsarbetet start togs ett beslut att i så hög grad som möjligt använda öppna data om företag. Syftet med det var att kunna generera en lista över företag i sektorn där enskilda företag kan identifieras och således kvalitetssäkras. Arbetet har inneburit en viss begräsning 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.

En automatiserad metod, som varit målet i detta arbete, har stora fördelar, men också sina 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. 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.

Ovan nämnda risker att felaktigt inkludera och exkludera företag ökar dessutom per definition vid en automatisk uppdatering av statistiken om

4 [MONA – SCB:s plattform för mikrodata](#)

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.

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.